Spatial modeling of the association between distance to hospital emergency care and severe anaemia among children aged $1-59$ months in Busia County

Mutinda, Moses Musau

Submitted in partial fulfilment of the requirements for the degree of Master of Science in Statistical Science of Strathmore University

Strathmore Institute of Mathematical Sciences

Strathmore University

Nairobi, Kenya

March 6, 2024

Declaration

I declare that this work has not been previously submitted and approved for award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Name: **Mutinda Moses Musau**

Signature: .....

Date: March 6, 2024

Approval

The thesis of Mutinda Moses Musau was reviewed and approved by the following:

Dr. Evans Otieno Omondi

Supervisor, March 6, 2024

Institute of Mathematical Sciences, Strathmore University.

Dr. Godfrey Madigu

Dean,

Institute of Mathematical Sciences, Strathmore University.

Dr. Bernard Shibwabo

Director,

Office of Graduate Studies, Strathmore University.

Abstract

Background: Access to Emergency Care (EC) services is a core component towards ending preventable paediatric deaths under the SDG 3.2. Physical access to EC is one of the facets of access that has been shown to be associated with health outcomes. Standard regression models often used to assess the association have key limitations including failure to adjust for either spatial heterogeneity in the risk of outcomes or spatial autocorrelation in outcome incidence. This study aimed to develop a Bayesian Model-Based geostatistical model to assess the association between physical access to EC services and severe anaemia among anaemic paediatric admissions in Busia County Hospital using the INLA-SPDE framework and compare changes in the observed association with results from the standard logistic regression.

Methods: Data from a hospital surveillance for paediatric admissions aged $1-59$ months who reside in a malaria endemic setting and were anaemic were assembled. Four models were fitted, two under the INLA-SPDE framework and two under the standard logistic regression framework. Physical access was defined as village travel times to the county hospital and adjustments for known confounders were done including spatial variations in *Plasmodium falciparum* Prevalence Rate (*Pf*PR) as the underlying driver of anaemia. Differences in the travel time coefficients were assessed across the models.

Results: In the developed model, INLA-SPDE model with spatially varying coefficient for *Pf*PR, the association between physical access to EC services and severe anaemia was significant only among admissions within $30-59$ minutes of travel time (AOR: 1.94, 95% CI:$1.18-3.08$) when compared to admissions within < 30 minutes of travel time to Busia County Hospital. In the standard logistic regression models and standard INLA-SPDE model, the risk of severe anaemia was associated with poor physical access across all other admissions in comparison to admissions within < 30 minutes of travel time. However, coefficient confidence intervals under the standard INLA-SPDE model were wider compared to those in the standard logistic regression models.

Conclusions: In assessing the association between physical access to EC services and health outcomes, it is vital to not only adjust for spatial heterogeneity in the underlying drivers of health outcomes, but also to appropriately model the association. Further, in the presence of spatial dependence, models should account for spatial autocorrelation so as not to underestimate standard errors.

KEY WORDS: Spatial heterogeneity, Spatial autocorrelation, INLA-SPDE, Anaemia, Malaria, Travel time

Contents

List of Figures vii

List of Tables viii

List of abbreviations ix

Acknowledgement x

Dedication xi

1 Introduction 1

1.1 Background to the Study 1

1.2 Statement of the Problem 3

1.3 Research Objectives 5

1.3.1 General Objective 5

1.3.2 Specific Objectives 5

1.4 Justification of the Study 5

1.5 Significance of the Study 6

2 Literature Review 7

2.1 Introduction 7

2.2 Models 7

2.2.1 Logistic Regression 7

2.2.2 Cox Proportional Hazard Regression 14

2.3 Gaps Identified 16

2.4 Current Research 17

3 Methodology **19**

3.1 Introduction . . . 19

3.2 Study Area . . . 19

3.3 Data Description . . . 22

3.3.1 Data Source: Hospital Surveillance . . . 22

3.3.2 Residence Details . . . 23

3.3.3 Outcome variable: Anaemia Diagnosis . . . 24

3.3.4 Physical Access to Emergency care: Travel time . . . 25

3.3.5 *Plasmodium falciparum* Prevalence Rate (*Pf*PR) . . . 28

3.3.6 Possible Confounders . . . 32

3.3.7 Inclusion and Exclusion Criteria . . . 34

3.4 Statistical analysis . . . 35

3.4.1 Logistic Regression model . . . 35

3.4.2 Spatial Models: Bayesian Inference . . . 36

3.4.3 Latent Gaussian models . . . 37

3.4.4 Inference with MCMC . . . 38

3.4.5 Inference with INLA . . . 39

3.5 Parameter Estimation . . . 42

3.5.1 Covariate Selection and model diagnostics . . . 42

3.5.2 Notations . . . 43

3.5.3 Standard Logistic Regression . . . 43

3.5.4 Evidence for Spatial Autocorrelation . . . 44

3.5.5 Standard INLA-SPDE model . . . 45

3.5.6 INLA-SPDE Model: Spatially Varying Coefficient . . . 45

3.6 Model Comparison . . . 46

4 Results and Interpretation **47**

4.1 Introduction . . . 47

4.2 Descriptive Results . . . 47

4.3 Covariate Selection and Model Diagnostics . . . 55

4.3.1 Covariate Selection 55
4.3.2 Model Diagnostics 55
4.3.3 Evidence for Spatial Autocorrelation 57
4.4 The association between physical access to EC services and Severe Anaemia 58

5 Discussions, Conclusions and Recommendations 63
5.1 Pre-amble 63
5.2 Discussion 63
5.2.1 Strengths and Limitations 65
5.2.2 Implications for Policy 65
5.2.3 Future Directions 66
5.3 Conclusion 66

Bibliography 67

Appendix A Previous Studies 76

Appendix B Travel Time Maps 83

Appendix C Ethics Approval 84

Appendix D Similarity Index 86

Appendix E R code 89

List of Figures

Figure 3.1: Study Area 20
Figure 3.2: Rainfall pattern in Busia County 21
Figure 3.3: Annual Spatial Distribution of *Pf*PR 31
Figure 3.4: Variogram Illustration (Source: Bohling (2005)) 44

Figure 4.1: Data Inclusion criteria 48
Figure 4.2: Travel Times by Age and Anaemia status. The values are the counts of admissions for the anaemia classes and the p-values are from wilcoxon test for medians 49
Figure 4.3: *Pf*PR by age and Anaemia status. Red dots represent the position of the mean, the values are the % mean *Pf*PR for the anaemia classes and the p-values are from wilcoxon test for means 50
Figure 4.4: Spatial Distribution of Anaemic Admissions 51
Figure 4.5: Hospital Incidence by Travel Time 52
Figure 4.6: Linearity Assumption 55
Figure 4.7: Influential Outliers Assumption 56
Figure 4.8: Variogram: Evidence for Spatial Dependence. 57
Figure 4.9: Effects size of Spatial Autocorrelation on Risk of severe anaemia . . 60
Figure 4.10: Effects size of Spatially varying *Pf*PR on Risk of severe anaemia . . 62

Figure B.1: Spatial Distribution of Travel time in Busia County 83

List of Tables

Table 3.1: Anaemia Definitions 24

Table 3.2: Land cover classifications, Associated speeds and Modes of Transport 27

Table 3.3: Possible Confounders 32

Table 4.1: Demographic Characteristics 53

Table 4.2: Admission and Residence Characteristics 54

Table 4.3: Multicollinearity Assumption 57

Table 4.4: Impact of Physical Access to EC on risk of severe anaemia 58

Table A.1: Studies evaluating the association between Physical Access to EC services and Community Health Outcomes 77

Table A.2: Studies evaluating association between Physical Access to EC services and in-hospital mortality 79

Table A.3: Studies evaluating association between Physical Access to EC services and disease severity. 80

List of abbreviations

AOR	Adjusted Odds Ratio	IPTp	Intermittent Preventive Treatment of malaria
BCRH	Busia County Referral Hospital	LLINs	Long Lasting Insecticide Treated Nets
CI	Confidence Interval	LM	Linear Models
CIN	Clinical Information Network	LMIC	Low and Middle Income Countries
EC	Emergency Care	MCMC	Monte Carlo Markov Chains
GAM	Generalized Additive Models	MoH	Ministry of Health
GAMM	Generalized Additive Mixed Models	PAR	Paediatric Admissions Record
GLM	Generalized Linear Models	*Pf*PR	*Plasmodium falciparum* Prevalence Rate
GLMM	Generalized Linear Mixed Models	REDCap	Research Electronic Data Capture
HB	Haemoglobin	SPDE	Stochastic Partial Differential Equation
HDSS	Health and Demographics Survey	VHT	Village Health Team
INLA	Integrated Nested Laplace Approximation	WHO	World Health Organization

Acknowledgement

First and above all, I am grateful to the Almighty God for His provision, grace, and for granting me good First and above all, I am grateful to the Almighty God for His provision, grace, and for granting me good health throughout the study period.

I am grateful to my mother, Rosemary Musau, for her immeasurable support throughout the study period.

I am grateful to Prof Bob Snow from the KEMRI-Wellcome Trust for the scholarship and his invaluable supervisory insights. In addition, to colleagues of the KWTRP who provided the data from Busia County Referral hospital that form part of continuous surveillance from the Clinical Investigation Network, notably Dr Sam Akech, Ms Cynthia Khazenzi and Ms Eda Mumo, and finally members of the Population and Health Information Surveillance Group at KWTRP, notably Prof Emelda Okiro and Dr Peter Macharia who provided comments on the draft thesis.

I am indebted to my supervisor, Dr. Evans Otieno Omondi for his availability, kind guidance, and support without which this thesis would not have been a success and my classmates for their support and encouragement.

Funding for the data assembly from Busia County Referral hospital and analysis was provided by Prof Bob Snow as part of the Wellcome Trust Principal Research Fellowship (212176/Z/18/Z), the Wellcome Trust to the Core Award for the East Africa Major Overseas Programme (203077/Z/16/Z) and support by the CDC Foundation, through funding from the World Health Organization for the RTS,S evaluation (project requisition number 2018/854999).

Dedication

This thesis is dedicated to God Almighty for giving me wisdom and good health.

Chapter 1

Introduction

1.1 Background to the Study

Preventable paediatric deaths continue to be a major public health concern (UN, 2015). In 2021, approximately five million deaths occurred among children under five years of age (UN-IGME, 2022). Sustainable Development Goal 3.2 (SDG 3.2) aims at ending preventable paediatrics deaths through the implementation of key lifesaving interventions. Among these, access to quality emergency care services plays a critical role, representing a key component in achieving international goals (WHO, 2017).

Emergency care (EC) services include rapid assessment or triage, diagnostics, prompt treatment, stabilization, referral, and close monitoring; all aimed at the prompt provision of urgent medical care to avert further severe disease progression, disability and death (Razzak and Kellermann, 2002). Studies examining the impact of various EC services consistently show an association between improved patient outcomes and the respective EC service (Clark et al., 2012; Molyneux et al., 2006; Robison et al., 2012; Sylverken et al., 2019). For instance, in the case of malaria, a major contributor to paediatric deaths in sub-Saharan Africa (SSA) (UN-IGME, 2022), EC services, such as prompt intravenous artesunate, have been shown to prevent progression to in-hospital mortality (Hadjilaou et al., 2023). Similarly, access to prompt blood transfusion has been linked to reduce poor outcomes in children with anaemia (Ackerman et al., 2020; Ippolito et al., 2022). In resource-limited settings, improved emergency referral systems to higher level facilities for specialized care has proven to be critical towards child survival (Fitzgerald et al., 2018). These studies collectively underscore the importance of promptly receiving EC services in the management of illnesses

amongst paediatric patients, potentially saving lives and reducing paediatric mortality burden in low-middle income countries (LMICs).

To actualize the benefits of EC services, at-risk populations must have access to these services. In many LMICs, hospitals and higher order facilities serve as the primary point for EC services (Ouma et al., 2018). Access to EC encompasses various dimensions including availability, affordability, acceptability, accommodation, and accessibility (Levesque et al., 2013). Availability refers to the presence of EC services that meet the demands or needs of the population. This includes factors such as facility density, personnel, equipment and the range of services provided, especially those related to EC services that the population requires. Affordability relates to the population's financial capacity to access and afford the necessary EC services. Acceptability examines the alignment between the population's expectations and actual delivery of EC services based on socio-demographics factors such as economic status, age and ethnicity. Accommodation focuses on the coordination and organization of EC services, such as operating hours, to ensure they meet the population's needs (Ouma et al., 2021). Lastly, accessibility, commonly referred to as physical access, refers to the ability of individuals seeking EC services to physically reach the points of service (healthcare facilities) from their point of need (Moturi et al., 2022; Ouma et al., 2021). Accessibility involves various factors including distance, travel time, transportation options, and the presence of physical barriers that may hinder or facilitate physical access to care.

The associations between health outcomes, EC services and access have prompted calls for research studies aimed at understanding the association between various access components and health outcomes (Aluisio et al., 2019; Calvello et al., 2013; Hansoti et al., 2017; Razzak et al., 2019). In the context of physical access to care, the focus of the current study, the goal has been to establish whether there exists an increased risk of adverse health outcomes among individuals with poor physical access to EC services (see Appendix A). Understanding the variations in the risk of adverse health outcomes with respect to physical access to EC services is important as it 1) identifies marginalized at-risk populations and 2) informs policy on the development of interventions to ensure that all at-risk populations can reach the points of service administration.

1.2 Statement of the Problem

Analyses assessing the association between physical access to EC services and health outcomes overlook spatial variations in the underlying drivers of event risk (Al-Taiar et al., 2008; Ippolito et al., 2018; Kadobera et al., 2012; Kahabuka et al., 2012; Karra et al., 2017; Kashima et al., 2012; Kazembe et al., 2006; Lohela et al., 2012; Magnani et al., 1996; Manongi et al., 2014; Moïsi et al., 2011; Mousa et al., 2020; Mpimbaza et al., 2017; Målqvist et al., 2010; Mutsigiri-Murewanhema et al., 2017; Noori et al., 2021; Quattrochi et al., 2020; Rees et al., 2016; Rutherford et al., 2009; Zoungrana et al., 2014) (see Appendix A). Spatial variations in underlying drivers of risk refers to instances where the mechanism causing the occurrence of health outcomes is not homogeneously distributed in space. As such, the health event is more likely to occur in individuals residing in high-risk areas compared to those in low-risk areas. Not accounting for spatial variations in the underlying risk drivers may result in under/over estimation in the association between physical access to EC services and health outcomes (Magnani et al., 1996; Manongi et al., 2014; Zoungrana et al., 2014)

In addition, most analyses assessing the association between physical access to EC services and health outcomes fail to account for spatial autocorrelation of health outcomes (Al-Taiar et al., 2008; Ippolito et al., 2018; Kadobera et al., 2012; Kahabuka et al., 2012; Karra et al., 2017; Kashima et al., 2012; Lohela et al., 2012; Magnani et al., 1996; Manongi et al., 2014; Mousa et al., 2020; Mpimbaza et al., 2017; Målqvist et al., 2010; Mutsigiri-Murewanhema et al., 2017; Noori et al., 2021; Quattrochi et al., 2020; Rees et al., 2016; Rutherford et al., 2009; Zoungrana et al., 2014) and in the few that do, spatial autocorrelation is adjusted for at large geographical scales (Kazembe et al., 2006; Moïsi et al., 2011). Spatial autocorrelation of health outcomes refers to the clustering of health outcomes in space such that individuals near other individuals already with the health outcome are more likely to experience the health outcome compared to those further away (Lawson, 2013). Therefore, spatial autocorrelation is a significant confounder in the risk of infection that should be accounted for in the analysis. Further, accounting for spatial autocorrelation at large geographical scales masks the local heterogeneities of disease risk that exist and should be done at a finer geographical scale.

The choice of statistical methods to investigate the association between physical access to EC services and health outcomes is subject to data availability in different formats; for individual level data, the standard logistic regression is the most utilized statistical method (see Appendix A). Accounting for variations in the underlying risk driver within the standard logistic regression framework would imply that the regression coefficients are constant across space; stationarity assumption in space (Gelfand et al., 2003). However, the stationarity assumption may not be appropriate in cases where the impact of the underlying driver on the outcome under study varies in space (Assunção, 2003; Assunção et al., 2002; Gelfand et al., 2003; Okango et al., 2016). In addition, in the presence of spatial dependence in the occurrence of health outcomes, the independence assumption of standard errors under the standard logistic regression model is violated and consequently, coefficient standard errors would be underestimated (Wilson and Lorenz, 2015).

Model-based geostatistics (MBG) provides an alternative in assessing the association between physical access to EC services on health outcomes as they have provisions to incorporate spatial autocorrelation and spatially varying coefficients (Diggle et al., 1998; Gelfand et al., 2003).

The current study proposes to use severe hospitalized anaemia among paediatrics in a malaria endemic setting to develop a Bayesian MBG model that incorporates spatial heterogeneity in disease risk and spatial autocorrelation of disease incidence in assessing the association between physical access to EC services and severe anaemia. This approach promises a more robust understanding of the relationship between physical access to EC services and health outcomes thus providing valuable insights into the impact of accessibility on health outcomes.

1.3 Research Objectives

1.3.1 General Objective

The **main objective** of this study is to develop a Bayesian MBG model to investigate the association between physical access to emergency care services and severe anaemia among children aged $1-59$ months in Busia County Referral Hospital, Western Kenya, using the Integrated Nested Laplace Approximation-Stochastic Partial Differential Equation (INLA-SPDE) framework.

1.3.2 Specific Objectives

1. To develop a Bayesian MBG model to assess the association between travel time and severe anaemia adjusting for known risk factors using INLA-SPDE.

2. Evaluate the potential modifying effect of spatial autocorrelation and varying levels of malaria transmission intensity on the association between physical access to EC services and the risk of severe anaemia.

1.4 Justification of the Study

Severe anaemia, characterized by low number of red blood cells (and consequently low oxygen concentrations) (WHO, 2011), continues to be a significant health concern with an estimated 60.2% of children < 5 years suffering from anaemia in Africa (Kassebaum et al., 2014; WHO, 2021). Various factors contribute to severe anaemia including nutrition status and infection (Chaparro and Suchdev, 2019; Kwambai et al., 2022; Loechl et al., 2023). In malaria endemic regions, malaria infection notably stands out as a major cause of anaemia (Loechl et al., 2023; White and Watson, 2018). Furthermore, paediatric hospital admissions with anaemia face an increased risk of readmission from severe anaemia or death within 6 months after discharge (Kwambai et al., 2022; Phiri et al., 2012). This heightened risk has

prompted recent calls for the implementation of post-discharge malaria chemoprevention to mitigate these adverse outcomes (Kwambai et al., 2018; Phiri et al., 2012). However, the projected health impact of the intervention will be lower where severe anaemia cases have poor access to hospital care as they need to access the hospital for subsequent monthly doses (Okell et al., 2023).

By developing an improved model on the association between physical access to EC services on severe anaemia, the output of this study will aid in informing the design of effective interventions to ensure none of the populations at-risk are marginalized.

1.5 Significance of the Study

The current project aims to provide insight on the role of physical access to EC on severe anaemia in a malaria endemic setting. The findings are expected to contribute valuable insights that can inform policy decisions aimed at optimizing outcomes in severe anaemia patients. This includes consideration for timely interventions such as blood transfusion on admission and the implementation of post-discharge chemoprevention for survivors.

Chapter 2

Literature Review

2.1 Introduction

This section provides a review of the statistical models that have been previously used to assess the association between physical access to EC services and disease related health outcomes on individual level data. The focus is on their application and limitations. In addition, the section concludes by outlining the gaps identified across all the studies and highlighting how the current study aims to overcome the identified gaps.

2.2 Models

The most commonly utilized models in studies assessing the association between physical access to EC services and health outcomes are logistic regression and cox proportional hazard regression (Appendix A). The following is a detailed description of how the models have been utilized across studies.

2.2.1 Logistic Regression

Logistic Regression is a statistical technique under Generalized Linear Models (GLMs) applied to binary outcomes (James et al., 2013). The binary outcome is assumed to come from a Bernoulli distribution and the goal of the logistic regression model is to predict the probability of event occurrence (James et al., 2013).

Magnani et al. (1996) evaluated the impact of a national initiative in rural Niger that was aimed at improving access to primary care through upgrading of dispensaries and establishment of village health teams (VHT) using a logistic regression model. Physical access for each village was defined as either being $\leq$ 5km from a dispensary, $>$ 5km from a dispensary but served by a VHT or $>$ 5km from a dispensary and not served by a VHT whereas children born (and their mortality outcomes) within 5 years prior to the survey were the outcome variables (Magnani et al., 1996). The study revealed that children living $\leq$ 5km from a dispensary had lower odds of mortality (OR: 0.68; 95% CI : $0.49 - 0.94$) compared to those $>$ 5km from a dispensary and not served by a VHT. In addition, risk of mortality for children residing $>$ 5km from a dispensary but served by a VHT was not significantly different from those $>$ 5km from a dispensary and not served by a VHT (OR: 0.80; 95% CI: $0.58 - 1.09$). The study, however, did not adjust for mortality risk factors that occurred during the study period including famine and measles outbreaks in Niger; mortalities from areas that were heavily impacted by such events may confound the observed findings.

In rural Gambia, Rees et al. (2016) investigated the risk factors to severe illness at presentation for diarrheal disease, lower respiratory infections and malaria among paediatrics using logistic regression. Some of the villages in the study were defined as "core villages" due to the existence of interventions including transportation incentives and presence of child welfare clinics providing, among other services, child vaccination services (Rees et al., 2016). Physical access, measured as distance to clinic, was found not to be associated with severe illness at presentation. However, the study revealed reduced risk of severe disease among paediatrics residing in "core villages" characterized by availability of free transport (OR: 0.557; 95% CI: $0.325 - 0.954$). Rees et al. (2016) posits that the results maybe confounded by the continued research/interventions in “core villages" that include health care awareness campaigns that may have resulted in reduced risk of illness due to improved health practices. Such findings illustrate the role of community characteristics in modifying the impact of similar exposures to an underlying driver risk on the occurrence of adverse health outcomes.

Manongi et al. (2014) used logistic regression to model risk of inpatient child mortality by travel time in an intense malaria transmission area. The study included day admissions aged

2 - 59 months with a fever or fever history and travel times from their village of residence. The findings were that children residing in villages whose travel time was > 3 hours to the health facility were at a high risk of mortality (OR: 2.23; 95% CI: 1.17 – 4.27) (Manongi et al., 2014). The limitations in the Manongi et al. (2014) study included the exclusion of night-time admissions which introduced bias as admissions outside working hours are more likely to be severe cases. In addition, Manongi et al. (2014) cites greater disease exposure in different areas as a significant confounder in their results that could not be accounted for due to lack of data.

A study in rural Tanzania by Kahabuka et al. (2012) utilized a logistic regression model to investigate the risk factors for severe disease for malaria, pneumonia and diarrhea. The study was a hospital-based cross-sectional design study with all participants being recruited at the outpatient department of the district hospital. The study established that the risk of severe disease was positively associated with travel time to the district hospital for severe pneumonia (OR: 1.6; 95% CI: 0.9 – 3.1) and severe diarrhea (OR: 3.9; 95% CI: 1.4 – 10.9) but not for severe malaria (OR: 0.9; 95% CI: 0.3 – 2.9). Although the study stratified the analysis by disease, disease specific risk determinants at areas of residence such as exposures to malaria transmission for malaria and hygiene levels for diarrhea were not included in the model.

A similar study by Zoungrana et al. (2014) used a logistic regression in a cross-sectional study during a high malaria transmission season to explore determinants of severe malaria among paediatric admissions in Burkina Faso. Physical access was assessed as both travel time and distance to the health facilities and in both cases, no association was found between physical access and severe malaria (Zoungrana et al., 2014). However, despite the study being conducted during a high malaria transmission season and results revealing that children residing in areas with high malaria intensity (such as those near water ponds) have higher chances of developing severe malaria (OR: 1.67; 95% CI:1.02 – 2.74), the study did not adjust for spatial variations in malaria transmission rates (Zoungrana et al., 2014).

Målqvist et al. (2010) conducted a study in Vietnam to understand the association between distance and early childhood mortality. The study utilized Euclidean distances as a measure

of physical access, recorded all live births and early childhood deaths that occurred during the study period and analyzed the risk of mortality using a logistic regression model. The study established increased risk of mortality for those residing > 1.2km away from a health facility (OR: 1.96, 95% CI: $1.40 - 2.75$) compared to those residing < 1.2km to a health facility (Målqvist et al., 2010). The study area covered by Målqvist et al. (2010) is described as highly mountainous, therefore, the use of Euclidean distances represents a crude measure of physical access. Further, the logistic regression model was adjusted for only maternal age at delivery and marital status; adjustments for variations in the risk of mortality at residences were not included in the model.

Mousa et al. (2020) conducted a systematic literature review and pooled analysis to investigate the impact of delayed treatment on severe malaria. In their study, a mixed-effects logistic regression (to account for random variations among studies) was utilized to investigate the effect of covariates, including physical access, on severe malaria (Mousa et al., 2020). The finding was that poor physical access, travel time or distance, was associated with increased risk of severe malaria. The study was, however, limited in terms of small sample sizes, not accounting for variations in the risk of malaria and not adjusting for the inclusion of participants with comorbidities such as HIV.

Other studies have utilized data from Demographic Health Surveys (DHS) to overcome the challenges of sample size. For example, Karra et al. (2017) used multivariable logistic regression to investigate the risk of neonatal mortality across 21 low- and middle-income countries (LMICs) from DHS data. Clusters within the DHS were used as places of residence and distances from each cluster to the nearest facility were used as a measure of physical access to EC services. The study established that children residing in clusters $\geq$ 2km from a health facility had higher risk of mortality (OR: 1.163; 95% CI: $1.020 - 1.327$) compared to those living within 1km of a health facility. The outcome, neonatal mortality, in the study includes all-cause neonatal mortality. As such, the results of the study are confounded by exposures to the different risks of neonatal mortality at both country and cluster levels; this was not accounted for in the analysis.

Kashima et al. (2012) investigated the association between early childhood mortality and proximity to health centres using a random intercept logistic regression. The study utilized data from a national DHS where neonatal and infant mortality estimates were outcomes of interest and distances between cluster locations and nearest health centres were measures of proximity (Kashima et al., 2012). The odds ratios were 1.36 (95% CI: $0.92-2.01$) and 1.42 (95% CI: $1.06-1.90$) for neonatal mortality and infant mortality respectively among residents $\geq$ 5km from health centres compared to those within $>1.5-3.0$km. The use of random intercept model at sub-national regions was an attempt to adjust for variations in disease exposures, however, the intercept at regions as opposed to areas of residence still mask spatial differences at finer scales.

Another study using DHS data was conducted by Lohela et al. (2012) in Zambia and Malawi aimed at investigating if there exists an association between distance to EC services and neonatal mortality. The study used national DHS data from the two countries and focusing on clusters in rural areas, developed a logistic regression model to assess the association between Euclidean distances from each cluster to delivery care on neonatal mortality (Lohela et al., 2012). The study's findings were that distance was not associated with neonatal mortality in Malawi (OR: 0.97; 95% CI: $0.58-1.60$) whereas in Zambia, longer distances were associated with reduced odds of mortality (OR 0.55; 95% CI: $0.35-0.87$). The findings were attributed to various limitations in the study including the use of clusters to approximate household residences, geoscrambling of DHS clusters and underreporting of mortality in remote areas due to difficulties in access by survey staff (Lohela et al., 2012). As such, Lohela et al. (2012) recommends the use of more accurate estimates of physical access and use of data from surveillance systems in which under-reporting of events is minimal.

Rutherford et al. (2009) utilized Health and Demographic Surveillance Survey (HDSS) data, that is a more accurate surveillance system compared to DHS, to investigate associations between all-cause paediatric mortality and access to health care services. In an attempt to account for confounding effects, the study utilized a 1:5 case-control design where matching was by age and sex and a conditional logistic regression was used for analysis. Physical access, both travel time and distance to nearest health care facility, were found not to be

associated with child mortality. The use of community all-cause mortality as an outcome is inclusive of deaths from causes not related to physical access or severe disease such as those from fatal accidents, burns, poisoning or animal bites which could confound the findings and Rutherford et al. (2009) suggests future studies should consider using other outcomes. Further, case-control matching was not based on residence, therefore, accounting for the spatial variations in underlying risk of mortality across the study area was not done.

Ippolito et al. (2018) utilized conditional logistic regression model to investigate the risk of in-hospital mortality among severe malaria admissions. The study design was a time-matched case control study with the goal of controlling for temporal confounders including difference in road conditions, availability of blood units among other hospital resources. Ippolito et al. (2018) revealed a 4% increase in the odds of mortality with every 1km of distance to the health facility. The limitations of the study were that the study did not control for co-infections that might have resulted in mortality among cases and case-control matching did not include areas of residence, thus variations in the underlying exposures to mortality at areas of residence was not adequately accommodated.

Mutsigiri-Murewanhema et al. (2017) conducted a $1:2$ unmatched case control study where cases and controls were recruited from similar areas to investigate factors associated with severe malaria using logistic regression. The study revealed that paediatrics residing in areas $\geq$ 10km from the nearest health facility had odds ratio of 14.35 (95% CI$= 1.30, -158.81$) for severe malaria compared to those in areas $<$ 10km to the nearest health facility. The key limitation of the study was that the minimum sample size necessary to conduct the $1:2$ unmatched case control study was not achieved (Mutsigiri-Murewanhema et al., 2017), which in part, explains the large CI observed in the odds ratio. In addition, using unmatched case-control design did not account for differences in exposures to malaria transmission among cases and controls as intended.

Similar to Mutsigiri-Murewanhema et al. (2017), Al-Taiar et al. (2008) conducted a case-control study to establish risk factors for the progression of malaria from mild malaria to severe malaria in a malaria endemic setting using a logistic regression model. Cases, paediatrics with severe malaria, were recruited from the main public hospital in the study area

whereas controls were recruited from other health facilities within the catchment of the main public hospital and matching was based on age. The findings by Al-Taiar et al. (2008) were that residing in areas $>$ 2km away from the nearest health centre was a risk factor for malaria progression to severe form. Al-Taiar et al. (2008) further notes that paediatrics residing in areas known to have high malaria transmission were less likely to develop severe malaria and attributes the finding to the complex interaction between malaria transmission intensities and acquired immunity over time. Such a finding illustrates not only the importance of adjusting for differences in exposure intensities but also the spatially varying effects on the populations at risk.

Mpimbaza et al. (2017) applied conditional logistic regression model in a case-control study to investigate risk factors for severe malaria among paediatrics in Jinja, Uganda. Matching was done by age, area of residence and time in an effort to reduce bias in confounding effect of age, variations in disease exposure and temporal confounders respectively ((Mpimbaza et al., 2017)). The findings were that distance to facilities of level III or higher was associated positively with severe malaria; distance to lower-level facilities was not associated with severe malaria. However, the study was limited by the enrollment of non-representative controls where cases were enrolled from a level VI hospital whereas controls were enrolled at level III and IV hospitals which introduced bias as higher-level hospitals are few (therefore longer distances) and handle more severe cases.

Moïsi et al. (2011) utilized a hospital-based surveillance to explore the association between physical access and child survival in Kilifi, Kenya, for children hospitalized with pneumonia or suspected meningitis using a logistic regression. The study recognized presence of spatial clustering of disease events and made adjustments using spatial bootstrapping for sublocations and fitted the model on 50 bootstraps (Moïsi et al., 2011). The study revealed increased odds of mortality with each hour increase in pedestrian travel times (OR: 1.12; 95% CI: $1.06 - 1.18$) and with each hour increase in half hour of vehicular travel times (OR: 1.10; 95% CI: $1.02 - 1.19$). Although the study used more accurate residence locations, villages, spatial clustering of disease was adjusted at a larger geographic area, sublocation, masking

differences that may exist at village level which is a finer geographical scale. Further, no adjustment was done for spatial variations in the risk of disease at the villages of residence.

Similarly, Kazembe et al. (2006) used a spatial logistic regression model under the Monte Carlo Markov Chain (MCMC) framework to adjust for the confounding effect of disease clustering in their study aimed at examining the association between physical access to EC services and malaria-related in-hospital paediatric mortality in Malawi. In the study, admissions were aggregated to wards administrative level which formed the unit of spatial analysis for both calculation of Euclidean distances to the health facility and spatial random effects and distance was binary with a cut-off of 5km (Kazembe et al., 2006). The study established that paediatric patients from wards $\leq$ 5km were at a reduced risk of mortality (OR: 0.78, 95% CI: $0.66 - 0.94$) compared to those residing in wards $\geq$ 5km from the hospital. However, the study had two key limitations. One, the study used wards as the unit of spatial analysis resulting in distance measures that are not representative of the distances between actual residences, villages, to the health facility. In addition, summarizing residence characteristics to ward level, which are relatively large geographic regions, masks local variations in outcome patterns/risk. Two, in their analysis, Kazembe et al. (2006) did not adjust for variations in malaria transmission in the study region despite results indicating increased admission rates during peak malaria transmission periods in the wet season. Furthermore, Kazembe et al. (2006) posits that the geographical disparities in mortality risk observed could be substantially explained by factors that influence the spread of malaria, thus showcasing the need to not only adjust for disease clustering (spatial autocorrelation) but also adjust for spatial variations in underlying risk of outcome.

2.2.2 Cox Proportional Hazard Regression

Cox Proportional Hazard Regression technique estimates the impact of a predictor(s) on the time to occurrence of an event, that is, how a given predictor influences the time for an event to occur (Miller, 2011). In the context of mortality, cox proportional hazard regression models are utilized to assess the effect of predictors on survival times. Specifically, cox regression

models estimate hazard ratio (ratios of event occurrence) across groups of individuals for each predictor variable (Miller, 2011).

Quattrochi et al. (2020) investigated the association between distance to health facilities and paediatric mortality using cox regression model. The study covered an 18-year period and the analysis was done to not only assess the association between physical access and mortality but also to assess if the impact changed over time with the establishment of more facilities. The finding was that there was increased mortality risk with increasing distances, however, mortality risk did not change over time with increased health facilities (Quattrochi et al., 2020). This was attributed to the fact that construction of facilities is not random implying that facilities may be in areas of lower risk thus confounding the effect of distance on mortality (Quattrochi et al., 2020).

Noori et al. (2021) also used cox models to investigate proximity to health facilities as a mortality risk factor. The study established that paediatrics residing in areas 40-60 minutes from inpatient facility had hazard ratio of 1.52 (95% CI: $1.13 - 2.06$) compared to those within 20 minutes of travel time, however, no association was established between proximity to outpatient facilities and mortality (Noori et al., 2021). The latter was attributed to the confounding effect of the strategic location of inpatient facilities in towns characterized by better living standards and thus populations in proximal areas may have lower risk of disease incidence (Noori et al., 2021).

Kadobera et al. (2012) utilized cox proportional hazard regression in their study to investigate the association between distance and child mortality in rural Tanzania in population-based cohort study. Using data from a DSS, the study included children aged < 5 years who were alive at the time of the survey and children who had died that would have been < 5 years at the time of survey. Networked distances to nearest health facilities were used to define physical access. Kadobera et al. (2012) established that children residing in areas $>$ 5km away from health facilities had higher mortality risk (HR:1.17; 95% CI: $1.02 - 1.38$). The study was limited to adjusting for variables only available in the DSS, therefore, no adjustments were made for variables on risk of underlying causes of mortality (Kadobera et al., 2012).

2.3 Gaps Identified

From the review of published literature on access to EC and health outcomes (Section 2.2; Appendix A), two key statistical limitations and one generic limitation stand out. The latter involves the use of varying definitions of physical access (Euclidean distances, cost distances, networked distances, and travel time) across studies to which, as highlighted in a systematic review by Rutherford et al. (2010), conflicting results on the association between physical access to emergency care and health outcomes can be attributed. The use of varying definitions of physical access is down to limitations in data availability. As for the statistical limitations, the following is a description of the gaps identified:

1) **Spatial Variations in underlying drivers**

   As highlighted in Section 2.2, the applications of standard regression techniques to assess the association between physical access to EC services and health outcomes often overlooked spatial variations in the underlying drivers of event risk. Spatial heterogeneity of disease risk acts as a confounder for adverse health outcomes, especially when the risk is higher among children residing in high-risk areas. The inverse care law further complicates this scenario, suggesting that with respect to certain diseases, areas with poor physical access may exhibit increased risk due to a higher disease burden in distant areas (Hart, 1971), a proposition furthered by evidence from Quattrochi et al. (2020) and Noori et al. (2021) that health facilities might be located in areas of lower risk.

   In addition, as demonstrated by Al-Taiar et al. (2008), the effect of the underlying risk drivers is heterogeneous across space. Furthermore, findings from Rees et al. (2016) show that the impact of similar levels of exposure to an underlying driver may vary across villages based on a community characteristics involving improved health practices. Such findings showcase the need for a model that not only adjusts for variations in underlying risk drivers, but also, does so in a manner that incorporates the spatially varying effect.

2) **Failure to account for spatial autocorrelation**

As described in Section 2.2, studies assessing the association between physical access to EC services and health outcomes did not account for spatial autocorrelation of disease events or did so over large geographical scales. Given the clustering patterns that have been reported for various illnesses such as anaemia (Endris et al., 2021; Robert et al., 2023), malaria (Bejon et al., 2014; Kamau et al., 2021) and pneumonia (Kazembe et al., 2007), the implication is that the risk of an individual experiencing an outcome is dependent on their proximity to individuals with the said outcome. Therefore, there is a need to account for spatial autocorrelation in incidence rates of illness in proximal areas at a finer geographical scale.

To address these limitations, the Integrated Nested Laplace Approximation-Stochastic Partial Differential Equation (INLA-SPDE) MBG framework can be utilized. In brief, INLA allows for the implementation of different statistical models including LMs, GLMs, GAMS, GAMMS under the Bayesian framework (Rue et al., 2009) and the SPDE component can be used to define both the spatial dependence of outcomes and the spatial variations in the association between exposure and outcomes in space (Lindgren et al., 2011). INLA-SPDE models have been widely utilized in disease mapping as in the case of malaria (Rue et al., 2009) and HIV (Musenge et al., 2013) and in mapping/delineating of hospital catchments (Alegana et al., 2020, 2012). Despite the foregoing applications of the MBG framework, INLA-SPDE has been underutilized in assessing the association between physical access to EC services and health outcomes. Note that although there exists a Monte-Carlo Markov Chain (MCMC) MBG framework, INLA-SPDE has the edge over MCMC framework for reasons described in detail in Chapter 3.

2.4 Current Research

From the reviewed literature, there exists a gap in the utility of Bayesian MBG models in assessing the association between physical access to EC services and health outcomes. No study, to the best of the author's knowledge, has quantified the association between physical

access to EC services and health outcomes using Bayesian MBG modeling techniques that account for both spatial autocorrelation of outcomes and spatial variations in underlying risk of disease. As such, the present objective is to formulate a spatial model under the INLA-SPDE MBG framework to assess the association between physical access to EC services and severe anaemia in a malaria endemic setting. Achieving this objective might provide novel methods to better understand the role of physical access to EC to guide policy related to targeting marginalized at-risk populations.

Chapter 3

Methodology

3.1 Introduction

This chapter outlines the context of the data and the statistical analysis methods that were utilized. The first section outlines the study area, source of data and how study variables were obtained/defined. The second section outlines the underlying concepts, justifications and development of the statistical analysis methods, concluding with model comparison metrics.

3.2 Study Area

The study area is in Busia County, which is located between latitude 0^0 45 north and longitude 34^0 25 east, covering an area of 1,694.5 square kilometres. Topographically, the majority of the county lies within Lake Victoria basin with an altitude range of 1,130 - 1,500 metres above sea level. Busia County borders three counties; Bungoma to the north, Kakamega to the east and Siaya to the south (Figure 3.1).

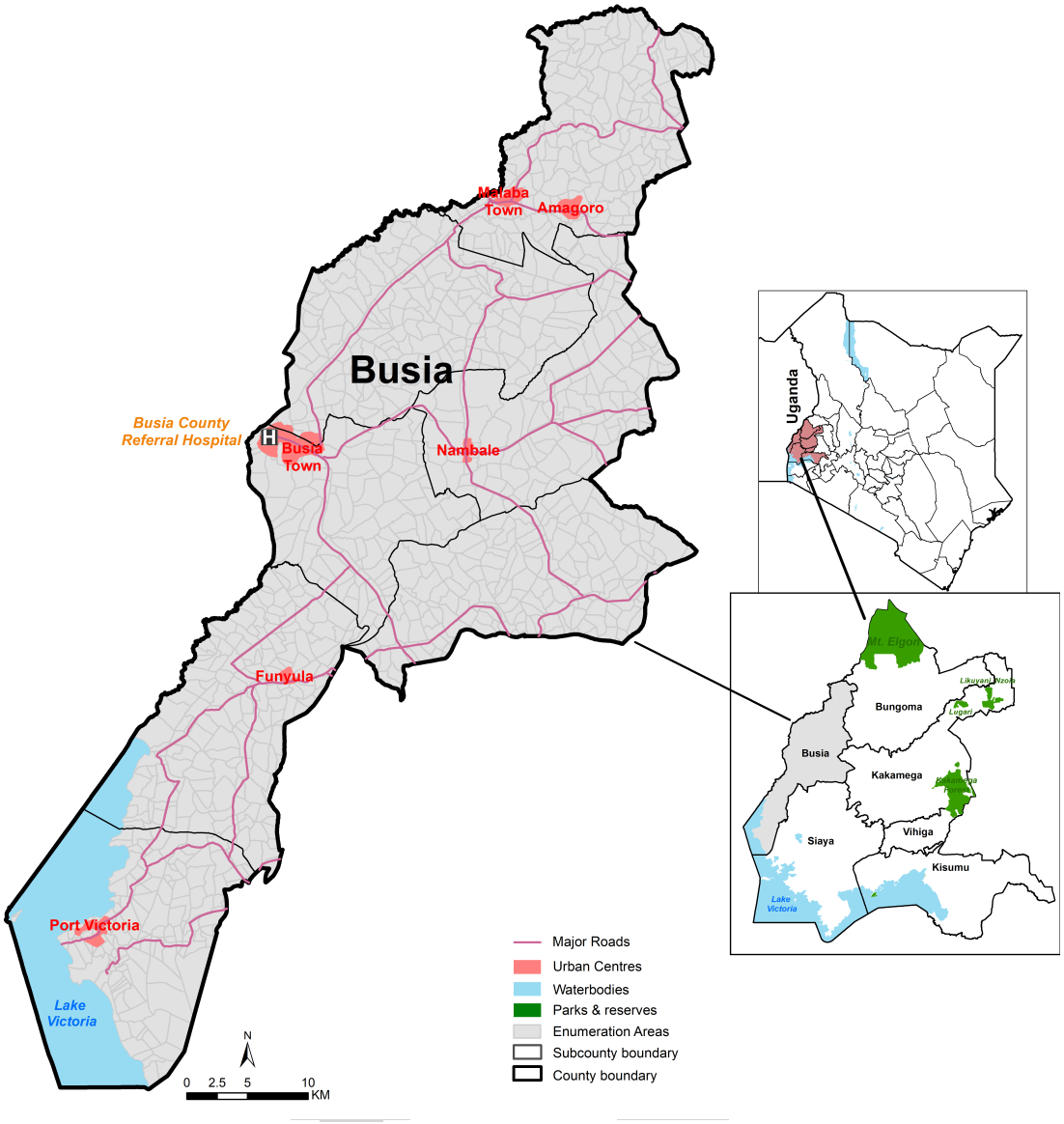

Figure 3.1: Study Area

In addition, Busia County serves as a crucial entry point into Uganda with two border crossings at the towns of Busia and Malaba and borders Uganda to the west. Administratively, Busia county is divided into 7 subcounties, 60 locations, 181 sub-locations and 989 enumeration areas. Most parts of the county are rural with six urban centres: Malaba town, Amagoro, Busia town, Nambale, Funyula and Port Victoria (Figure 3.1) (Macharia et al., 2017).

Busia County experiences bi-annual rainy seasons, with the long rains occurring from March to May and the short rains between August and October (Figure 3.2). Areas in close proximity

to Lake Victoria, notably Budalangi constituency, are prone to flooding during heavy rainfall as a result of the lake overflowing and/or bursting of the feeder rivers

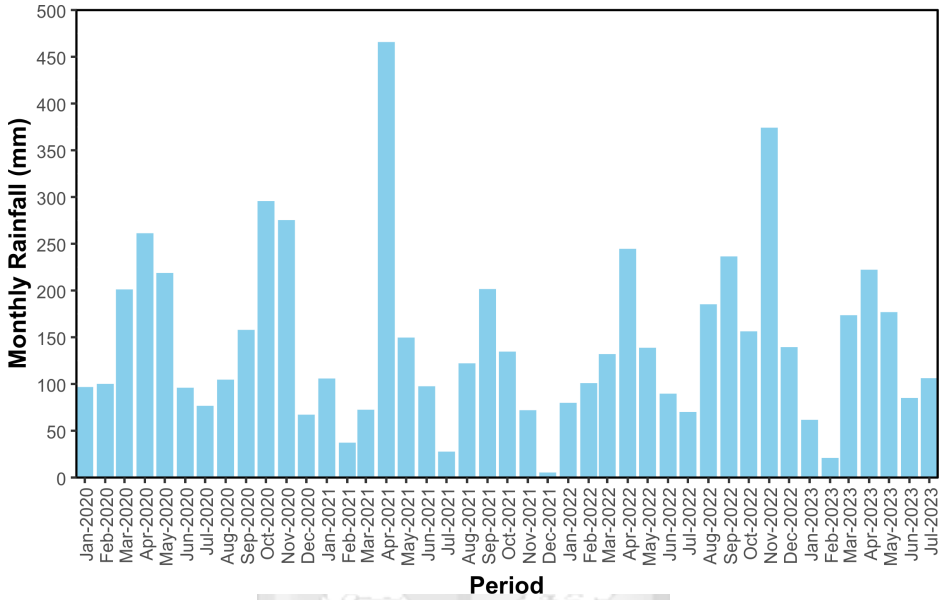

Figure 3.2: Rainfall pattern in Busia County

The population size in Busia County is approximately 893,681 people with Luhya and Teso ethnic groups being the most dominant (KNBS, 2019). In urban centres, the population is concentrated whereas in rural areas the population is sparsely distributed. Crop agriculture and animal farming account for majority of the land use in the county with agriculture, fishing, and trade (mostly in the border towns) being the main economic activities. Busia County Human Development Index (HDI) is 0.43 (lower than the national HDI of 0.52) and literacy level stands at 75.3% among the 15 years and above population (Busia-CIDP, 2018). Road is the main mode of transportation in the county with a road network of 1,600 km comprising of tarmacked surface, gravel surface and earth surface roads; the latter are often impassable during the rainy season. Busia county also has one railway station and two ports at Lake Victoria (Busia-CIDP, 2018).

Busia County is malaria endemic with an estimated *Plasmodium falciparum* prevalence rate (*Pf*PR) of 30% (Alegana et al., 2021). In recent years, the county has been subject to various malaria control interventions including routine distribution of Long-Lasting Insecticide

Treated Nets (LLINs) (Suiyanka et al., 2021), administrations of intermittent preventive treatment of malaria (IPTp) during pregnancy and the recently launched malaria vaccine (Asante et al., ND). In addition, Busia County has a high facility density whereby $>$ 90% of the population are within 1-h travel time to a public health facility (Moturi et al., 2022).

3.3 Data Description

3.3.1 Data Source: Hospital Surveillance

Busia County Referral Hospital is a level IV health facility located at longitude 34.1045 and latitude 0.4604 (Figure 3.1). According to the Kenya Health Sector Strategic Plan (KHSSP) 2018, level IV hospitals serve as referral points to all other lower levels within the given county and/or nearby counties for severe health cases or cases that require more specialized attention in terms of personnel, treatment and/or equipment (MoH, 2018; Ouma et al., 2018). Busia County referral hospital offers a wide range of health services including but not limited to, comprehensive inpatient diagnostic, medical and surgical care, accident and emergency unit, blood transfusion services, specialized outpatient services including habilitative and rehabilitative care and reproductive health services and accelerating referrals to higher level facilities.

Busia County Referral Hospital was enrolled to be part of the Clinical Information Network (CIN); a collaborative initiative between the KEMRI-Wellcome Trust Research Program, The Ministry of Health, The Kenya Pediatric Association, and a set of Kenyan county referral hospitals with the aim of standardizing in-hospital documentation of routine paediatric data for effective passive surveillance and promotion of evidence-based interventions amongst paediatric admissions (Ayieko et al., 2016; Tuti et al., 2016). Surveillance was enhanced to improve coverage of haemoglobin and malaria testing as part of the RTS,S malaria vaccine study in 2019 (Akech et al., 2019; Asante et al., ND). The CIN project supports the implementation of standardized admission and discharge record forms, provides a data clerk and a basic computer for data abstraction from the physical forms onto Research

Electronic Data Capture (REDCap) in the paediatric wards across the participating hospitals. In brief, patient details at admission, during hospitalization and at discharge (or death) are documented by on duty nurses/clinicians attending to the patients using the standardized data collection form, Paediatric Admissions Record (PAR). A trained data clerk, at discharge (or death) of the patient, then abstracts data from the patient file constituting of the PAR and other medical notes on to REDCap. The abstracted data elements include patients' demographics, residence details, anthropometric measurements, medical history, clinical examinations, laboratory tests ordered and corresponding results, prescribed treatments and details on the final discharge diagnosis. The CIN data collection team performs bimonthly data quality checks by re-entering 10 random records and comparing the values in the fields of the re-entered records to the same records as entered by the data clerk using R scripts. A concordance score is calculated and communicated to the clerk on areas of improvement (Amboko et al., 2016; Ayieko et al., 2016; Tuti et al., 2016).

For the current work, secondary data from the CIN database on paediatric admissions in Busia County Referral hospital was extracted for the period between January 2020 and July 2023. The period was defined based on improved data collection on haemoglobin count resulting from the implementation of the RTS,S/AS01 program, where investigation of haemoglobin concentration is done for all paediatric admissions.

3.3.2 Residence Details

To obtain the villages of residence for all admissions, the goal was to match the recorded residence details of each admission to an enumeration area (EA). The Kenya National Bureau of Statistics (KNBS) defines an enumeration area (EA) as a region containing, on average, 100 households, but this number could range from 50 to 149 depending on the area's population density, topography, and/or size. An EA may be a village, a collection of villages, or a section of a village (KNBS, 2019). The 2009 EA shapefile was obtained from KNBS and constituted of EA attributes including name, code and longitude and latitude coordinates. The area of the EAs ranged from 0.01 - 12 square kilometres. Some of the EAs

within the shapefile were designated a suffix i.e. “A", “B", or “C" indicating the subdivisions of a village - such EAs were re-configured and merged.

For each admission, details on patient residence obtained at admission including county, sub-county, location, sublocation, village, nearest health facility, nearest markets and nearest school were used to identify the village of residence. Geocoordinates of the villages were coded to match the updated 2009 EA shapefile. Geocoding was done by the Geographic Information System team under the Population Health Unit of the KEMRI-Welcome Research Programme.

3.3.3 Outcome variable: Anaemia Diagnosis

Haemoglobin (HB) concentrations are affected by elevation above sea level; individuals at altitudes above 1000m tend to have increased levels of HB concentration to compensate the low oxygen levels at such altitudes (Sullivan et al., 2008). The WHO recommends adjusting HB levels for such individuals to avoid underestimating anaemia prevalence (Sullivan et al., 2008). Therefore, the HB concentration for each admission was adjusted based on the following equation as recommended in Sullivan et al. (2008).

$$\textbf{Adjustment Factor} = -0.032 \times (\text{altitude} \times 0.0032808) + 0.022 \times (\text{altitude} \times 0.0032808)^2$$

The above value was subtracted from individual HB measures to obtain altitude-adjusted HB concentration. Using the altitude-adjusted HB concentration, anaemia diagnosis was defined based on WHO guidelines (WHO, 2011) as shown in Table 3.1 below.

Table 3.1: Anaemia Definitions

Anaemia Classes	**Definition used in current work**
Non-Anaemic	HB $\geq$ 11 g/dl
Mild Anaemia	5 $\leq$ HB $<$ 11 g/dl
Severe Anaemia	HB $<$ 5 g/dl

The sequence of anaemia disease progression is moving from non-anaemic to mild anaemia and finally progressing from mild anaemia to severe anaemia which is the life-threatening form of anaemia (Chaparro and Suchdev, 2019). The focus of this study is on severe anaemia, therefore, the comparison is between mild anaemia admissions and severe anaemia admissions to establish whether physical access to EC services plays a role in an individual progressing from mild anaemia status to severe anaemia status. The comparison between severe anaemia admissions and non-anaemic admissions is not appropriate since the risk factors of moving from non-anaemic to mild anaemia and the risk factors for progressing from mild anaemia to severe anaemia are not necessarily the same (Chaparro and Suchdev, 2019). Therefore, such a comparison would result in the compounding of two different levels of risk factors among severe anaemia admissions; risk of moving from non-anaemic to mild anaemia and risk of progressing from mild anaemia to severe anaemia.

Admissions incidence per 1000 persons for each EA by anaemia diagnosis was calculated as;

$$A_{ij} = \frac{\text{Number of Admissions in EA}_{ij} \times 1000}{\text{Under 5 Population in EA}_i}$$

where A_{ij} is the admissions incidence per 1000 persons in EA_i for anaemia diagnosis $j = \{\text{Mild Anaemia}, \text{Severe Anaemia}\}$ and the denominator is the total population per EA aged $1-59$ months for the years $2020-2022$ retrieved from WorldPop (https://www.worldpop.org/); the study period is upto 2023, however, population estimates for 2023 are not yet available. WorldPop models population estimates using random forest estimation technique based on country census data while adjusting for raster datasets associated with population such as land cover types, road networks, water ways, protected areas and locations of facilities resulting in a 100×100 m gridded population density (Stevens et al., 2015). WorldPop avails population estimates by sex (male and female) and age (5-year grouping).

3.3.4 Physical Access to Emergency care: Travel time

Physical access to care in this study was defined as the time taken to travel from the patient's residence, EA, to Busia County Referral hospital. To calculate these travel times, a modeling

approach was employed using AccessMod software (alpha version 5.8.0) (Ray and Ebener, 2008). The modeling process utilized auxiliary datasets detailed as follows.

a) *Roads Network*: Road networks received from Kenya Roads Board (KRB) 2021 were cleaned and reclassified into five classes; international and national trunk roads, primary and secondary roads, minor and government roads, settlement roads and rural access farm and unclassified roads (Table 3.2).

b) *Land Use/Cover*: Given that not all areas are covered by roads or footpaths, satellite-derived information from the European Space Agency (ESA) Sentinel-2 imagery 2021 was used to classify the nature of the underlying geographic space covered by patients on their way to hospital in the absence of roads. The 10m $\times$10m land cover obtained from ESA had eight classes: tree cover, shrubland, grassland, cropland, built-up areas, bare and sparse vegetation, waterbodies and herbaceous wetlands.

c) *Travel Barriers*: Rivers, national parks and game reserves were considered as barriers to travel. Although lakes are present in the region, lakes were not included as barriers since water transport in the form of boats is present. Data on barriers to transport were obtained from world database on protected areas.

d) *Digital Elevation model (DEM)*: Both walking and cycling speeds are significantly impacted by varying slopes of the land. To ensure the variations in the speeds are accounted for in the estimation of travel time, the digital elevation model was obtained from ALOS PALSAR at 12.5m $\times$12.5m spatial resolution.

e) *Travel Scenario*: The travel scenario adopted was walking followed by motorcycle and then a vehicle transport. That is, an individual (1) walks from home to hospital if no motorised road/water is close by (2) walks to nearest bus stop/docking station to take motorised transport to the health facility.

Table 3.2: Land cover classifications, Associated speeds and Modes of Transport

Land Cover Category	**Speed (Km/Hr)**	**Mode of Transport**
International and national trunk roads	65.00	Motorized
Primary and Secondary roads	40.00	Motorized
Minor and government roads	25.00	Motorized
Settlement roads	20.00	Motorized
Rural access farm and unclassified roads	20.00	Motorized
Tree Cover	2.50	Walking
Shrubland	5.00	Walking
Grassland	3.50	Walking
Cropland	3.25	Walking
Built-Up	5.00	Walking
Bare and Sparse Vegetation	5.00	Walking
Waterbodies	8.50	Motorized
Herbaceous Wetland	2.00	Walking

The "Accessibility module" in AccessMod software (version 5.6.0) was implemented to model travel time to Busia County referral hospital (https://www.accessmod.org/). The algorithm in AccessMod models travel time using the terrain-based least cost path distance calculation and is widely applied in the computation of spatial access metrics (Bouanchaud et al., 2022). To model travel time,

a) a friction raster surface was created using the road, land cover and travel barriers datasets described above under the "merge land cover" module in AccessMod.

b) Motorized and walking speeds were assigned to the different road classes and land cover types as shown in Table 3.2 based on previous literature (Joseph et al., 2020; Macharia et al., 2017). The study assumed that a care-seeker would walk above land cover classes then use a motorcycle above the minor road classes (Rural access farm and unclassified roads, Settlement roads and Minor and government roads) after-which

they can take a vehicle on the major road classes (Primary and Secondary roads and International and national trunk roads).

c) DEM was used to account for changing walking speeds due to variations in slope according to Tobler's formulation; an exponential function that describes how human walking speed varies with slope (Ray and Ebener, 2008).

d) Travel time was then be computed, i.e., AccessMod calculated the travel time towards the hospital. The output from AccessMod was a raster surface of the travel time of each pixel to the hospital (Appendix B: Figure B.1).

The travel time for each EA was extracted using the extract function under the terra package in R version 4.2.1. The extract function requires three arguments; the raster data to extract from, the extent of the extraction and the function to extract the values. The travel time raster from AccessMod and the updated 2009 EA shapefile were the respective inputs - travel times were extracted as the average values of each EA (Appendix B: Figure B.1). Each admission was then assigned the travel time corresponding to their EA of residence.

Admissions incidence per 1000 persons by travel time for each anaemia diagnosis was calculated as;

$$T_{ij} = \frac{\text{Number of Admissions within T}_{ij} \times 1000}{\text{Under 5 Population in T}_i}$$

where T_{ij} is the admissions incidence per 1000 persons in travel time band i for anaemia diagnosis $j = \{\text{Mild Anaemia}, \text{Severe Anaemia}\}$ and the denominator is the total population per EA aged $1-59$ months for the years $2020-2022$ retrieved from WorldPop.

3.3.5 *Plasmodium falciparum* Prevalence Rate (*Pf*PR)

Severe anaemia is mainly attributed to malaria in malaria endemic areas (White and Watson, 2018). Therefore, to assess the association between physical access to EC services and severe anaemia in malaria endemic areas, adjusting for malaria transmission rates is vital. *Plasmodium falciparum* Prevalence Rate (*Pf*PR), characterized as the prevalence of blood

stage asexual infection in the community (Drakeley et al., 2017; Smith et al., 2007), is a commonly used and established method of estimating the community's risk of malaria (Drakeley et al., 2017; Snow et al., 2017). In Kenya, *Pf*PR is routinely modeled based on malaria prevalence data using a Bayesian spatial-temporal model adjusting for covariates associated with *Pf*PR including population and environmental factors (precipitation, the Enhanced Vegetation Index, Temperature Suitability Index, Night-time light, and aridity) (Alegana et al., 2021; Macharia et al., 2018; Noor et al., 2014). Specifically, Alegana et al. (2021) assembled data on 1) malaria prevalence from malaria surveys between 2010-2020 and 2) long-term averages of environmental covariates known to be associated with malaria infection prevalence. Using the assembled datasets, Alegana et al. (2021) built a Bayesian space-time hierarchical geostatistical model to create annual smooth surfaces for age standardized *Pf*PR estimates.

Various studies have been undertaken to investigate how the historical changes in *Pf*PR levels relate to the changes in the occurrence of malaria-related health outcomes in different ages and generally, the studies have demonstrated a complex relationship between *Pf*PR levels, age and occurrence of severe health outcomes (Kamau et al., 2022; Okiro et al., 2009; Paton et al., 2021; Snow and Marsh, 2002; Snow et al., 1997). In brief, extended exposure to high *Pf*PR increases the frequency of malaria infection at young ages resulting in severe health outcomes and among those who survive, immunity against severe outcomes is developed. The other side of the spectrum is that extended exposure to low *Pf*PR reduces the frequency of malaria infection at young ages and consequently delayed development of immunity to severe outcomes which occur at older ages (Okiro et al., 2009; Snow and Marsh, 2002).

Despite this phenomenon, Okiro et al. (2009) also reports exceptions to the described pattern of *Pf*PR where, in some locations, disease burden remained greater at younger ages despite significant reductions in *Pf*PR. Furthermore, Snow and Marsh (2002) posit that there exist considerable variations in the occurrence of severe outcomes within the same malaria transmission level, which are attributable to 1) differences in the frequency of infection due to the adoption of malaria prevention interventions such as LLINs that aim to reduce the Entomological Inoculation Rate (EIR) irrespective of the transmission level and 2) intrinsic

differences among households/communities regarding their behaviors and/or attitudes in the context of better health practices which have been described to mediate the efficacy of malaria interventions (Loechl et al., 2023; Rees et al., 2016). Therefore, to appropriately adjust for *Pf*PR when assessing the association between physical access to EC services and severe anaemia, the differences in the impact of the same *Pf*PR level on severe outcomes in different geographical areas should be incorporated.

Given the complex relationship between *Pf*PR, age and severe outcomes, this study used the annual *Pf*PR for each EA in the 4 years before the study, $2017-2020$ (Figure 3.3) obtained from the analysis by Alegana et al. (2021). *Pf*PR was extracted from annual *Pf*PR 1km $\times$1km raster files as the average values for each EA based on the updated 2009 EA shapefile in R version 4.2.1 using extract function under the terra package. For each EA, the annual *Pf*PR was obtained. In addition, given the use of year-specific surveys, the annual *Pf*PR estimates by Alegana et al. (2021) are representative of the malaria risk in the given year, therefore, to represent malaria transmission rates that are in line with the potential acquired immunity that develops with exposure to malaria transmission over time among admissions aged $1-59$ months, and to capture the spatial-temporal changes of *Pf*PR, admissions were assigned to cumulative average *Pf*PR based on their ages such that 4 year olds were assigned their EA average *Pf*PR for the four years preceding the study period, those age 3 years their EA 3 year average *Pf*PR for the 3 years preceding the study and so on.

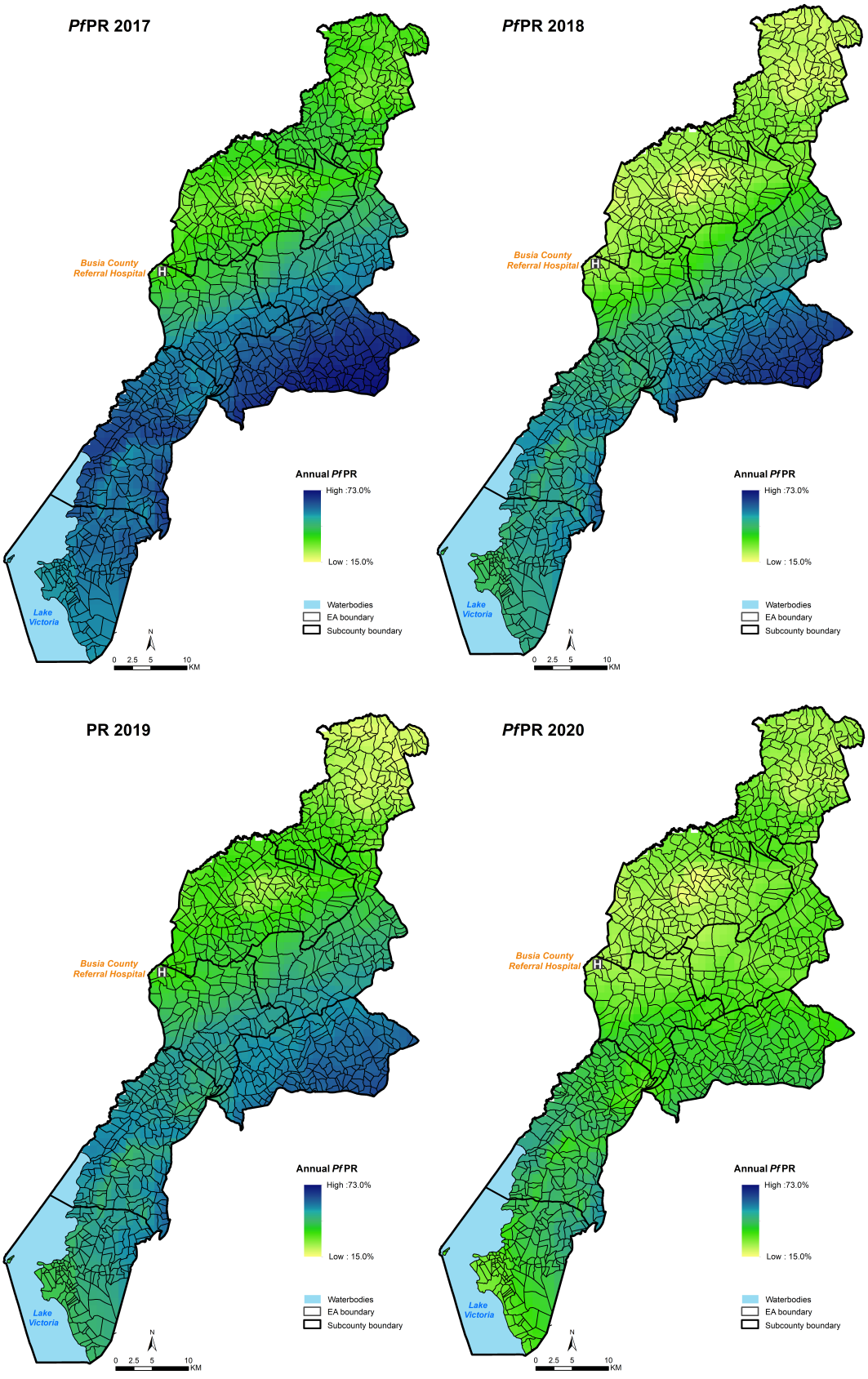

Figure 3.3: Annual Spatial Distribution of *Pf*PR

3.3.6 Possible Confounders

Factors/Variables that are known to be confounders for physical access to care and/or anaemia at both individual and residence levels were obtained/calculated from the data as outlined in Table 3.3.

Table 3.3: Possible Confounders

Variable	Associations	Source	Study Definition
Age	Occurrence of severe outcomes has been described to vary by age and malaria transmission intensities (Kamau et al., 2022; Okiro et al., 2009; Paton et al., 2021; Snow and Marsh, 2002).	Data - Age in years	Categorical: <1 year (1-11 months) 1 year 2 years 3 years 4 years
Malaria Diagnosis	Malaria is associated to causing severe anaemia in malaria endemic settings (White and Watson, 2018)	Data - Primary Diagnosis Defined as having a malaria diagnosis in the absence of underlying conditions	Categorical: Yes, No
Underlying Condition [a]	Presence of an underlying condition that is associated with (or causes) anaemia might increase the risk of severe anaemia (Brittenham et al., 2023; Calis et al., 2008)	Data - Underlying conditions	Categorical: None, SCD, HIV, Other[a]
Nutrition Status [a]	Nutrition status has been shown to be a contributor to severe anaemia (Brittenham et al., 2023)	Calculated: Based on missingness WFA MFA if missing Weight	Categorical: Well nourished (-1 <Z-score) Mildly Malnourished (-3 <Z-score <-2) Moderately Malnourished (-3 <Z-score <-2) Severely Malnourished (Z-score <= -3)
***Pf*Pr**	see subsection 3.3.5	Calculated field: see subsection 3.3.5	Continuous

Gender	Studies have demonstrated gender bias in care seeking patterns (Manongi et al., 2014).	Data - Child sex	Categorical: Male, Female
Good Vaccinees [c]	Vaccination history is a reflection of general health awareness and utilization of health care facilities.	Calculated: If vaccinated for BCG, Penta 1 and Penta 3	Categorical: Yes, No
Referral Admission	Referrals from lower level facilities are likely to be more severe	Data - Referral	Categorical: Yes, No
Admission Day	Admissions that occur n weekdays are more likely to be severe compared to those on weekends; caregivers prioritize income generating taks during weekdays and therefore only seek care in severe cases.	Calculated field: based on date of admission	Categorical: Weekday, Weekend
Season	Ease of transportation differs in rainy and wet seasons.	Calculated field - based on date of admission	Categorical: Wet (March, April, May,August, September, October), Dry (All other months)
Year of Admission	To capture any changes in the risk of severe anaemia in the different contexts of the Covid-19 pandemic	Calculated field - based on date of admission	Categorical: 2020, 2021, 2022, 2023
Urban Index [d]	Rural and urban areas differ in terms of ease of access to EC services including available transportation options.	Calculated field: based Night time light (NTL data has been shown to be a proxy indicator of both economic status (Ebener et al., 2005; Mellander et al., 2015; Noor et al., 2008) and urbanicity (Yang et al., 2019))	Categorical: Urban (NTL >0) Rural (NTL == 0)

[a] "Other" category included various underlying conditions including Burkitts Lymphoma, Cerebral palsy, Congenital abnormality, Diabetes, Epilepsy, Genetic abnormality, Malignancy/neoplasm, Nephritic/Nephrotic Syndrome, Tuberculosis

[b] Nutrition status was determined using z-scores based on WHO guidelines (WHO, 2009) as follows; WFA (Weight for Age) and MFA (MUAC for Age) for admissions missing weight

[c] Good vacciness - the selected vaccines are freely administered and are part of the recommended data quality assessment tools by WHO (WHO, 2015)

[d] The vaerage NTL values from the 2019 satelite image by the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) (https://eogdata.mines.edu/products/dmsp/) was extracted for each village.

The nature of the CIN surveillance is that the data is recorded from patient files post-discharge in the absence of both the patient and caregiver. As such, household/caregiver attributes that may be of importance in the context of physical access to EC services such as caregiver education level, wealth index or caregiver marital status were not available. In addition, other variables such as length of illness that were available in the dataset had significant missingness and were excluded in the analysis.

3.3.7 Inclusion and Exclusion Criteria

This study included children aged $1-59$ months who were residents of Busia county. Given that Busia County Referral Hospital is a level IV health facility, it admits patients from diverse geographical locations, including neighboring Uganda. Inclusion criteria also involves admissions with available data on haemoglobin (HB) results.

Exclusions were applied to remove children under one month due to the challenge of distinguishing between admissions from the newborn unit and those from home and children >59 months. Admissions discharged on the same day while alive were excluded, as they were likely represent cases admitted for observation or awaiting further investigation results. However, those who died on arrival were retained for analysis. Duplicate records were identified and excluded from the study. Additionally, cases related to accidents/trauma, surgery and poisoning were excluded as these conditions may not be directly linked to the place of residence and might serve as underlying causes of admission. Further, admissions without HB result and those defined as non-anaemic were excluded.

3.4 Statistical analysis

The objective was to formulate a Bayesian MBG model using the INLA-SPDE framework to assess the association between physical access to EC services and severe anaemia. In addition, given that previous works utilize the standard logistic regression model to assess the association between physical access to EC services and adverse health outcomes (Appendix A; subsection 2.2.1), the second objective was to compare the coefficient estimates under the standard logistic regression model and those under the developed INLA-SPDE model. As such, two approaches were considered; standard logistic regression and INLA-SPDE. In this section, we describe the underlying concepts of both approaches and in addition, we highlight the properties of the MCMC spatial framework and the reasons for choosing INLA-SPDE. The section concludes with how parameter estimation was done in the study under the logistic regression and INLA-SPDE frameworks.

3.4.1 Logistic Regression model

Logistic Regression models are widely utilized in the study of binary outcomes. Let $E = \{0,1\}$ be the set of outcomes where 0 indicates that the event has not occurred and 1 denotes that the event under study has occurred and let X be a set of covariates. The linear additive model for the probability of event occurrence $p(E=1|X)$ is given as

$$p(E=1|X) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \varepsilon_i \tag{3.1}$$

where β_0 is the intercept, β_i are covariate coefficients and x_i are covariates and ε_i is the error term. From Equation 3.1, it is possible to obtain probabilities outside the $[0,1]$ threshold. Consequently, the logistic regression uses the *logit* function to model odds ratios to ensure the probabilities in additive model are within the expected range (between 0 and 1) as shown below

$$\text{logit}\left(\frac{p(E=1|X)}{1-p(E=1|X)}\right) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \varepsilon_i \tag{3.2}$$

3.4.2 Spatial Models: Bayesian Inference

Bayesian inference is a class of statistical inference that utilizes Bayes' theorem in the estimation of parameters (Lesaffre and Lawson, 2012). Specifically, Bayesian inference techniques combines what is already known about the parameters in question (prior information) and the observed data (likelihood function) to estimate posterior probability distributions of parameters as shown in equation Equation 3.3.

$$p(\theta|X) = \frac{p(X|\theta) \times p(\theta)}{p(X)} = \frac{p(X|\theta) \times p(\theta)}{\int p(X|\theta) \times p(\theta)} \propto p(X|\theta) \times p(\theta) \tag{3.3}$$

where:

- $p(\theta|X)$ is the posterior probability distribution.
- $p(\theta)$ is the prior distribution that defines the knowledge/beliefs about the parameters under estimation.
- $p(X|\theta)$ is the likelihood function of the data that defines the likelihood of the data given the parameters and is given as $p(\theta) = \prod_i f(x_i|\theta)$.
- $\int p(X|\theta) \times p(\theta)$ is the proportionality constant.

From the definition and goal of Bayesian inference, it is evident that the parameters, say θ_i's, being estimated are treated as random variables, hence, the estimation is of posterior distribution of the estimates (Lesaffre and Lawson, 2012). The foregoing is what differentiates Bayesian inference from classical statistics as the latter treats the unknown parameters, θ_i's, as fixed quantities (Lesaffre and Lawson, 2012). .

Like classical statistics, Bayesian inference is applied to various model structures one of which is structured additive regression models (Rue et al., 2009). Structured additive regression models are a class of models in which, say for a response variable y_i, the distribution of y_i is assumed to be from an exponential family of distributions and there exists a function, say $g(.)$ linking μ_i, mean, to an additive predictor, say η_i, that additively accounts for covariate

effects (Rue et al., 2009) as depicted in equation Equation 3.4.

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{k=1}^{m} f_k(z_{kj}) + \varepsilon_i \tag{3.4}$$

where β_0 is the intercept, β_i's are the covariate coefficients, x_i is a vector of covariates, f_k is a function on covariates z and ε_i are unstructured terms. Structured additive regression models have been applied extensively owing to the different forms that $f_k(.)$ can assume including splines, random walks, temporal effects, and spatial random effects e.g., under the Besag-York Mollié model (Rue et al., 2009).

One key step in Bayesian inference is the selection of prior distributions that constitute of parameters defining their form and a distribution of the said parameters. Parameters in a prior distribution are known as hyperparameters and their distribution, hyperprior distribution (Lesaffre and Lawson, 2012).

3.4.3 Latent Gaussian models

Latent Gaussian models are a subclass of structured additive regression models in which a Gaussian prior is assigned to all the latent variables - $\{\beta_0, \beta_i, f_k(.), \varepsilon_i\}$ in equation Equation 3.4 (Rue et al., 2009). Let $\boldsymbol{\phi} = \{\beta_0, \beta_i, f_k(.), \varepsilon_i\}$ be the latent field. If ϕ satisfies the assumption of conditional independence under the Markov properties, ϕ is referred to as a Gaussian Markov Random Field (GMRF) (Rue and Held, 2005). GMRF aids in providing fast inference thanks to its inherent sparse precision matrix on which quicker numerical methods can be applied (Rue and Held, 2005).

Let $p(\phi|\theta; \theta = \{\theta_k\})$ be the gaussian (multivariate normal) prior of ϕ, that is, $p(\phi|\theta) \sim N\left(0, Q(\theta)^{-1}\right)$, where θ is the hyperparameter and $Q(\theta)^{-1}$ is the precision matrix. Given the conditionally independent response variable $y = \{y_i; i \in I\}$ and Bayes theorem, the joint

posterior distribution is given as:

$$\begin{aligned}\boldsymbol{p(\phi,\theta|y)} &\propto p(\theta)\times p(\phi|\theta)\times\prod_{i\in I}p(y_i|\phi,\theta)\\ &\propto p(\theta)\times|Q(\theta)|^{(1/2)}\times exp\left[-\frac{1}{2}\phi^T Q(\theta)\phi+\sum_{i\in I}log\left\{p(y_i|\phi_i,\theta)\right\}\right]\end{aligned}$$

The goal therefore becomes the estimation of posterior marginals $p(\phi_i|y)$ and $p(\theta_k|y)$ expressed as;

$$\boldsymbol{p(\phi_i|y)} = \int p(\phi|\theta,y)\times p(\theta|y)\,d\theta \tag{3.5}$$

$$\boldsymbol{p(\theta_k|y)} = \int p(\theta_i|y)\,d\theta_{\_k} \tag{3.6}$$

3.4.4 Inference with MCMC

Inference on latent Gaussian models can be done using Markov Chain Monte Carlo methods (MCMC). As opposed to evaluating the complex integrals in Equation 3.5 and Equation 3.6, MCMC methods work by sampling from the posterior distributions based on the Markovian property; $\boldsymbol{p\left(\theta^{t+1}|\,\theta^t,\theta^{t-1},\theta^{t-2},\ldots,\theta^1\right)=p(\theta^{t+1}|\,\theta^t)}$ (the future state is dependent only on the current state and not on past states) (Lesaffre and Lawson, 2012). Specifically, MCMC methods obtain samples from the posterior distribution using Markov chains from which the quantity of interest, say mean, is estimated (Lesaffre and Lawson, 2012). The MCMC method is backed by the Strong Law of Large Numbers for convergence and ergodic theorem for dependent samples (Lesaffre and Lawson, 2012). Gibbs sampler and Metropolis Hastings algorithm are the two common sampling techniques under MCMC methods and are discussed in detail elsewhere (Lesaffre and Lawson, 2012).

Despite their popularity, MCMC methods have the following limitations. One, convergence of the algorithm is slow or in some instances, might not occur altogether which is attributed to the large number of simulations (samples) required for accurate estimates and/or complexities in the choice of initial values for the Markov Chains (Brooks et al., 2011; Lesaffre and

Lawson, 2012). Two, to counter dependence in the Markov chains and improve convergence, thinning technique is often applied. However, the technique results in higher Monte Carlo error translating to poor estimates (Lesaffre and Lawson, 2012). Finally, MCMC methods are characterized by high computational cost; for MCMC samplers to compute precise posterior marginals, they require large number of iterations which can take hours or even days (Rue et al., 2009).

3.4.5 Inference with INLA

The limitations of MCMC methods necessitated the development of the Integrated Nested Laplace Approximation (INLA) technique (Rue et al., 2009). INLA is a deterministic technique that analytically evaluates the integrals in Equation 3.5 and Equation 3.6 in the following steps:

Step 1: Approximating $\boldsymbol{p(\theta|y)}$ - Hyperparameter Posterior Marginals

$$p(\theta|y) \propto \frac{p(\theta) \times p(\phi|\theta) \times p(y|\phi,\theta)}{p(\phi|\theta,y)} \tag{3.7}$$

The denominator above is evaluated by gaussian approximation as

$$\begin{aligned} p(\phi|y,\theta) &\propto exp\left[-\frac{1}{2}\phi^T Q(\theta)\phi + \sum_{i\in I} log\left\{p(y_i|\phi_i,\theta)\right\}\right] \\ &= (2\pi)^{-\frac{n}{2}} \left|P(\theta)\right|^{\frac{1}{2}} exp\left(-\frac{1}{2}\left(\phi-\mu(\theta)\right)P(\theta)\left(\phi-\mu(\theta)\right)\right) \\ &= \tilde{p}_G(\phi|\theta,y) \end{aligned} \tag{3.8}$$

where $\mu(\theta)$ is the location of the mode, $P(\theta) = Q(\theta) + diag(c(\theta))$ and $c(\theta)$ is a vector of the negative second derivatives from the log-likelihood evaluated at the mode (Rue et al., 2017).

Equation 3.7 is then estimated as

$$\tilde{p}(\theta|y) \propto \frac{p(\theta) \times p(\phi|\theta) \times p(y|\phi,\theta)}{\tilde{p}_G(\phi|\theta,y)}\Bigg|_{\phi=\phi*(\theta)} \tag{3.9}$$

where $\phi*(\theta)$ is the mode, which is located using quasi-Newton algorithm (Rue et al., 2009).

The negative Hessian Matrix (Covariance matrix for θ) is then computed at $\boldsymbol{\phi*(\theta)}$ for normalization to simplify subsequent numerical integrations. A grid search is performed on $log\{\tilde{p}(\theta|y)\}$ to detect the probability mass and then numerical integration is performed on $\tilde{p}(\theta|y)$ to obtain posterior marginals for each θ_k.

Step 2: Approximating $\boldsymbol{p(\phi_i|\theta,y)}$ - Posterior Marginals for the Latent Field

For selected values of θ_k from step 1 above, $\boldsymbol{p(\phi_i|\theta,y)}$ can be estimated using either gaussian approximation based on $\tilde{p}_G(\phi|\theta,y)$, however, errors may exist due to lack of skewness (Rue et al., 2009). Using laplace approximation overcomes the challenge by evaluating

$$\boldsymbol{\tilde{p}_{LA}(\phi_i|\theta,y)} \propto \frac{p(\phi,\theta,y)}{\tilde{p}_{GG}(\phi_i|\theta,y)}\Bigg|_{\phi_i=\phi_i^*(\phi_i,\theta)} \tag{3.10}$$

where $\tilde{p}_{GG}$ is the Gaussian approximation at the mode, $\phi_i^*(\phi_i,\theta)$. Note that $\tilde{p}_{GG}$ is computed for each ϕ_i and θ making it computationally expensive.

The simplified Laplace approximations, $\tilde{p}_{SLA}(\phi_i|\theta,y)$, that performs the Taylor Series expansion on $\tilde{p}_{LA}$ around $\phi_i = \mu_i(\theta)$ is recommended (Rue et al., 2009). The simplified Laplace approximation overcomes the challenges of the Gaussian approximation and is faster than the laplace approximation.

Step 3: Numerical Integration for $\boldsymbol{\tilde{p}(\phi_i|y)}$

Equation 3.5 is estimated by integrating the combined outputs from steps 2 and 3 above over all integration points θ_k and corresponding area weights Δ_k

$$\begin{aligned}\tilde{p}(\phi_i|y) &\approx \int \tilde{p}(\phi_i|\theta,y) \times \tilde{p}(\theta|y)\, d\theta \\ &\approx \sum_k \tilde{p}(\phi_i|\theta_k,y) \times \tilde{p}(\theta_k|y)\, \Delta_k \end{aligned} \tag{3.11}$$

Mathematical Implementation of INLA-SPDE model

The SPDE approach is implemented in INLA by first representing a continuous Gaussian field as a discretized Gaussian Markov Random Field (GMRF) through a stochastic partial differential equation (SPDE) under the Matèrn covariance model (Lindgren et al., 2011). The Matèrn covariance model is utilized to describe the spatial correlation between locations. For two locations, say s, t, the Matèrn covariance is defined as;

$$Cov(s,t) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\kappa h)^{\nu} K_{\nu}(\kappa h) \tag{3.12}$$

where h is the distance between s and t, ν is the smoothness parameter, κ is the range parameter, σ^2 is the marginal variance, K_{ν} is the modified Bessel function of the second kind, and $\Gamma(\cdot)$ denotes the gamma function.

Given a Gaussian field, $\mathscr{X}(t)$, with the Matèrn covariance, the SPDE is of the form

$$\tau(\kappa^2 - \Delta)^{\alpha/2}(\mathscr{X}(t)) = \mathscr{W}(t) \tag{3.13}$$

where Δ is the Laplacian, τ and κ are scale and range parameters, α controls the smoothness, $\mathscr{X}(t)$ is the Gaussian field, and $\mathscr{W}(t)$ is Gaussian white noise (Lindgren et al., 2011).

The additive regression model, Equation 3.4, can, under latent Gaussian models in INLA-SPDE, be defined in vector form as

$$g(\mu(s_i)) = \eta_i(s_i) = X(s_i)^T \beta + \omega(s_i) + \varepsilon(s_i) \tag{3.14}$$

where $g(\mu(s_i))$ is the link function at location s_i, $X(s_i)^T$ is the vector of covariates at s_i and β is the vector of coefficients for the covariates, $\omega(s_i)$ is the gaussian spatial random effect to capture spatial autocorrelation in the model and is defined under the Matèrn covariance structure. $\varepsilon(s_i)$ is an iid gaussian random effect denoting the non-spatial random variation in the model.

3.5 Parameter Estimation

This section highlights how the above statistical modeling frameworks were implemented in the study.

3.5.1 Covariate Selection and model diagnostics

A multivariable logistic regression, of the form Equation 3.2, was fit. Note that all covariates including travel time were modelled as categorical variables as defined in Table 3.3. *Pf*PR was modeled as a continuous variable. The significance of confounders (Table 3.3) was determined based on AIC from stepwise selection (both directions) or importance from literature. Multicollinearity in the selected confounders was assessed using the Variance inflation factor (VIF). VIF is a statistical tool used to evaluate the existence of correlation between predictor variables in a regression model by quantifying the extent to which Multicollinearity inflates the variance of aregression coefficient (James et al., 2013). For each predictor, the interpretation of VIF values is as follows; VIF $= 1$ implies no correlation exists between the predictor and other predictors, $1 <$ VIF ≥ 5 implies the existence of moderate correlation between the predictor variable and other predictors and VIF > 5 indicates the presence of high correlation (James et al., 2013).

The linearity assumption for *Pf*PR was assessed using the component vs component-plus-residual plot where component refers to the product of *Pf*PR and its coefficient and residual refers to the working residuals in GLMs. Finally, the presence of influential outliers in the dataset was assessed using absolute standard residuals from the model.

3.5.2 Notations

The following are the notations used across the parameter estimation models

- $\boldsymbol{c = \{0,1\}}$ define anaemia status where 0 denotes mild anaemia and 1 denotes severe anaemia
- $\boldsymbol{s_k}$; $k = 1,2,\ldots,m$ denote the residence (village) locations for admissions
- $\boldsymbol{Y_i}$; $i = 1,2,\ldots,n$ be the observations (individuals in the study)
- $\boldsymbol{Y_{ikc}}$ is observation i in location k with anaemia status C
- $\boldsymbol{x_j}$ is a vector of covariates including travel time
- $\boldsymbol{\phi}$ represent malaria transmission and W the coefficient
- $\omega(s_k)$ is the spatial random effect with the Matèrn covariance structure
- $\boldsymbol{\varepsilon_i}$ is the non-spatial random variation

3.5.3 Standard Logistic Regression

To emulate previous works that assessed the association between physical access to EC services and health outcomes, the standard logistic regression model was fitted adjusting for covariates resulting from the covariate selection process. Specifically, a logistic regression model was fit as represented in Equation 3.15.

$$\text{logit}(Y_{ikc}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \varepsilon_i \tag{3.15}$$

The vector x_j is inclusive of all selected covariates excluding the underlying driver of severe anaemia; *Pf*PR.

To illustrate the implications of adjusting for the underlying risk driver under the standard logistic regression, the covariate *Pf*PR was added to Equation 3.15 as shown below.

$$\text{logit}(Y_{ikc}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + W\,\phi + \varepsilon_i \tag{3.16}$$

Both the standard logistic regression models were implemented under the INLA framework with non-informative priors to ensure comparability with the INLA-SPDE geostatistical models.

3.5.4 Evidence for Spatial Autocorrelation

Prior to fitting spatial models under the INLA-SPDE framework to assess association between physical access to EC services and severe anaemia, the presence of spatial autocorrelation was investigated using a variogram. A variogram is a standard tool for assessing presence of spatial autocorrelation (Bohling, 2005). A variogram is a plot of semi variance against distance (Figure 3.4) that demonstrates how different the observations are at a given distance.

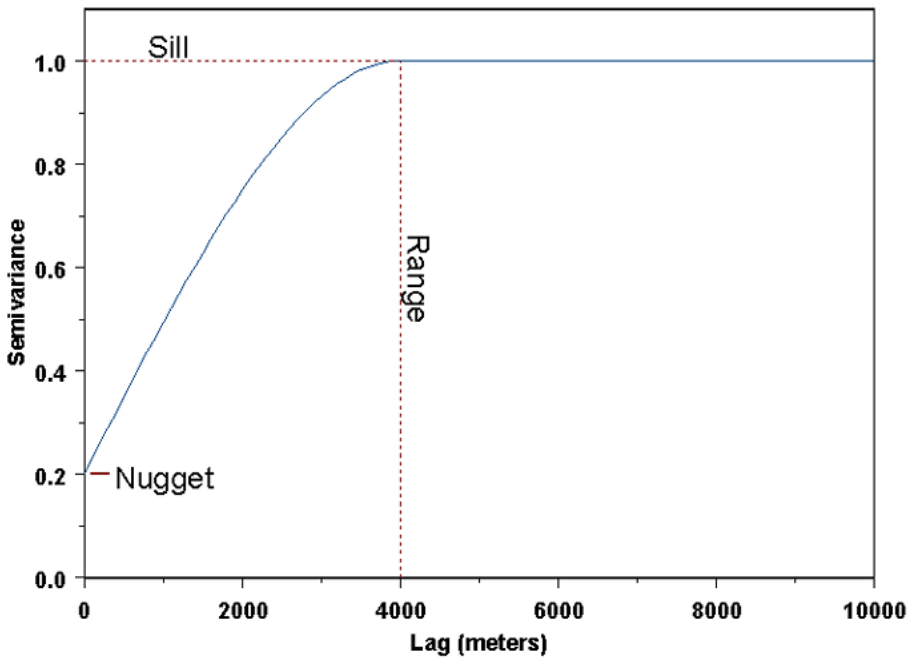

Figure 3.4: Variogram Illustration (Source: Bohling (2005))

A variogram has three key characteristics; a) sill - the semi variance value at which the variogram levels off, b) range - distance at which the semi variance levels off; autocorrelation is said to be equal to zero beyond the range and c) nugget - indicates existence of additional variation beyond the one accounted for by spatial autocorrelation and included covariates (Bohling, 2005). In Figure 3.4, the semivariance increases with increasing distance implying data points within smaller distances are correlated up to a range of $4,000$ metres.

3.5.5 Standard INLA-SPDE model

To investigate the association between physical access to EC services and severe anaemia using INLA-SPDE, the standard INLA-SPDE geostatistical model was fitted by adding the spatial random effect component under Matérn covariance structure, $\omega(s_i)$, to the standard logistic regression model as shown in Equation 3.17 below.

$$\text{logit}(Y_{ikc}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + W\phi + \omega(s_i) + \varepsilon_i \tag{3.17}$$

The inclusion of spatial random effect component is to adjust for the presence of spatially structured unmeasured variation not explained in the set of covariates in the model (Assunção, 2003)

3.5.6 INLA-SPDE Model: Spatially Varying Coefficient

As described in subsection 3.3.5, there exist differences in the occurrence of severe outcomes within the same level of *Pf*PR (Loechl et al., 2023; Rees et al., 2016; Snow and Marsh, 2002). Such differences are not captured in the standard INLA-SPDE model (equation 3.17); the effect of *Pf*PR, W, on severe anaemia is assumed to be the same in all areas such that the difference in the *Pf*PR coefficient in each village, $W(s_k)$, and the mean estimate, $W = E[W(s_k)]$, is incorporated as a random effect to the error terms (Assunção, 2003; Assunção et al., 2002). However, the goal is to utilize the spatial structure to directly model the variations in $W(s_k)$ to account for how the impact of *Pf*PR on severe anaemia varies in

space (Assunção, 2003). As such, the *Pf*PR covariate was modelled in a spatially varying manner as shown in Equation 3.18 below.

$$\text{logit}(Y_{ikc}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \tilde{W}\,\phi(s_k) + \omega(s_i) + \varepsilon_i \tag{3.18}$$

where $\tilde{W} = W + W(s_k)$ is the random spatial adjustment to the coefficient W with respect to location s_k (Gelfand et al., 2003).

3.6 Model Comparison

The purpose of fitting the four models was to observe the modifying effects brought about by adjusting for spatial autocorrelation and adjusting for both spatial autocorrelation and spatially varying effect of underlying driver of health outcomes. Therefore, changes in the coefficients of physical access, measured as travel time, were assessed across the four models to portray how different the findings would have been in previous studies in the absence of the identified limitations.

Chapter 4

Results and Interpretation

4.1 Introduction

This chapter outlines the findings from the statistical analysis and is organized as follows; descriptive statistics, covariate selection and model diagnostics and the association between physical access to EC services and severe anaemia.

4.2 Descriptive Results

A total of $8,082$ admission records were extracted for the period January 2020 – July 2023. Of these records, the age of 6 admissions could not be determined and were excluded, 16 admissions aged less than 28 days were excluded as they could not be distinguished from admissions from the new-born unit and $1,860$ admissions aged > 59 months were excluded as the focus was on children at higher risk of severe anaemia; $1 - 59$ months old. A total of $1,279$ admissions were excluded based on their residence details as either missing (n$=$ 356) or not residing in Busia county (n$=$ 923). Admissions discharged on the same day of admission while alive (n $=$ 194), were excluded as the distinction between admissions admitted for observation and those admitted for continued treatment due to the seriousness of the ailment could not be made. The remaining $4,368$ admission records were investigated further and 7 were found to be duplicate records and 589 records had underlying conditions not related to access to EC and were excluded. Finally, 280 admissions were missing HB results and 702 admissions were non-anemic, both were excluded (Figure 4.1).

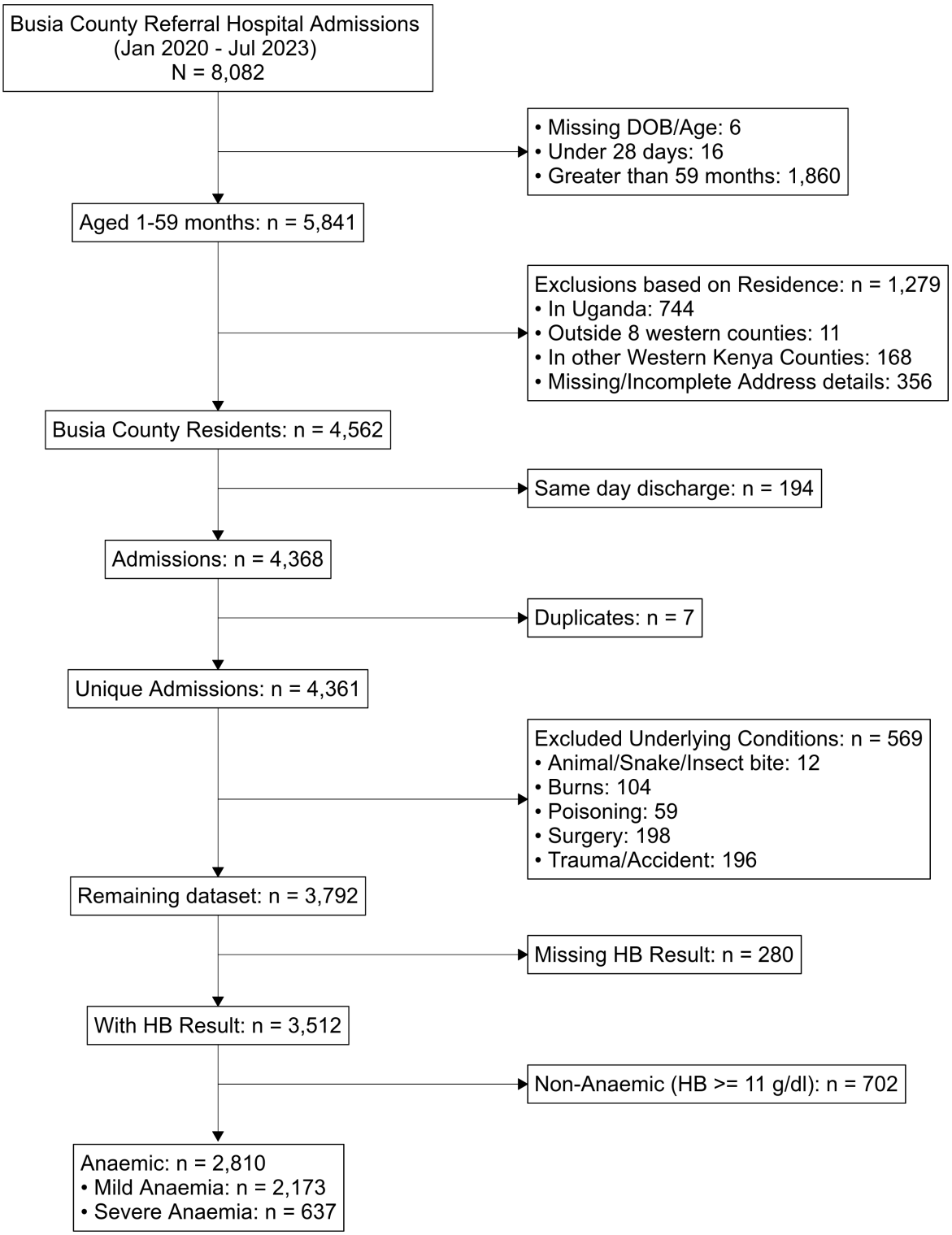

Figure 4.1: Data Inclusion criteria

Of the 2,810 anaemic admissions aged $1-59$ months residing in Busia County, 2,173 (77%) had mild anaemia and 637 (23%) had severe anaemia. Overall, severe anaemia admissions had higher median travel time of 36 (min,max: 2,117) minutes compared to mild anaemia admissions with 17 (min,max: 2,193) minutes (p-value: < 0.001) (Figure 4.2). Similarly, severe anaemia admissions had significantly higher travel times compared to mild anaemia admissions across all the age groups (Figure 4.2) ranging from 34 vs 15 minutes among those aged < 1 year (1-11 months) to 41 vs 19 minutes among those aged 3 years.

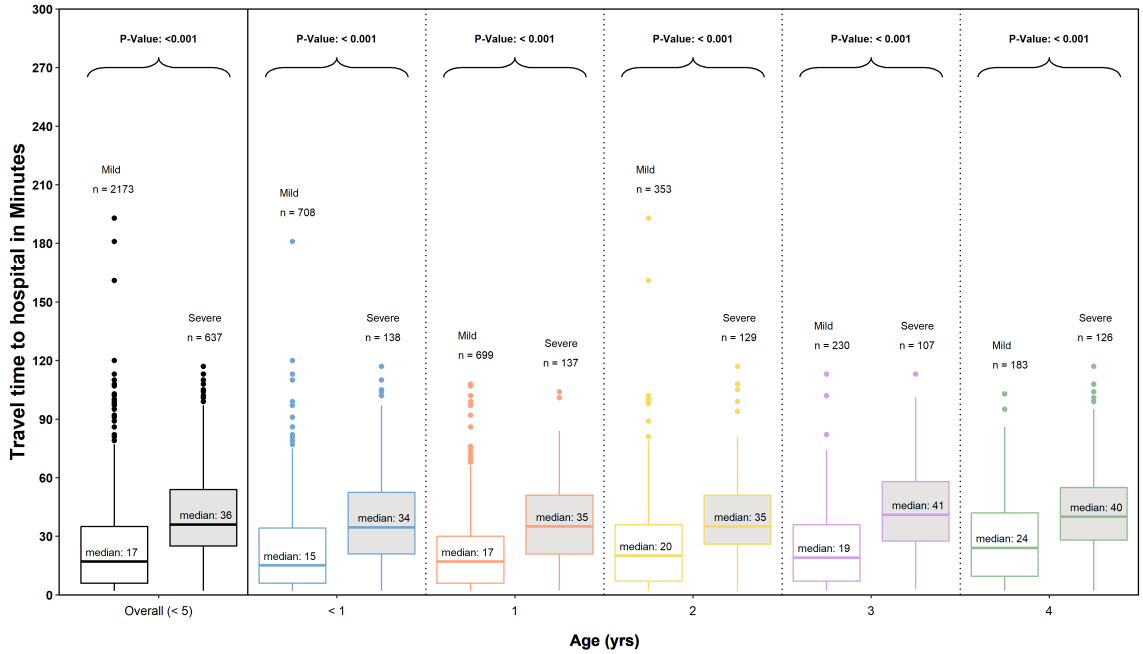

Figure 4.2: Travel Times by Age and Anaemia status. The values are the counts of admissions for the anaemia classes and the p-values are from wilcoxon test for medians

In addition, severe anaemia admissions resided in areas with significantly higher *Pf*PR compared to mild anaemia admissions across all the age groups (Figure 4.3). The overall mean *Pf*PR for mild anaemia and severe anaemia admissions were 30.1% and 34.6% respectively (p-value: < 0.001). The same trend of severe anaemia admissions residing in EAs with higher *Pf*PR is observed across all ages (Figure 4.3).

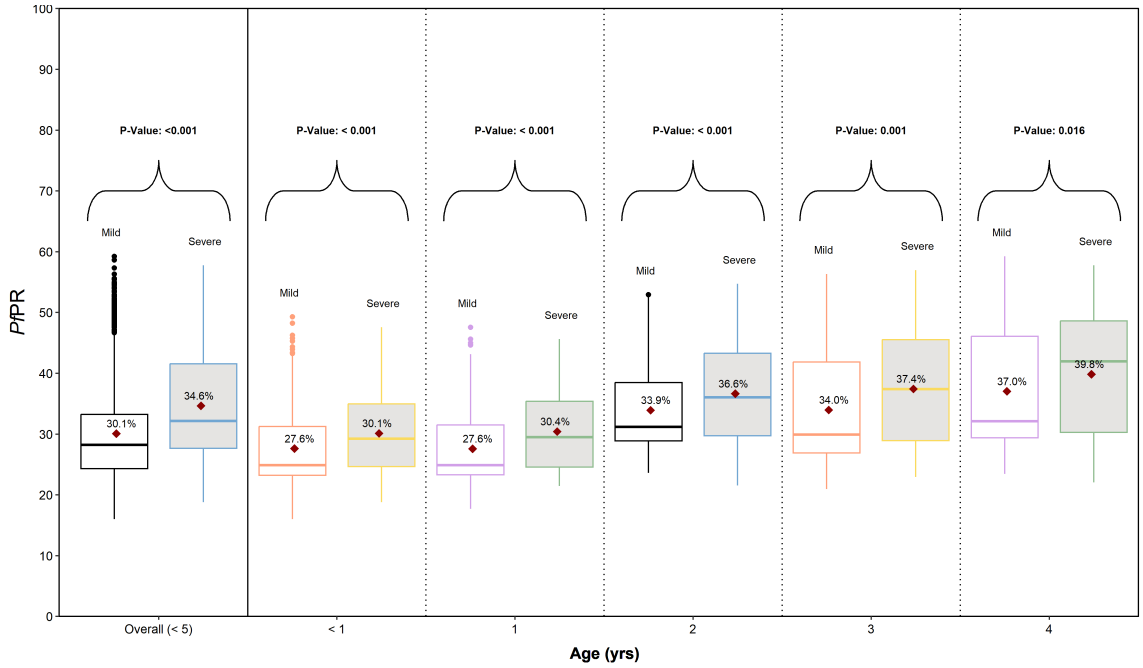

Figure 4.3: *Pf*PR by age and Anaemia status. Red dots represent the position of the mean, the values are the % mean *Pf*PR for the anaemia classes and the p-values are from wilcoxon test for means

A difference in the spatial patterns of mild and severe anaemia admissions incidence was observed (Figure 4.4). The median (range) EA admissions incidence per 1000 persons for mild anaemia was 4.91 $(0.62, 237.56)$ whereas for severe anaemia, the median (range) was 3.68 $(0.21, 38.60)$. As shown in Figure 4.4 (panel A), EAs proximal to Busia County referral Hospital have high incidence (darker shades) of mild anaemia admissions and a decreasing trend in incidence is observed in distant EAs. In contrast, there is no decreasing trend in incidence among severe anaemia admissions for proximal and distant EAs (Figure 4.4 panel B).

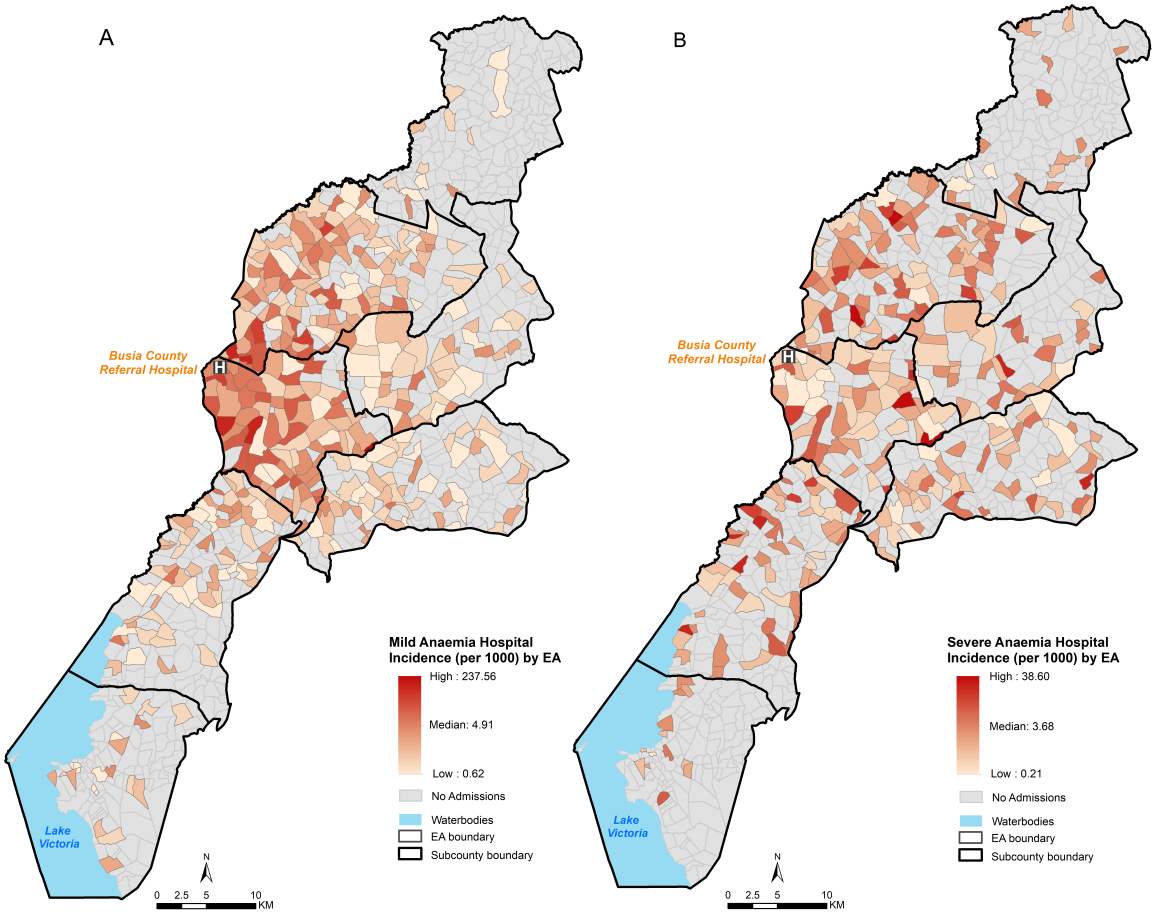

Figure 4.4: Spatial Distribution of Anaemic Admissions

In addition. a distance decay pattern was observed for hospital incidence (per 1000) among mild anaemia admissions but not among severe anaemia admissions (Figure 4.5). Specifically, at 20 minutes, a decline is observed in hospital incidence (per 1000 persons) among mild anaemia admissions from $\approx$ 22 persons per 1000 to $\approx$ 5 persons per 1000 at 35 minutes of travel time, the decreasing trend continues to higher travel times. For severe anaemia admissions, hospital incidence (per 1000 persons) increases from 0 to $\approx$ 35 minutes of travel time and remains relatively constant afterwards (Figure 4.5).

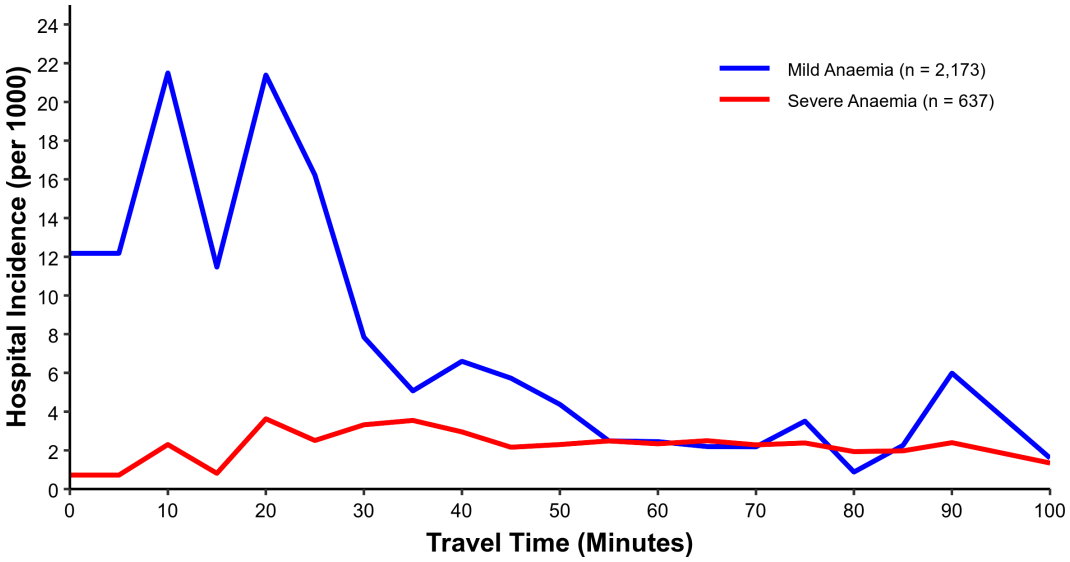

Figure 4.5: Hospital Incidence by Travel Time

Table 4.1 shows the demographic characteristics across the two anaemia status definitions. On travel time classes, 70.0% of mild anaemia admissions were within 30 minutes travel time to Busia County Referral Hospital compared to 34.7% of severe anaemia admissions (p-value: < 0.001). The majority of admissions were aged < 1 year for both mild (32.6%) and severe (21.7%) anaemic admissions. The proportion of admissions with a malaria diagnosis was significantly higher among severe anaemia admissions at 53.5% compared to 33.4% among mild anaemia admissions (p-value: < 0.001). In addition, the proportion of admissions without an underlying condition was significantly lower among severe anaemia admissions, 74.6% compared to 83.6% of mild anaemia admissions (p-value: < 0.001). Severe malnourishment was more common among mild anaemia admissions, 13.1% compared to 8.2% among severe anaemia admissions (p-value: 0.005). There were no differences among the anaemia classes by gender with females constituting the lower proportion at 41.4% and 42.7% for mild and severe admissions respectively. On vaccination history, the proportion of admissions vaccinated for BCG, Penta 1 and Penta 3 was high in both anaemia classes with severe anaemia admissions having a significantly higher proportion at 95.1% compared to 92.9% in mild admissions (p-value: 0.043).

Table 4.1: Demographic Characteristics

Characteristic	Mild (5 <= HB < 11) N = 2173 (77%)[1]	95% CI[2]	Severe (HB < 5) N = 637 (23%)[1]	95% CI[2]	p-value[3]
Travel Time Classes (mins)					<0.001
< 30	1,522 (70.0%)	68%, 72%	221 (34.7%)	31%, 39%	
30 - 59	530 (24.4%)	23%, 26%	303 (47.6%)	44%, 52%	
60 - 89	90 (4.1%)	3.4%, 5.1%	87 (13.7%)	11%, 17%	
>= 90	31 (1.4%)	0.99%, 2.0%	26 (4.1%)	2.7%, 6.0%	
Age Categories (yrs)					<0.001
< 1	708 (32.6%)	31%, 35%	138 (21.7%)	19%, 25%	
1	699 (32.2%)	30%, 34%	137 (21.5%)	18%, 25%	
2	353 (16.2%)	15%, 18%	129 (20.3%)	17%, 24%	
3	230 (10.6%)	9.3%, 12%	107 (16.8%)	14%, 20%	
4	183 (8.4%)	7.3%, 9.7%	126 (19.8%)	17%, 23%	
Malaria Diagnosis					<0.001
No	1,448 (66.6%)	65%, 69%	296 (46.5%)	43%, 50%	
Yes	725 (33.4%)	31%, 35%	341 (53.5%)	50%, 57%	
Underlying Condition					<0.001
None	1,816 (83.6%)	82%, 85%	475 (74.6%)	71%, 78%	
SCD	214 (9.8%)	8.6%, 11%	151 (23.7%)	20%, 27%	
HIV	27 (1.2%)	0.84%, 1.8%	3 (0.5%)	0.12%, 1.5%	
Other	116 (5.3%)	4.4%, 6.4%	8 (1.3%)	0.59%, 2.6%	
Nutrition Status					0.005
Well nourished	1,206 (55.5%)	53%, 58%	356 (55.9%)	52%, 60%	
Mildly Malnourished	458 (21.1%)	19%, 23%	157 (24.6%)	21%, 28%	
Moderately Malnourished	223 (10.3%)	9.0%, 12%	71 (11.1%)	8.9%, 14%	
Severely Malnourished	284 (13.1%)	12%, 15%	52 (8.2%)	6.2%, 11%	
Unknown	2 (0.1%)	0.02%, 0.37%	1 (0.2%)	0.01%, 1.0%	
Gender					0.6
Female	901 (41.5%)	39%, 44%	272 (42.7%)	39%, 47%	
Male	1,272 (58.5%)	56%, 61%	365 (57.3%)	53%, 61%	
Good Vaccinees (BCG, Penta 1 & Penta 3)					0.043
Yes	2,018 (92.9%)	92%, 94%	606 (95.1%)	93%, 97%	
No	155 (7.1%)	6.1%, 8.3%	31 (4.9%)	3.4%, 6.9%	

[1] *n (%)*

[2] *95% Confidence Interval for Proportions from Wilson test*

[3] *Pearson's Chi-squared test; Fischer's exact test*

Table 4.2 shows a summary of admission and residence characteristics. The proportion of referral admissions was greater among severely anaemic admissions at 46.3% compared to 25.9% among mild anaemic admissions (p-value: < 0.001). In addition, severe anaemia

admissions had a higher proportion of admissions during weekdays, 81.9% compared to 76.9% among mild anaemia admissions (p-value: 0.007).

Table 4.2: Admission and Residence Characteristics

Characteristic	**Mild (5 <= HB < 11)** N = 2173 (77%)[1]	95% CI[2]	**Severe (HB < 5)** N = 637 (23%)[1]	95% CI[2]	**p-value[3]**
Referral Admission					<0.001
No	1,609 (74.1%)	72%, 76%	341 (53.7%)	50%, 58%	
Yes	562 (25.9%)	24%, 28%	294 (46.3%)	42%, 50%	
Missing	2		2		
Admission Day					0.007
Weekday	1,671 (76.9%)	75%, 79%	522 (81.9%)	79%, 85%	
Weekend	502 (23.1%)	21%, 25%	115 (18.1%)	15%, 21%	
Season					0.2
Dry	1,037 (47.7%)	46%, 50%	321 (50.4%)	46%, 54%	
Wet	1,136 (52.3%)	50%, 54%	316 (49.6%)	46%, 54%	
Year of Admission					<0.001
2020	447 (20.6%)	19%, 22%	188 (29.5%)	26%, 33%	
2021	513 (23.6%)	22%, 25%	150 (23.5%)	20%, 27%	
2022	770 (35.4%)	33%, 37%	182 (28.6%)	25%, 32%	
2023	443 (20.4%)	19%, 22%	117 (18.4%)	15%, 22%	
***Pf*PR** mean(SD)	30 (7)		35 (9)		<0.001
Urban Index					<0.001
Rural	1,347 (62.0%)	60%, 64%	556 (87.3%)	84%, 90%	
Urban	826 (38.0%)	36%, 40%	81 (12.7%)	10%, 16%	

[1] *n (%)*

[2] *95% Confidence Interval for Proportions from Wilson test*

[3] *Pearson's Chi-squared test; Wilcoxon rank sum test*

Conversely, the proportion of severe anaemia admissions during the wet season was lower at 49.6% compared to mild anaemia admissions at 52.3% (p-value: 0.2). There were variations in the annual admission trends across the study period whereby for mild anaemia admissions, 2022 observed the highest admission numbers at 35.4% whereas for severe anaemia admissions, the highest proportion was in 2020 at 29.5% (p-value: < 0.001). In addition, a lower proportion of mild anaemia admissions resided in rural villages, 62.0%, compared to severe anaemia admissions where 87.3% of admissions were from rural villages (p-value: < 0.001).

4.3 Covariate Selection and Model Diagnostics

4.3.1 Covariate Selection

A total of 7 confounders were found to be statistically significant based on AIC using stepwise selection (both directions); age categories, malaria diagnosis, underlying conditions, referral admissions, admission day (weekday/weekend), year of admission and urban index. Gender, vaccination history (Good vaccinees), nutrition status and season were not statistically significant but were included in all the models based on their importance from literature.

4.3.2 Model Diagnostics

Linearity assumption for *Pf*PR was satisfied as shown in the plot of *Pf*PR component against its component-plus-residual (Figure 4.6).

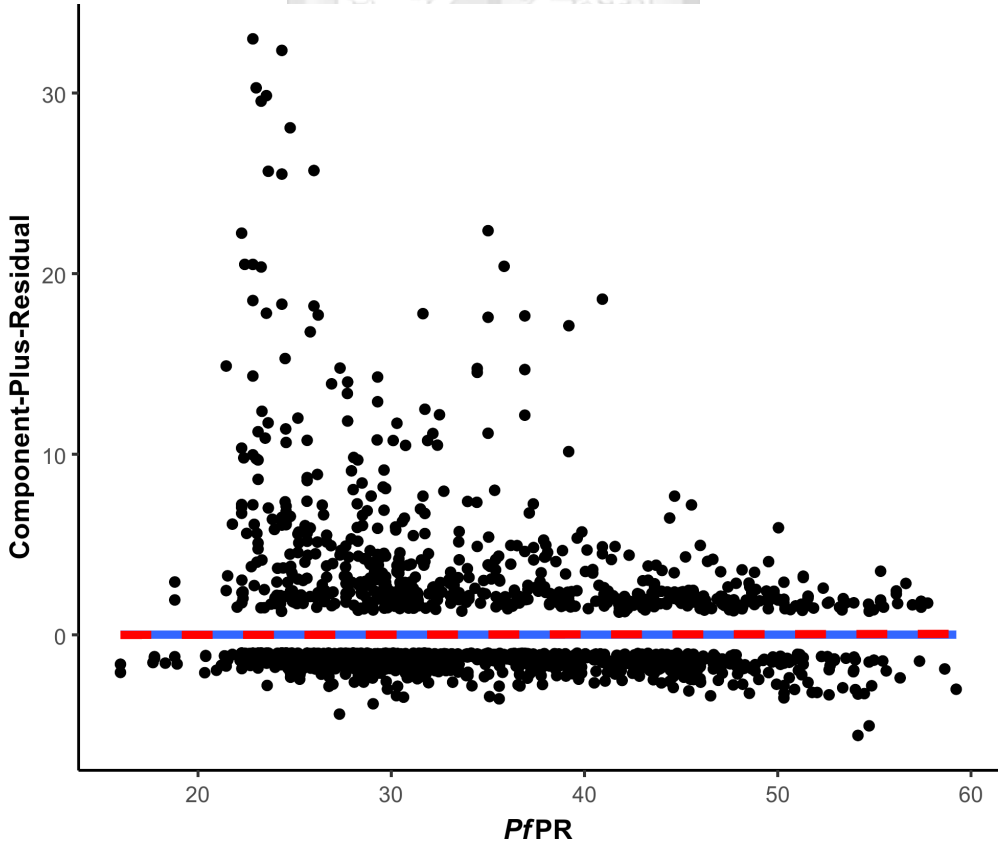

Figure 4.6: Linearity Assumption

The dashed red line represents a linear fit and the blue line is the smoothed conditional mean; linearity assumption is met since the two lines are close to each other.

Using a cut off of 3 absolute standard residuals, no influential outliers were present in the dataset (Figure 4.7), i.e., none of the data points have absolute standard residuals greater than 3.

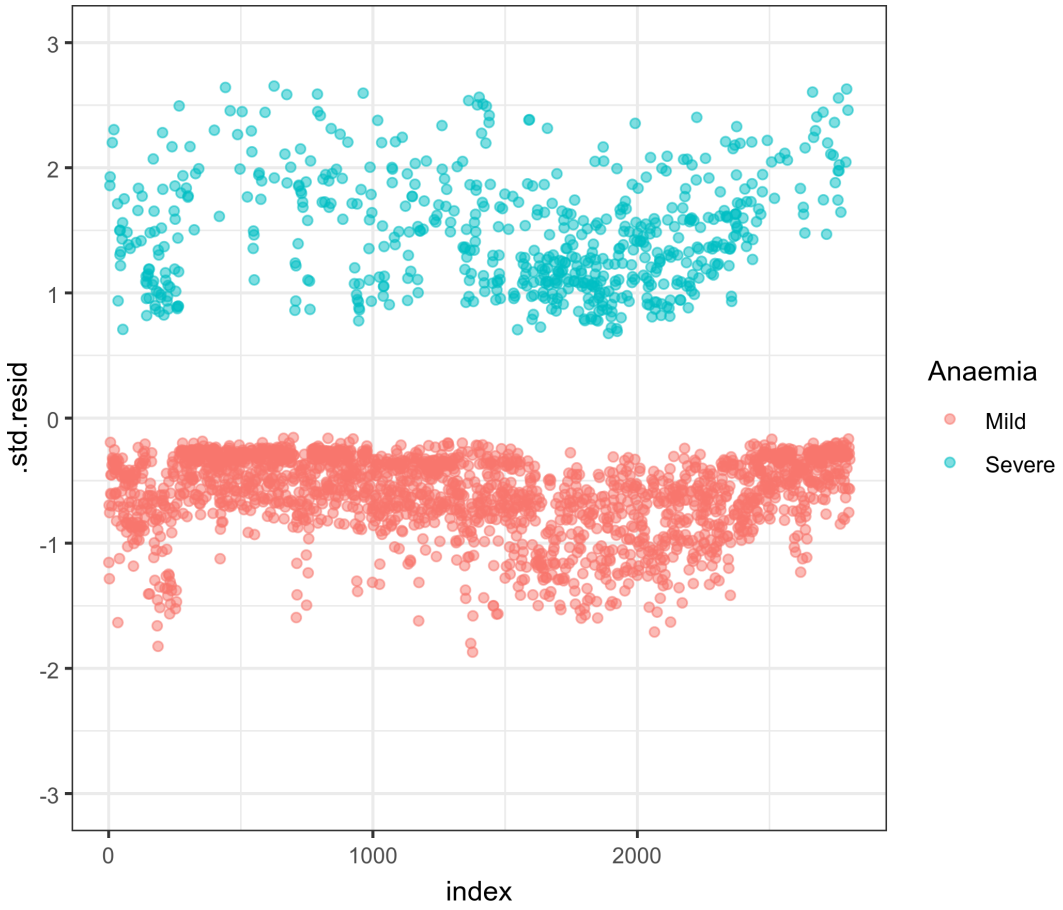

Figure 4.7: Influential Outliers Assumption

As described under subsection 3.5.1, the Variance Inflation Factor (VIF) was used to detct the presence of multicollinearity such that VIF $= 1$ implies no correlation exists between a predictor and other predictors, $1 <$ VIF ≥ 5 implies the existence of moderate correlation between a predictor variable and other predictors and VIF > 5 indicates the presence of high correlation. Table 4.3 shows VIF values for each of the variables. All VIF values across the variables lie between 1 and 5, suggesting the existence of moderate mutlicollinearity (Table 4.3)

Table 4.3: Multicollinearity Assumption

Variable	VIF
Travel Time	1.55
Age Categories	1.55
Malaria Diagnosis	1.55
Underlying Condition	1.62
Nutrition Status	1.13
Gender	1.01
Good Vaccinees (BCG, Penta 1 and Penta 3)	1.13
Referral Admission	1.17
Admission Day	1.01
Season	1.01
Year of Admission	1.06
Urban Index	1.52
***Pf*PR**	1.64

4.3.3 Evidence for Spatial Autocorrelation

The existence of spatial dependence was investigated using a variogram (Figure 4.8).

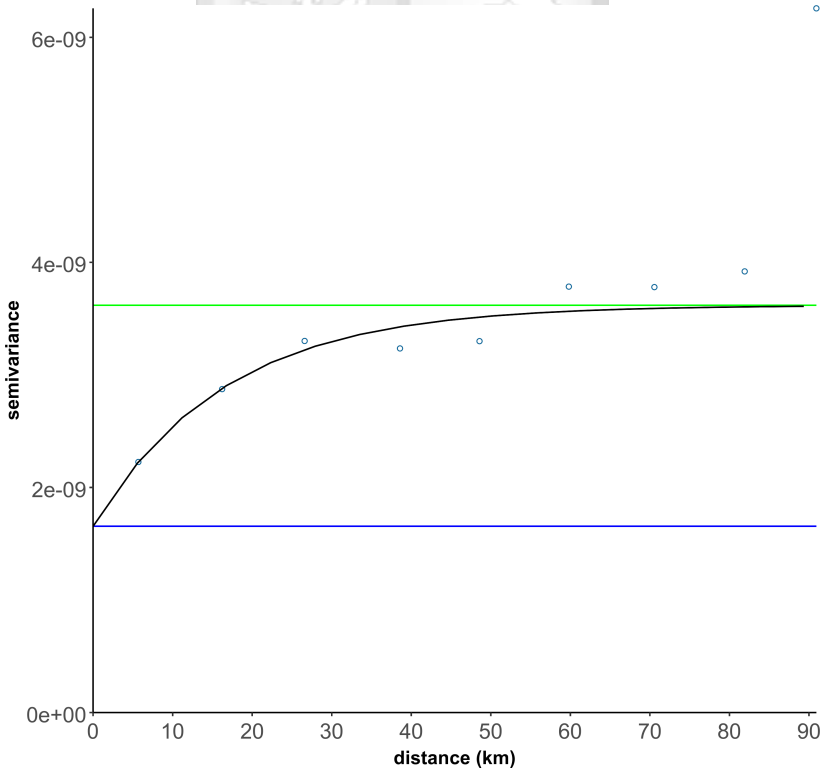

Figure 4.8: Variogram: Evidence for Spatial Dependence.

From Figure 4.8, there exists spatial dependence up to a range of ≈ 60 km, thus, the spatial models are justified. The y-intercept for the blue and green lines represent the nugget and sill values respectively as described in (Chapter 3,subsection 3.5.4)

4.4 The association between physical access to EC services and Severe Anaemia

Each of the four models assessed the association between physical access to EC services, measured as travel time, and the risk of severe anaemia among admissions aged 1-59 months while adjusting for other known risk factors. Table 4.4 shows the resulting Adjusted Odds Ratio (AOR) for severe anaemia over mild anaemia from each of the models.

Table 4.4: Impact of Physical Access to EC on risk of severe anaemia

Travel Time (mins)	**Model 1** AOR (95% CI)	**Model 2** AOR (95% CI)	**Model 3** AOR (95% CI)	**Model 4** AOR (95% CI)
< 30	Ref	Ref	Ref	Ref
$30 - 59$	2.62 (2.05 – 3.34)	2.62 (2.04 – 3.35)	2.68 (1.87 – 3.83)	1.94 (1.18 – 3.08)
$60 - 89$	3.85 (2.64 – 5.61)	3.84 (2.60 – 5.67)	3.76 (2.27 – 6.17)	2.09 (0.89 – 4.54)
≥ 90	3.82 (2.09 – 6.99)	3.82 (2.08 – 7.01)	3.93 (1.85 – 8.32)	2.76 (0.73 – 10.24)

Model 1: Standard Logistic Regression adjusted for all confounders except *Pf*PR.

Model 2: Standard Logistic Regression adjusted for all confounders including *Pf*PR.

Model 3: Standard INLA-SPDE model adjusted for all confounders including *Pf*PR and spatial autocorrelation.

Model 4: INLA-SPDE Model with spatially varying *Pf*PR adjusted for all confounders and spatial autocorrelation.

In the standard logistic regression model not adjusted for *Pf*PR, model 1, the risk of severe anaemia increased with increasing travel time from the hospital (Table 4.4). Admissions residing in EAs within $30-59$ minutes to Busia County referral hospital were 2.62 times more likely to develop severe anaemia in comparison to those residing <30 minutes of travel time (95% CI: $2.05-3.34$). Similarly, admissions within $60-89$ minutes and ≥ 90 minutes of travel time to the hospital had higher odds of severe anaemia, AOR: 3.85 (95% CI: $2.64-5.61$) and AOR: 3.82 (95% CI: $2.09-6.99$) respectively, compared to admissions within ≤ 30 minutes of travel time (Table 4.4).

The association between physical access to EC services and severe anaemia in the standard logistic regression model adjusted for *Pf*PR, model 2, was similar to that of standard logistic regression model 1 that did not adjust for *Pf*PR; the risk of severe anaemia was higher among admissions residing within $30-59$ minutes (AOR: 2.62, 95% CI: $2.04-3.35$), $60-89$ minutes (AOR: 3.84, 95% CI: $2.60-5.67$) and ≥ 90 minutes (AOR: 3.82, 95% CI: $2.08-7.01$) compared to those living in EAs within 30 minutes of travel time to Busia county referral hospital (Table 4.4).

In the standard INLA-SPDE model, model 3, the risk of severe anaemia increased with increasing travel time for admissions residing within $30-59$ minutes (AOR: 2.68, 95% CI: $1.87-3.83$), $60-89$ minutes (AOR: 3.76, 95% CI: $2.27-6.17$) and ≥ 90 minutes (AOR: 3.93, 95% CI: $1.85-8.32$) compared to those living in EAs within 30 minutes of travel time to Busia county referral hospital (Table 4.4). However, the effect size of travel time changes on average by $\approx 2.5\%$. In addition, the confidence intervals in INLA-SPDE model 3 were wider with an average increase of 36%.

Finally, in INLA-SPDE model with spatially varying coefficient for *Pf*PR, model 4, the risk of severe anaemia with physical access was significant only among admissions residing within $30-59$ minutes of travel time (AOR: 1.94, 95% CI:.$1.18-3.08$). Although the signals for admissions residing within $60-89$ minutes and >90 minutes of travel time indicate increased risk of severe anaemia, the effect was not significant; AOR: 2.18 (95% CI: $0.99-4.63$) and AOR: 2.74 (95% CI: $0.75-9.58$) respectively.

Figure 4.9 shows the spatial variation in the risk of severe anaemia attributed to the existence of spatial autocorrelation. Depending on the influence of spatial autocorrelation, spatial variations in the risk of severe anaemia can be in the form of reduced risk (protective effect), increased risk or no effect.

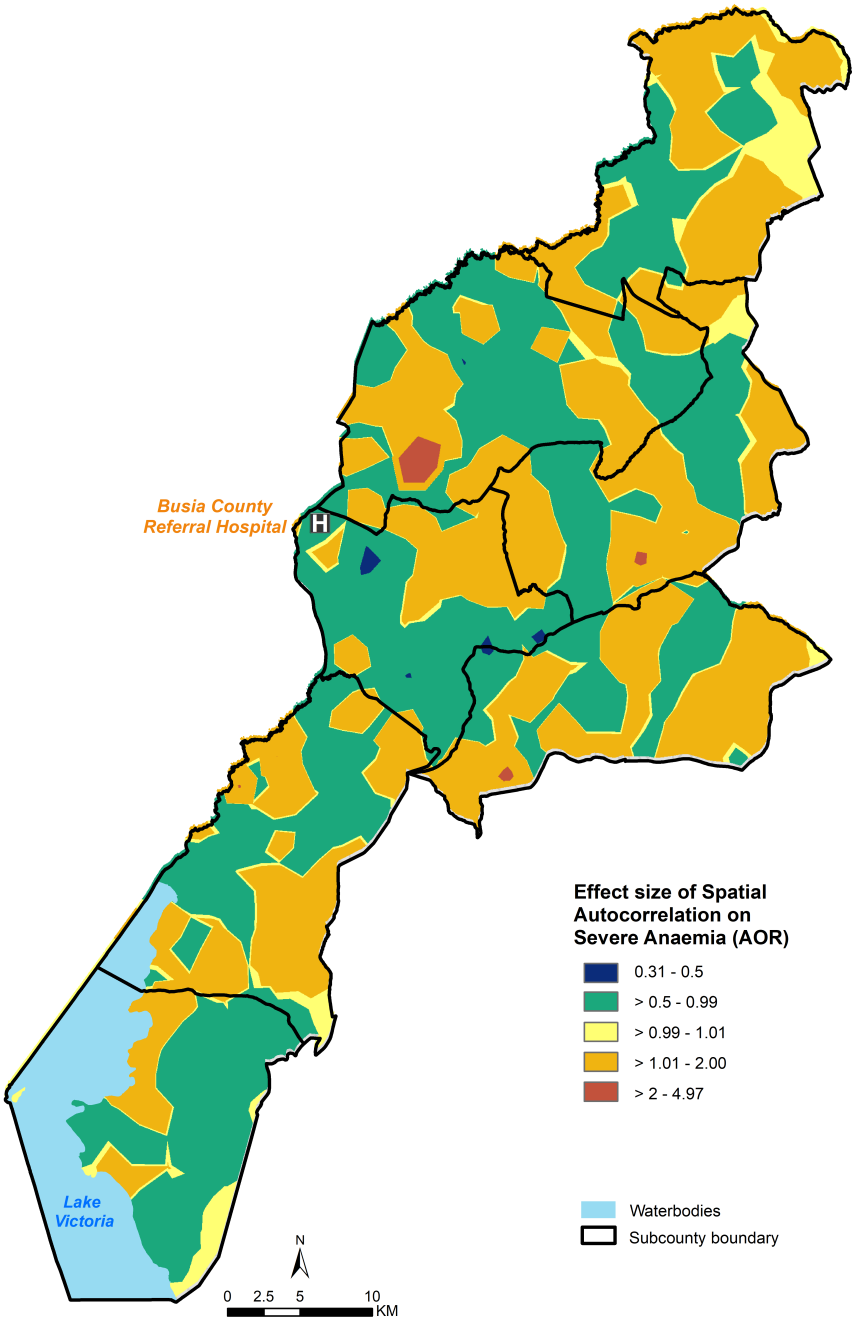

Figure 4.9: Effects size of Spatial Autocorrelation on Risk of severe anaemia

Specifically, in Figure 4.9, the range of the effect size of spatial autocorrelation on the Adjusted Odds Ratio (AOR) for severe anaemia over mild anaemia is $0.31 - 4.87$. The blue and green areas in Figure 4.9 represent areas in which spatial autocorrelation offers a protective effect on severe anaemia with AOR of ≤ 0.99. On the other hand, areas shaded brown and dark brown in Figure 4.9 represent areas in which spatial autocorrelation increases the risk of severe anaemia with AOR ≥ 1.01. Finally, areas shaded yellow represents areas in which spatial autocorrelation has no effect on the risk of severe anaemia.

Figure 4.10 shows the spatial variation in the risk of severe anaemia attributed to spatially varying *Pf*PR. The impact of *Pf*PR on the risk of severe anaemia varied in space with a range of $0.97 - 1.04$ implying that a) *Pf*PR was associated with higher risk of severe anaemia in some areas (areas shaded brown in Figure 4.10), b) *Pf*PR was associated with lower risk of severe anaemia - protective effect (areas shaded green in Figure 4.10) and 3) PfPR was not associated with the risk of severe anaemia (yellow areas in Figure 4.10).

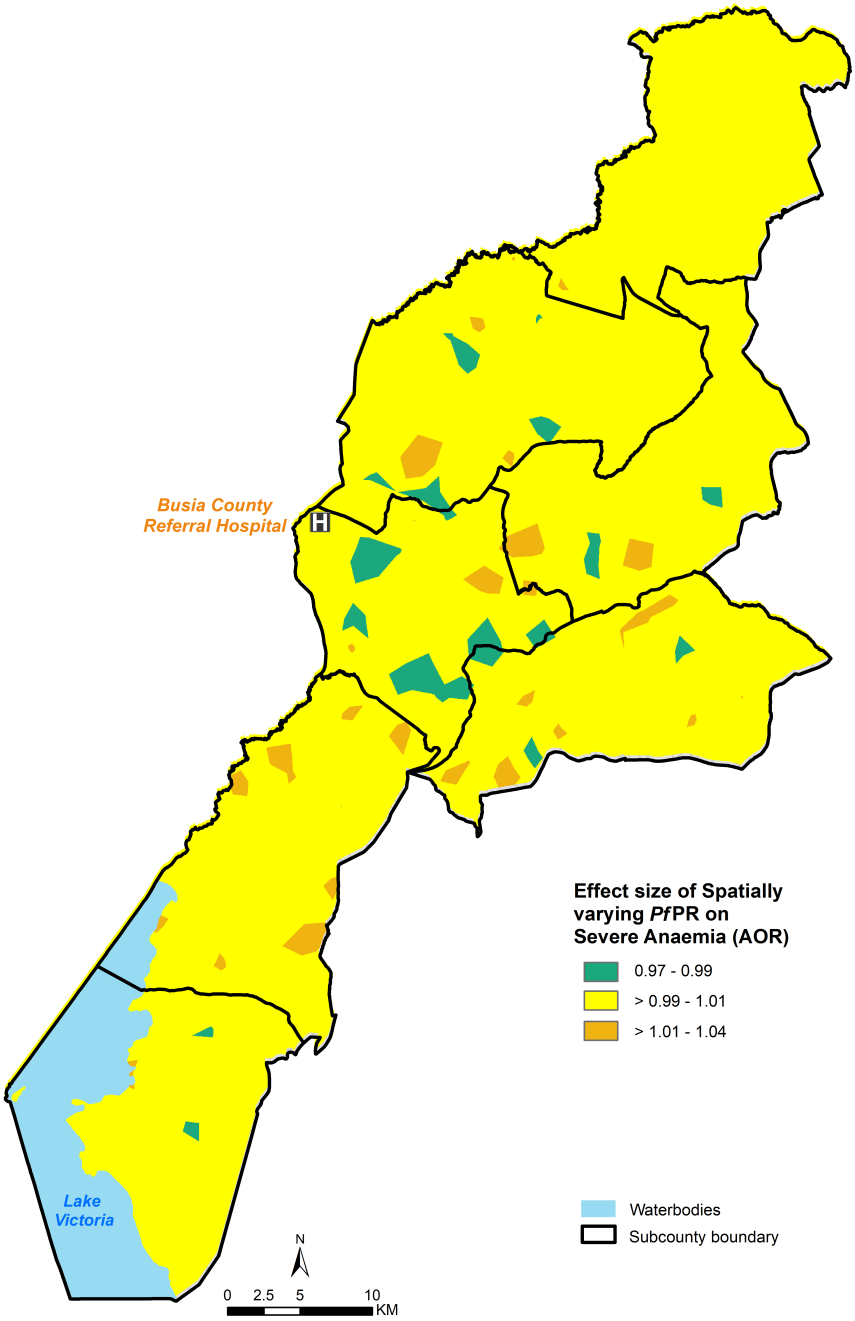

Figure 4.10: Effects size of Spatially varying *Pf*PR on Risk of severe anaemia

Chapter 5

Discussions, Conclusions and Recommendations

5.1 Pre-amble

The objective of this study was to develop a Bayesian MBG model to investigate the association between physical access to EC services and severe anaemia among children aged 1-59 months in Busia County Referral Hospital, Western Kenya using the IINLA-SPDE framework. In addition, the study aimed to assess the modifying effects of spatial autocorrelation and spatially varying effects on the association between physical access to EC services and severe anaemia.

5.2 Discussion

In this study, four model variations were fitted to assess the association between physical access to EC services and severe anaemia; two models under the standard logistic regression framework and two models under the INLA-SPDE framework.

Estimates on the association between physical access to EC services and severe anaemia in the two models under logistic regression indicate increased risk of severe anaemia with increasing travel time from the hospital; a similar finding by (Al-Taiar et al., 2008; Mpimbaza et al., 2017; Mutsigiri-Murewanhema et al., 2017) who also utilized logistic regression on severe malaria. In addition, there were no significant differences in the estimates across the two models under the logistic regression framework implying that including PfPR in the

standard logistic regression model under the stationarity assumption did not represent the true relationship between the underlying driver and the health outcome and thus did not have any modifying effect on the association between physical access to EC services and severe anaemia.

On the other hand, estimates on the association between physical access to EC services and severe anaemia under the INLA-SPDE framework were different from models under the standard logistic regression framework and between themselves. Estimates from the standard INLA-SPDE model indicate an increasing risk of severe anaemia with increasing travel time from the hospital, however, the estimates are different from those under logistic regression models implying a modifying effect of spatial autocorrelation on the association between physical access to EC services and health outcomes as suggested by (Moïsi et al., 2011). Further, the wider confidence intervals under standard INLA-SPDE model are indicative of underestimations of standard errors in the standard logistic regression framework models as they ignore the existence of spatial autocorrelation (Wilson and Lorenz, 2015).

In the INLA-SPDE model with spatially varying coefficients, only admissions within $30-59$ minutes of travel time are at higher risk of severe anaemia compared to those within 30 minutes. These findings indicate the modifying effect of spatially varying underlying drivers of health outcomes and as suggested by (Magnani et al., 1996; Manongi et al., 2014; Zoungrana et al., 2014); the association between physical access to EC services and health outcomes may be over/underestimated if underlying drivers are not accounted for in the models. In addition, similar to findings by (Moïsi et al., 2011), travel time was not associated with increased risk of severe anaemia for admissions residing ≥ 60 minutes from the hospital which is attributable to the high facility density in Busia County as $> 90\%$ of the population in Busia County are within 1-h travel time to a public health facility (Moturi et al., 2022). Finally, the results in the INLA-SPDE model with spatially varying coefficients are indicative of the implication of directly modelling the spatial variations in the impact of the underlying risk driver on health outcomes as opposed to including the spatial variations as additional random effects.

5.2.1 Strengths and Limitations

One strength of this study is the use of villages as the unit of spatial analysis. Village locations represent a fine geographical scale which overcomes the challenge of masking local heterogeneities in diseases risk that is brought about by analyzing data at large spatial units. Secondly, the study defines physical access as village travel time to hospital, which is modeled using WHO's recommended tool. This overcomes the generic gap of varying definitions of physical access and provides a more accurate representation of the difficulty in movement from places of residence to points of care. Thirdly, the outcome used in the analysis, severe anaemia, is distinctly defined; it is not based on syndromic diagnosis by the clinician which might have some bias. Finally, the study utilizes the same dataset and implements the different statistical techniques under the same framework, which allows for the direct comparison of estimates across the statistical techniques.

The current study, however, is not without limitations. First, data on admissions from other competing facilities in the area was not available for analysis. As such, the study could not assess hospital competition to EC services which has been shown to be an important factor to consider in assessing physical access to EC services (Chen et al., 2020; Mumo et al., 2023). In addition, the nature of the surveillance is that data is abstracted at discharge in the absence of both the patient and caregiver which limits the dataset in availing household characteristics such as wealth status which might be significant factors in influencing physical access to EC services.

5.2.2 Implications for Policy

The association between physical access to EC services and severe anaemia remains to be significant even in the context of high facility density. In Kenya, the standard of physical access is defined as 1-hour (MoH, 2013), however, the study has demonstrated increased risk of severe anaemia among admissions within $30-59$ minutes of travel time suggesting that the 1-hour recommendation for physical access to EC services may not be appropriate for severe anaemia. Consequently, policies on improving transportation options and/or incentivizing

transport to healthcare facilities can be implemented to reduce the travel times to EC services. Regarding new and existing interventions such as post-discharge malaria chemoprevention in malaria endemic areas, the implementation ought to consider physical access to dispensation points by the population at risk to improve the projected success of the intervention. Lastly, the observed differences in the effect of *Pf*PR on severe anaemia across the study area imply there exist variations in health practices towards malaria treatment/prevention between communities. As such, periodic health campaigns on improved health practices should be localized to address community-specific challenges/attitudes that may be derailing the progress in the fight against malaria.

5.2.3 Future Directions

The current study has demonstrated the application of a Bayesian MBG model under the INLA-SPDE framework for individual level data. However, health surveillance data is often readily available to the public in form of aggregated data at sub-national levels. Future studies can therefore extend the work by applying the framework to aggregated datasets.

5.3 Conclusion

Spatial autocorrelation and spatial variations in the underlying risk drivers are significant confounders on the association between physical access to EC services and health outcomes. Therefore, geostatistical models should be considered over standard regression techniques in assessing the association between physical access to EC services and health outcomes as they can be modeled to account for these confounders. In addition, when the effect of the underlying drivers on the health outcome varies spatially, the spatially varying relationship should be modelled directly and not under the stationarity assumption.

There exists a significant association between physical access to EC services and severe anaemia, however, the association is attenuated at longer travel times in the presence of high facility density.

Bibliography

Ackerman, H., Ayestaran, A., Olola, C. H. O., Jallow, M., Agbenyega, T., Bojang, K., Roberts, D. J., Krishna, S., Kremsner, P. G., Newton, C. R., Taylor, T., Valim, C., and Casals-Pascual, C. (2020). The association of blood transfusion with outcome among african children hospitalized with plasmodium falciparum malaria: A prospective multicentre observational study hhs public access. *Lancet Haematol*, 7:789–797.

Akech, S., Chepkirui, M., Ogero, M., Agweyu, A., Irimu, G., English, M., and Snow, R. W. (2019). The clinical profile of severe pediatric malaria in an area targeted for routine rts, s/as01 malaria vaccination in western kenya. *Clinical Infectious Diseases*, 71(2):372–380.

Al-Taiar, A., Jaffar, S., Assabri, A., Al-Habori, M., Azazy, A., Al-Gabri, A., Al-Ganadi, M., Attal, B., and Whitty, C. J. (2008). Who develops severe malaria? impact of access to healthcare, socio-economic and environmental factors on children in yemen: A case-control study. *Tropical Medicine and International Health*, 13:762–770.

Alegana, V. A., Khazenzi, C., Akech, S. O., and Snow, R. W. (2020). Estimating hospital catchments from in-patient admission records: a spatial statistical approach applied to malaria. *Scientific Reports*, 10(1).

Alegana, V. A., Macharia, P. M., Muchiri, S., Mumo, E., Oyugi, E., Kamau, A., Chacky, F., Thawer, S., Molteni, F., Rutazanna, D., Maiteki-Sebuguzi, C., Gonahasa, S., Noor, A. M., and Snow, R. W. (2021). Plasmodium falciparum parasite prevalence in east africa: Updating data for malaria stratification. *PLOS Global Public Health*, 1:e0000014.

Alegana, V. A., Wright, J. A., Pentrina, U., Noor, A. M., Snow, R. W., and Atkinson, P. M. (2012). Spatial modelling of healthcare utilisation for treatment of fever in namibia. *International Journal of Health Geographics*, 11(1):6.

Aluisio, A. R., Waheed, S., Cameron, P., Hess, J., Jacob, S. T., Kissoon, N., Levine, A. C., Mian, A., Ramlakhan, S., Sawe, H. R., and Razzak, J. (2019). Clinical emergency care research in low-income and middle-income countries: opportunities and challenges. *BMJ Global Health*, 4(Suppl 6):e001289.

Amboko, B. I., Ayieko, P., Ogero, M., Julius, T., Irimu, G., and English, M. (2016). Malaria investigation and treatment of children admitted to county hospitals in western kenya. *Malaria Journal*, 15.

Asante, K. P., Mathanga, D. P., Milligan, P., Akech, S., Oduro, A. R., Mwapasa, V., Moore, K., Kwambai, T. K., Hamel, M. J., Gyan, T., Westercamp, N., Kapito-Tembo, A., Njuguna, P., Ansong, D., Kariuki, S., Mvalo, T., Snell, P., Schellenberg, D., Welaga, P., Otieno, L., Chimala, A., Afari, E. A., Bejon, P., Maleta, K., Agbenyega, T., Snow, R. W., Zulu, M., Chinkhumba, J., Samuels, A. M., and Malaria Vaccine Programme Evaluation Partners (N.D.). Feasibility, safety, and impact of the rts, s/as01 malaria vaccine when implemented through national immunisation programmes: Evaluation of cluster-randomized introduction of the vaccine in ghana, kenya, and malawi. unpublished.

Assunção, R. M. (2003). Space varying coefficient models for small area data. *Environmetrics*, 14(5):453–473.

Assunção, R. M., Potter, J. E., and Cavenaghi, S. M. (2002). A bayesian space varying parameter model applied to estimating fertility schedules. *Statistics in Medicine*, 21(14):2057–2075.

Ayieko, P., Ogero, M., Makone, B., Julius, T., Mbevi, G., Nyachiro, W., Nyamai, R., Were, F., Githanga, D., Irimu, G., English, M., Tuti, T., Gathara, D., Gachau, S., N'gar, S. N., Aduro, N., Mutai, B., Mutai, L., Emadau, C., Mutiso, C., Nzioki, C., Chabi, M., Kanyingi, F., Mithamo, A., Kuria, M., Otido, S., Kariuki, A., Njiiri, P., Inginia, R., Musabi, M., Kigen, B., Ochieng, G., and Thuranira, L. (2016). Characteristics of admissions and variations in the use of basic investigations, treatments and outcomes in kenyan hospitals within a new clinical information network. *Archives of Disease in Childhood*, 101:223–229.

Bejon, P., Williams, T. N., Nyundo, C., Hay, S. I., Benz, D., Gething, P. W., Otiende, M., Peshu, J., Bashraheil, M., Greenhouse, B., Bousema, T., Bauni, E., Marsh, K., Smith, D. L., and Borrmann, S. (2014). A micro-epidemiological analysis of febrile malaria in coastal kenya showing hotspots within hotspots. *Elife*, 3.

Bohling, G. (2005). Introduction to geostatistics and variogram analysis. *Kansas geological survey*, 1(10):1–20.

Bouanchaud, P., Macharia, P. M., Demise, E. G., and Nakimuli, D. (2022). Comparing modelled with self-reported travel time and the used versus the nearest facility: modelling geographic accessibility to family planning outlets in kenya. *BMJ Global Health*, 7:e008366.

Brittenham, G. M., Moir-Meyer, G., Abuga, K. M., Datta-Mitra, A., Cerami, C., Green, R., Pasricha, S.-R., and Atkinson, S. H. (2023). Biology of anemia: A public health perspective. *The Journal of Nutrition*, 153:S7–S28.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC.

Busia-CIDP (2018). Busia county integrated development plan 2018-2022.

Calis, J. C., Phiri, K. S., Faragher, E. B., Brabin, B. J., Bates, I., Cuevas, L. E., de Haan, R. J., Phiri, A. I., Malange, P., Khoka, M., Hulshof, P. J., van Lieshout, L., Beld, M. G., Teo, Y. Y., Rockett, K. A., Richardson, A., Kwiatkowski, D. P., Molyneux, M. E., and van Hensbroek, M. B. (2008). Severe anemia in malawian children. *New England Journal of Medicine*, 358(9):888–899.

Calvello, E. J. B., Broccoli, M., Risko, N., Theodosis, C., Totten, V. Y., Radeos, M. S., Seidenberg, P., and Wallis, L. (2013). Emergency care and health systems: Consensus-based recommendations and future research priorities. *Academic Emergency Medicine*, 20(12):1278–1288.

Chaparro, C. M. and Suchdev, P. S. (2019). Anemia epidemiology, pathophysiology, and etiology in low-and middle-income countries. *Annals of the New York Academy of Sciences*, 1450(1):15–31.

Chen, B. Y., Cheng, X.-P., Kwan, M.-P., and Schwanen, T. (2020). Evaluating spatial accessibility to healthcare services under travel time uncertainty: A reliability-based floating catchment area approach. *Journal of Transport Geography*, 87:102794.

Clark, M., Spry, E., Daoh, K., Baion, D., and Skordis-Worrall, J. (2012). Reductions in inpatient mortality following interventions to improve emergency hospital care in freetown, sierra leone. *PLoS ONE*, 7(9):e41458.

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(3):299–350.

Drakeley, C., Abdulla, S., Agnandji, S. T., Fernandes, J. F., Kremsner, P., Lell, B., Mewono, L., Bache, B. E., Mihayo, M. G., Juma, O., Tanner, M., Tahita, M. C., Tinto, H., Diallo, S., Lompo, P., D'Alessandro, U., Ogutu, B., Otieno, L., Otieno, S., Otieno, W., Oyieko, J., Asante, K. P., Dery, D. B.-E., Adjei, G., Adeniji, E., Atibilla, D., Owusu-Agyei, S., Greenwood, B., Gesase, S., Lusingu, J., Mahende, C., Mongi, R., Segeja, M., Adjei, S., Agbenyega, T., Agyekum, A., Ansong, D., Bawa, J. T., Boateng, H. O., Dandalo, L., Escamilla, V., Hoffman, I., Maenje, P., Martinson, F., Carter, T., Leboulleux, D., Kaslow, D. C., Usuf, E., Pirçon, J.-Y., and Bahmanyar, E. R. (2017). Longitudinal estimation of plasmodium falciparum prevalence in relation to malaria prevention measures in six sub-saharan african countries. *Malaria Journal*, 16(1).

Ebener, S., Murray, C., Tandon, A., and Elvidge, C. C. (2005). From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics*, 4(1):5.

Endris, B. S., Dinant, G.-J., Gebreyesus, S. H., and Spigt, M. (2021). Geospatial inequality of anaemia among children in ethiopia. *Geospat. Health*, 16.

Fitzgerald, E., Mlotha-Mitole, R., Ciccone, E. J., Tilly, A. E., Montijo, J. M., Lang, H.-J., and Eckerle, M. (2018). A pediatric death audit in a large referral hospital in malawi. *BMC Pediatr.*, 18.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.

Hadjilaou, A., Brandi, J., Mathias, M. A. R., Friese, and Jacobs, T. (2023). Pathogenetic mechanisms and treatment targets in cerebral malaria. *Nat. Rev. Neurol.*

Hansoti, B., Aluisio, A. R., Barry, M. A., Davey, K., Lentz, B. A., Modi, P., Newberry, J. A., Patel, M. H., Smith, T. A., Vinograd, A. M., and Levine, A. C. (2017). Global health and emergency care: Defining clinical research priorities. *Academic Emergency Medicine*, 24(6):742–753.

Hart, J. T. (1971). The inverse care law. *Lancet*, 1:405–412.

Ippolito, M. M., Kabuya, J.-B. B., Hauser, M., Kamavu, L. K., Banda, P. M., Yanek, L. R., Malik, R., Mulenga, M., Bailey, J. A., Chongwe, G., Louis, T. A., Shapiro, T. A., Moss, W. J., Southern, and of Excellence for Malaria Research, C. A. I. C. (2022). Whole blood transfusion for severe malarial anemia in a high plasmodium falciparum transmission setting. *Clin. Infect. Dis.*, 75:1893–1902.

Ippolito, M. M., Kamavu, L. K., Kabuya, J. B., Tente, C., Chileshe, E., Wapachole, M., Thuma, P. E., Muleba, M., Chaponda, M., Mulenga, M., and Moss, W. J. (2018). Risk factors for mortality in children hospitalized with severe malaria in northern zambia: A

retrospective case-control study. *American Journal of Tropical Medicine and Hygiene*, 98:1699–1704.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer New York.

Joseph, N. K., Macharia, P. M., Ouma, P. O., Mumo, J., Jalang'o, R., Wagacha, P. W., Achieng, V. O., Ndung'u, E., Okoth, P., niz, M. M., Guigoz, Y., Panciera, R., Nicolas, E. A. R., and Okiro (2020). Spatial access inequities and childhood immunisation uptake in kenya. *BMC Public Health*, 20:1407.

Kadobera, D., Sartorius, B., Masanja, H., Mathew, A., and Waiswa, P. (2012). The effect of distance to formal health facility on childhood mortality in rural tanzania, 2005-2007. *Global health action*, 5:1–9.

Kahabuka, C., Kvåle, G., and Hinderaker, S. G. (2012). Factors associated with severe disease from malaria, pneumonia and diarrhea among children in rural tanzania - a hospital-based cross-sectional study. *BMC Infectious Diseases*, 12.

Kamau, A., Mtanje, G., Mataza, C., Bejon, P., and Snow, R. W. (2021). Spatial-temporal clustering of malaria using routinely collected health facility data on the kenyan coast. *Malar. J.*, 20:227.

Kamau, A., Paton, R. S., Akech, S., Mpimbaza, A., Khazenzi, C., Ogero, M., Mumo, E., Alegana, V. A., Agweyu, A., Mturi, N., Mohammed, S., Bigogo, G., Audi, A., Kapisi, J., Sserwanga, A., Namuganga, J. F., Kariuki, S., Otieno, N. A., Nyawanda, B. O., Olotu, A., Salim, N., Athuman, T., Abdulla, S., Mohamed, A. F., Mtove, G., Reyburn, H., Gupta, S., Lourenço, J., Bejon, P., and Snow, R. W. (2022). Malaria hospitalisation in east africa: age, phenotype and transmission intensity. *BMC Medicine*, 20(1).

Karra, M., Fink, G., and Canning, D. (2017). Facility distance and child mortality: A multi-country study of health facility access, service utilization, and child health outcomes. *International Journal of Epidemiology*, 46:817–826.

Kashima, S., Suzuki, E., Okayasu, T., Louis, R., Eboshida, A., and Subramanian, S. V. (2012). Association between proximity to a health center and early childhood mortality in madagascar. *PLoS ONE*, 7.

Kassebaum, N. J., Jasrasaria, R., Naghavi, M., Wulf, S. K., Johns, N., Lozano, R., Regan, M., Weatherall, D., Chou, D. P., Eisele, T. P., Flaxman, S. R., Pullan, R. L., Brooker, S. J., and Murray, C. J. L. (2014). A systematic analysis of global anemia burden from 1990 to 2010. *Blood*, 123(5):615–624.

Kazembe, L. N., Kleinschmidt, I., and Sharp, B. L. (2006). Patterns of malaria-related hospital admissions and mortality among malawian children: An example of spatial modelling of hospital register data. *Malaria Journal*, 5.

Kazembe, L. N., Muula, A. S., Appleton, C. C., and Kleinschmidt, I. (2007). Modelling the effect of malaria endemicity on spatial variations in childhood fever, diarrhoea and pneumonia in malawi. *Int. J. Health Geogr.*, 6:33.

KNBS (2019). 2019 kenya population and housing census: Volume ii.

Kwambai, T. K., Dhabangi, A., Idro, R., Opoka, R., Kariuki, S., Samuels, A. M., Desai, M., Hensbroek, M. B. V., John, C. C., Robberstad, B., Wang, D., Phiri, K., and Kuile, F. O. T. (2018). Malaria chemoprevention with monthly dihydroartemisinin-piperaquine for the post-discharge management of severe anaemia in children aged less than 5 years in uganda and kenya: Study protocol for a multi-centre, two-arm, randomised, placebo-controlled, superiority trial. *Trials*, 19.

Kwambai, T. K., Mori, A. T., Nevitt, S., van Eijk, A. M., Samuels, A. M., Robberstad, B., Phiri, K. S., and ter Kuile, F. O. (2022). Post-discharge morbidity and mortality in children admitted with severe anaemia and other health conditions in malaria-endemic settings in africa: a systematic review and meta-analysis. *The Lancet Child and Adolescent Health*, 6:474–483.

Lawson, A. B. (2013). *Statistical methods in spatial epidemiology*. John Wiley & Sons.

Lesaffre, E. and Lawson, A. A. B. (2012). *Bayesian biostatistics*. John Wiley I& Sons, Ltd.

Levesque, J.-F., Harris, M. F., and Russell, G. (2013). Patient-centred access to health care: conceptualising access at the interface of health systems and populations. *Int. J. Equity Health*, 12:18.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach.

Loechl, C. U., Datta-Mitra, A., Fenlason, L., Green, R., Hackl, L., Itzkowitz, L., Koso-Thomas, M., Moorthy, D., Owino, V. O., Pachón, H., Stoffel, N., Zimmerman, M. B., and Raiten, D. J. (2023). Approaches to address the anemia challenge. *The Journal of Nutrition*, 153:S42–S59.

Lohela, T. J., Campbell, O. M., and Gabrysch, S. (2012). Distance to care, facility delivery and early neonatal mortality in malawi and zambia. *PLoS ONE*, 7.

Macharia, P. M., Giorgi, E., Noor, A. M., Waqo, E., Kiptui, R., Okiro, E. A., and Snow, R. W. (2018). Spatio-temporal analysis of plasmodium falciparum prevalence to understand the past and chart the future of malaria control in kenya. *Malaria Journal*, 17(1).

Macharia, P. M., Odera, P. A., Snow, R. W., and Noor, A. M. (2017). Spatial models for the rational allocation of routinely distributed bed nets to public health facilities in western kenya. *Malar. J.*, 16.

Magnani, R. J., Rice, J. C., Mock, N. B., Abdoh, A. A., Mercer, D. M., and Tankari, K. (1996). The impact of primary health care services on under-five mortality in rural niger. *International Journal of Epidemiology*, 25(3):568–577.

Manongi, R., Mtei, F., Mtove, G., Nadjm, B., Muro, F., Alegana, V., Noor, A. M., Todd, J., and Reyburn, H. (2014). Inpatient child mortality by travel time to hospital in a rural area of tanzania. *Trop. Med. Int. Health*, 19:555–562.

Mellander, C., Lobo, J., Stolarick, K., and Matheson, Z. (2015). Night-time light data: A good proxy measure for economic activity? *PLOS ONE*, 10(10):e0139779.

Miller, Jr, R. G. (2011). *Survival Analysis*. Wiley Classics Library. Wiley-Interscience, Newy York, 2 edition.

MoH (2013). Kenya health sector strategic and investment plan (khssp) july 2013-june 2017.

MoH (2018). Kenya health sector strategic plan.

Moïsi, J. C., Nokes, D. J., Gatakaa, H., Williams, T. N., Bauni, E., Levine, O. S., and Scott, J. A. G. (2011). Sensitivity of hospital-based surveillance for severe disease: a geographic information system analysis of access to care in kilifi district, kenya. *Bull. World Health Organ.*, 89(2):102–111.

Molyneux, E., Ahmad, S., and Robertson, A. (2006). Improved triage and emergency care for children reduces inpatient mortality in a resource-constrained setting. *Bull World Health Organ*, 84:314–319.

Moturi, A. K., Suiyanka, L., Mumo, E., Snow, R. W., Okiro, E. A., and Macharia, P. M. (2022). Geographic accessibility to public and private health facilities in kenya in 2021: An updated geocoded inventory and spatial analysis. *Front. Public Health*, 10:1002975.

Mousa, A., Al-Taiar, A., Anstey, N. M., Badaut, C., Barber, B. E., Bassat, Q., Challenger, J. D., Cunnington, A. J., Datta, D., Drakeley, C., Ghani, A. C., Gordeuk, V. R., Grigg, M. J., Hugo, P., John, C. C., Mayor, A., Migot-Nabias, F., Opoka, R. O., Pasvol, G., Rees, C., Reyburn, H., Riley, E. M., Shah, B. N., Sitoe, A., Sutherland, C. J., Thuma, P. E., Unger, S. A., Viwami, F., Walther, M., Whitty, C. J., William, T., and Okell, L. C. (2020). The impact of delayed treatment of uncomplicated p. falciparum malaria on progression to severe malaria: A systematic review and a pooled multicentre individual-patient meta-analysis. *PLoS Medicine*, 17.

Mpimbaza, A., Ndeezi, G., Katahoire, A., Rosenthal, P. J., and Karamagi, C. (2017). Demographic, socioeconomic, and geographic factors leading to severe malaria and delayed care seeking in ugandan children: A case-control study. *American Journal of Tropical Medicine and Hygiene*, 97:1513–1523.

Målqvist, M., Sohel, N., Do, T. T., Eriksson, L., and Persson, L.-o. (2010). Distance decay in delivery care utilisation associated with neonatal mortality. a case referent study in northern vietnam. *BMC Public Health*, 10(1).

Mumo, E., Agutu, N. O., Moturi, A. K., Cherono, A., Muchiri, S. K., Snow, R. W., and Alegana, V. A. (2023). Geographic accessibility and hospital competition for emergency blood transfusion services in bungoma, western kenya. *International Journal of Health Geographics*, 22(1).

Musenge, E., Chirwa, T. F., Kahn, K., and Vounatsou, P. (2013). Bayesian analysis of zero inflated spatiotemporal hiv/tb child mortality data through the inla and spde approaches: Applied to data observed between 1992 and 2010 in rural north east south africa. *International Journal of Applied Earth Observation and Geoinformation*, 22:86–98.

Mutsigiri-Murewanhema, F., Mafaune, P. T., Shambira, G., Juru, T., Bangure, D., Mungati, M., Gombe, N. T., and Tshimanga, M. (2017). Factors associated with severe malaria among children below ten years in mutasa and nyanga districts, zimbabwe, 2014-2015. *Pan African Medical Journal*, 27.

Noor, A. M., Alegana, V. A., Gething, P. W., Tatem, A. J., and Snow, R. W. (2008). Using remotely sensed night-time light as a proxy for poverty in africa. *Population Health Metrics*, 6(1).

Noor, A. M., Kinyoki, D. K., Mundia, C. W., Kabaria, C. W., Mutua, J. W., Alegana, V. A., Fall, I. S., and Snow, R. W. (2014). The changing risk of plasmodium falciparum malaria infection in africa: 2000–10: a spatial and temporal analysis of transmission intensity. *The Lancet*, 383(9930):1739–1747.

Noori, N., Derra, K., Valea, I., Oron, A. P., Welgo, A., Rouamba, T., Boua, P. R., Somé, A. M., Rouamba, E., Wenger, E., Sorgho, H., Tinto, H., and Ouédraogo, A. L. (2021). Patterns of child mortality in rural area of burkina faso: evidence from the nanoro health and demographic surveillance system (hdss). *BMC Public Health*, 21.

Okango, E., Mwambi, H., and Ngesa, O. (2016). Spatial modeling of hiv and hsv-2 among women in kenya with spatially varying coefficients. *BMC Public Health*, 16(1).

Okell, L. C., Kwambai, T. K., Dhabangi, A., Khairallah, C., Nkosi-Gondwe, T., Winskill, P., Opoka, R., Mousa, A., Kühl, M.-J., Lucas, T. C. D., Challenger, J. D., Idro, R., Weiss, D. J., Cairns, M., ter Kuile, F. O., Phiri, K., Robberstad, B., and Mori, A. T. (2023). Projected health impact of post-discharge malaria chemoprevention among children with severe malarial anaemia in africa. *Nature Communications*, 14(1).

Okiro, E. A., Al-Taiar, A., Reyburn, H., Idro, R., Berkley, J. A., and Snow, R. W. (2009). Age patterns of severe paediatric malaria and their relationship to plasmodium falciparum transmission intensity. *Malaria Journal*, 8(1).

Ouma, P., Macharia, P. M., Okiro, E., and Alegana, V. (2021). *Methods of Measuring Spatial Accessibility to Health Care in Uganda*, pages 77–90. Springer International Publishing.

Ouma, P. O., Maina, J., Thuranira, P. N., Macharia, P. M., Alegana, V. A., English, M., Okiro, E. A., and Snow, R. W. (2018). Access to emergency hospital care provided by the public sector in sub-saharan africa in 2015: a geocoded inventory and spatial analysis. *The Lancet Global Health*, 6:e342–e350.

Paton, R. S., Kamau, A., Akech, S., Agweyu, A., Ogero, M., Mwandawiro, C., Mturi, N., Mohammed, S., Mpimbaza, A., Kariuki, S., Otieno, N. A., Nyawanda, B. O., Mohamed, A. F., Mtove, G., Reyburn, H., Gupta, S., Bejon, P., Lourenço, J., and Snow, R. W. (2021). Malaria infection and severe disease risks in africa. *Science*, 373(6557):926–931.

Phiri, K., Esan, M., van Hensbroek, M. B., Khairallah, C., Faragher, B., and ter Kuile, F. O. (2012). Intermittent preventive therapy for malaria with monthly artemether-lumefantrine for the post-discharge management of severe anaemia in children aged 4-59 months in southern malawi: a multicentre, randomised, placebo-controlled trial. *www.thelancet.com/infection*, 12:191–200.

Quattrochi, J. P., Hill, K., Salomon, J. A., and Castro, M. C. (2020). The effects of changes in distance to nearest health facility on under-5 mortality and health care utilization in rural malawi, 1980-1998. *BMC health services research*, 20:899.

Ray, N. and Ebener, S. (2008). Accessmod 3.0: computing geographic coverage and accessibility to health care services using anisotropic movement of patients. *Int. J. Health Geogr.*, 7:63.

Razzak, J., Beecroft, B., Brown, J., Hargarten, S., and Anand, N. (2019). Emergency care research as a global health priority: key scientific opportunities and challenges. *BMJ Global Health*, 4(Suppl 6):e001486.

Razzak, J. A. and Kellermann, A. L. (2002). Emergency medical care in developing countries: is it worthwhile? *Bull World Health Organ*, 80:900–905.

Rees, C. P., Hawkesworth, S., Moore, S. E., Dondeh, B. L., and Unger, S. A. (2016). Factors affecting access to healthcare: An observational study of children under 5 years of age presenting to a rural gambian primary healthcare centre. *PLoS ONE*, 11.

Robert, B. N., Cherono, A., Mumo, E., Mwandawiro, C., Okoyo, C., Gichuki, P. M., Blanford, J. l., Snow, R. W., and Okiro, E. A. (2023). Spatial variation and clustering of anaemia prevalence in school-aged children in western kenya. *PLOS ONE*, 18(11):e0282382.

Robison, J. A., Ahmad, Z. P., Nosek, C. A., Durand, C., Namathanga, A., Milazi, R., Thomas, A., Soprano, J. V., Mwansambo, C., Kazembe, P. N., and Torrey, S. B. (2012). Decreased pediatric hospital mortality after an intervention to improve emergency care in lilongwe, malawi. *Pediatrics*, 130(3):e676–e682.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields : theory and applications*. Chapman I& Hall/CRC.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with inla: A review. *Annual Review of Statistics and Its Application*, 4(1):395–421.

Rutherford, M. E., Dockerty, J. D., Jasseh, M., Howie, S. R., Herbison, P., Jeffries, D. J., Leach, M., Stevens, W., Mulholland, K., Adegbola, R. A., and Hill, P. C. (2009). Access to health care and mortality of children under 5 years of age in the gambia: A case-control study. *Bulletin of the World Health Organization*, 87:216–224.

Rutherford, M. E., Mulholland, K., and Hill, P. C. (2010). How access to health care relates to under-five mortality in sub-saharan africa: Systematic review.

Smith, D. L., Guerra, C. A., Snow, R. W., and Hay, S. I. (2007). Standardizing estimates of the plasmodium falciparum parasite rate. *Malaria Journal*, 6(1).

Snow, R. and Marsh, K. (2002). *The consequences of reducing transmission of Plasmodium falciparum in Africa*, page 235–264. Elsevier.

Snow, R. W., Omumbo, J. A., Lowe, B., Molyneux, C. S., Obiero, J.-O., Palmer, A., Weber, M. W., Pinder, M., Nahlen, B., Obonyo, C., Newbold, C., Gupta, S., and Marsh, K. (1997). Relation between severe malaria morbidity in children and level of plasmodium falciparum transmission in africa. *The Lancet*, 349(9066):1650–1654.

Snow, R. W., Sartorius, B., Kyalo, D., Maina, J., Amratia, P., Mundia, C. W., Bejon, P., and Noor, A. M. (2017). The prevalence of plasmodium falciparum in sub-saharan africa since 1900. *Nature*, 550(7677):515–518.

Stevens, F. R., Gaughan, A. E., Linard, C., and Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLOS ONE*, 10(2):e0107042.

Suiyanka, L., Alegana, V. A., and Snow, R. W. (2021). Insecticide-treated net distribution in western kenya: impacts related to covid-19 and health worker strikes. *International Health*, 14(5):537–539.

Sullivan, K. M., Mei, Z., Laurence, I. G.-S., and Parvanta (2008). Haemoglobin adjustments to define anaemia. *Trop. Med. Int. Health*, 13:1267–1271.

Sylverken, J., Robison, J. A., Osei-Akoto, A., Nguah, S. B., Addo-Yobo, E., Balch, A., Bolte, R., and Ansong, D. (2019). Decreased mortality after establishing a pediatric emergency unit at an urban referral hospital in ghana. *Pediatric Emergency Care*, 37(7):e391–e395.

Tuti, T., Bitok, M., Paton, C., Makone, B., Malla, L., Muinga, N., Gathara, D., and English, M. (2016). Innovating to enhance clinical data management using non-commercial and open source solutions across a multi-center network supporting inpatient pediatric care and research in kenya. *Journal of the American Medical Informatics Association*, 23:184–192.

UN (2015). Transforming our world: the 2030 agenda for sustainable development united nations united nations transforming our world: the 2030 agenda for sustainable development. *United Nations*.

UN-IGME (2022). Levels i& trends in child mortality : report 2022.

White, M. and Watson, J. (2018). Age, exposure and immunity. *Elife*, 7.

WHO (2009). *WHO child growth standards: Growth velocity based on weight length and head circumference*. World Health Organization, Genève, Switzerland.

WHO (2011). Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Accessed: 2023-12-13.

WHO (2015). Guide to the health facility data quality report card. 2014. Accessed: 2024-01-08.

WHO (2017). Health in the 2030 agenda for sustainable development.

WHO (2021). Global anaemia estimates 2021.

Wilson, J. R. and Lorenz, K. A. (2015). *Modeling binary correlated responses using SAS, SPSS and R*, volume 9. Springer.

Yang, C., Yu, B., Chen, Z., Song, W., Zhou, Y., Li, X., and Wu, J. (2019). A spatial-socioeconomic urban development status curve from npp-viirs nighttime light data. *Remote Sensing*, 11(20):2398.

Zoungrana, A., Chou, Y.-J., and Pu, C. (2014). Socioeconomic and environment determinants as predictors of severe malaria in children under 5 years of age admitted in two hospitals in koudougou district, burkina faso: A cross sectional study. *Acta Tropica*, 139:109–114.

Appendix A

Previous Studies

This section highlights the different studies found in literature evaluating the impact of physical access to EC on health outcomes and the identified gaps.

Table A.1: Studies evaluating the association between Physical Access to EC services and Community Health Outcomes

Author	**Country (District)**	Metric	Outcome	Statistical Method	Covariates	Findings	Gap(s)
Magnani et al. (1996)	Niger	Euclidean Distance	Child Mortality	Logistic regression	Maternal literacy, Maternal age	Lower odds of mortality among children $\leq$ 5km from a dispensary compared to those $>$ 5km	Did not adjust for underlying risk factors present in the study - measles and famine
Målqvist et al. (2010)	Vietnam (Quang Ninh)	Euclidean Distance	Neonatal mortality	Logistic regression	Maternal age and status at delivery	Increased neonatal mortality risk with increasing distance	Use of Euclidean distances especially in mountainous region Not adjusting for underlying driver of mortality Missing other possible confounders
Karra et al. (2017)	LMICs	Distance (DHS service availability module)	Neonatal mortality Facility Delivery ANC utilization	Logistic Regression	Sex, multiple births, age, household wealth index, mother's education, birth order, maternal age, marital status and urban/rural	Decreased incidence by travel time. Increased risk of mortality with increasing travel and distance. Decrease odds of facility delivery with increasing distance	Do not account for the underlying drivers of neonatal mortality at cluster locations
Kashima et al. (2012)	Madagascar	Euclidean Distance	Early childhood mortality	Logistic regression	birth order, the type of the nearest health, existence of reference hospital within 30 km, wealth, maternal education, religion, maternal smoking, maternal age at birth, and birth spacing	Large distances are a risk factor for early childhood mortality	Use of cluster locations as points of residences Exlcusion of some health centres (new ones or those with missing coordinates) Not adjusting for underlying driver of mortality - suggested to be maternal conditions

Lohela et al. (2012)	Malawi and Zambia	Euclidean Distance	Early neonatal mortality	Logistic regression	mens opinion on female autonomy, ethnicity, partners occupation, partners education, wanted pregnancy, siblings under 7 years old in household, estimate of newborn size (by mother), media use (men & women), womens mobility & financial autonomy, language, multiple pregnancy, occupation, marital status, age at birth, modern attitudes, mens modern attitudes, exposure to health programmes in the media, sex of index child.	Distance had an impact in Zambia only	Use of cluster locations as points of residences Different risk levels at different clusters were not adjusted for.
Rutherford et al. (2009)	Gambia	Travel time and Euclidean distance	All-cause paediatric mortality	Conditional Logistic Regression	readily available money, vitamin A status, vaccination status	No association between travel time (or distance) and mortality	Not adjusting for underlying risk in their case-control matching
Quattrochi et al. (2020)	Malawi	Euclidean Distances	Under 5 mortality	Cox proportional hazards regression	Brith order, multiple births, mother age and education, year and month	Greater distances associated with higher mortality	Inclusion of closed facilities during the study period (18 years) Variations in underlying drivers was not accounted for.
Noori et al. (2021)	Burkina Faso (Nanoro)	Travel time (MAP friction surface)	Child Mortality	Cox proportional hazards regression	Age and season	Increasing distance to an inpatient facility increased the odds of child mortality (no effect was observed for distance to outpatient facilities)	Missing confounders on socio-economic factors and underlying drivers of mortality

Kadobera et al. (2012)	Tanzania (Ifakara)	Networked Distance	Child Mortality	Cox proportional hazards regression	age of child, gender, mother's age, mother's education, mother's age, death of mother, death of preceding sibling, and district of residence	Longer distances were associated with higher mortality.	No adjustments for variables on underlying causes of mortality.

Table A.2: Studies evaluating association between Physical Access to EC services and in-hospital mortality

Author	**Country (District)**	Metric	Outcome	Statistical Method	Covariates	Findings	Gap(s)
Manongi et al. (2014)	Tanzania (Muheza)	Travel time - estimated by adjusted linear distances and travel scenario	Inpatient Child mortality (Under 5 years)	Poisson regression Logistic regression	Age, sex, mother's education, no. of ill days pre-admission, medication within 48 h of admission	Increased mortality risk for $>$ 3 hours travel time	The study exclusions such admissions outside working hours may have introduced bias as such case may be more severe. Do not adjust for varying rates of disease exposure.
Ippolito et al. (2018)	Zambia (Nchelenge)	Distance - provided by MoH	In-hospital mortality	Conditional Logistic Regression (case control: 1 - 1 matching)	Age, Gender	Cases travelled longer distances than controls. Deaths on arrival had more severe anaemia and came from distant villages in comparison to those who died in wards	Unable to exclude admissions with co-infections Did not adjust for malaria transmission rates No details on how the provided distances by MoH are measured

Moïsi et al. (2011)	Kenya (Kilifi)	Travel time (Pedestrian and Vehicular) - cost-distance algorithm	Hospital incidence and hospital mortality (Severe pneumonia, meningitis, all cause) (Under 5 years)	Log-linear model Logistic regression	Sex, ethnicity, season, maternal education, migrant status	Decreased incidence by travel time. Increased risk of mortality with increasing travel time. Disease severity decreased the effect of travel time	Socio-demographics were measured at sublocation level
Kazembe et al. (2006)	Malawi (Zomba)	Distance - Euclidean	Hospital incidence and hospital mortality (Under 14 years)	Logistic regression	Age, sex, admission Day (weekday/weekend), Season, referral, length of stay, Treatment	Hospital incidence was higher in infants and lower in 5-9 years age group. Greater distances had higher odds of death. Use of wards for residence as opposed to villages.	Did not adjust for malaria transmission rates.

Table A.3: Studies evaluating association between Physical Access to EC services and disease severity.

Author	**Country (District)**	Metric	Outcome	Statistical Method	Covariates	Findings	Gap(s)
Rees et al. (2016)	Gambia (Lower River)	Euclidean Distance	Delayed Presentation, Severe illness (diarrheal, lower respiratory, malaria)	Logistic Regression	Child age, only child, dead sibling, no. of maternal siblings, birth order, mother's age	No association	No adjustments for socioeconomic factors. No adjustments for different levels of risk in different villages

Kahabuka et al. (2012)	Tanzania (Muhenza)	Travel time	Severe Malaria	Logistic Regression	Age, sex, caretaker's education, caretaker's socio-economic status, no. of own alive children, use of nearer facility	No association between travel time to hospital with severe malaria	Inclusion of severe comorbidities (diarrhea and pneumonia) Transmission was not accounted for.
Zoungrana et al. (2014)	Burkina Faso	Travel time and Distance	Severe Malaria	Logistic Regression	Parent age, gender, father's education, polygamy, ethnicity, father religion, area of residence, household wealth, transport means, malaria knowledge, residing beside water pond	No association between travel time (or distance) and severe malaria	Confounding from variations in malaria transmission was not accounted for
Mousa et al. (2020)	LMICs	Travel time	Severe Malaria. Delayed Care seeking	Mixed-effects Logistic Regression	Age, illness duration	1 hour of travel time was associated with increased odds of severe malaria	Failed to adjust for transmission intensity when assessing impact of distance on severe malaria (was adjusted for assessing impact of delayed treatment on severe malaria)
Mutsigiri-Murewanhema et al. (2017)	Zimbabwe (Mutasa and Nyanga)	Distance to nearest facility	Severe Malaria	Logistic Regression	None: Bivariate Analysis	Distance of $\geq$ 10 km was a risk factor for severe malaria	Lack of a multi-variate analysis.
Al-Taiar et al. (2008)	Yemen (Taiz)	Distance to nearest facility	Severe Malaria	Unconditional logistic regression	Age, Presence of water stream, Presence of water pump, Burning mosquito coils	Distance of $\geq$ 2 km was a risk factor for severe malaria	Malaria prevalence not adjusted for

Mpimbaza et al. (2017)	Uganda (Jinja)	Distance to nearest facility (stratified by facility level) Severe Malaria.	Delayed Care seeking	Conditional Logistic Regression	Delay to seek care, mother autonomy, age, caretaker employment status, household head's years of education, socio economic position, number of under 5 children in household, gametocytemia, danger symptoms on day1, drug shop as 1st response	Risk of both Severe Malaria and delayed care seeking increased with increasing distance from nearest facilities of level III or higher	Malaria prevalence not adjusted for. Cases were enrolled from a level VI hospital whereas controls were enrolled at level III and IV hospitals - introduced bias as higher-level hospitals are few (therefore longer distances) and handle more severe cases.

Appendix B

Travel Time Maps

The following are map outputs for Travel time.

Figure B.1 shows the travel times to BCRH; panel A is the raster output from Accessmod and panel B is the extracted average travel time for each EA.

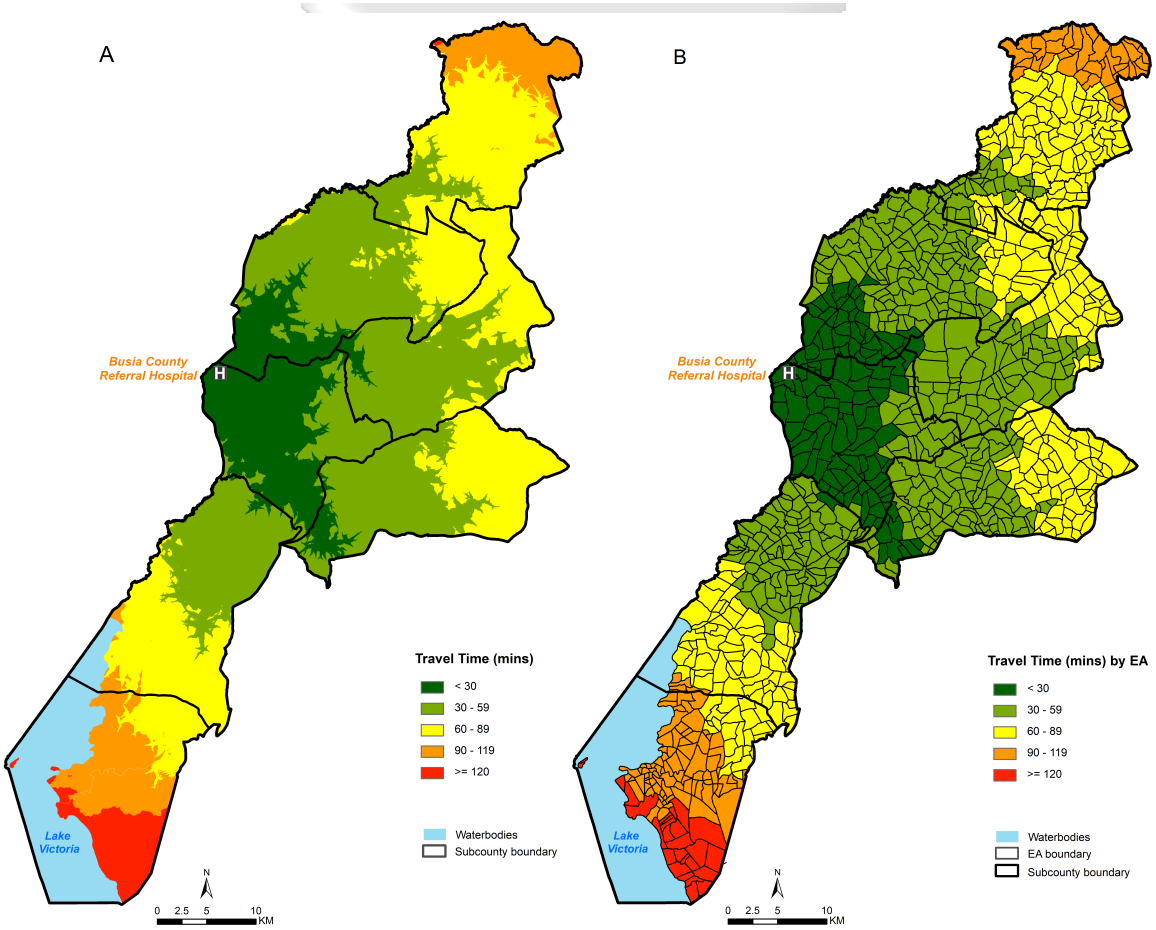

Figure B.1: Spatial Distribution of Travel time in Busia County

Appendix C

Ethics Approval

26th February 2024

Mr Mutinda Moses,
moses.mutinda@strathmore.edu

Dear Mr Mutinda,

RE: Impact of Distance to Hospital Emergency Care on Severe Anaemia in Busia County: A Spatial Modelling Approach

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** research proposal. Your application reference number is **SU-ISERC2009/24.** The approval period is from **26th February 2024 to 25th February 2025.**
This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

STRATHMORE UNIVERSITY INSTITUTIONAL
SCIENTIFIC AND ETHICAL REVIEW COMMITTEE
(SU-ISERC)
27-Feb-2024
Email:ethicsreview@strathmore.edu
P.O BOX 59857-00200
NAIROBI-KENYA

Appendix D

Similarity Index

Moses_Musau_Thesis__(050424).pdf

ORIGINALITY REPORT

7%	7%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Strathmore University Student Paper	1%
2	rdrr.io Internet Source	1%
3	www.ncbi.nlm.nih.gov Internet Source	1%
4	export.arxiv.org Internet Source	<1%
5	researchonline.lshtm.ac.uk Internet Source	<1%
6	su-plus.strathmore.edu Internet Source	<1%
7	dash.harvard.edu Internet Source	<1%
8	oro.open.ac.uk Internet Source	<1%
9	hdl.handle.net Internet Source	<1%

10	etheses.whiterose.ac.uk Internet Source	<1 %
11	stars.library.ucf.edu Internet Source	<1 %
12	worldwidescience.org Internet Source	<1 %
13	Submitted to University of Edinburgh Student Paper	<1 %
14	Submitted to Associatie K.U.Leuven Student Paper	<1 %
15	etd.aau.edu.et Internet Source	<1 %
16	ihi.eprints.org Internet Source	<1 %
17	Submitted to University of Southampton Student Paper	<1 %
18	Xiaoqian Sun. "Bayesian spatial models with repeated measurements: with application to the herbaceous data analysis", Statistical Methods and Applications, 03/06/2009 Publication	<1 %
19	www.biorxiv.org Internet Source	<1 %
20	docplayer.net Internet Source	<1 %

21	erepository.uonbi.ac.ke Internet Source	<1 %
22	cdr.lib.unc.edu Internet Source	<1 %
23	est.ufmg.br Internet Source	<1 %
24	erepository.uonbi.ac.ke:8080 Internet Source	<1 %
25	onlinelibrary.wiley.com Internet Source	<1 %
26	André Python, Janine B. Illian, Charlotte M. Jones-Todd, Marta Blangiardo. "A Bayesian approach to modelling subnational spatial dynamics of worldwide non-state terrorism, 2010-2016", Journal of the Royal Statistical Society: Series A (Statistics in Society), 2018 Publication	<1 %

Exclude quotes	On	Exclude matches	< 25 words
Exclude bibliography	On		

Appendix E

R code

```
#=================================================================
#==== PART 1: COVARIATE SELECTION AND MODEL DIAGNOSTICS ===
#=================================================================
## Loading packages
pacman::p_load(caret, pROC, gtsummary, tidyverse)

## Loading Data
busia_cin_data <- read.csv( "Busia_County_Anaemic_Admissions_
    Descriptives_U5.csv",
check.names = F,
na.strings = "")

## Variable definition
# severity2 - 0 and 1 for mild and severe anaemia respectively
# tt2 - travel time categories
# age_cat - age categories
# child_sex - Male and Female
# mal_diag - Malaria diagnosis; yes, no
# nourishment_cat - Nourishment status
# und_cat - Underlying conditions
# good_vacc - Good vaccinees
# referal - Referral status
# day_week - Day of admission; Weekday and weekend
# season - Rainy and dry
# yy_adm - Year of admission
```

```
25        # ntl_urbn - Urban index
26        # avg_pfpr1 - PfPR
27        #-----------------------------------------------------------------
28        ## Fit model with all covariates
29        model_full <- glm(severity2 ~ tt2 + age_cat + child_sex +
30                      mal_diag + nourishment_cat + und_cat +
31                      good_vacc + referal +  day_week +
32                      season + yy_adm + ntl_urbn + avg_pfpr1,
33                      family = binomial("logit"),
34                      data = busia_cin_data)
35        #-----------------------------------------------------------------
36        ## check importance of other variables - stepwise regression
37        step(model_full, direction = "both", trace = 1)
38
39        ## selected model
40        selected_model <- glm(severity ~ tt2 + age_cat + child_sex +
41                            mal_diag + nourishment_cat + und_cat +
42                          good_vacc + referal + day_week +
43                          season + yy_adm +  ntl_urbn + avg_pfpr1+,
44                            family = binomial("logit"),
45                            data = busia_cin_data)
46
47        #=================================================================
48        #==== 1.2: MODEL DIAGNOSTICS
49
50        ## Linearity assumption
51        l_plot <- busia_cin_data %>%
52               mutate(comp_res = coef(selected_model)["avg_pfpr1"]*avg_
                      pfpr1 +
53               residuals(selected_model, type = "working")) %>%
54               ggplot(aes(x = avg_pfpr1, y = comp_res)) +
55               geom_point() + geom_smooth(se = F, linewidth = 1.5)+
```

```
56              geom_smooth(color = "red", method = "lm",
57                                  linetype = 2, se = F, linewidth =
                                        1.5) +
58              labs(x = "*Pf*PR", y = "Component-Plus-Residual")+
59              theme_classic()+
60              theme(
61                      axis.title=element_text(#size=14,face="bold"),
62                      plot.margin = unit(c(0.2, 0.2, 0.2, 0.2), "cm"),
63                      panel.background = element_rect(fill='white'),
64                      axis.title.x = ggtext::element_markdown()
65              );l_plot
66
67      #-----------------------------------------------------------------
68      ## Outliers
69      # Extract model results
70      model.data <- broom::augment(selected_model) %>%
71              mutate(index = 1:n(),
72              Anaemia= factor(severity2, labels = c("Mild","Severe"))
73              )
74      #create plot
75      outlier_plot <- ggplot(model.data, aes(index, .std.resid)) +
76              geom_point(aes(color =Anaemia), alpha = .5) +
77              scale_y_continuous(limits = c(-3,3),breaks = seq(-3,3,1))+
78              theme_bw();outlier_plot
79
80      #-----------------------------------------------------------------
81      ## Multicollinearity
82      vif_df <- car::vif(selected_model) %>%
83              as.data.frame() %>%
84              mutate_if(is.numeric, round, digits = 2) %>%
85              rownames_to_column();vif_df
86
```

```
#-------------------------------------------------------------------
## Spatial Dependence
#create necessary variables
busia_cin_data2 <- busia_cin_data %>%
       mutate(severity = severity2,
       cn = 1:length(tt))%>%
       rename(Long = EA_long, Lat = EA_Lat)

cluster_df <- busia_cin_data2
#define model formula for spatial residuals
formula <- severity ~ tt2 + age_cat + child_sex +
                      mal_diag + nourishment_cat + und_cat +
                      good_vacc + referal + day_week +
                      season + yy_adm +  ntl_urbn + avg_pfpr1 +
                      f(cn, model = "iid")

#inla call
inla.mod <- inla(formula, data = data.frame(cluster_df),
       family = "binomial",
      control.compute = list(dic = TRUE, waic = TRUE),
      control.predictor = list(compute = TRUE, link=1))
#Extract the residuals
cluster_df$residuals <- inla.mod$summary.random$cn$mean
#Calculate empirical variogram using the residuals
vgm1 <- variogram(residuals~1, data = cluster_df,
              ~Long+Lat, cutoff = 1.5)
#Fit a variogram model - Matern covariance
foo <- fit.variogram(vgm1, vgm("Mat", fit.kappa = TRUE))
# dataframe for plotting
vline_df <- variogramLine(vgm(psill = foo$psill[2], "Mat",
       range = foo$range[2],  kappa = foo$kappa[2],
       nugget = foo$psill[1]), 10)
```

```
#varigram plot
variogram_plot1 <- ggplot(data = vgm1,
                mapping = aes(x = dist, y = gamma))+
        labs(x = "distance␣(km)", y = "semivariance")+
        geom_point(shape = 1, color ="#0D6599")+
        scale_y_continuous(  breaks = seq(0,y_max,20*10^-10),
                limits = c(0,y_max))+
        scale_x_continuous( breaks = seq(0,x_max,0.09),
                labels = round(seq(0,x_max,0.09)*111.11),
        limits = c(0,x_max))+
        coord_cartesian(expand = FALSE, clip = "off")+
        geom_hline(yintercept = foo$psill[1], color = "blue" )+
        geom_hline(yintercept = foo$psill[1] + foo$psill[2],
                color = "green" )+
        geom_line(data = vline_df,
                mapping = aes(x = dist, y = gamma), type = 'l')+
        theme_classic()+
        theme(
                legend.position = "none",
                axis.text=element_text(size=16),
                axis.title=element_text(size=14,face="bold"),
                plot.margin = unit(c(0.2, 0.2, 0.2, 0.2), "cm"),
                panel.background = element_rect(fill='white')
                );variogram_plot1

#=================================================================
#==== PART 2: CREATING SPATIAL OBJECTS AND MODEL FITTING ==
#=================================================================

## Loading Packages
pacman::p_load(INLA, sf, terra, readxl, maps, tmap,
        tmaptools, tidyverse)
```

```
#set seed
set.seed(152950)
#=========================================================================
#==== 2.1: CREATING SPATIAL OBJECTS

#define Projected coordinate system for Busia County
projMercator<-"+proj=utm +zone=37 +south +a=6378249.145 +rf
    =293.465 +towgs84=-160,-6,-302,0,0,0,0 +units=km +no_defs +
    type=crs"

#load study area shapefile - Busia county
study_area_sf <- st_read("Busia_County2.shp")%>%
       st_make_valid() %>% st_transform(projMercator)

#obtain all sampled villages and create a shapefile
sampled_eas <- busia_cin_data %>%
       st_as_sf(coords = c("EA_long","EA_Lat"),
                     crs = st_crs(wk_eas_sf), remove = TRUE) %>%
              st_transform(projMercator) %>%
              mutate(
              easting = st_coordinates(.)[,1],
              northing = st_coordinates(.)[,2]
              );dim(sampled_eas)

##Buid Mesh
hull <- inla.nonconvex.hull (
       points = as.matrix(st_coordinates(sampled_eas)),
       convex = 10, concave = 18)
mesh <- inla.mesh.2d(
       boundary = hull, max.edge = c(0.5 , 3),
       cutoff= 2.5 , offset= c( 1, 3))

```

```
## Define SPDE on the mesh
spde <- inla.spde2.matern(mesh = mesh, alpha = 2)

## create index sets for spatial dependence and PfPR
indexs <- inla.spde.make.index("spatial.field", spde$n.spde)
pfpr_idx <- inla.spde.make.index(name = "pfpr",
       n.spde = spde$n.spde)

##create a matrix of coordinates
all_coords <- sampled_eas %>%
       st_coordinates() %>% as.matrix()
##define projector matrices
sampled_matrix <- inla.spde.make.A(mesh = mesh,
       loc = all_coords);dim(sampled_matrix)
A_pfpr <- inla.spde.make.A(mesh = mesh,
       loc = all_coords,
       weights = sampled_eas$avg_pfpr1 );dim(A_pfpr)

#stack all spatial objects
stk.e <- inla.stack(
       tag = "est",
       data =
       #response variable
       list(severity2 =as.vector(busia_cin_data$severity2)),
       #fixed and random effects
       effects = list(data.frame(
              b0 = rep(1, nrow(busia_cin_data)),
              tt2 = busia_cin_data$tt2,
              age_cat = busia_cin_data$age_cat,
              child_sex = busia_cin_data$child_sex,
              avg_pfpr1 = busia_cin_data$avg_pfpr1,
              ntl_urbn = busia_cin_data$ntl_urbn,
```

```
                    season = busia_cin_data$season,
                    referal = busia_cin_data$referal,
                    und_cat = busia_cin_data$und_cat,
                    mal_diag = busia_cin_data$mal_diag,
                    yy_adm = busia_cin_data$yy_adm,
                    nourishment_cat = busia_cin_data$nourishment_cat
                    day_week = busia_cin_data$day_week,
                    good_vacc = busia_cin_data$good_vacc,
                    EA_code1 = busia_cin_data$EA_code),
                    s = indexs,
                    pfpr = pfpr_idx ),
             A = list(1, #for non-spatial terms
                    sampled_matrix,
                    A_pfpr ))

#=================================================================
#==== 2.2: MODEL FITTING

## Model 1: Standard Logistic Regression adjusted for all
    confounders except PfPR.
#define formula
formula1 <- severity2 ~ severity ~ tt2 + age_cat + child_sex +
        mal_diag + nourishment_cat + und_cat +
        good_vacc + referal + day_week +
        season + yy_adm +  ntl_urbn

#Execute INLA call
model1 <- inla(formula1, data=busia_cin_data,
        family="binomial",
        control.compute=list(dic=TRUE,mlik=TRUE,cpo=TRUE),
        control.predictor = list(compute = TRUE, link=1))
```

```
## Model Summaries
model_summary1 <- summary(model1)
#exponentiate to obtain odds ratios
exp(model_summary1$fixed[,c(1,3,5)])

#-----------------------------------------------------------------
## Model 2: Standard Logistic Regression adjusted for all
    confounders including PfPR.

#define formula
formula2 <- severity ~ tt2 + age_cat + child_sex +
       mal_diag + nourishment_cat + und_cat +
       good_vacc + referal + day_week +
       season + yy_adm +  ntl_urbn + avg_pfpr1

#Execute INLA call
model2 <- inla(formula2, data=busia_cin_data,
       family="binomial",
       control.compute=list(dic=TRUE,mlik=TRUE,cpo=TRUE),
       control.predictor = list(compute = TRUE, link=1))

## Model Summaries
model_summary2 <- summary(model2)
#exponentiate to obtain odds ratios
exp(model_summary2$fixed[,c(1,3,5)])

#-----------------------------------------------------------------
## Model 3: Standard INLA-SPDE model adjusted for all confounders
    including PfPR and spatial autocorrelation.

#define formula
formula3 <- severity ~ tt2 + age_cat + child_sex +
```

```
274             mal_diag + nourishment_cat + und_cat +
275             good_vacc + referal + day_week +
276             season + yy_adm +  ntl_urbn + avg_pfpr1 +
277         #spatial autocorrelation
278         f(spatial.field, model = spde)
279
280         #Execute INLA call
281         model3 <- inla(formula3, data=inla.stack.data(stk.e),
282                 family="binomial",
283                 control.predictor=list(A=inla.stack.A(stk.e),
284                 compute = FALSE, link=1),
285                 control.compute=list(dic=TRUE,mlik=TRUE,cpo=TRUE,
286                         config = TRUE))
287
288         ## Model Summaries
289         model_summary3 <- summary(model3)
290         #exponentiate to obtain odds ratios
291         exp(model_summary3$fixed[,c(1,3,5)])
292
293
294         #-------------------------------------------------------------
295         ## Model 4: INLA-SPDE Model with spatially varying PfPR adjusted
                for all confounders and spatial autocorrelation.
296
297         #define formula
298         formula4 <- severity ~ tt2 + age_cat + child_sex +
299         mal_diag + nourishment_cat + und_cat +
300         good_vacc + referal + day_week +
301         season + yy_adm +  ntl_urbn +
302         #spatial autocorrelation
303         f(spatial.field, model = spde)
304         #spatially varying effect
```

```
305         f(pfpr, model = spde)
306
307         #Execute INLA call
308         model4 <- inla(formula4, data=inla.stack.data(stk.e),
309                 family="binomial",
310                 control.predictor=list(A=inla.stack.A(stk.e),
311                 compute = FALSE, link=1),
312                 control.compute=list(dic=TRUE,mlik=TRUE,cpo=TRUE,
313                         config = TRUE))
314
315         ## Model Summaries
316         model_summary4 <- summary(model4)
317         #exponentiate to obtain odds ratios
318         exp(model_summary4$fixed[,c(1,3,5)])
319
320         #------------------------------------------------------------------
321         ## Obtaining Effect size of spatial autocorrelation
322         spatial.field <- data.frame(
323                 mean= exp(model3$summary.random$spatial.field$mean))
324
325         #mapping grid
326         pred_grid <- as.matrix(inla.mesh.project(mesh_proj,spatial.field))

327         # create a raster surface
328         out_stk <- rast()
329         mean_j <- cbind(expand.grid(x = mesh_proj$x, y = mesh_proj $y) ,
330         Z = c(matrix(pred_grid[, 1] , grd_dims[1] )))
331         mean_j <- rast(mean_j, crs = projMercator)
332         out_j <- c(mean_j)
333         terra::add(out_stk) <- out_j
334         names(out_stk) <- c("spatial.field_mean")
335         out_stk <- terra::mask(out_stk, study_area_sf, touches = FALSE)
```

```
        writeRaster(out_stk, "spatial.field_mean.tif", overwrite=TRUE)

        #------------------------------------------------------------------
        ## Obtaining Effect size of spatially varying PfPR
        pfpr <- data.frame(
               mean= exp(model4$summary.random$pfpr$mean))

        #mapping grid
        pred_grid <- as.matrix(inla.mesh.project(mesh_proj,pfpr))
        # create a raster surface
        out_stk <- rast()
        mean_j <- cbind(expand.grid(x = mesh_proj$x, y = mesh_proj $y) ,
        Z = c(matrix(pred_grid[, 1] , grd_dims[1] )))
        mean_j <- rast(mean_j, crs = projMercator)
        out_j <- c(mean_j)
        terra::add(out_stk) <- out_j
        names(out_stk) <- c("pfpr_mean")
        out_stk <- terra::mask(out_stk, study_area_sf, touches = FALSE)
        writeRaster(out_stk , pfpr_mean.tif", overwrite=TRUE)
```

Maps were visualized using ArcGIS version 10.8.2