

https://dspace.org/

Malaria morbidity in under-five children in Malawi: A spatial-temporal Bayesian model

Chirwa, Vitumbiko Chijere

2021
https://hdl.handle.net/10539/33591

Downloaded from a DSpace repository. https://dspace.org/

Malaria morbidity in under-five children in Malawi: A spatial-temporal Bayesian model



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

FACULTY OF HEALTH SCIENCES
SCHOOL OF PUBLIC HEALTH

By

Vitumbiko Chijere Chirwa
(Student Number: 1756102)

Supervisor 1: Prof Eustasius Musenge (PhD)
Supervisor 2: Dr Kennedy Otwombe (PhD)

A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand School of Public Health in partial fulfilment of the requirements for the degree of Master of Science in Epidemiology in the field of Biostatistics

October 2021

DECLARATION

I, **Vitumbiko Chijere Chirwa (Student No: 1756102),** declare that this research report: ***Malaria morbidity in under-five children in Malawi: A spatial-temporal Bayesian model*** is my own original work; and that all other sources that I have used in this report have been acknowledged by explicitly indicating references. This report is being submitted for the degree of Master of Science in Epidemiology in the field of Biostatistics at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.



Vitumbiko Chijere Chirwa

Date: 14th October 2021

DEDICATION

This research report is dedicated to:

- Violet Kamasumbi Chijere Chirwa my wife
- The entire Chijere Chirwa family

ABSTRACT

Introduction

Malaria disease is one of the public health challenges in Malawi, with under-five children being the most affected. It is a leading cause of morbidity, hospital admission, and mortality in children under five years of age. To achieve the "Malaria –Free Malawi" goal, there is a need to know the specific geographical areas which still have the highest burden of malaria despite the passing of time. This study aimed at determining the malaria prevalence trends, the spatial distribution and clustering of malaria, and determining the spatial-temporal effects on malaria morbidity in under-five children in Malawi for the years 2010, 2012, 2014, and 2017.

Methodology

This study used data from the cross-sectional 2010, 2012, 2014, and 2017 Malawi Malaria Indicator Surveys (MMIS). Malaria Prevalence trends, at district level, were explored by using Line graphs in Stata. Spatial distribution of Malaria was explored by using Choropleth maps in ArcGIS. Spatial autocorrelation of Malaria was explored using the Moran's I index and the Getis-Ord G_i^* statistic in ArcGIS software; while spatial clustering of malaria was explored using the Kulldorf spatial scan statistic in SatScan software. Summary statistics of continuous variables were done using survey weighted means, while categorical variables were summarized using survey weighted frequencies with the associated proportions/percentages. For each survey year, determining the factors associated with malaria was done by using Multilevel Logistic Regression in Stata, of which Odds Ratios together with 95% Confidence Intervals were used to report the results. In determining the spatial-temporal effects on malaria, seven Negative Binomial models were fit; of which the first four models only included the cluster spatial random effect, while the last three models included both the spatial and temporal random effects. The model with the lowest DIC was chosen as the best fitting model. All the Bayesian models were fit using the INLA method.

Results

1758 children (from the 2010 MMIS), 2112 children (from the 2012 MMIS), 1928 children (from the 2014 MMIS), and 2305 children (from the 2017 MMIS) were included in the analysis. There was a general declining trend in malaria prevalence at national level (43% in 2010, 27% in 2012, 33% in 2014, and 24% in 2017). Most districts in the central region still had high malaria prevalence as compared to other districts in the other regions. Significant spatial positive autocorrelation of malaria prevalence values was observed in 2010 (I=0.044, p=0.021) and in 2012 (I=0.074, p<0.001). Most high malaria values were clustered in the central region and the south-eastern parts of Malawi, as per the results from the Kulldorf scan statistic. The Bayesian spatial-temporal random effects Negative Binomial model with the interaction term was chosen as the best fitting model because it had the lowest DIC value (DIC= 1839.70). The significant factors associated with low malaria risk in under-five children were: a child living in a rich house [RR=0.52, 95%Cr.I=(0.35,0.75)], a child whose mother attained Secondary (or higher) education level [RR=0.38, 95%Cr.I=(0.20,0.71)], and a child living in clusters with higher altitude [RR=0.97, 95%Cr.I=(0.94,0.99)]. The significant factors associated with high malaria risk were: Age (in months) of a child, and a child living in a rural area [RR=1.48, 95%Cr.I=(1.10,1.99)]. Spatial and temporal effects values were greater than 0, and most spatial malaria heterogeneity was explained by the structured spatial random effects.

Conclusions

There was a general decline in under-five malaria in Malawi, although malaria burden is still high in most areas of the central region and the south-eastern areas of Malawi. The significant predictors of malaria were: Age of the child, Place of residence, Wealth Index, Mother's education, and Cluster altitude. Bayesian Spatial-temporal models provide better fitting models in modelling under-five malaria morbidity in Malawi over time.

ACKNOWLEDGMENTS

My heartfelt appreciation goes to the Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) for offering me this once-in-a-lifetime scholarship to undertake the MSc Epidemiology (in the field of Biostatistics) at the University of the Wiwatersrand. Particularly, I am very grateful to Dr Pascalia Munyewende and Ms Gloria Bowes for their tireless work in handling any matters relating to the SSACAB scholarship.

I am very thankful to my dear supervisors Prof. Eustasius Musenge and Dr. Kennedy Otwombe for their professional as well as personal advice, encouragement and guidance throughout this thesis work. You always saw my potential even when I doubted myself, as you kept motivating me to do my best. Many thanks for instilling in me a desire to deepen my knowledge in Bayesian and Spatial analysis methods.

My appreciation also goes to the MEASURE DHS team and Mr Austin Gumbo of the Malawi Ministry of Health for granting me access to MMIS data for my research report.

I would also like to thank all the Biostatistics Lecturers (Prof E. Musenge, Prof J. Levin, Prof T. Chirwa, Dr. I. Maposa, Prof. N.B Kandala, Dr. K. Otwombe) and all the academic staff of the Department of Epidemiology and Biostatistics, School of Public Health, University of Witwatersrand, for the valuable, practical lectures as well as the support.

Many thanks to my fellow 2018 MSc Biostatistics classmates: Abdallah Kabwaago, Abel Mapfumo, Adatia Chivafa, Iimbasazethu Ngalo, Masedi Menyantsoe, and Samuel Oladokun.

To my lovely wife Violet, and the entire Chijere Chirwa, thank you for always telling me the *"Carpe Diem"* (seize the moment) principle and the *"the hard way, the only way"* phrase.

Lastly thanks be to God, for always reminding me that "without Him I can do nothing" (John 15:5), and "whatever my hand finds to do, I should do with all my might" (Ecclesiastes 9:10).

TABLE OF CONTENTS

DECLARATION ..... **i**

DEDICATION ..... **ii**

ABSTRACT ..... **iii**

ACKNOWLEDGMENTS ..... **v**

TABLE OF CONTENTS ..... **vi**

LIST OF FIGURES ..... **viii**

LIST OF TABLES ..... **ix**

ABBREVIATIONS ..... **x**

CHAPTER 1: INTRODUCTION **1**

1.1 Background **1**

1.1.1 **Malaria burden** 1

1.1.2 **Bayesian Spatial-temporal analysis in Epidemiological studies** 2

1.2 Literature review **3**

1.3 Problem Statement **8**

1.4 Justification of the study **8**

1.5 Research question **9**

1.6 Aim of the study **9**

1.7 Objectives of the study **9**

CHAPTER 2: METHODOLOGY **11**

2.1 Study Area and Setting **11**

2.2 Study Design **13**

2.3 Study Population **13**

2.4 Sampling Design, Sample size and Power Calculation **13**

2.5 Data Sources **14**

2.6 Study Variables **15**

2.7 Data Management **16**

2.8 Data Analysis **17**

2.8.1 Trends, Spatial Distribution, and Spatial Autocorrelation and Clustering of Malaria 17

2.8.2 Factors associated with malaria morbidity in 2010, 2012, 2014, 2017 20

2.8.3 Spatial-Temporal Effects on Malaria Morbidity 22

2.9 Ethical Considerations 27

CHAPTER 3: RESULTS 29

3.1 Trends, Spatial Distribution, Spatial Autocorrelation and Clustering of Malaria 29

3.1.1 Trends of Malaria Prevalence 29
3.1.2 Spatial Distribution of Malaria 31
3.1.3 Spatial Autocorrelation of Malaria 33

3.2 Factors associated with malaria morbidity in 2010, 2012, 2014, 2017 40

3.2.1 Descriptive Statistics for the Demographic, Socio-economic, Environmental, and Vector-control factors 40
3.2.2 Multilevel Logistic Regression Results for 2010, 2012, 2014, and 2017 MMIS 45

3.3 Spatial-temporal random effects on under-five Malaria morbidity 50

3.3.1 Model Fit Results 50

CHAPTER 4: DISCUSSIONS 59

4.1 Spatial-temporal Random Effects on Malaria variation 59

4.2 Factors associated with Malaria Risk in under-five children 60

4.2.1 Demographic Factors 60
4.2.2 Socioeconomic Factors 60
4.2.3 Environmental Factors 62
4.2.4 Vector-control Factors 62

4.3 Predictive Malaria Risk map 63

4.4 Strengths and Limitations of the Study 63

4.4.1 Strengths of the study 63
4.4.2 Limitations of the study 64

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS 66

5.1 Conclusions 66

5.2 Recommendations and Scope for Future Research 66

5.2.1 Recommendations to Policy Makers and Modellers 66
5.2.2 Scope for Future Research 67

REFERENCES 68

APPENDICES I

Appendix 1 I

Appendix 2: II

Appendix 3: III

Appendix 4: IV

Appendix 5: V

Appendix 6: VI

LIST OF FIGURES

Figure 1-1: A conceptual framework of factors associated with the Malaria among under-five children.....6

Figure 2- 1: Map of Malawi (source: MDHS 2015-16 report).....12

Figure 3-1: Malaria prevalence trends for Malawian districts, by region 31

Figure 3-2: Spatial distribution of Malaria at cluster level for 2010, 2012, 2014, and 2017 .32

Figure 3-3: Hotspot and Coldspot analysis of Malaria clustering for 2010, 2012, 2014, and 2017.....35

Figure 3-4: SatScan spatial scanning of Malaria clustering for 2010, 2012, 2014, and 2017 39

Figure 3-5: Non-linear effect of age (in months) on Malaria 54

Figure 3-6: Modelled Malaria hotspots (red shade) and colds pots (blue shade) in clusters, overall results from the Bayesian spatial-temporal model.....57

Figure 3-7: Temporal random effects distribution from the space-time interaction Bayesian Negative Binomial spatial-temporal model 57

Figure 3-8: Posterior density plots of means for fixed effects of the Negative Binomial spatial-temporal model.....58

LIST OF TABLES

Table 2. 1: List of independent variables used in this study 16

Table 2. 2: Mean and Variance of Malaria counts per year 23

Table 3. 1: Global Moran's I indices for 2010, 2012, 2014, and 2017 33

Table 3. 2: Spatial clustering of Malaria in Malawi in 2010, 2012, 2014, 2017 using SatScan 36

Table 3. 3: Descriptive statistics for the Demographic, Socio-Economic, Environmental and Vector-control characteristics 43

Table 3. 4: Multilevel logistic regression results for 2010, 2012, 2014, and 2017 MMIS 48

Table 3. 5: Non-Spatial, Spatial, and Spatial-temporal Multivariable Negative Binomial Models.....51

ABBREVIATIONS

AR	Autoregressive
BYM	Besag-York Mollie model
CHSU	Community Health Science Unit
C.I	Confidence Interval
Cr.I	Credible Interval
CSPro	Censuses and Surveys Processing
DHS	Demographic and Health Survey
GLMM	Generalized Linear Mixed Model
i.i.d	Independent and identically distributed
INLA	Integrated Nested Laplace Approximation
ITN	Insecticide Treated Net
MDHS	Malawi Demographic and Health Survey
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
MMIS	Malawi Malaria Indicator Survey
NB	Negative Binomial
NMCP	National Malaria Control Program
OR	Odds Ratio
RR	Relative Risk
SES	Socio-Economic Status
WHO	World Health Organization

CHAPTER 1: INTRODUCTION

This chapter presents, firstly, the background of under-five malaria burden in Malawi, as well as a general overview of Bayesian Spatial-temporal analysis in Epidemiological studies. It then presents a review of the published literature on the various factors associated with under-five malaria, as well as the use of spatial-temporal methods in analysing under-five malaria in Malawi. The problem statement and Justification of the study are also presented. The chapter ends with the aim and objectives of this study.

1.1 Background

1.1.1 Malaria burden

Malaria remains one of the significant global health problems. It was estimated that in 2010, there were 237 million malaria cases globally, while in 2015 there were 211 million malaria cases and 216 million malaria cases in 2016 (WHO, 2017).There is a general decline in the incidence rate of malaria by 18% globally from 76 to 63 cases per 1000 population at risk, between 2010 and 2016 (WHO, 2017).

Most malaria cases occur in the African continent, especially the Sub-Saharan African region. It is documented in the 2017 World Health Organisation (WHO) report that out of 91 countries worldwide which reported indigenous malaria cases in 2016, 15 countries were from Sub-Saharan Africa; and these 15 countries accounted for 80% of the global malaria burden. In the WHO African region, malaria incidence has decreased by 20% between 2010 and 2016 (WHO, 2017).

In Malawi, as one of the countries in sub-Saharan Africa, malaria remains a public health challenge where an estimated 4 million cases occur each year (NMCP and ICF, 2018). Of the general population in Malawi, children under the age of five years old are the most affected (Lowe, Chirombo and Tompkins, 2013; NMCP and ICF, 2018). It is a leading cause of

morbidity, hospital admission, and mortality in children under five years (Zgambo, Mbakaya and Kalembo, 2017; NMCP and ICF, 2018). Evidence from reports of the Malawi Malaria Indicator Surveys (MMIS) that were done in 2010, 2012, 2014 and 2017, malaria prevalence in under-five children based on Microscopy tests were estimated to be 43.3 in 2010 (NMCP, 2011), 28% in 2012 (NMCP and ICF, 2012); 33% in 2014 (NMCP and ICF, 2014); and 24% in 2017 (NMCP and ICF, 2018). There is a general decline, at national level, in the prevalence of under-five malaria in children in Malawi.

The Malawi government, together with other national and international stakeholders, have put malaria in the 2017-2022 Malawi Health Sector Strategic Plan as one of the priority problems to be addressed. The government has, in the draft National Malaria Strategic Plan 2017–2022, stressed the goal of scaling up malaria interventions to reduce morbidity and mortality by 50% in 2022 in order to ensure that there is progress towards achieving the national vision of a malaria-free Malawi (NMCP and ICF, 2018). To achieve such a goal, there is a need to know the specific geographical areas which still have the highest burden of malaria despite the passing of time. Determining the spatial and temporal distribution of malaria can help public health policymakers to identify those areas with high malaria burden in order to target resources towards those areas in order to reduce the burden (Lowe, Chirombo and Tompkins, 2013).

1.1.2 Bayesian Spatial-temporal analysis in Epidemiological studies

The analysis of disease prevalence, incidence, and mortality over space and time has considerably gained more focus due to the growing demand for reliable disease mapping (Torabi, 2017). Spatial and spatial-temporal modelling have increased in popularity in the last couple of decades in epidemiological research due to advances in computational methods and statistical methodology, and availability of spatially-referenced data (Arab, 2015). In Sub-

Saharan Africa, Spatial-temporal Bayesian models have been utilised in different studies especially on malaria and other infectious diseases. The advantage of using Bayesian methods is that they take into account prior information on the parameters used in the models. Furthermore, Bayesian methods offer a better way by allowing concurrent modelling of spatial-temporal autocorrelation while still estimating the usual fixed effects (Lawson, 2008; Manda *et al.*, 2015; Mariella and Tarantino, 2016).

1.2 Literature review

1.2.1 Factors associated with Malaria morbidity

Various studies have investigated various factors associated with malaria morbidity. Such factors are commonly grouped into Demographic factors, Socio-economic factors, environmental/geographical factors, climatic factors, and Vector-control factors (Dzinjalamala, 2009; Montosi *et al.*, 2012; Chirombo, Lowe and Kazembe, 2014; Shimaponda, 2015).

1.2.1.1 Demographic and socioeconomic factors

Demographic and socioeconomic factors that are commonly documented to be directly or indirectly associated with malaria are Age, place of residence, knowledge of malaria, education, and wealth index. Evidence from the literature shows that children who are under five years of age are the most vulnerable to malaria (Dzinjalamala, 2009; Chirombo, Lowe and Kazembe, 2014).

Place of residence and wealth index are related to malaria in the sense that less privileged families tend to live in rural areas than urban areas, and such families lack essential resources to help them curb malaria infections in their households. Studies that were done in countries like Tanzania (Njau *et al.*, 2006), Ethiopia (Graves *et al.*, 2009), and Indonesia (Dale *et al.*,

2005), showed that households with higher wealth index had lower malaria risk than the households with middle or low wealth index.

The Malawi Malaria Indicator Survey reports document the importance of mothers' education and knowledge of malaria as factors that may help in reducing malaria morbidity (NMCP and ICF, 2018). Education status is related to knowledge about malaria prevention and control among the population. In recent years, more emphasis has been put on the idea that improving knowledge about malaria in communities will lead to better use of interventions (Lowe, Chirombo and Tompkins, 2013).

1.2.1.2 Environmental/geographical factors

Among the notable environmental/geographical factors are geographical altitude and vegetation. High altitude areas tend to have lower malaria prevalence than low altitude areas (Kazembe *et al.*, 2006; Dzinjalamala, 2009; Alegana *et al.*, 2013). With respect to vegetation, deforestation may result in a change in the landscape such that gullies may be formed. Hence when rain comes, water tends to fill these gullies which results in having stagnant water. This provides a good breeding ground for mosquitoes, hence an increased likelihood of malaria infections.

1.2.1.3 Vector control factors

Vector control factors that are highly recommended by the World Health Organisation (WHO) are the use of insecticide-treated nets (ITNs) and Indoor Residual Spraying (IRS) (Mabaso, Sharp and Lengeler, 2004). The use of ITNs has been documented in various studies as the most effective method of reducing and preventing malaria in children who are less than five years of age (Lowe, Chirombo and Tompkins, 2013; Chirombo, Lowe and Kazembe, 2014).

1.2.1.4 Climatic factors

The most important climatic factors that are associated with malaria are temperature and rainfall (Montosi *et al.*, 2012). Mosquitoes tend to breed faster in temperatures that are 21° Celsius or higher, hence resulting in increased malaria transmission rate (Dzinjalamala, 2009). In cooler temperatures, there is less mosquito breeding thus resulting in less malaria transmission rate. Concerning rainfall, various studies have shown that areas with frequent rainfall have higher malaria transmissions. Rain leaves pools of stagnant water hence providing a more conducive breeding site for mosquitoes (Chitunhu and Musenge, 2016).

From above description of various demographic, socioeconomic, environmental, geographical, climatic, and vector-control factors associated with malaria, **Figure 1.1** below shows a conceptual framework that illustrates the possible associations between malaria episodes in households and these factors. The variables in the framework are shown as reviewed above in the literature. For this study, the variables: *Enhanced Vegetation Index, Indoor Residual Spraying*, and all the climatic factors variables were not included in the analysis as these were not available in the actual data for all the four MMISs.

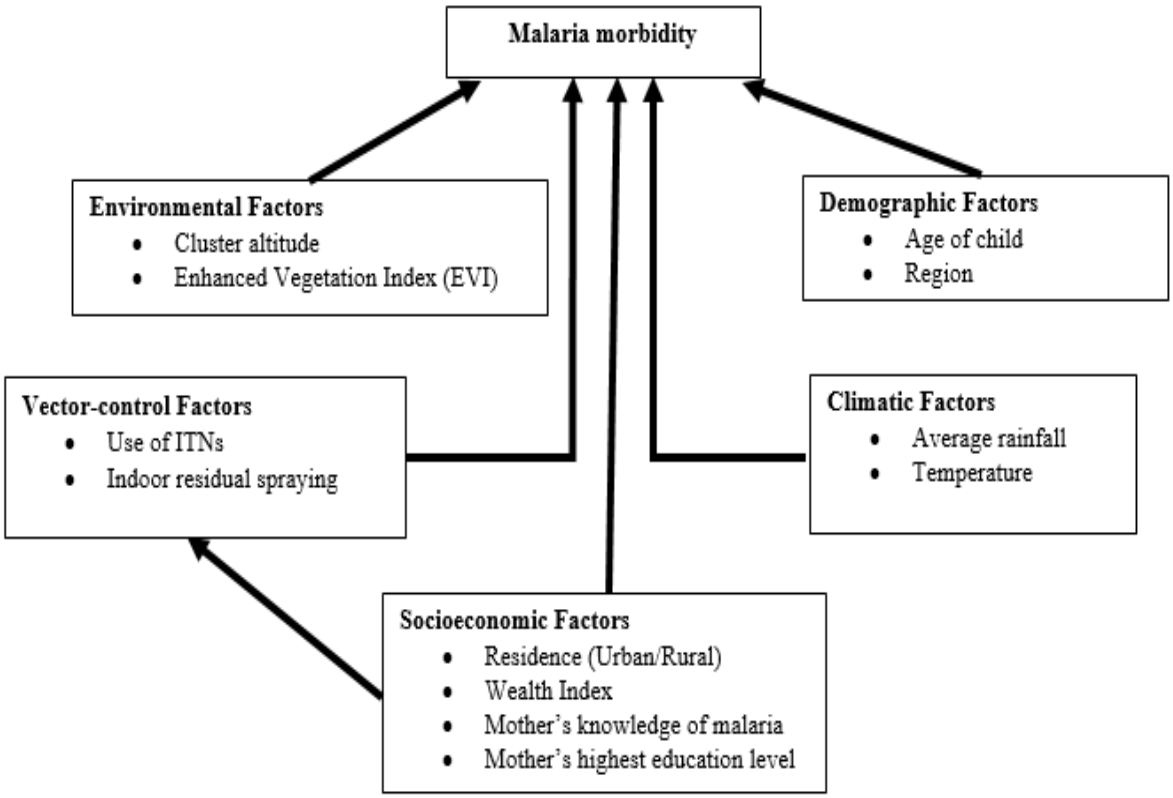


Figure 1-1: A conceptual framework of factors associated with the Malaria among under-five children

1.2.2 Spatial-temporal modelling in Epidemiological studies

As a discipline, Epidemiology incorporates the inter-links between the person, place and time in the distribution of diseases in a population (Musenge, Vounatsou and Kahn, 2011). There have been advances in computational methods and statistical methodology, and the availability of spatially-referenced data. As such, with such advances, spatial and spatial-temporal modelling have increased in popularity in the last couple of decades in epidemiological research (Arab, 2015).

Spatial and spatial-temporal disease mapping models are commonly used to estimate disease risk, in determining geographical areas and periods with high risk, and time trends of prevalence or incidence of a disease. Some of the examples of studies on diseases that have applied spatial and spatial-temporal methods in modelling disease patterns include:

Malnutrition in children (Wambui and Musenge, 2019), HIV/TB child mortality (Musenge, 2014), lung cancer (Knorr-Held and Besag, 1998), Anaemia (Soares Magalhães *et al.*, 2013), and Malaria (Lowe, Chirombo and Tompkins, 2013; Chirombo, Lowe and Kazembe, 2014; Noor *et al.*, 2014), just to mention a few.

1.2.3 Spatial-temporal modelling of Malaria in Malawi

Lowe *et.al* (2013) modelled the association between several climatic, geographic and socio-economic determinants and malaria incidence at district-level in Malawi using an age-stratified spatial-temporal dataset of malaria cases from July 2004 to June 2011. Under a hierarchical Bayesian framework that accounted for spatial correlation, they used a Generalized Linear Mixed Model (GLMM) that included both structured and unstructured spatial and temporal random effects. The study conclusion was that it is important to take into account spatial and temporal variations and correlation between districts.

Chipeta *et al.*, (2019) used a spatio-temporal Geostatistical model to predict annual malaria risk between 2010 and 2017 for children aged 2–10 years. Parameter estimation was done using the Monte Carlo maximum likelihood methods. The study found out that Malawi has made substantial progress in reducing the prevalence of malaria, but that the declining prevalence was not equal across the country.

To the best knowledge of the researcher of this current study, the studies done by Lowe *et al* (2013) and Chipeta *et.al* (2019) are the two only studies done in Malawi that utilised spatial-temporal methods in modelling malaria morbidity. There is no study in Malawi (thus far) that has specifically examined the **malaria morbidity in under-five children** while taking into account spatial-temporal variations at smaller geographical areas like cluster level.

In Malawi, most malaria studies have only utilised the spatial analyses only (like those done by Kazembe *et al.*, 2006; Chirombo, Lowe and Kazembe, 2014; Chitunhu and Musenge, 2016). This study sought to supplement to the current knowledge existing as per the studies done by Lowe et al and Chipeta et.al. It is, therefore, the aim of this study to model spatial-temporal variations of malaria morbidity in under-five children in Malawi for the years 2010, 2012, 2014, and 2017. This study will focus on cluster level geographical areas.

1.3 Problem Statement

Much as the malaria prevalence at national level has declined from 43% in 2010 to 24% as of 2017, under-five malaria is still high and remains one of the major public health problems (Chipeta *et al.*, 2019). This shows that, Malawi as a nation, is still far from making the national vision to have a Malaria-Free Malawi a reality.

Although most Epidemiological studies have used spatial analysis in determining and mapping malaria hotspots in Malawi, most of these studies of under-five malaria morbidity have focussed at one point in time; these studies do not show how the spatial patterns of malaria morbidity are changing over time. Little is known whether, over time, there has been an increase or decrease in under-five malaria morbidity in such malaria hotspot areas. There is need to ascertain which specific areas are still contributing to the high malaria burden.

1.4 Justification of the study

There is a need to understand how under-five malaria morbidity is changing in space and over time in order to know which specific geographical areas are hotspots or cold-spots of malaria morbidity. This implies that there is also a need to consider models that incorporate spatial and

temporal random effects in modelling malaria morbidity, while also taking into account the effects of particular risk factors. A better understanding of the spatial-temporal variations of malaria morbidity in under-five children in Malawi is essential for policy makers and policy implementers to think and develop significant interventions to help lessen under-five malaria morbidity by prioritising those specific geographical areas which still have high malaria burden despite the passing of time. The outcomes of this study, therefore, may help contribute to such a better understanding of variations of malaria morbidity over time by adding to the already existing knowledge pool on this matter.

1.5 Research question

What are the spatial-temporal effects on malaria morbidity in Malawian under-five children between 2010, 2012, 2014 and 2017?

1.6 Aim of the study

This study aimed at determining the spatial-temporal effects on malaria morbidity in under-five children in Malawi for the years 2010, 2012, 2014, and 2017.

1.7 Objectives of the study

1) To describe the Malaria prevalence trends and the spatial distribution and clustering of Malaria in under-five children for the years 2010, 2012, 2014, and 2017.
2) To determine the demographic, socio-economic, environmental, and vector-control factors associated with malaria morbidity in all the survey years of 2010, 2012, 2014, and 2017

3) To determine the spatial-temporal random effects on under-five malaria morbidity after adjusting for demographic, socio-economic, environmental, and vector-control factors

CHAPTER 2: METHODOLOGY

This chapter presents the methodological approach of this study. It presents: The study area and setting, the study design, the study population, sample size and power calculation, data sources, study variables, data management, data analysis methods, and the ethical considerations.

2.1 Study Area and Setting

The country of study was Malawi. Geographically, Malawi is a landlocked country south of the equator in sub-Saharan Africa. Zambia borders it in the west, Mozambique in the east, south and southwest, and Tanzania in the north. It has a total land area of 118,484 km^2. Administratively, the country is divided into three regions, namely: Northern region, Central region, and Southern region. The country has 28 districts (as shown in **Figure 2.1** below) that are distributed as follows: 6 districts in the Northern region, 9 districts in the Central region, and 13 districts in the South region (NSO and ICF, 2011). For this study, Likoma district (in the northern region), and Neno district (in the Southern region) were not included in the analysis, as these districts had no data except for the year 2017 only.

Economically, Malawi is one of the poorest countries in the world with a per capita GDP of US$350 (World Bank, 2018).



Figure 2- 1: Map of Malawi (source: MDHS 2015-16 report)

2.2 Study Design

This study was a secondary data analysis based on cross-sectional Malawi Malaria Indicator Surveys (MMIS) that were done in 2010, 2012, 2014, and 2017.

2.3 Study Population

The study population consisted of all the under-five (6-59 month-old) children in all 2010, 2012, 2014, and 2017 MMIS sampled households in the sampled clusters of the Malawian districts.

The eligibility criteria were as follows:

a) Inclusion criteria

- All children between 6 and 59 months' old
- All children who had a malaria test result as either positive or negative

b) Exclusion criteria

- All children who were not in the 6-59-month age category, or those whose ages were not recorded
- All children whose malaria test result was not recorded or had an indeterminate malaria test result
-

2.4 Sampling Design, Sample size and Power Calculation

The original MMISs utilised a two-stage sampling design. The first stage involved the selection of sample points (clusters). For 2010, 2012 and 2014 surveys, 140 clusters were selected; while for the 2017 survey, 150 clusters were selected. In the second stage, a systematic sampling approach was used to select 25 households per EA. For 2010 and 2012 MIS, 3500 households were sampled; for 2014 MIS, 3501 households were sampled; and for 2017 MIS, 3750 households were sampled. All children in the selected households were tested for malaria, with consent from parents/guardians.

Since this study used all the available data of the sample sizes from the 2010, 2012, 2014, and 2017 MMIS, sample size calculation was not done. In terms of determining the Statistical Power of this current study, separate power estimations were done for each survey year using 0.05 alpha-level together with the respective sample sizes (1758 in 2010, 2112 in 2012, 1928 in 2014, and 2305 in 2017) and malaria prevalence for each year (43% in 2010, 27% in 2012, 33% in 2014, and 24% in 2017). The power estimations were done in Stata 15 using the commands:

For 2010: `power oneproportion 0.50 0.43, test(binomial) n(1758) onesided`

For 2012: `power oneproportion 0.50 0.27, test(binomial) n(2112) onesided`

For 2014: `power oneproportion 0.50 0.33, test(binomial) n(1928) onesided`

For 2017: `power oneproportion 0.50 0.24, test(binomial) n(2305) onesided`

The power estimations for each year specific was more that 90%. The table below shows the details of the power estimations.

Year	Sample size	Prevalence (%)	Alpha-level	Power Estimation
2010	1758	43%	0.05	100%
2012	2112	27%	0.05	100%
2014	1928	33%	0.05	100%
2017	2305	24%	0.05	100%

2.5 Data Sources

2.5.1 Malaria data

This study used data from all the above-mentioned Malawi Malaria Indicators Surveys of 2010, 2012, 2014, and 2017. The 2010 Malaria data was obtained from the Malawi Ministry of Health upon requesting for access and use the data. The 2012, 2014, and 2017 Malaria data were

obtained from the Measure Demographic and Health Survey (DHS) program website (https://dhsprogram.com/data/available-datasets.cfm) after requesting for permission to access, download, and use the data.

2.5.2 Shape Files

The Shape files used in this study were obtained freely from the DIVA GIS website (https://www.diva-gis.org/gdata). Of main interest, *admin-level 2* (cluster level) shape file was used.

2.6 Study Variables

2.6.1 Outcome variable

The outcome from the original Malaria Indicator surveys was a binary outcome; whether a child had malaria or not (using the ***microscopic Malaria blood smears test*** as this is documented elsewhere that it provides a more accurate result (Zgambo, Mbakaya and Kalembo, 2017)). For this study, depending on the study objective, the outcome variables were as follows:

- For the second objective of this study, the outcome variable was whether a child had malaria or not (for individual child level analysis).
- For the third objective (the main aim of this study), the outcome variable was the number of under-five children who tested positive for malaria aggregated at cluster level (for cluster level analysis). This variable was a count, a non-negative integer. This variable was derived from the binary malaria outcome variable of whether a child was malaria positive or not.

2.6.2 Independent variables

With respect to the reviewed literature above, as well as the available variables in the datasets, the following variables were considered in this study as shown in **Table 2.1** below:

Table 2. 1: List of independent variables used in this study

Variable name	Variable Type	As defined in original survey	As defined in this study
Age of child	Numeric	*In months*	*Same as in original survey*
Region	Categorical	• *North* • *Central* • *South*	*Same as in original survey*
Place of Residence	Categorical	• *Urban* • *Rural*	*Same as in original survey*
Wealth Index	Categorical	• *Poorest* • *Poorer* • *Middle* • *Richer* • *Richest*	• *Poor* • *Middle* • *Rich*
Mother's highest education level	Categorical	• *No education* • *Primary* • *Secondary* • *Higher*	• *None* • *Primary* • *Secondary/higher*
Mother heard about Malaria	Categorical	• *No* • *Yes*	*Same as in original survey*
Cluster altitude	Numeric	*In metres*	*Per 100 metres*
Have ITN**	Categorical	• *No* • *Yes*	• *No net* • *Yes (not ITN)* • *Yes (ITN)*

**for this study, the ITN variable was combined with whether the child had a net or not, which resulted into the three categories

2.7 Data Management

2.7.1 Malaria data

In the original 2010, 2012, 2014, and 2017 MMIS, Censuses and Surveys Processing (CSPro) system was used for data editing, cleaning, weighting, checking any inconsistencies, and tabulation. All cleaned data were then stored in CSPro file format and other formats like in Stata format.

For this study, the datasets were checked for any inconsistencies, so as to make sure all the categorical variables had the same coding. All the Data merging, cleaning, and recoding of

categorical variables were done in Stata™ version 15. The cleaned datasets (individual child level data for individual level analysis as per Objective 2, and cluster-aggregated data for cluster level analysis as per objective 3) were then saved in Stata format.

2.7.2 Merging Malaria data with Shape files

Each of the respective survey-year dataset was merged, using ArcGIS software, with the Malawi admin-level two shape file by using the functions: *Join, Union, and intersect.* This was done by linking the cluster variable as well as the longitude and latitude variables in the malaria datasets with the particular cluster numbering format in the shape file. Maps showing malaria spatial distribution were then drawn after the merging process.

2.8 Data Analysis

2.8.1 Trends, Spatial Distribution, and Spatial Autocorrelation and Clustering of Malaria

2.8.1.1 Malaria Prevalence Trend analysis and Statistics

For each district, malaria prevalence was calculated as the number of malaria positive divided by the number of children who tested for malaria, for each year. To explore the trend of malaria prevalence at district level, line graphs were plotted using Stata™ version 15. The line graphs were plotted based on four time points representing the four survey years (2010, 2012, 2014, and 2017).

2.8.1.2 Spatial Distribution of Malaria

Using ArcGIS software, choropleth maps were used to show the descriptive spatial distribution of under-five malaria prevalence across the sampled clusters of the study area (Malawi), respective of each of the four survey years.

2.8.1.3 Spatial Autocorrelation and Clustering of Malaria

To determine the spatial autocorrelation (presence or absence of clustering) of under-five malaria, the Global Moran's I index, Getis-Ord Gi* Hotspot analysis, and the Kulldorff spatial scan statistic were used.

a) Global Moran's I index

This was used determine whether, in overall, under-five malaria was clustered, dispersed or randomly distributed across the study area. The Moran's I index formula is as follows:

$$I = \frac{n}{W} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \quad \text{.....Equation 2.1}$$

Where:

- n is the total number of geographical areas
- $w_{i,j} = \frac{1}{D_{i,j}}$is the is the spatial weight between locations i and j, $D_{i,j}$is the distance between locations i and j
- $W = \sum_{i=1}^{n} \sum_{j \neq j}^{n} w_{i,j}$ is the aggregate of all spatial weights
- $z_i = x_i - \bar{x}$, the deviation of malaria prevalence at location i from its mean
- z_j the deviation malaria prevalence from a sub-location j

The Moran's I index statistic gives a range of values between -1 and 1 (Moran, 1950). This statistic was interpreted as follows (Ord and Getis, 2001):

i) A positive value close to 1 indicates positive autocorrelation which implied that the malaria distribution was clustered, that is, geographical areas with similar values of malaria tended to cluster together.

ii) A negative value close to -1 indicates negative spatial autocorrelation which implied dispersed malaria distribution, that is, geographical areas with similar values of Malaria tended to scatter across the study area.

iii) A value of 0 (or close to 0) indicates no spatial autocorrelation implying random distribution of malaria in the study area.

The null hypothesis was that there was no spatial autocorrelation of malaria, while the alternative hypothesis is that there is spatial autocorrelation of malaria. If the p-value associated with the Moran's I value was less than a significance level of 0.05, the null hypothesis was rejected and the conclusion was that there was evidence of clustering in the malaria distribution in the study area. This was done in ArcGIS ArcMap version 10.5.

b) *Getis-Ord Gi* Hotspot analysis*

This was used to explore the degree of concentration of high values (hot spots) and low values (cold spots) of malaria so as to ascertain which clusters were statistically significant hotspots or cold spots of malaria. For a given geographical area *i*, the G_i^* statistic gives a z-score using the following mathematical formula (Ord and Getis, 1995):

$$G_i^* = \frac{\sum_{j=i}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2}{n-1}}} \quad \text{..... } \textbf{\textit{Equation 2.2}}$$

Where:

- n is the total number of geographical areas
- $w_{i,j} = \frac{1}{D_{i,j}}$is the is the spatial weight between locations i and j, $D_{i,j}$is the distance between locations i and j
- x_j is the Malaria prevalence value at location j
- $\bar{X}$ is the mean of malaria prevalence,
- S is the standard deviation of malaria prevalence

Since this is a z-score, then for each particular geographical area, this statistic gives a value which denotes whether such an area is a significant hot spot or cold spot relative to its neighbouring areas (Ord and Getis, 1995).

i) A statistically significant (p-value<0.05) positive z-score implied clustering of high values of malaria prevalence.

ii) A statistically significant (p-value<0.05) negative z-score implied clustering of low values of malaria prevalence.

iii) A z-score close zero (p-value>0.05) implied no significant clustering.

This Hotspot/Coldspot analysis was also done in ArcGIS ArcMap version 10.5.

c) Kulldorff spatial scan statistic

For purely spatial clustering, this statistic, in SatScan™ version 9.6 software, uses geographical area-based aggregated data to estimate spatial clustering of the outcome of interest. In this case, SaTScan™ uses a discrete Poisson-based model, where the number of events in a geographical location is Poisson-distributed, according to a known underlying population at risk (Kulldorff *et al.*, 2006). The software then uses the statistic to impose a circular window/circle on the study area map and calculates the likelihood of observing the events inside and outside the study area. This method creates an infinite number of distinct geographical circles with different sets of neighbouring data locations within them. The window with the maximum likelihood is termed as "the most likely cluster", that is, the cluster least likely to be due to chance (Kulldorff *et al.*, 2006).

For this study, SatScan™ software was used to ascertain local clustering of malaria by drawing "malaria clusters/windows" on the study area map. The identified clusters were then classified into primary and secondary so long as they had a significant p-value. The circular windows with insignificant p-value were classified as insignificant. Maps were then drawn in ArcGIS ArcMap version 10.5.

2.8.2 Factors associated with malaria morbidity in 2010, 2012, 2014, 2017

2.8.2.1 Descriptive Statistics for the Demographic, Socio-economic, Environmental, and Vector-control factors

Taking into consideration that this study used original MMIS which involved a two-stage complex survey design, descriptive statistics for this study took into account the sampling weights. As such, projected sample totals for each survey year were calculated.

For (numeric) continuous variables, summary descriptive statistics were done by using weighted means with their respective confidence intervals. For categorical variables, weighted frequencies and the associated weighted percentages (proportions) were done. These descriptive statistics were done respective of each year of the MMIS, and then were sub-categorized according to malaria outcome.

2.8.2.2 Demographic, Socio-economic, Environmental, and Vector-control factors associated with Malaria in 2010, 2012, 2014, and 2017

In order to achieve Objective 2 of this research, Multilevel Logistic regression model in Stata version 15 was used which accounted for cluster random effects for each of the four survey years. This was done at individual child level data because the outcome variable of interest was a binary outcome: whether a child had malaria or not. Mathematically, the model was defined as follows:

Let

$$X_{ic} = \begin{cases} 1 & \text{had malaria} \\ 0 & \text{no malaria} \end{cases}$$

for child i in cluster c. Then

$$X_{ic} \sim Bernoulli\ (p_{ic}),$$

and the likelihood function follows a Binomial distribution given as:

$$L\left(X_{ict}=x_{ic}\mid p_{ic}\right)=\binom{n_{ic}}{x_{ic}}p_{ic}^{\,x_{ic}}\left(1-p_{ic}\right)^{n_{ic}-x_{ic}}\ \ldots\ldots\ldots\ldots\ldots\ldots\ \textbf{\textit{Equation 2.3}}$$

where

- n_{ic} is the number of children tested for malaria in a household in cluster *c*
- p_{ic} the probability of a child having malaria.

The Multilevel logistic regression model is as follows:

$$log_e\left(\frac{p_{ic}}{1-p_{ic}}\right)=\underset{\sim}{Z}_{ic}^{T}\underset{\sim}{\beta}+\phi_c+\varepsilon_{ic},\ \ldots\ldots\ldots\ldots\ \textbf{\textit{Equation 2.4}}$$

Where

- $\underset{\sim}{Z}_{ic}^{T}$ is a vector of covariates
- $\underset{\sim}{\beta}$ is a vector of coefficients of the covariates
- ϕ_c is the random cluster effects
- ε_{ic} are unexplained errors

The multilevel modelling approach used here takes into account that people who lived in the same survey cluster may have had some common characteristics, hence may not be independent of each other.

Since this research used survey data, sampling weights were accounted for in the Multilevel Logistic modelling so as to take into consideration the two-stage complex survey design of the original MMIS.

2.8.3 Spatial-Temporal Effects on Malaria Morbidity

In order to achieve Objective 3, the data were first aggregated at cluster level with respect to each year. Numeric variables were aggregated as means, while the categorical variables were

aggregated as proportions per each category. Data aggregation was done in Stata version 15. The outcome variable of interest, in this case, was the number of malaria positive cases (a count variable) in the survey clusters. **Table 2.2** below gives a summary of the mean and variance of the malaria counts.

Table 2. 2: Mean and Variance of Malaria counts per year

Year	Mean count of Malaria	Variance
2010	5.83	26.18
2012	5.14	18.19
2014	5.46	30.85
2017	4.48	14.83
Overall	**5.26**	**22.75**

Since the outcome was malaria counts and the variance of the counts was higher than the mean counts for each year and overall (hence implying over-dispersed data), the Negative Binomial (NB) spatial-temporal model was used (Musenge, Vounatsou and Kahn, 2011). The Bayesian modelling approached was used than the Maximum Likelihood Estimation (MLE) approach because the Bayesian methods have the capacity to incorporate both the structured and the unstructured random effects, plus the fact that prior information can be used (Lawson, 2008; Chitunhu and Musenge, 2016; Wambui and Musenge, 2019).

2.8.3.1 Bayesian Spatial-Temporal Negative Binomial Model Specification

Let Y_{ct} be the number of children who were tested positive for malaria, and μ_{ct} be the mean number of children tested positive for malaria, in cluster c at time point t , for $c = 1, 2, 3, ..., n_t$; $t = 1$ (2010), 2 (2012), 3 (2014), 4 (2017).

Then:

$$Y_{ct} \sim NegBin\,(\mu_{ct}, r)$$

And:

$$P(Y = y_{ct} \mid \mu_{ct}, r) = \binom{y_{ct} + r - 1}{r - 1} \left(\frac{\mu_{ct}}{r + \mu_{ct}} \right)^{y_{ct}} \left(\frac{r}{r + \mu_{ct}} \right)^{r} \quad \text{..... Equation 2.5}$$

$$= \frac{\Gamma(y_{ct} + r)}{\Gamma(y_{ct} + 1)\Gamma(r)} \left(\frac{\mu_{ct}}{r + \mu_{ct}} \right)^{y_{ct}} \left(\frac{r}{r + \mu_{ct}} \right)^{r} \quad \text{..... Equation 2.6}$$

Where r is the over-dispersion parameter.

The spatial-temporal Generalized Linear Mixed Model (GLMM) with unstructured and structured spatial effects as well as temporal effects to estimate the relative risk μ_{ct} was fitted as:

$$log_e(\mu_{ct}) = log_e(N_{ct}) + \underset{\sim}{X}_{ct}^{T}\underset{\sim}{\beta} + f(age_{ct}) + \underset{\sim}{\phi}_c + \underset{\sim}{\varphi}_c + \underset{\sim}{\psi}_t + \underset{\sim}{\varepsilon}_{ct} \text{ ..... Equation 2.7}$$

Where:

- N_{ct} is the offset variable representing the number of children (Chitunhu and Musenge, 2016) in a particular cluster at time *t*,
- $\underset{\sim}{X}_{ct}^{T}$ is a vector of covariates,
- $\underset{\sim}{\beta}$ is a vector of coefficients of the covariates,
- $f(age_{ct})$ represents a function of age (in months) modelled with a random walk model of order 2 (Lindgren and Rue, 2008). The age variable was assumed to have a non-linear effect on the Malaria outcome.
- $\underset{\sim}{\phi}_c$ is the unstructured random effects,
- $\underset{\sim}{\varphi}_c$ is the structured random effects,
- $\underset{\sim}{\psi}_t$ is the temporal random effects,
- $\underset{\sim}{\varepsilon}_{ct}$ are unexplained errors.

Bayesian spatial-temporal modelling approach was used to investigate structured and unstructured spatial random effects as well as temporal random effects on malaria morbidity after adjusting for other covariates of interest (Brezger, Kneib and Lang, 2012). All the

posterior estimates were displayed with their corresponding 95% credible intervals. The hierarchical Bayesian framework for the spatial-temporal model was formulated as follows:

$$Posterior(parameters \mid data) \propto Likelihood(data \mid parameter) \times priors(parameters)$$

The Posterior likelihood for the Malaria data is:

$$L\left\{\Theta \equiv \left(\underset{\sim}{\beta}, \underset{\sim}{\phi_c}, \underset{\sim}{\varphi_c}, \underset{\sim}{\psi_t}\right) \mid y_{ct}\right\} = \prod_{c=1}^{n_t} \prod_{t=1}^{4} P\left(y_{ct} \mid \Theta\right) \times P\left(\underset{\sim}{\beta}, \underset{\sim}{\phi_c}, \underset{\sim}{\varphi_c}, \underset{\sim}{\psi_t}\right) \ldots\ldots\ldots \textbf{\textit{Equation 2.8}}$$

And the full conditional for this model is:

$$P\left(\Theta \mid y_{ct}\right) \propto L\left(y_{ct} \mid \Theta\right) \times P\left(\underset{\sim}{\beta}\right) \times P\left(\underset{\sim}{\phi_c} \mid \tau_h\right) \times P\left(\underset{\sim}{\varphi_c} \mid \tau_j\right) \times P\left(\underset{\sim}{\psi_t} \mid \tau_k\right) \ldots\ldots\ldots\ldots \textbf{\textit{Equation 2.9}}$$

The prior distributions were specified as follows:

- The Beta coefficients $\underset{\sim}{\beta} \overset{i.i.d}{\sim} N\left(\mu_\beta, \sigma_\beta^2\right)$
- The unstructured random effects $\underset{\sim}{\phi_c} \overset{i.i.d}{\sim} N\left(0, \frac{1}{\tau_h}\right)$, and $\tau_h \sim Gamma\left(\alpha_h, \beta_h\right); \alpha_h > 0, \beta_h > 0$
- The structured random effects $\underset{\sim}{\varphi_c}$ was modeled as a Conditional Auto Regressive (CAR) prior (Banerjee S, Carlin B.P. and Gelfand A.E., 2004; Lawson, 2008):

$$\underset{\sim}{\varphi_c} \sim CAR\left(\tau_j\right) \sim N\left(\frac{\varphi_k}{\bar{\varphi}_k}, \frac{1}{\tau_j m_k}\right) \ldots\ldots\ldots \textbf{\textit{Equation 2.10}}$$

  And the likelihood of the neighbouring sub-locations is given as (Chitunhu and Musenge, 2016):

$$p\left(\varphi_k \mid \tau_j\right) \propto \exp\left\{-\frac{\tau_j}{2} \sum_{k=1}^{n_t} w_{kk'} \left(\varphi_{k'} - \varphi_k\right)^2\right\} \ldots\ldots\ldots\ldots \textbf{\textit{Equation 2.11}}$$

  where $w_{kk'}$ denotes the adjacency matrix and kk' shows that the sub location k' is a neighbour of sub location k and m_k is the number of neigbours for sublocation k. $\tau_j \sim Gamma\left(\alpha_j, \beta_j\right); \alpha_j > 0, \beta_j > 0$

- The temporal effects $\underset{\sim}{\psi}_t$ were modeled as first order Autoregressive (Mabaso *et al.*, 2006; Blangiardo *et al.*, 2013) model, given as

$$\psi_t = f_{temp}(x_t) \sim AR(1), \text{ with first term prior } x_t = \rho x_{t-1} + \varepsilon_t,\ x_t \sim N\left(0, \frac{1}{\tau_e(1-\rho^2)}\right)$$

The modelling of a combination of the unstructured and the structured random effects priors (hence a convolution prior) was done using the Besag-York-Mollie (BYM) model (Rue, Martino and Chopin, 2009).

This study was fully Bayesian, and as such, inference was made by sampling from the posterior distribution. To estimate the posterior marginal distribution of the parameters of interest, the Integrated Nested Laplace Approximation (INLA) method was used rather than the more common Markov Chain Monte Carlo (MCMC) approach. INLA is capable of approximating the posterior distribution with high accuracy at a much faster computational rate than MCMC (Blangiardo *et al.*, 2013).

2.8.3.2 Model selection and goodness of fit

For this research, 7 models were built:

a) A multilevel model (in Stata 15) with unstructured cluster random effects only
b) A Bayesian model with unstructured spatial random effects only
c) A Bayesian model with structured spatial random effects only
d) A Bayesian model with both the unstructured and structured spatial random effects
e) A Bayesian Spatial-temporal model with ustructured and structured random effects. The temporal random effects were modeled as *i.i.d*

f) A Bayesian Spatial-temporal model with ustructured and structured random effects. The temporal random effects were modeled as *AR1*

g) A Bayesian Spatial-temporal model with ustructured and structured random effects. The temporal random effects were modeled as *AR1* and there was an interaction term of the spatial and the temporal random effects

For model comparison, the Deviance Information Criterion (DIC) was used. DIC is defined as the sum of the posterior mean of the deviance $D(\theta \mid y)$ and the effective number of parameters p_D: $DIC = E\left[D(\theta \mid y)\right] + p_D$, and p_D is the difference in the posterior mean deviance and the deviance evaluated at the posterior mean of the parameters: $p_D = E\left[D(\theta \mid y)\right] - D\left(E\left[\theta \mid y\right]\right)$.

The interpretations is that if the effective number of parameters is smaller, then the model is more parsimonious, and the smaller the DIC the better the model fit. Most parsimonious model may not always the best model (Musenge *et al.*, 2013).

The convergence of the chosen model was shown using posterior mean density plots for the fixed effects.

Except for the Multilevel modelling done in Stata 15, the rest of the model building and running was done in RStudio using the INLA package.

2.9 Ethical Considerations

This research involved secondary data analysis which used data from the original 2010, 2012, 2014, and 2017 MMIS. For the 2010 MMIS, since the data was not available on the Measure DHS website, a separate application for permission to access and use the data was made to the Community Health Science Unit (CHSU) of the Malawi Ministry of Health; and the approval was granted as per the approval letter shown in **Appendix 4**. Approval to use the 2012, 2014, and 2017 MMIS data was obtained from the measure DHS website (letter of approval shown in **Appendix 5**). Prior to the commencement of this research, an application for ethics approval

for this study's protocol was presented to the University of the Witwatersrand's Human Research Ethics Committee. Ethics approval was then granted, as shown by a copy of the clearance certificate number **M181003** attached in **Appendix 3**.

In the original surveys, the data was anonymised such that individual participants' names or any household identifying information were kept anonymous. For this study, confidentiality and participant anonymity were ensured by keeping the data as extracted from the original data sources, and storing the data in the researcher's laptop with a secure password.

CHAPTER 3: RESULTS

This chapter presents the results of the analysis done in this study. The first section (3.1) presents the trends, spatial distribution, spatial autocorrelation and clustering of Malaria. The second section (3.2) presents the factors associated with malaria morbidity for 2010, 2012, 2014, and 2017. The last section (3.3) presents the spatial-temporal random effects on Malaria by showing results from various models built.

3.1 Trends, Spatial Distribution, Spatial Autocorrelation and Clustering of Malaria

3.1.1 Trends of Malaria Prevalence

At national level, the Malaria prevalence (after accounting for sampling weights) was 43.1% in 2010; 27.7% in 2012, 33.3% in 2014; and 24.6% in 2017. The malaria prevalence trends at regional level were as follows. For Northern Region, the prevalence trend was: 21.9% (2010), 19.8% (2012), 28.5% (2014), and 11.8% (2017). For Central Region, the prevalence trend was: 50.3% (2010), 34.3% (2012), 36.3% (2014), and 26.1% (2017). For Southern Region, the prevalence trend was as follows: 42.0% (2010), 23.9% (2012), 32.5% (2014), and 26.1% (2017).

At district level (Likoma, Mwanza, Neno excluded from the analysis), there was an overall general decreasing trend of Malaria in almost all districts as shown in **Figure 3-1**. It should be noted, however, that malaria prevalence in almost all Northern Region districts (except for Nkhata-bay) was generally lower (mostly below 20%) for all the survey years, as compared to the other districts in the Central and Southern Regions. Some districts in the Central Region (Dowa, Kasungu, Lilongwe) and in the Southern Region (Machinga, Mangochi) had a slower decreasing trend of malaria prevalence (**Figure 3-1b** and **Figure 3-1c**); while some districts in the Central Region (Nkhotakota) and Southern Region (Chiradzulu, Mulanje, and Zomba) had

initially a decreasing trend then an increasing trend of malaria prevalence (**Figure 3-1b** and **Figure 3-1c**).

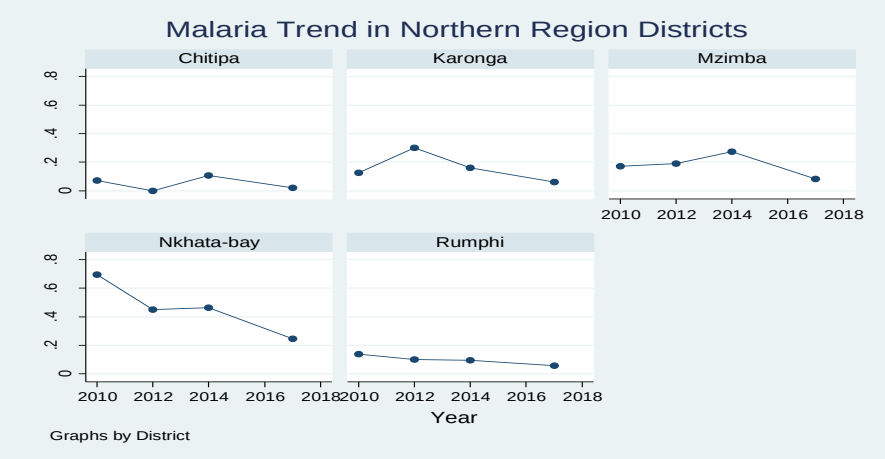


Figure 3-1a: Northern Region districts

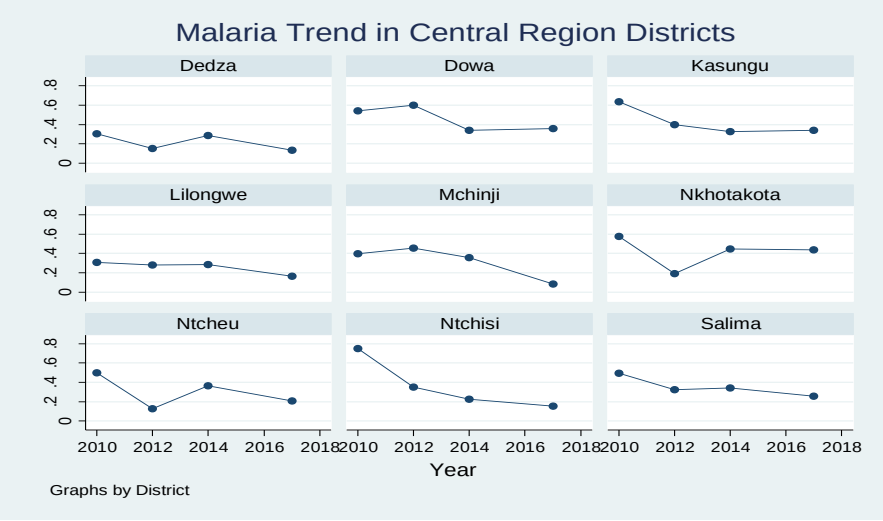


Figure 3-1b: Central Region districts

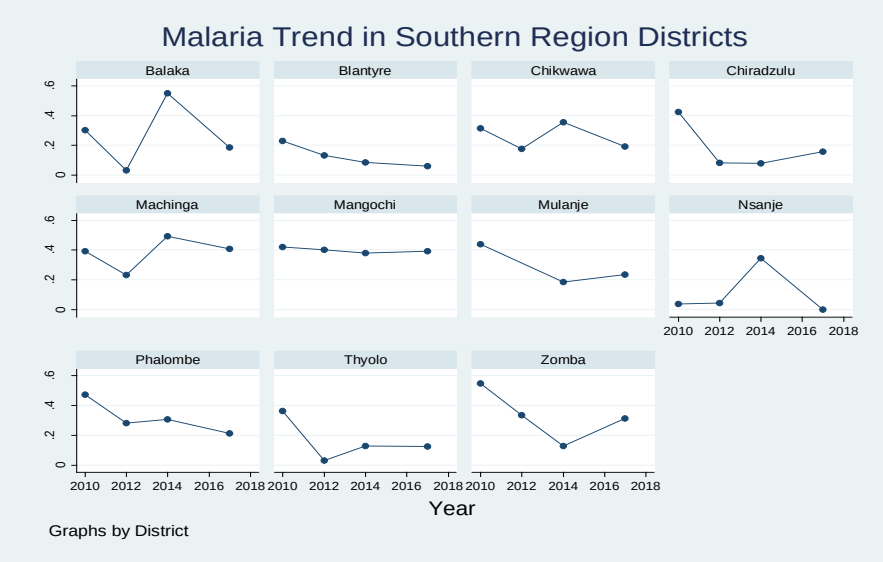


Figure 3-1c: Southern Region districts

Figure 3-1: Malaria prevalence trends for Malawian districts, by region

3.1.2 Spatial Distribution of Malaria

Figure 3-2 displays the spatial distribution of malaria prevalence at cluster level for the 2010, 2012, 2014, and 2017 survey years. The red dots represent the malaria prevalence; the smaller the dot, the lower the prevalence while the bigger the dot, the higher the prevalence. As the figure shows, almost all clusters in the Northern region clusters had a generally lower malaria prevalence as compared to most clusters in the Central region and Southern region, for all the survey years. The figure further shows that the Central region is the most affected region in as far as malaria is concerned, regardless of the survey year.

Malaria prevalence in 2010 (**Figure 3-2a**) was much higher especially in most of the Central and Southern region clusters (mostly ranging from 50% to 80%) as compared to the 2012 malaria prevalence (**Figure 3-2b**, mostly between 20% and 50%), 2014 malaria prevalence (**Figure 3-2c**, mostly between 20% and 50% with a few clusters above 75%), and 2017 malaria prevalence (**Figure 3-2d**, mostly around 20% with about 5 clusters above 50%). This then shows that, at cluster level, there was a general decreasing trend of malaria prevalence across the survey years.

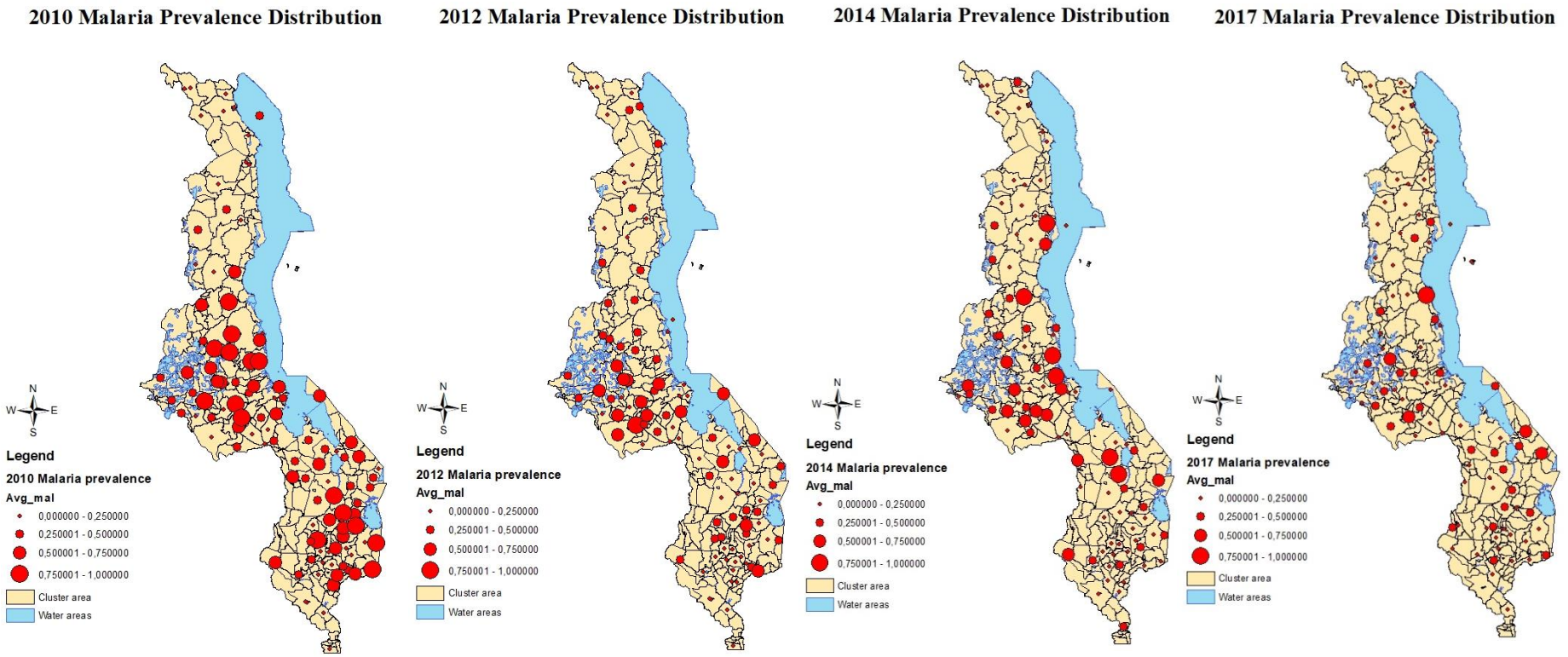


Figure 3-2a: 2010 Malaria prevalence **Figure 3-2b**: 2012 Malaria prevalence **Figure 3-2c**: 2014 Malaria prevalence **Figure 3-2d**: 2017 Malaria prevalence

Figure 3-2: Spatial distribution of malaria at cluster level for 2010, 2012, 2014, and 2017

3.1.3 Spatial Autocorrelation of Malaria

3.1.3.1 Global Moran's I index

The Spatial Autocorrelation (Global Moran's I) tool in ArcGIS software was used to ascertain if there was significant/insignificant global spatial clustering of malaria for the four years in Malawi. **Table 3.1** gives summary statistics of the results of the Global Moran's I spatial autocorrelation analysis. As the table shows, significant positive Moran's I spatial autocorrelation indices were observed only in 2010 (I=0.044; p=0.021) and 2012 (I=0.074; p<0.001), hence signifying that there was spatial clustering of malaria prevalence values. On the other hand, insignificant positive Moran's I spatial autocorrelation indices for 2014 (I=0.022; p=0.209) and 2017 (I=0.006; p=0.601) were observed, implying a random spatial distribution of malaria prevalence values.

Table 3. 1: Global Moran's I indices for 2010, 2012, 2014, and 2017

Year	Moran's I Statistic	P-value	Interpretation
2010	0.044	0.021*	Presence of spatial clustering
2012	0.074	<0.001**	Presence of spatial clustering
2014	0.022	0.209^{NS}	Absence of spatial clustering
2017	0.006	0.601^{NS}	Absence of spatial clustering

***** significant
** highly significant
NS: Not Significant

It has been documented that Moran's I statistics have a limitation in identifying the degree to which high and low values cluster together (Ord and Getis, 2001). As such, for this particular study, other spatial autocorrelation techniques such as the Getis and Ord G statistic in ArcGIS software and the Kulldorff spatial scan statistic in SaTScan software were used to further investigate the degree of further local clustering in the study area for all the four survey years.

3.1.3.2 Getis-Ord Gi Hotspot analysis*

Using the **Getis-Ord Gi* Hotspot analysis** tool in ArcGIS software, a hotspot and coldspot analysis was done. The dots signify the degree of statistical significance of clustering of high/low values for malaria. The darker the red dots, the higher the degree of confidence for high values hence signifying hotspots, while the darker the blue dots the higher the degree of confidence for low values hence signifying coldspots of malaria.

As shown in **Figure 3-3**, most of the statistically significant hotspots (most of them at 99% confidence level, signifying maximum spatial clustering of high values of malaria) were observed in the Central region, seconded by the Southern region; while more significant coldspots were observed in the Northern region and the furthest southern part of Malawi. The figure further shows that in 2010 (**Figure 3-3a**) there was a more closely spatial clustering of hotspots, but after the successive survey years, the pattern shows more random distribution of the hotspots (especially in 2017, as shown in **Figure 3-3d**).

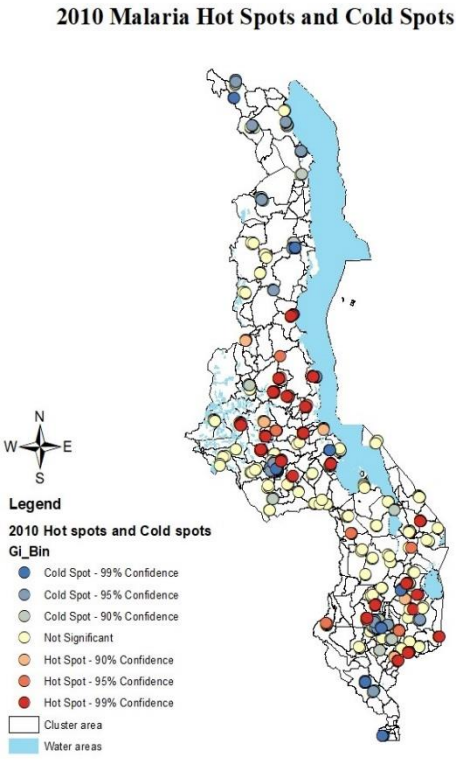


Figure 3-3a: 2010 hotspots/coldspots

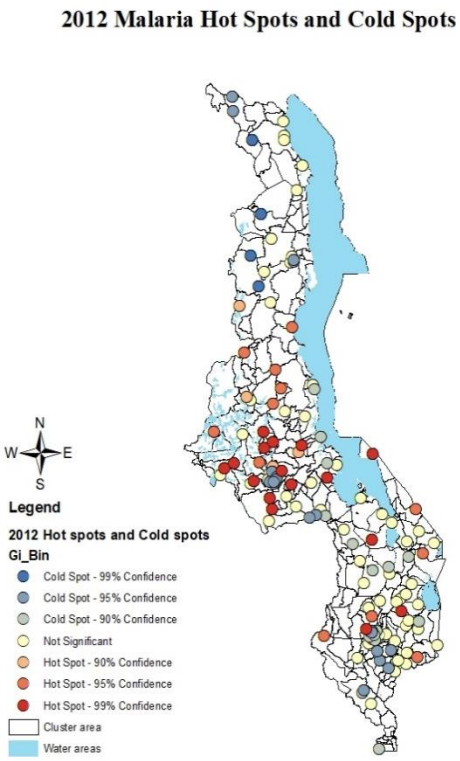


Figure 3-3b: 2012 hotspots/coldspots

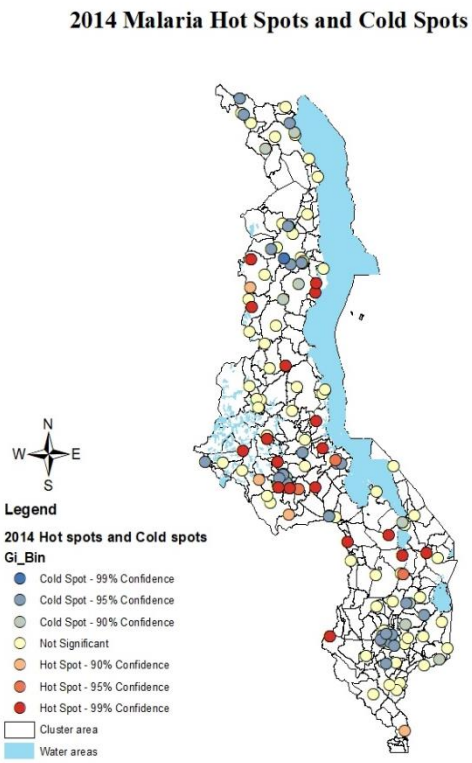


Figure 3-3c: 2014 hotspots/coldspots

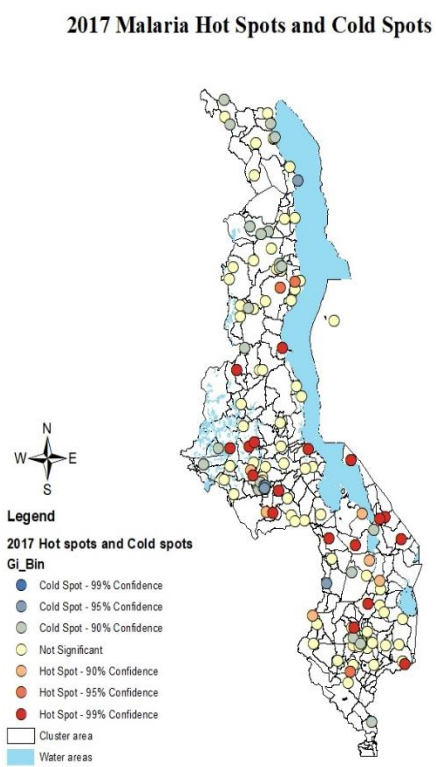


Figure 3-3d: 2017 hotspots/coldspots

Figure 3-3: Hotspot and Coldspot analysis of Malaria clustering for 2010, 2012, 2014, and 2017

3.1.3.3 Malaria Cluster Analysis using SatScan

Table 3.2 below shows the results of the SatScan spatial scan analysis of malaria across Malawi for 2010, 2012, 2014, and 2017 survey years.

Table 3. 2: Spatial clustering of malaria in Malawi in 2010, 2012, 2014, 2017 using SatScan

Year	Cluster	Circle Radius	No. of Locations	Pop	Observed	Expected	RR	P-value	Hot/Cold spot
2010	1	119.45 km	26	342	208	123.73	2.01	<0.001	Hot
	2	160.49 km	12	201	21	72.72	0.26	<0.001	Cold
	3	18.92 km	6	198	27	71.63	0.35	<0.001	Cold
	4	73.14 km	4	48	3	17.37	0.17	0.016	Cold
	5	0 km	1	157	29	56.80	0.49	0.017	Cold
	6	20.87 km	5	53	39	19.17	2.10	0.029	Hot
2012	1	76.45 km	14	325	21	80.63	0.23	<0.001	Cold
	2	109.22 km	24	418	181	103.71	2.14	<0.001	Hot
	3	45.24 km	3	50	0	12.41	0	<0.001	Cold
	4	25.40 km	5	72	42	17.86	2.47	<0.001	Hot
	5	0 km	1	45	0	11.16	0	0.001	Cold
	6	41.29 km	5	90	5	22.33	0.22	0.004	Cold
	7	0 km	1	23	16	5.71	2.86	0.142	Hot
	8	33.23 km	5	72	6	17.86	0.33	0.307	Cold
	9	24.13 km	4	72	30	17.86	1.72	0.876	Hot
2014	1	49.66 km	15	333	32	88.60	0.32	<0.001	Cold
	2	37.29 km	4	107	72	28.47	2.78	<0.001	Hot
	3	181.03 km	17	385	45	102.44	0.39	<0.001	Cold
	4	125.70 km	21	395	157	105.10	1.71	<0.001	Hot
	5	0 km	1	142	13	37.78	0.33	0.001	Cold
	6	11.84 km	2	39	28	10.38	2.80	0.002	Hot
	7	37.48 km	5	97	48	25.81	1.95	0.027	Hot
	8	4.36 km	2	26	0	6.92	0	0.073	Cold
	9	0 km	1	44	22	11.71	1.92	0.833	Hot
2017	1	138.61 km	25	565	167	97.64	2.23	<0.001	Hot
	2	124.66 km	15	458	27	79.15	0.29	<0.001	Cold
	3	16.91 km	2	187	5	32.31	0.14	<0.001	Cold
	4	0 km	1	236	11	40.78	0.25	<0.001	Cold
	5	40.12 km	7	129	48	22.29	2.31	<0.001	Hot
	6	0 km	1	12	10	2.07	4.92	0.023	Hot

For 2010, two statistically significant hotspot malaria clusters (with RR>1 and p<0.05) and four statistically significant coldspot malaria clusters (with RR<1 and p<0.05) were observed. The primary cluster was a hotspot cluster (detected in the Central region, as shown in **Figure 3-4a**) of a radius of 119.45 km with total of 26 geographical locations and had an associated RR=2.01. This implies that the locations in the primary cluster had 2.01 times the risk of having high malaria prevalence as compared to the other locations. The hotspot clusters were observed

in the Central region and Southern region while the coldspot clusters were observed in the Northern and the furthest Southern parts of Malawi, as shown in **Figure 3-4a**.

For 2012, two statistically significant hotspot malaria clusters (with RR>1 and p<0.05), two statistically insignificant hotspot malaria clusters (with RR>1 but p>>0.05), four statistically significant coldspot malaria clusters (with RR<1 and p<<0.05), and one statistically insignificant coldspot malaria cluster (with RR<1 but p>>0.05) were observed. The primary cluster was a coldspot cluster (detected in the furthest part of the Southern region, as shown in **Figure 3-4b**) of a radius of 76.45 km with total of 14 geographical locations and had an associated RR=0.23. This implies that the locations in this primary cluster had 77% less risk of having high malaria prevalence as compared to the other locations. The hotspot clusters were observed in the Central region and Southern region; while the coldspot clusters were observed in the furthest part of Northern region, some parts in the Central region, and the furthest Southern parts of malawi, as shown in **Figure 3-4b**.

For 2014, four statistically significant hotspot malaria clusters (with RR>1 and p<0.05), one statistically insignificant hotspot malaria clusters (with RR>1 but p>>0.05), three statistically significant coldspot malaria clusters (with RR<1 and p<<0.05), and one statistically insignificant coldspot malaria cluster (with RR<1 but p>>0.05) were observed. The primary cluster was a coldspot cluster (detected in the Southern region, as shown in **Figure 3-4c**) of a radius of 49.66 km with total of 15 geographical locations and had an associated RR=0.32. This implies that the locations in this primary cluster had 68% less risk of having high malaria prevalence as compared to the other locations. The hotspot clusters were observed in the Central region and Southern region; while the coldspot clusters were observed in the Northern part and the Southern part of Malawi, as shown in **Figure 3-4c**.

For 2017, three statistically significant hotspot malaria clusters (with RR>1 and p<0.05) and three statistically significant coldspot malaria clusters (with RR<1 and p<<0.05) were observed. The primary cluster was a hotspot cluster (detected in the South-eastern region, as shown in **Figure 3-4d**) of a radius of 138.61 km with total of 25 geographical locations and had an associated RR=2.23. This implies that the locations in the primary cluster had 2.23 times the risk of having high malaria prevalence as compared to the other locations. The hotspot clusters were observed in the Central region and South-eastern region; while the coldspot clusters were observed in the Northern part and the Southern part of Malawi, as shown in **Figure 3-4d**.

Generally, **Figure 3-4** shows that, for each survey year, malaria hotspots were in the Central region and the South-eastern region geographical areas, while the malaria coldspots were observed in the Northern Region and the furthest geographical areas in the Southern region.

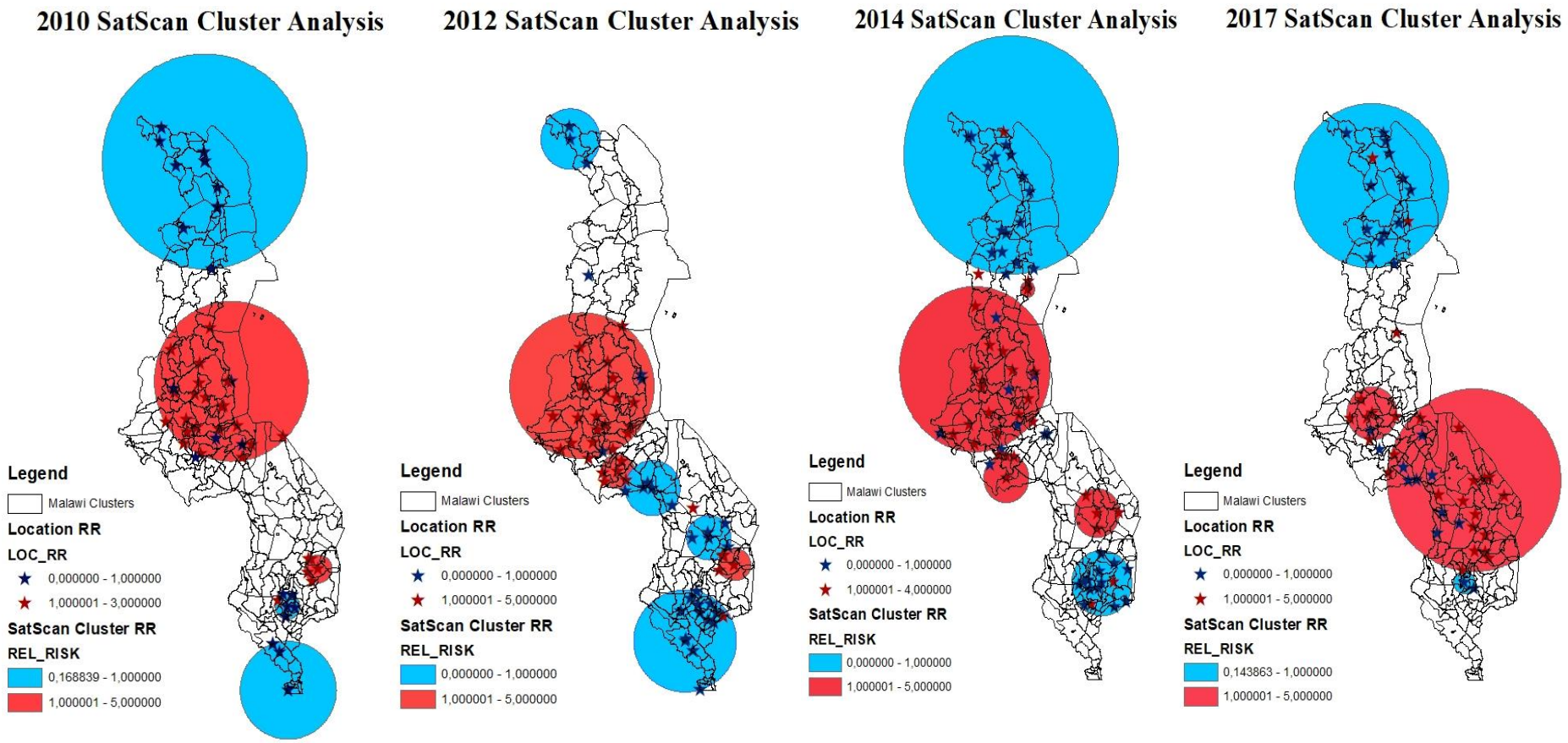


Figure 3-4a: 2010 Malaria clusters **Figure 3-4b**: 2012 Malaria clusters **Figure 3-4c**: 2014 Malaria clusters **Figure 3-4d**: 2017 Malaria clusters

Figure 3-4: SatScan spatial scanning of malaria clustering for 2010, 2012, 2014, and 2017

3.2 Factors associated with malaria morbidity in 2010, 2012, 2014, 2017

For this research, the number of under-five children that were included in this study are as follows: 1758 (1822 after accounting for sampling weights) children from the 2010 MIS; 2112 children from the 2012 MIS, 1928 children from the 2014 MIS; and 2305 children from the 2017 MIS. **Table 3.3** displays the details of weighted descriptive statistics (after accounting for sampling weights) for the Demographic, Socio-economic, Environmental, and Vector-control factors for study participants.

3.2.1 Descriptive Statistics for the Demographic, Socio-economic, Environmental, and Vector-control factors

3.2.1.1 Descriptive Statistics for the Demographic Factors

The overall mean age (in months) was 31.5 in 2010, 32.2 in 2012, 31.8 in 2014, and 33.9 in 2017. The mean age was higher for the children who tested positive for malaria than those who tested negative for malaria, in all of the survey years (33.2 versus 30.2 in 2010; 34.8 versus 31.2 in 2012; 35.0 versus 30.2 in 2014; and 38.6 versus 32.4 in 2017)

As regards to Region of residence, in overall, more children were sampled from the Southern Region (931 representing 51.1% in 2010; 920 representing 43.6% in 2012, 820 representing 42.5% in 2014; and 1074 representing 46.8% in 2017), seconded by the Central Region, and lastly the Northern Region. Among the children who tested positive for malaria, more came from the Southern region in 2010 [391(49.8%)]; in 2012 and 2014, more came from the central region [307(52.4%) and 269(42.0%) respectively]; and in 2017, more came from the Southern region [280(49.7%)].

3.2.1.2 Descriptive Statistics for the Socio-economic Factors

In terms of place of residence, in overall for all the four survey years, more children were from the rural arears [1606(88.2%) in 2010, 1838(87.1%) in 2012, 1666(86.4%) in 2014, and 1993(86.4%) in 2017]. Amongst the malaria positive children, regardless of the survey year,

more children still were from the rural areas [754(96.0%) in 2010, 558(95.3%) in 2012, 612(95.6%) in 2014, and 552(97.8%) in 2017].

In terms of Wealth Index, in overall for all the four survey years, more children were from the Poor families [787(43.2%) in 2010, 951(45.0%) in 2012, 859(44.5%) in 2014, and 1079(46.8%) in 2017]. Among the malaria positive children, the highest proportions of them were from the Poor families [411(52.4%) in 2010, 323(55.2%) in 2012, 373(58.2%) in 2014, and 360(63.8%) in 2017].

As regards to Mother's education level, for all the survey years overall, more children had mothers whose education level was up to Primary school level [891(67.8%) in 2010, 1212(64.6%) in 2012, 1134(64.8%) in 2014, and 1363(67.8%)]. Among the malaria positive children, the highest proportions, in all the survey years, were from those whose mothers had reached up to Primary school level [400(70.6%) in 2010, 354(68.3%) in 2012, 388(69.4%) in 2014, and 339(71.9%) in 2017].

As regards to whether a child's mother heard about malaria, overall, more children had mothers who heard about malaria disease [1275(96.9%) in 2010, 1158(92.8%) in 2012, 1318(92.6%) in 2014, and 1564(95.6%) in 2017]. Amongst the malaria positive children, for all the survey years, the highest proportions were among the children whose mothers heard about the malaria disease [554(97.6%) in 2010, 299(87.3%) in 2012, 408(87.9%) in 2014, and 360(94.2%) in 2017].

3.2.1.3 Descriptive Statistics for the Environmental Factors

Except for 2014 survey year, the mean Cluster altitude (per 100 meters) was lower among the children who tested malaria positive than that among the children who tested malaria negative (8.11 versus 8.69 in 2010, 8.56 versus 9.41 in 2014, and 8.30 versus 8.79 in 2017].

3.2.1.4 Descriptive Statistics for the Vector-control factors

In terms of whether the household had Insecticide Treated Nets (ITNs), the highest proportion of children came from household which had ITNs [1017(55.8%) in 2010, 1160(61.0%) in 2012, 1270(73.4%) in 2014, and 1510(81.2%) in 2017]. Among the malaria positive children, the highest proportion of them came from households which had ITNs [416(53.0%) in 2010, 293(56.6%) in 2012, 383(67.6%) in 2014, and 330(76.1%) in 2017].

Table 3. 3: Descriptive statistics for the Demographic, Socio-Economic, Environmental and Vector-control characteristics

	2010			2012			2014			2017		
Variable	**Positive** n = 785(43%)	**Negative** n = 1037(57%)	**Overall** n = 1822	**Positive** n = 585(28%)	**Negative** n = 1527(72%)	**Overall** n = 2112	**Positive** n = 640(33%)	**Negative** n = 1288(67%)	**Overall** n = 1928	**Positive** n = 564(24%)	**Negative** n = 1741(76%)	**Overall** n = 2305
Demographic												
Age(months)	33.2 (31.9,34.5)	30.2 (29.2,31.1)	31.5 (30.8,32.1)	34.8 (33.7,35.9)	31.2 (30.5,32.0)	32.2 (31.6,32.9)	35.0 (33.7,36.3)	30.2 (29.0,31.5)	31.8 (30.9,32.8)	38.6 (37.0,40.3)	32.4 (31.5,33.3)	33.9 (33.2,34.6)
Region												
North	42(5.4%)	150(14.5%)	192(10.5%)	59(10.0%)	239(15.7%)	298(14.1%)	105(16.4%)	263(20.4%)	368(19.1%)	29(5.1%)	219(12.6%)	248(10.8%)
Central	352(44.8%)	347(33.5%)	699(38.4%)	307(52.4%)	588(38.5%)	894(42.3%)	269(42.0%)	471(36.6%)	740(38.4%)	255(45.2%)	722(41.5%)	977(42.4%)
South	391(49.8%)	540(52.0%)	931(51.1%)	220(37.6%)	700(45.8%)	920(43.6%)	266(41.6%)	554(43.0%)	820(42.5%)	280(49.7%)	800(45.9%)	1080(46.8%)
Socio-Economic												
Residence												
Urban	31(4.0%)	185(17.8%)	216(11.8%)	27(4.7%)	247(16.1%)	274(12.9%)	28(4.4%)	234(18.2%)	262(13.6%)	12(2.2%)	300(17.2%)	312(13.6%)
Rural	754(96.0%)	852(82.2%)	1606(88.2%)	558(95.3%)	1280(83.9%)	1838(87.1%)	612(95.6%)	1054(81.8%)	1666(86.4%)	552(97.8%)	1441(82.8%)	1993(86.4%)
Wealth Index												
Poor	411(52.4%)	376(36.2%)	787(43.2%)	323(55.2%)	628(41.1%)	951(45.0%)	373(58.2%)	486(37.8%)	859(44.5%)	360(63.8%)	719(41.3%)	1079(46.8%)
Middle	207(26.3%)	235(22.7%)	442(24.3%)	131(22.4%)	290(19.0%)	421(20.0%)	134(20.9%)	248(19.2%)	382(19.8%)	106(18.9%)	335(19.2%)	441(19.1%)
Rich	167(21.3%)	426(41.1%)	593(32.5%)	131(22.4%)	609(39.9%)	740(35.0%)	133(20.9%)	554(43.0%)	687(35.7%)	98(17.3%)	687(39.5%)	785(34.1%)
Mother's education												
None	126(22.3%)	129(17.3%)	256(19.4%)	133(25.7%)	259(19.1%)	392(20.9%)	130(23.3%)	181(15.2%)	311(17.8%)	103(22.0%)	199(12.9%)	302(15.0%)
Primary	400(70.6%)	491(65.7%)	891(67.8%)	354(68.3%)	858(63.2%)	1212(64.6%)	388(69.4%)	746(62.6%)	1134(64.8%)	339(71.9%)	1025(66.5%)	1363(67.8%)
Secondary/higher	41(7.1%)	127(17.0%)	168(12.8%)	31(6.0%)	240(17.7%)	271(14.5%)	41(7.3%)	264(22.2%)	305(17.4%)	29(6.1%)	316(20.6%)	345(17.2%)
Mother heard about Malaria												

No	13(2.4%)	27(3.6%)	40(3.1%)	43(12.7%)	47(5.2%)	90(7.2%)	56(12.1%)	49(5.1%)	105(7.4%)	22(5.8%)	50(4.0%)	72(4.4%)
Yes	554(97.6%)	721(96.4%)	1275(96.9%)	299(87.3%)	859(94.8%)	1158(92.8%)	408(87.9%)	910(94.9%)	1318(92.6%)	360(94.2%)	1204(96.0%)	1564(95.6%)
Environmental												
Cluster altitude (per 100 metres)	8.11 (7.50,8.73)	8.69 (7.96,9.42)	8.46 (7.85,9.07)	8.93 (8.33,9.54)	8.79 (8.05,9.54)	8.83 (8,19,9.47)	8.56 (7.61,9.52)	9.41 (8.67,10.14)	9.13 (8.38,9.87)	8.30 (7.45,9.15)	8.79 (7.91,9.66)	8.66 (7.86,9.47)
Vector Control												
Have ITN												
None	340(43.4%)	400(38.6%)	740(40.6%)	197(38.1%)	444(32.1%)	641(33.8%)	182(32.1%)	265(22.8%)	447(25.8%)	102(23.5%)	213(15.0%)	315(17.0%)
Yes (not ITN)	29(3.6%)	36(3.5%)	65(3.6%)	28(5.3%)	71(5.2%)	99(5.2%)	2(0.3%)	11(1.0%)	13(0.8%)	2(0.4%)	33(2.3%)	35(1.8%)
Yes (ITN)	416(53.0%)	601(57.9%)	1017(55.8%)	293(56.6%)	867(62.7%)	1160(61.0%)	383(67.6%)	887(76.2%)	1270(73.4%)	330(76.1%)	1181(82.7%)	1510(81.2%)

3.2.2 Multilevel Logistic Regression Results for 2010, 2012, 2014, and 2017 MMIS

For each of the four survey years, a weighted Multilevel Logistic Regression model was fitted. **Table 3.4** below shows the results for each model.

3.2.2.1 Demographic Factors associated with under-five Malaria

In each of the survey years, age (in months) was significantly associated with under-five malaria, much as in 2010 the age variable was marginally significant. A one-month increase in the child's age resulted into a corresponding increase in the odds of a child having malaria.

Children who lived in the Central region had higher odds of being malaria positive than those who lived in the (reference category) Northern region, much as this was only statistically significant for 2010 [OR= 2.99, 95% C.I=(1.24,7.19)] and 2012 [OR=2.89, 95%C.I=(1.04,7.96)]. Except for 2014 [OR=0.45, 95% C.I=(0.11,1.72)], children who lived in the Southern region had higher odds of having malaria than those in the Northern Region but this was not statistically significant [**2010**: OR=1.42, 95%C.I=(0.54,3.73); **2012**: OR=1.40, 95%C.I=(0.46,4.23); **2017**: OR=1.19, 95%C.I=(0.55,2.59)].

3.2.2.2 Socio-economic Factors associated with under-five Malaria

In terms of Place of Residence, the odds of having malaria for children who lived in rural areas were at least 3 times more than for those that lived in urban areas; and this was statistically significant for 2010, 2014, and 2017. For 2012, the odds of having malaria for children who lived in rural areas was 2.17 times more than those who lived in the urban areas, but this was not statistically significant [OR=2.17, 95%C.I=(0.83,5.68)].

In terms of Wealth Index for a household, the general trend in all the survey years was that the odds of a child having malaria decreased as the wealth index level increased. Specifically, children who lived in "Rich" households were, respective of each survey year, 42% [in 2010:

OR=0.58, 95%C.I=(0.36,0.94)], 53% [in 2012: OR=0.47, 95%C.I=(0.28,0.81)], 49% [in 2014: OR=0.51, 95%C.I=(0.27,1.00)], and 57%[in 2017: OR=0.43, 95%C.I=(0.21,0.89)] less likely to be malaria positive that children who lived in "Poor" households.

Generally, the odds of a child having malaria decreased with increase in the level of education of the mother, for all the four survey years. With respect to each survey year, the children whose mothers had a Secondary or higher education level were 51% (in 2010), 61% (in 2012), 35% (in 2014), and 65% (in 2017) less likely to have malaria than children whose mothers had no education at all; and the odds ratios were statistically significant for 2010 [OR=0.49, 95%C.I=(0.27,0.88)], 2012 [OR=0.39, 95%C.I=(0.16,0.97)], and 2017 [OR=0.35, 95%C.I=(0.17,0.71)].

The odds of a child having malaria, in 2010, was 1.27 times more (95%C.I=0.62,2.64) for children whose mothers heard about malaria than for children whose mothers never heard about malaria. For the other three survey years (2012, 2014, and 2017), the odds of a child having malaria was lower for children whose mothers heard about malaria than for those children whose mothers never heard about malaria. In all the survey years, the association between whether the mother heard about malaria and a child having malaria was not statistically significant (all p-values>>0.05), though for 2014, the association was marginally significant [OR=0.58, p=0.092].

3.2.2.3 Environmental Factors associated with under-five Malaria

In terms of Cluster Altitude (per 100 metres), the general trend observed was that a 100 metres increase in altitude resulted in a general decrease in the odds of a child having malaria. However, this association was only statistically significant in 2010 [OR=0.89, 95%C.I=(0.80,0.99)] and 2014 [OR=0.85, 95%C.I=(0.74,0.98)] MMIS, but such was not the case for the 2012 and 2017 MMIS.

3.2.2.4 Vector-control factors associated with under-five Malaria

For children who had bed nets which were ITNs for sleeping, the odds for them having malaria were lower as compared to those who had no bed nets at all [**2010**: OR=0.91, 95%C.I=(0.60,1.41); **2012**: OR=0.84, 95%C.I=(0.53,1.34); **2014**: OR=0.86, 95%C.I=(0.47,1.57); **2017**: OR=0.55, 95%C.I=(0.28,1.09)]. The effect of having a net on malaria was marginally statistically significant only in 2017 (p-value=0.087), while for the first three survey years, the effect was not statistically significant (p-value>>0.05).

Table 3. 4: Multilevel logistic regression results for 2010, 2012, 2014, and 2017 MMIS

	2010		2012		2014		2017	
	OR(95% C.I)	p-value	OR(95% C.I)	p-value	OR(95% C.I)	p-value	OR(95% C.I)	p-value
Variable								
Demographic								
Age(months)	1.01(1.00,1.02)	0.065	1.02(1.01,1.03)	<0.001	1.02(1.01,1.03)	<0.001	1.03(1.02,1.05)	<0.001
Region								
North	Ref		Ref		Ref		Ref	
Central	2.99(1.24,7.19)	0.015	2.89(1.04,7.96)	0.041	1.54(0.60,3.96)	0.365	1.60(0.79,3.26)	0.189
South	1.42(0.54,3.73)	0.475	1.40(0.46,4.23)	0.551	0.45(0.11,1.72)	0.243	1.19(0.55,2.59)	0.653
Socio-Economic								
Residence								
Urban	Ref		Ref		Ref		Ref	
Rural	3.43(1.81,6.54)	<0.001	2.17(0.83,5.68)	0.111	3.58(1.31,9.80)	0.014	3.42(1.57,7.47)	0.002
Wealth Index								
Poor	Ref		Ref		Ref		Ref	
Middle	0.88(0.56,1.38)	0.569	1.15(0.69,1.92)	0.575	0.87(0.44,1.70)	0.679	0.50(0.28,0.89)	0.019
Rich	0.58(0.36,0.94)	0.026	0.47(0.28,0.81)	0.006	0.51(0.27,1.00)	0.050	0.43(0.21,0.89)	0.023
Mother's education								
None	ref		ref		Ref		ref	
Primary	0.86(0.56,1.32)	0.497	1.00(0.60,1.67)	0.987	1.09(0.57,2.08)	0.798	0.61(0.36,1.03)	0.065
Secondary/higher	0.49(0.27,0.88)	0.019	0.39(0.16,0.97)	0.044	0.65(0.28,1.55)	0.328	0.35(0.17,0.71)	0.004

Mother heard about Malaria								
No	Ref		Ref		Ref		Ref	
Yes	1.27(0.62,2.64)	0.507	0.86(0.32,2.24)	0.752	0.58(0.31,1.09)	0.092	0.74(0.30,1.77)	0.490
Environmental								
Cluster altitude (per 100m)	0.89(0.80,0.99)	0.047	1.00(0.88,1.14)	0.972	0.85(0.74,0.98)	0.021	0.93(0.84,1.02)	0.143
Vector Control								
Have net								
No	Ref		Ref		Ref		Ref	
Yes (not ITN)	2.24(0.79,6.37)	0.128	0.87(0.35,2.14)	0.758	empty	-	empty	
Yes (ITN)	0.91(0.60,1.41)	0.691	0.84(0.53,1.34)	0.476	0.86(0.47,1.57)	0.625	0.55(0.28,1.09)	0.087
Variance (cluster)	1.02(0.58,1.80)		1.55(0.89.2.71)		1.54(0.74,3.14)		1.04(0.58,1.86)	

3.3 Spatial-temporal random effects on under-five Malaria morbidity

As outlined in the methodology **Section 2.8.3.2,** seven Negative Binomial models were fit, adjusting for unstructured spatial random effects, structured spatial random effects, and spatial-temporal random effects. One model was fitted using the Mixed Effects Maximum Likelihood approach, of which the results are presented with Relative Risk (RR) together with the 95% Confidence Intervals; while the last six models were fitted using the Bayesian approach, and the results are presented with Relative Risk together with 95% Credible Intervals. **Table 3.5** below displays the results of all the fitted models.

3.3.1 Model Fit Results

The Spatial-Temporal model with Spatial (both structured and unstructured), temporal (AR1), and interaction-term random effects was chosen as the best fitting model because it had the lowest DIC value (**DIC=1839.70**, effective number of parameters $\mathbf{P_D=100.70}$). The results explained below are mainly based on this model, but the sixth model (spatial-temporal model with AR1 temporal term without the interaction term) is also considered as this model was also considered more parsimonious because it had a smaller number of the effective number of parameters.

Table 3. 5: Non-Spatial, Spatial, and Spatial-temporal Multivariable Negative Binomial Models

	Multilevel Model (MLE)		**Bayesian Unstructured random effects model (INLA)**	**Bayesian Structured Spatial random effects model (INLA)**	**Bayesian Unstructured and Structured Spatial random effects model (INLA)**	**Spatial (structured, unstructured)-temporal (i.i.d) random effects model**	**Spatial (structured, unstructured)-temporal (AR1) random effects model**	**Spatial (structured, unstructured)-temporal (AR1), interaction random effects model**
Variable	RR(95% C.I)	p-value	RR(95% Cr.I)	RR(95% Cr.I)	RR(95% Cr.I)	RR(95% Cr.I)	RR(95% Cr.I)	RR(95% Cr.I)
Demographic								
Age(months)	1.01(0.99,1.03)	0.420	-	-	-	-	-	-
Region								
North	Ref		Ref	Ref	Ref	Ref	Ref	Ref
Central	1.53(1.20,1.93)	<0.01	1.59(1.21,2.10)	1.19(0.62,2.24)	1.21(0.65,2.23)	1.19(0.65,2.16)	1.19(0.65,2.19)	1.06(0.66,1.68)
South	1.13(0.88,1.44)	0.332	1.14(0.86,1.50)	1.41(0.59,3.37)	1.41(0.60,3.28)	1.45(0.64,3.28)	1.45(0.64,3.33)	1.08(0.60,1.95)
Socio-Economic								
Residence								
Urban	Ref		Ref	Ref	Ref	Ref	Ref	Ref
Rural	1.44(1.04,2.00)	0.027	1.41(1.00,1.99)	1.50(1.07,2.11)	1.50(1.07,2.10)	1.63(1.18,2.27)	1.63(1.17,2.28)	1.48(1.10,1.99)
Wealth Index								
Poor	Ref		Ref	Ref	Ref	Ref	Ref	Ref
Middle	1.10(0.60,2.02)	0.758	1.28(0.72,2.25)	1.29(0.73,2.28)	1.30(0.74,2.30)	1.13(0.67,1.91)	1.13(0.66,1.93)	1.07(0.64,1.78)
Rich	0.50 (0.32,0.79)	0.003	0.53(0.35,0.82)	0.54(0.35,0.82)	0.54(0.35,0.83)	0.54(0.36,0.80)	0.54(0.36,0.80)	0.52(0.35,0.75)
Mother's education								
None	Ref		Ref	Ref	Ref	Ref	Ref	Ref
Primary	0.57(0.35,0.92)	0.022	0.59(0.38,0.92)	0.58(0.37,0.91)	0.58(0.37,0.90)	0.63(0.42,0.95)	0.63(0.42,0.94)	0.63(0.43,0.91)
Secondary/higher	0.25(0.12,0.55)	<0.01	0.27(0.13,0.56)	0.30(0.14,0.61)	0.30(0.14,0.61)	0.37(0.19,0.72)	0.36(0.19,0.72)	0.38(0.20,0.71)
Mother heard about								
Malaria								
No	Ref		Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.82(0.43,1.56)	0.553	0.84(0.48,1.48)	0.72(0.41,1.28)	0.73(0.42,1.28)	0.64(0.39,1.07)	0.64(0.39,1.07)	0.69(0.45,1.06)

Environmental								
Cluster altitude	0.97(0.93,1.00)	0.040	0.96(0.94,0.99)	0.95(0.92,0.97)	0.95(0.92,0.97)	0.97(0.94,0.99)	0.97(0.94,0.99)	0.97(0.94,0.99)
Vector Control								
Have net								
None	Ref		Ref	Ref	Ref	Ref	Ref	Ref
Yes (not ITN)	1.30(0.36,4.66)	0.685	1.08(0.32,3.58)	0.85(0.24,2.89)	0.87(0.26,2.92)	1.55(0.48,4.84)	1.56(0.48,4.94)	1.56(0.49,4.87)
Yes (ITN)	0.87(0.60,1.27)	0.481	0.85(0.60,1.22)	0.80(0.56,1.15)	0.80(0.56,1.15)	1.21(0.83,1.76)	1.20(0.83,1.76)	1.16(0.81,1.65)
Over-dispersion parameter	0.17(0.11,0.29)		0.15(0.10,0.23)	0.17(0.16,0.172)	0.19(0.15,0.30)	0.08(0.04,0.15)	0.09(0.08,0.10)	0.08(0.04,0.22)
LR test VS neg bin. Model	χ^2 =13.51	<0.001	-	-	-			
ϕ_c (Unstructured effect)	0.15(0.08,0.28)		0.15(0.10,0.24)	-	0.02(0.01,0.023)	0.03(0.01,0.05)	0.05(0.04,0.06)	0.03(0.01,0.12)
φ_c (Structured effect)	-		-	0.37(0.30,0.42)	0.25(0.18,0.33)	0.32(0.22,0.67)	0.23(0.21,0.26)	0.24(0.14,0.61)
ψ_t (Temporal effect, i.i.d)	-				-	0.04(0.02,0.08)	-	
(Temporal effect, AR1)							0.04(0.03,0.05)	$5.44e^{-5}$($1.50e^{-5}$,$7.93e^{-5}$)
Interaction term								$2.30e^{-4}$($4.51e^{-5}$,0.05)
DIC (pD)	**-**		**1918.97(82.85)**	**1900.95(69.03)**	**1900.92(73.54)**	**1847.62(87.32)**	**1845.39(86.69)**	**1839.70(100.70)**

3.3.1.1 Demographic Factors associated with under-five Malaria

The age of the child was modelled with the assumption that it had a non-linear effect on the malaria outcome. **Figure 3.5** shows the effect of age on malaria, for all the six Bayesian models. In general, the increase in age had a corresponding increasing effect on malaria; but the last model showed clearly that age had a non-linear effect. As age increased, its effect on malaria increased up until from 25 months to 40 months the effect was a bit constant; but the effect then increased in magnitude from 40 months onwards.

Children who lived in the central region and southern region had a high relative risk of having malaria than those who lived in the northern region, but the effect was not statistically significant [Central: RR=1.06, 95%Cr.I=(0.66,1.68); South: RR=1.08, 95%Cr.I=(0.60,1.95)]

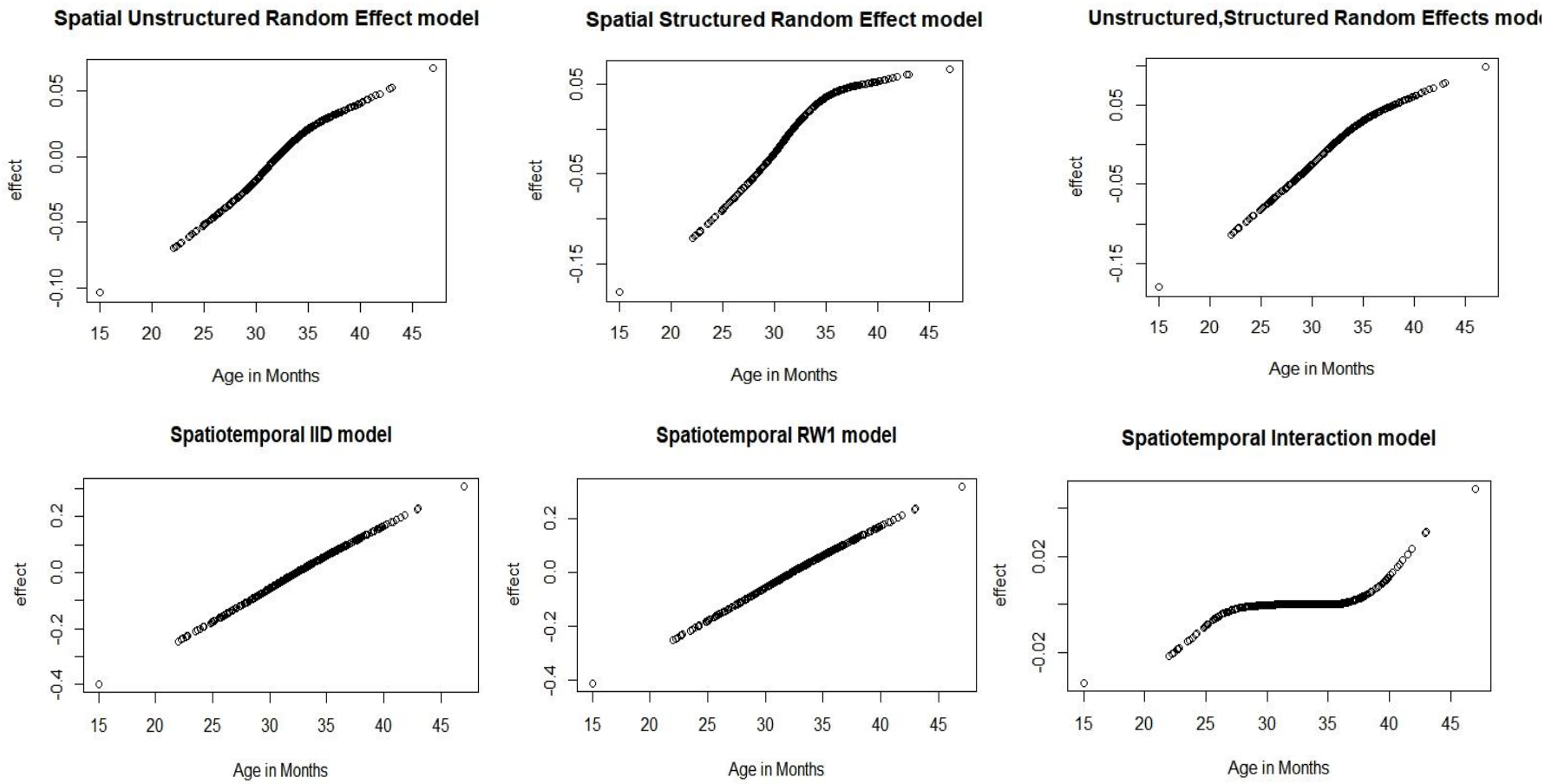


Figure 3-5: Non-linear effect of age (in months) on malaria

3.3.1.2 Socioeconomic Factors associated with under-five Malaria

Children who lived in rural areas were more prone to have malaria relative to those who lived in urban areas, and this effect was statistically significant [RR=1.48, 95%Cr.I=(1.10,1.99)].

The risk of a child having malaria decreased with increase in the wealth index of a household. More specifically, children who lived in households that were "rich" had 48% less risk of having malaria than those who lived in "poor" households, and this was statistically significant [RR=0.52, 95%Cr.I=(0.35,0.75)].

Children whose mothers had Primary school as their highest level of education had 37% less risk [RR=0.63, 95%Cr.I=(0.43,0.91)] of having malaria than those children whose mothers had no education at all. For children whose mothers had Secondary or higher education as the highest level, had 62% less risk [RR=0.38, 95%Cr.I=(0.20,0.71)] of having malaria than the children whose mothers had no any educational attainment.

Children whose mothers had ever heard about malaria had a 31% less risk [RR=0.69, 95%Cr.I=(0.45,1.06)] of having malaria than the children whose mothers had never heard anything about malaria, but the effect was not statistically significant.

3.3.1.3 Environmental Factors associated with under-five Malaria

A 100 meters increase in the cluster altitude had a decreasing effect on the risk of a child having malaria, and the effect was statistically significant [RR=0.97, 95%Cr.I=(0.94,0.99)]

3.3.1.4 Vector-control Factors associated with under-five Malaria

A unit increase in the proportion of children who had bed nets but not ITN had 56% increase in the risk of having malaria [RR=1.56, 95%Cr.I=(0.49,4.87), while children who had bed nets that were ITN had 16% increase in the risk of having malaria [RR=1.16, 95%Cr.I=(0.81,1.65)].

The interesting fact is that the effect decreased as a child owned ITN bed nets, though not statistically significant.

3.3.1.5 Spatial-Temporal Effects

For the fourth (spatial only), fifth, sixth and seventh (spatial-temporal) models, both the unstructured and the structured random effects were significantly different from 0, hence showing that variations in malaria could be explained by spatial differences. However, the structured random effects had higher values than the unstructured random effects, implying that spatial differences could be explained more by treating the spatial random effect as structured.

As for the temporal random effects, both the fifth and sixth models had very similar values (0.04) despite having different DIC values. However, the temporal effect was very close to 0 for the last model, when an interaction term of spatial-temporal was included in the model.

Figure 3.6 is a map display of the results from the overall modelling, after adjusting for all the factors associated with malaria. The red shades indicate a higher risk (hot spots) of having malaria, while the blue shades indicate a lower risk of having malaria. More hot spots were observed in the central region areas and some parts of the south-eastern region areas.

Figure 3.7 shows the temporal risk maps that explain the differences in malaria risk in terms of temporal random effects. As the map shows, there was not much of a pronounced temporal random effects differences in most of the study clusters.

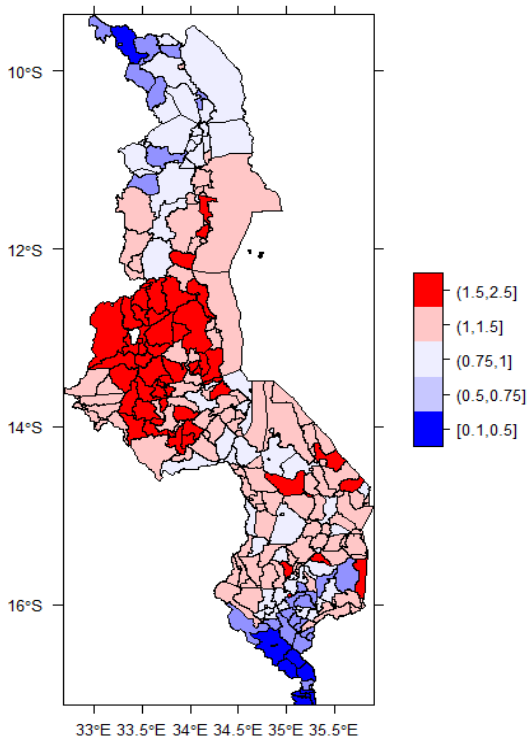


Figure 3-6: Modelled Malaria hotspots (red shade) and colds pots (blue shade) in clusters, overall results from the Bayesian spatial-temporal model

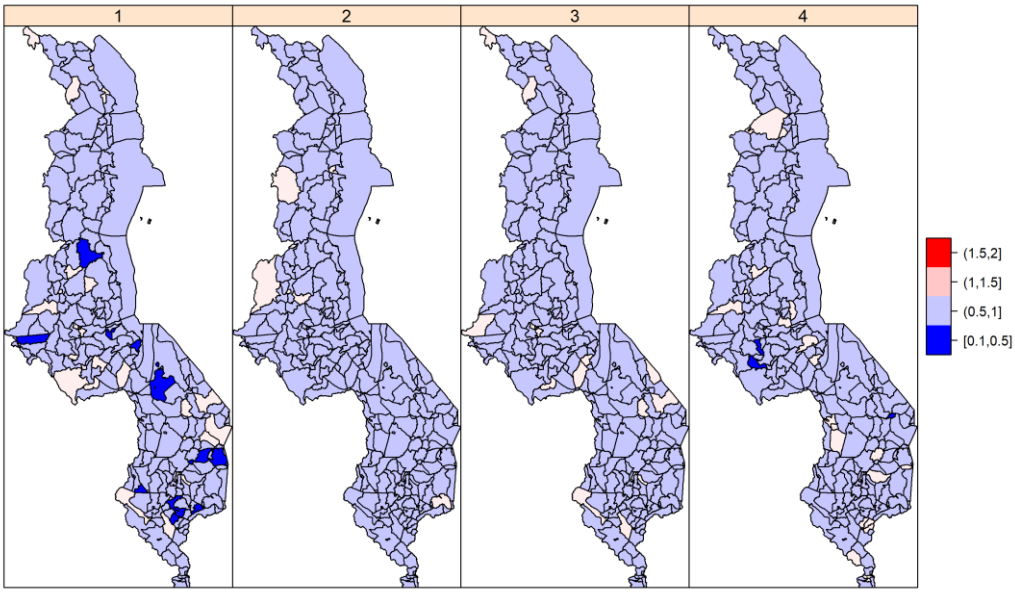


Figure 3-7: Temporal random effects distribution from the space-time interaction Bayesian Negative Binomial spatial-temporal model

3.3.1.6 Model convergence for the fixed effects

Figure 3.8 shows posterior density plots as a way of confirming model convergence of the fixed effects parameters. The symmetrical pattern shown in the figure for the estimated fixed effects parameters implies that the model captured the posterior parameter values well.

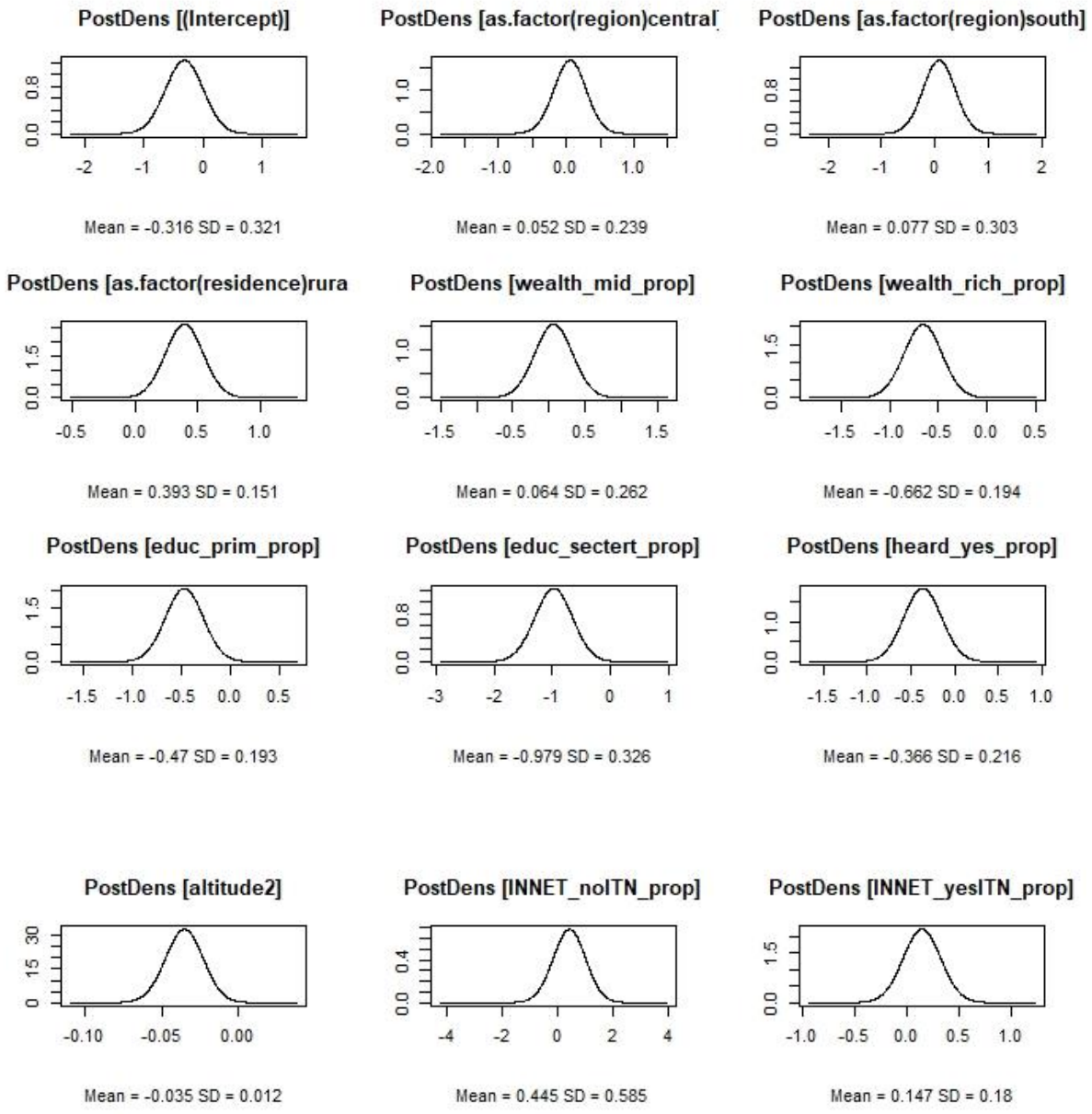


Figure 3-8: Posterior density plots of means for fixed effects of the Negative Binomial spatial-temporal model

CHAPTER 4: DISCUSSIONS

This chapter discusses the key findings of this study by comparing these findings with the literature for other similar studies that were done in Malawi and elsewhere. The strengths and limitations of the study are also discussed.

The main aim of the study was to determine the spatial-temporal random effects on under-five malaria morbidity for the years 2010, 2012, 2014, and 2017, after adjusting for demographic, socio-economic, environmental, and vector-control factors. The discussions presented in this chapter are mainly based on the Spatial-temporal model results.

4.1 Spatial-temporal Random Effects on Malaria variation

This study found that heterogeneity in under-five malaria morbidity could be mainly better explained when spatial and temporal random effects are incorporated in the models. The spatial and temporal random effects values, especially in the fifth and sixth models, were above zero (0), while in the last model, the temporal random effect value was very close to zero when the interaction term was introduced in the model. Furthermore, the structured spatial random effects contributed more to the spatial heterogeneity of malaria than the unstructured spatial random effects. These findings are similar to what Lowe *et.al* (2013) found in their study.

It is also interesting to note that when the structured spatial random effects as well as the temporal random effects are introduced in the Bayesian Negative Binomial models, the Relative Risk for the variable "Have net" shifted from below 1 to above 1. This shows that incorporating these random effects may change the effect of some variables on Malaria morbidity. This then shows the importance of accounting for spatial and temporal random

effects in explaining under-five malaria morbidity in Malawi, as also evidenced in Lowe *et.al* 2013.

4.2 Factors associated with Malaria Risk in under-five children

4.2.1 Demographic Factors

This study found that increase in age (in months) of a child had a corresponding increasing effect on the risk of a child having malaria. This well coincides with findings from studies done in Malawi by Zgambo *et.al* (2017), Chirombo *et.al* (2014), as well as a study done in Uganda by Roberts and Matthews, (2016). This phenomenon may be due to the fact that, during the first 6 months or so, a child may be breastfed and hence may have maternal immunity which may help fight against malaria infection (Chirombo, Lowe and Kazembe, 2014). Furthermore, most younger children in their early years are closely monitored and take care by their mothers, and as such, are more likely to be protected from being infected from various diseases such as malaria. Such children are more likely to sleep in bed nets as these share the same bed with their mothers (Zgambo, Mbakaya and Kalembo, 2017).

The results of this study further showed that children who lived in the central region were at a higher risk of having malaria, much as the effect was not statistically significant in the spatial-temporal models. Such may be the case because the central region of Malawi is largely composed of inland plains and have temperatures which provide more conducive environment for mosquitoes to breed, hence resulting to high cases of malaria in the region (Kazembe *et al.*, 2006; Chirombo, Lowe and Kazembe, 2014; Zgambo, Mbakaya and Kalembo, 2017).

4.2.2 Socioeconomic Factors

Place of residence was found to be a highly statistically significant variable associated with the risk of a child having malaria. Children who lived in rural areas had a relatively higher risk of testing positive for malaria as compared to their counterparts living in urban areas. Studies

done in Malawi (Kazembe *et al.*, 2006; Lowe, Chirombo and Tompkins, 2013; Chirombo, Lowe and Kazembe, 2014; Chitunhu and Musenge, 2016), as well as studies done elsewhere like in Uganda by (Yeka *et al.*, 2012), support this finding. One reason that may help explain this finding is that individuals living in rural areas are usually less privileged economically than their urban counterparts, and as such, they may have very limited access to health services that may help curb malaria infection in children. Another reason can be that people living in rural areas may have less exposure to various modern media that give information about malaria, hence such people may not fully know how to combat the disease.

Wealth Index was also found to be a highly significant factor associated with malaria risk. The risk of a child having malaria decreased with increase in the level of Wealth Index of the household. This is so because wealthier people are in a better position to afford better health services and interventions that may help lessen malaria burden on their children. A study done by (Njau *et al.*, 2006) found that highly economically-advantaged people were significantly at a lower risk to be parasitaemic, and were significantly more likely to acquire anti-malarial medication, than those in the middle or poor income groups. This current study's finding confirms previous results from other studies that showed that high malaria burden is associated with poverty (Clouston, Yukich and Anglewicz, 2015; Chitunhu and Musenge, 2016; Ssempiira *et al.*, 2017).

Mother's highest education level was found to be a significant factor affecting the risk of malaria. The results of this study showed that the higher the education level of the mother, the lower the risk of a child to test positive for malaria. This finding is in agreement with results reported from studies done in Malawi (Chitunhu and Musenge, 2016; Zgambo, Mbakaya and Kalembo, 2017) and in Nigeria (Fana *et al.*, 2015). The explanation of this finding can be that

highly educated mothers are more likely to have knowledge about malaria and its preventive methods, hence implying reduced risk of Malaria among children who have such mothers.

The risk of a child having malaria was lower for children whose mothers heard about malaria, even though this was not statistically significant. A similar study done by Lowe et.al (2013) showed that if more emphasis can be put on the idea that improving knowledge about malaria in communities, this may result in better use of interventions hence leading to reduced malaria risk.

4.2.3 Environmental Factors

Cluster altitude was found to be a significant variable associated with the risk of malaria in under-fives. The results of the study showed that the higher the altitude, the lower the malaria risk. This may be the case since higher altitudes experience lower temperatures as well as being less wet which make the malaria transmission slower because mosquitoes may not have adequate and suitable breeding grounds (Dzinjalamala, 2009).

4.2.4 Vector-control Factors

This study's spatial-temporal model results showed that the risk of having malaria in children was actually higher for those that had bed nets, but then there was a decreasing malaria risk trend between children who had bed nets but not ITN and those who had bed nets which were ITN. This finding was statistically not significant. Zgambo, Mbakaya and Kalembo, (2017) found that there is underutilization of ITNs in Malawi, as some people take such nets and use them for fishing, and hence they may not have used the ITNs for the intended malaria-prevention purposes. This then may partially help to explain such a high risk of malaria among under-fives with respect to uptake of ITN intervention. However, most studies done in Malawi like that of Chirombo, Lowe and Kazembe, (2014); Chitunhu and Musenge, (2016), and other

studies done elsewhere (Deressa, Hailemariam and Ali, 2007; Mmbando *et al.*, 2011) have shown that having ITN is associated with less risk of a child having malaria.

4.3 Predictive Malaria Risk map

The predictive risk map in **Figure 3-6** showed that modelled malaria hotspots were mostly in the central region clusters and south-eastern region clusters. This result fully agrees with findings of the predictive malaria risk maps from the studies done by Kazembe *et al.*, (2006); Chirombo, Lowe and Kazembe, (2014); Chitunhu and Musenge, (2016). As per the study done by Chirombo and others (2014), the central region is mainly covered by large portions of inland plain land and low-altitude areas especially those areas along lake Malawi.

4.4 Strengths and Limitations of the Study

4.4.1 Strengths of the study

The overall strengths of this study were the following:

1) The study used nationally representative data from MMIS which had 90% response, and as such, the results may be generalised to the whole Malawi nation under-five children
2) This study used data from four time points (2010, 2012, 2014, and 2017), and thus it was possible to make comparisons on changes in under-five malaria prevalence trends across the survey years.
3) This study used Bayesian modelling approach which helps in incorporating prior information, as well as categorizing spatial random effects as structured random effects and unstructured random effects. This helped to explain variations in malaria morbidity spatially rather than treating such random effects in a non-spatial way.

4) Most studies done in Malawi have results generated at district-level which may be prone to overgeneralization to such "large geographical" administrative level areas. This study used cluster-level data which are smaller in geographical size, hence accounting for malaria risk differences at a finer geographical scale.
5) The spatial-temporal model results from this study helped in generating predictive risk maps, which may help policy makers and health professionals to devise and implement interventions that may target specific malaria high risk areas.

4.4.2 Limitations of the study

The overall limitations of this study were the following:

1) Some important environmental and climatic factors that are related to malaria risk such as Enhanced Vegetation Index (EVI), rainfall, temperature were not included in the analysis of this study because these variables were not available in the MMIS for all years
2) The data used in this study were from the original cross sectional MMIS, and as such, causal associations could not be established between variables and malaria outcome.
3) Some clusters which may have been sampled in a particular survey year were not sampled in some other survey year. Hence, it was difficult to create a shape file that may cater such anomalies. As such, the overall Admin-level 2 shape file was used which contained all the 256 clusters. This resulted in over-smoothing of the model estimated which may in turn result in biased cluster-level analysis.
4) The data for this study were from four time points only and these time points were spaced in two-three years apart. Hence, more pronounced temporal trends could not be fully observed. Furthermore, this limitation hindered the incorporation of Time Series analysis methods which cater for seasonality trends. Malaria risk depends also on the seasons of the year, so this phenomenon could not be explored in this study.

5) Missing data especially for the variables: Mother's education, whether Mother heard about Malaria, and Use of nets. The analysis done in this study used the data "as it was", hence no methods such as multiple imputation were used to cater the problem of missing data.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

In terms of the overall trends of malaria prevalence, there is a general decrease in malaria prevalence at national level. However, at cluster level, malaria is still high in most parts of the central region and some south-eastern parts of Malawi.

The significant predictors of malaria as per the spatial-temporal model were found to be: Age of the child, Place of residence, Wealth Index, Mother's education, and Cluster altitude.

In modelling under-five malaria morbidity in Malawi over time, Bayesian Spatial-temporal models provide better fitting models than models that do not account for spatial and temporal random effects.

5.2 Recommendations and Scope for Future Research

5.2.1 Recommendations to Policy Makers and Modellers

- More targeted malaria interventions to help reduce malaria morbidity should be concentrated in the central region, areas along Lake Malawi, and areas of the south-eastern part of Malawi
- There is need to scale up the distribution of ITNs, as well as fully implementing IRS initiative at national level.
- More programs need to be initiated to improve the Socio-Economic Status (SES) especially to people with low economic status. This may in turn help improve the livelihoods of such people, hence leading to them having easier access malaria prevention interventions which may in turn result in reduced under-five malaria morbidity

- There is need to implement a consistent and sustainable malaria monitoring systems at smaller geographical areas across the whole country of Malawi that routinely collect data throughout the year. This will help monitor the trends of malaria throughout the year in such smaller geographical areas, hence providing a picture of whether substantial progress is made in reducing malaria morbidity in under-five children
- There is need to intensify the use of Bayesian spatial-temporal modelling approaches in understanding the spatial and temporal variations of under-five malaria in Malawi.
- Incorporation of predictive risk maps in modelling under-five malaria burden should also be intensified as these may help policy makers to pinpoint which areas need more scaling up of malaria interventions

5.2.2 Scope for Future Research

- There is need for studies that incorporate Time Series analysis methods to fully understand especially the temporal and seasonality trends of malaria
- Future spatial-temporal studies in Malawi may need to explore more on the climatic factors associated with malaria, to further substantiate the work done by Lowe and others (2013).
- Further under-five malaria studies are needed that may incorporate missing data analysis techniques so as to reduce biased modelling and interpretation of results.

It was the hope of the researcher of this study that the findings of this research may be significantly useful to national policy makers and policy implementers in their efforts of combat under-five malaria in Malawi, so that the goal of "Malaria-Free Malawi" may be achieved.

REFERENCES

Alegana, V. A. *et al.* (2013) 'Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial-temporal models', *Spatial and Spatio-temporal Epidemiology*. Elsevier, 7, pp. 25–36. doi: 10.1016/j.sste.2013.09.001.

Arab, A. (2015) 'Spatial and spatio-temporal models for modeling epidemiological data with excess zeros', *International Journal of Environmental Research and Public Health*. MDPI AG, pp. 10536–10548. doi: 10.3390/ijerph120910536.

Banerjee S, Carlin B.P. and Gelfand A.E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. 2nd edn. Available at: https://www.routledge.com/Hierarchical-Modeling-and-Analysis-for-Spatial-Data/Banerjee-Carlin-Gelfand/p/book/9781439819173 (Accessed: 29 May 2021).

Blangiardo, M. *et al.* (2013) 'Spatial and spatio-temporal models with R-INLA', *Spatial and Spatio-temporal Epidemiology*, pp. 33–49. doi: 10.1016/j.sste.2012.12.001.

Brezger, A., Kneib, T. and Lang, S. (2012) 'BayesX: analysing Bayesian structured additive regression models'. Available at: http://hdl.handle.net/10419/31014www.econstor.eu (Accessed: 29 May 2021).

Chipeta, M. G. *et al.* (2019) 'Geostatistical analysis of Malawi's changing malaria transmission from 2010 to 2017 [version 2; peer review: 2 approved, 1 approved with reservations]', *Wellcome Open Research*. F1000 Research Ltd, 4. doi: 10.12688/wellcomeopenres.15193.2.

Chirombo, J., Lowe, R. and Kazembe, L. (2014) 'Using structured additive regression models to estimate risk factors of malaria: Analysis of 2010 Malawi malaria indicator survey data', *PLoS ONE*. Public Library of Science, 9(7), p. e101116. doi: 10.1371/journal.pone.0101116.

Chitunhu, S. and Musenge, E. (2016) 'Spatial and socio-economic effects on malaria morbidity in children under 5 years in Malawi in 2012', *Spatial and Spatio-temporal Epidemiology*, 16, pp. 21–33. doi: 10.1016/j.sste.2015.11.001.

Clouston, S. A. P., Yukich, J. and Anglewicz, P. (2015) 'Social inequalities in malaria knowledge, prevention and prevalence among children under 5 years old and women aged 15-49 in Madagascar', *Malaria Journal*. BioMed Central, 14(1), p. 499. doi: 10.1186/s12936-015-1010-y.

Dale, P. *et al.* (2005) 'Malaria in Indonesia: A summary of recent research into its environmental relationships', *Southeast Asian Journal of Tropical Medicine and Public Health*. SEAMEO TROPMED Network, pp. 1–13. Available at: https://europepmc.org/article/med/15906636 (Accessed: 29 May 2021).

Deressa, W., Hailemariam, D. and Ali, A. (2007) 'Economic costs of epidemic malaria to households in rural Ethiopia', *Tropical Medicine and International Health*. John Wiley & Sons, Ltd, 12(10), pp. 1148–1156. doi: 10.1111/j.1365-3156.2007.01901.x.

Dzinjalamala, F. (2009) 'Epidemiology of malaria in Malawi', *Epidemiology of Malawi*, p. 21.

Fana, S. A. *et al.* (2015) 'Prevalence and risk factors associated with malaria infection among pregnant women in a semi-urban community of north-western Nigeria', *Infectious Diseases of Poverty*. BioMed Central Ltd., 4(1). doi: 10.1186/s40249-015-0054-0.

Graves, P. M. *et al.* (2009) 'Individual, household and environmental risk factors for malaria infection in Amhara, Oromia and SNNP regions of Ethiopia', *Transactions of the Royal Society of Tropical Medicine and Hygiene*. Trans R Soc Trop Med Hyg, 103(12), pp. 1211–1220. doi: 10.1016/j.trstmh.2008.11.016.

Kazembe, L. *et al.* (2006) 'Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data', *International Journal of Health Geographics*, 5.

Knorr-Held, L. and Besag, J. (1998) 'Modelling risk from a disease in time and space', *Statistics in Medicine*, 17(18), pp. 2045–2060. doi: 10.1002/(SICI)1097-0258(19980930)17:18<2045::AID-SIM943>3.0.CO;2-P.

Kulldorff, M. *et al.* (2006) 'An elliptic spatial scan statistic', *Statistics in Medicine*. John Wiley & Sons, Ltd, 25(22), pp. 3929–3943. doi: 10.1002/sim.2490.

Lawson, A. B. (2008) *Bayesian Disease Mapping*, *Bayesian Disease Mapping*. Chapman and Hall/CRC. doi: 10.1201/9781584888413.

Lindgren, F. and Rue, H. (2008) 'On the second-order random walk model for irregular locations', *Scandinavian Journal of Statistics*. John Wiley & Sons, Ltd, 35(4), pp. 691–700. doi: 10.1111/j.1467-9469.2008.00610.x.

Lowe, R., Chirombo, J. and Tompkins, A. M. (2013) 'Relative importance of climatic, geographic and socio-economic determinants of malaria in Malawi', *Malaria Journal*. BioMed Central, 12(1), p. 416. doi: 10.1186/1475-2875-12-416.

Mabaso, M. L. H. *et al.* (2006) 'Spatio-temporal analysis of the role of climate in inter-annual variation of malaria incidence in Zimbabwe', *International Journal of Health Geographics*. BioMed Central, 5(1), pp. 1–9. doi: 10.1186/1476-072X-5-20.

Mabaso, M. L. H., Sharp, B. and Lengeler, C. (2004) 'Historical review of malarial control in southern African with emphasis on the use of indoor residual house-spraying', *Tropical Medicine and International Health*. Trop Med Int Health, pp. 846–856. doi: 10.1111/j.1365-3156.2004.01263.x.

Manda, S. *et al.* (2015) 'Mapping HIV prevalence using population and antenatal sentinel-based HIV surveys: A multi-stage approach', *Population Health Metrics*. BioMed Central Ltd., 13(1), pp. 1–15. doi: 10.1186/s12963-015-0055-z.

Mariella, L. and Tarantino, M. (2016) 'Spatial Temporal Conditional Auto-Regressive Model: A New Autoregressive Matrix', *Austrian Journal of Statistics*. Austrian Statistical Society, 39(3), pp. 223–244. doi: 10.17713/ajs.v39i3.246.

Mmbando, B. P. *et al.* (2011) 'Spatial variation and socio-economic determinants of Plasmodium falciparum infection in northeastern Tanzania', *Malaria Journal*. BioMed Central, 10(1), pp. 1–9. doi: 10.1186/1475-2875-10-145.

Montosi, E. *et al.* (2012) 'An ecohydrological model of malaria outbreaks', *Hydrology and Earth System Sciences*, 16(8), pp. 2759–2769. doi: 10.5194/hess-16-2759-2012.

Moran, P. A. P. (1950) 'Notes on Continuous Stochastic Phenomena', *Biometrika*. JSTOR, 37(1/2), p. 17. doi: 10.2307/2332142.

Musenge, E. *et al.* (2013) 'The contribution of spatial analysis to understanding HIV/TB mortality in children: a structural equation modelling approach.', *Global health action*, 6, p.

19266. doi: 10.3402/gha.v6i0.19266.

Musenge, E. (2014) 'Modelling spatiotemporal patterns of childhood HIV/TB related mortality and malnutrition: applications to Agincourt data in rural South Africa', *undefined*.

Musenge, E., Vounatsou, P. and Kahn, K. (2011) 'Space-time confounding adjusted determinants of child HIV/TB mortality for large zero-inflated data in rural South Africa', *Spatial and Spatio-temporal Epidemiology*. Spat Spatiotemporal Epidemiol, pp. 205–217. doi: 10.1016/j.sste.2011.07.001.

Njau, J. D. *et al.* (2006) 'Fever treatment and household wealth: The challenge posed for rolling out combination therapy for malaria', *Tropical Medicine and International Health*. John Wiley & Sons, Ltd, pp. 299–313. doi: 10.1111/j.1365-3156.2006.01569.x.

NMCP (2011) *Malawi Malaria Indicator Survey (MIS) 2010*. Lilongwe. Available at: https://files.givewell.org/files/DWDA 2009/AMF/Malawi_MIS_2010_Final.pdf.

NMCP and ICF (2012) *Malawi Malaria Indicator Survey (MIS) 2012*. Lilongwe. Available at: https://dhsprogram.com/publications/publication-MIS13-MIS-Final-Reports.cfm.

NMCP and ICF (2014) *Malawi Malaria Indicator Survey (MIS) 2014*. Lilongwe. Available at: https://dhsprogram.com/publications/publication-mis18-mis-final-reports.cfm.

NMCP and ICF (2018) *Malawi Malaria Indicator Survey (MIS) 2017*. Lilongwe. Available at: https://dhsprogram.com/publications/publication-mis28-mis-final-reports.cfm (Accessed: 29 May 2021).

Noor, A. M. *et al.* (2014) 'The changing risk of Plasmodium falciparum malaria infection in Africa: 2000-10: A spatial and temporal analysis of transmission intensity', *The Lancet*. Elsevier B.V., 383(9930), pp. 1739–1747. doi: 10.1016/S0140-6736(13)62566-0.

NSO and ICF (2011) *Malawi Demographic and Health Survey 2010*. Zomba. Available at: http://www.measuredhs.com. (Accessed: 29 May 2021).

Ord, J. K. and Getis, A. (1995) 'Local Spatial Autocorrelation Statistics: Distributional Issues and an Application', *Geographical Analysis*. John Wiley & Sons, Ltd, 27(4), pp. 286–306. doi: 10.1111/j.1538-4632.1995.tb00912.x.

Ord, J. K. and Getis, A. (2001) 'Testing for local spatial autocorrelation in the presence of global autocorrelation', *Journal of Regional Science*. John Wiley & Sons, Ltd, 41(3), pp. 411–432. doi: 10.1111/0022-4146.00224.

Roberts, D. and Matthews, G. (2016) 'Risk factors of malaria in children under the age of five years old in Uganda', *Malaria Journal*. BioMed Central, 15(1), p. 246. doi: 10.1186/s12936-016-1290-x.

Rue, H., Martino, S. and Chopin, N. (2009) 'Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations', *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 71(2), pp. 319–392. doi: 10.1111/j.1467-9868.2008.00700.x.

Shimaponda, N. M. (2015) *Socio-Economic and Eco-Environmental Determinants of Malaria in Four Malaria Endemic Provinces of Zambia*. Available at: https://researchspace.ukzn.ac.za/handle/10413/14382 (Accessed: 29 May 2021).

Soares Magalhães, R. J. *et al.* (2013) 'Role of malnutrition and parasite infections in the

spatial variation in children's anaemia risk in northern Angola', *Geospatial health*. University of Naples Federico II, 7(2), p. 341. doi: 10.4081/gh.2013.91.

Ssempiira, J. *et al.* (2017) 'The contribution of malaria control interventions on spatio-temporal changes of parasitaemia risk in Uganda during 2009-2014', *Parasites and Vectors*. BioMed Central Ltd., 10(1), pp. 1–13. doi: 10.1186/s13071-017-2393-0.

Torabi, M. (2017) 'Zero-inflated spatio-temporal models for disease mapping', *Biometrical Journal*. Wiley-VCH Verlag, 59(3), pp. 430–444. doi: 10.1002/bimj.201600120.

Wambui, K. M. and Musenge, E. (2019) 'A space-time analysis of recurrent malnutrition-related hospitalisations in Kilifi, Kenya for children under-5 years', *BMC Nutrition*. BioMed Central, 5(1), pp. 1–9. doi: 10.1186/s40795-019-0296-5.

WHO (2017) *World malaria report 2017*. World Health Organization. Available at: http://www.who.int/malaria/publications/world-malaria-report-2017/en/ (Accessed: 1 September 2018).

World Bank (2018) *The World Bank Data: Malawi*. Available at: https://data.worldbank.org/country/MW (Accessed: 29 May 2021).

Yeka, A. *et al.* (2012) 'Malaria in Uganda: Challenges to control on the long road to elimination. I. Epidemiology and current control efforts', *Acta Tropica*. NIH Public Access, 121(3), pp. 184–195. doi: 10.1016/j.actatropica.2011.03.004.

Zgambo, M., Mbakaya, B. C. and Kalembo, F. W. (2017) 'Prevalence and factors associated with malaria parasitaemia in children under the age of five years in Malawi: A comparison study of the 2012 and 2014 Malaria Indicator Surveys (MISs)', *PLoS ONE*, 12(4). doi: 10.1371/journal.pone.0175537.

APPENDICES

Appendix 1: Plagiarism declaration report



PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I, **Vitumbiko Chijere Chirwa**, (Student number: **1756102**) am a student registered for the degree of **MSc Epidemiology (Biostatistics)** in the academic year **2021** .

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: _____ Date: **29th May 2021**

Appendix 2: Turnitin Report



1756102:Vitu_thesis_for_turn_it_in_May_2021.pdf

ORIGINALITY REPORT

12%	9%	9%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ibssusan2019.samrc.ac.za Internet Source	2%
2	hdl.handle.net Internet Source	1%
3	malariajournal.biomedcentral.com Internet Source	1%
4	www.researchsquare.com Internet Source	1%
5	wellcomeopenresearch.org Internet Source	<1%
6	orbi.uliege.be Internet Source	<1%
7	www.mdpi.com Internet Source	<1%
8	digitalcommons.usu.edu Internet Source	<1%
9	Simangaliso Chitunhu, Eustasius Musenge. "Spatial and socio-economic effects on malaria morbidity in children under 5years in	<1%

Appendix 3: Ethics Approval Certificate

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R14/49 Mr VC Chirwa

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M181003

NAME: (Principal Investigator)	Mr VC Chirwa
DEPARTMENT:	School of Public Health Medical School University
PROJECT TITLE:	Malaria morbidity in under-five children in Malawi: a spatial-temporal Bayesian model
DATE CONSIDERED:	26/10/2018
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Professor E Musenge
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	28/11/2018

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on 3rd floor, Phillip V Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.
I/We fully understand the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to resubmit to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date of the meeting when the study was initially reviewed. In this case, the study was initially reviewed in October and will therefore reports and re-certification will be due early in the month of October each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature _____ Date _____

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix 4: 2010 MMIS data authorization

Telegrams: MINMED, Lilongwe

Telephone: 01 789 400
Fax:

Communications should be addressed to:
Secretary for Health



Reply please quote Ref Med.....
Ministry of Health
P.O. Box 30377,
Lilongwe 3,
MALAWI

REF: CHSU/MIS/1/10

05th November 2018

To: Vitumbiko C. Chirwa
University of the Witwatersrand
Parktown 2193
South Africa

Dear Vitumbiko C. Chirwa,

AUTHORIZATION TO USE 2010 MALAWI MALARIA INDICATOR SURVEY (MMIS) DATASET

Further to your request letter dated 30th October 2018, I am pleased to inform you that you have been granted permission to access and use the 2010 MMIS data for your MSc degree research titled **Malaria morbidity in under-five children in Malawi: A spatial-temporal Bayesian model**.

In accordance with our ethics, all MMIS data should be treated as confidential. As such, you are required not to make any effort to identify any household or individual respondent interviewed in the survey. To further necessitate the confidentiality of the data, you will be given the MMIS data with anonymised unique identifiers. There are no names of individual participants or households in the data.

Please be informed that you are required to use the 2010 MMIS data only for the intended purpose of your MSc degree research. Should there be any further enquiries, please contact the undersigned.

Sincerely,

Austin Gumbo
Monitoring and Evaluation Head
Cell: +265 888 319 504/+265 999 071 871,
Email: aagumbo@yahoo.co.uk,
Skype: austin.gumbo2

Appendix 5: 2012, 2014, 2017 MMIS data authorization



Jun 25, 2018

Vitumbiko Chirwa
University of the Witwatersrand
South Africa
Phone: +27 833259569
Email: cvchijere@gmail.com
Request Date: 06/25/2018

Dear Vitumbiko Chirwa:

This is to confirm that you are approved to use the following Survey Datasets for your registered research paper titled: "Master of Science in Epidemiology, field of Biostatistics":

Malawi

To access the datasets, please login at: https://www.dhsprogram.com/data/dataset_admin/login_main.cfm. The user name is the registered email address, and the password is the one selected during registration.

The IRB-approved procedures for DHS public-use datasets do not in any way allow respondents, households, or sample communities to be identified. There are no names of individuals or household addresses in the data files. The geographic identifiers only go down to the regional level (where regions are typically very large geographical areas encompassing several states/provinces). Each enumeration area (Primary Sampling Unit) has a PSU number in the data file, but the PSU numbers do not have any labels to indicate their names or locations. In surveys that collect GIS coordinates in the field, the coordinates are only for the enumeration area (EA) as a whole, and not for individual households, and the measured coordinates are randomly displaced within a large geographic area so that specific enumeration areas cannot be identified.

The DHS Data may be used only for the purpose of statistical reporting and analysis, and only for your registered research. To use the data for another purpose, a new research project must be registered. All DHS data should be treated as confidential, and no effort should be made to identify any household or individual respondent interviewed in the survey. Please reference the complete terms of use at: https://dhsprogram.com/Data/terms-of-use.cfm.

The data must not be passed on to other researchers without the written consent of DHS. Users are required to submit an electronic copy (pdf) of any reports/publications resulting from using the DHS data files to: archive@dhsprogram.com.

Sincerely,

Bridgette Wellington

Bridgette Wellington
Data Archivist
The Demographic and Health Surveys (DHS) Program

Appendix 6: Analysis codes

A6.1 Multilevel Logistic Regression models for each year (Stata)

```
use "C:\Users\Vitumbiko C. Chirwa\Documents\MSC EPI
(BIOSTATISTICS\Thesis_issues\datasets\data analysis
thesis\MODELLING\FINAL_MALARIA_USE_MODELLING.dta"


svyset ID_2, strata(strata) vce(linearized)
singleunit(missing) || hhid, weight(sampweight)


svy linearized : melogit mal agemonths i.region i.residence
i.wealth i.motheredu   i.heardmalaria altitude2 i.ITNNET  if
year==1, || ID_2:, covariance(unstructured) or


svy linearized : melogit mal agemonths i.region i.residence
i.wealth i.motheredu   i.heardmalaria altitude2 i.ITNNET  if
year==2, || ID_2:, covariance(unstructured) or


svy linearized : melogit mal agemonths i.region i.residence
i.wealth i.motheredu   i.heardmalaria altitude2 i.ITNNET  if
year==3, || ID_2:, covariance(unstructured) or


svy linearized : melogit mal agemonths i.region i.residence
i.wealth i.motheredu   i.heardmalaria altitude2 i.ITNNET  if
year==4, || ID_2:, covariance(unstructured) or
```

A6.2 Multilevel Negative Binomial Regression model (Stata)

```
saveold "C:\Users\Vitumbiko C. Chirwa\Documents\MSC EPI
(BIOSTATISTICS\Thesis_issues\datasets\data analysis
thesis\MODELLING\FINAL_COLLAPSED_DATA.dta", replace


menbreg mal agemonths i.region i.residence wealth_mid_prop
wealth_rich_prop educ_prim_prop educ_sectert_prop
heard_yes_prop altitude2 INNET_noITN_prop INNET_yesITN_prop ,
exposure(numchildren) || ID_2:, irr
```

A6.3 Bayesian Negative Binomial Regression model (R-INLA)

```
########SETTING UP PACKAGES#######
###INSTALL PACKAGES
install.packages("rgeos")
install.packages("maptools")
install.packages("shapefiles")
install.packages("shp2graph")
install.packages("choroplethr")
install.packages("choroplethrMaps")
install.packages("ggplot2")
install.packages("spdep")
install.packages("rgdal")
install.packages("stringi")
install.packages("stringr")
install.packages("spam")
install.packages("XML")
install.packages("acs")
install.packages("fields")
install.packages("jpeg")
install.packages("checkmate")

require(spData)
require(sf)
require("spam")
require("igraph")
require("maps")
require("stringr")
require("XML")
require("acs")


library(rgeos)
library(maptools)
library(ggplot2)
library(INLA)
library(spdep)
library(foreign)
library(sp)
library(rgdal)
library(lattice)
library(shapefiles)

## coloring the spplot
library(colorspace)
library(coda)
library(fields)
library(shp2graph)
library(choroplethr)
library(choroplethrMaps)
library(fillmap)
library(RColorBrewer)


names (inla.models()$likelihood)

#######READ MAP######
set.seed(123)
setwd("~/MSC EPI (BIOSTATISTICS/Thesis_issues/datasets/data analysis
thesis/MODELLING")
Malawi <- readOGR("~/MSC EPI (BIOSTATISTICS/Thesis_issues/datasets/data
analysis thesis/MODELLING","MWI_adm2")
#Malawi<- readShapePoly("MWI_adm2.shp", IDvar="ID_2",
proj4string=CRS("+proj=longlat +ellps=WGS84"))

plot(Malawi, border="black", axes=TRUE, las=1)
str(Malawi, 2)
nbmala<-poly2nb(Malawi)
nb2INLA(file="Malawi.graph",nbmala)
```

```
data<-read.dta("FINAL_COLLAPSED_DATA.dta")
data<-
cbind(data,cluster.unstruct=as.numeric(data$ID_2),cluster.struct=as.numeri
c(data$ID_2))
attach(data)
View(data)


m = get("inla.models", INLA:::inla.get.inlaEnv())
m$latent$rw2$min.diff = NULL
assign("inla.models", m, INLA:::inla.get.inlaEnv())

##########MODEL BUILDING###########
#1) MODEL 1: no random effects
norand<-mal~as.factor(region)+as.factor(residence)+altitude2+
wealth_mid_prop+wealth_rich_prop+INNET_noITN_prop+INNET_yesITN_prop+heard_
yes_prop+ educ_prim_prop+educ_sectert_prop+f(agemonths,model="rw2")
resultnorand<-
inla(norand,family="nbinomial",data=data,control.compute=list(dic=TRUE,cpo
=TRUE),offset = log(numchildtested))
summary(resultnorand)
resultnorand$dic$dic;resultnorand$dic$p.eff
betas = resultnorand$summary.fixed
plot(resultnorand)
exp(betas)
age=resultnorand$summary.random$agemonths

plot(resultnorand$summary.random$agemonths$ID,resultnorand$summary.random$
agemonths[,5],xlab="Age in Months",ylab="effect")

#2) MODEL 2: SPATIAL ONLY-UNSTRUCTURED
spatunstr<-
mal~f(agemonths,model="rw2")+as.factor(region)+as.factor(residence)+wealth
_mid_prop+wealth_rich_prop+
educ_prim_prop+educ_sectert_prop+heard_yes_prop+altitude2+INNET_noITN_prop
+INNET_yesITN_prop+f(ID_2,model="iid",prior = "normal", param = c(0,
0.001), initial = 1)
resultspatunstr<-
inla(spatunstr,family="nbinomial",data=data,control.compute=list(dic=TRUE,
cpo=TRUE),offset = log(numchildtested))
summary(resultspatunstr)
resultspatunstr$dic$dic;resultspatunstr$dic$p.eff
betas = resultspatunstr$summary.fixed
exp(betas)
plot(resultspatunstr)
age=resultspatunstr$summary.random$agemonths
plot(resultspatunstr$summary.random$agemonths$ID,resultspatunstr$summary.r
andom$agemonths[,5],xlab="Age in Months",ylab="effect", main="Spatial
Unstructured Random Effect model")


#3) MODEL 3: SPATIAL STRUCTURED##
spatstruc<-
mal~f(agemonths,model="rw2")+as.factor(region)+as.factor(residence)+wealth
_mid_prop+wealth_rich_prop+
educ_prim_prop+educ_sectert_prop+heard_yes_prop+altitude2+INNET_noITN_prop
+INNET_yesITN_prop+f(ID_2,model="besag",graph="Malawi.graph")
resultspatstruc<-
inla(spatstruc,family="nbinomial",data=data,control.compute=list(dic=TRUE,
cpo=TRUE), offset = log(numchildtested))
summary(resultspatstruc)
resultspatstruc$dic$dic;resultspatstruc$dic$p.eff
psi= resultspatstruc$summary.fixed
exp(psi)
plot(resultspatstruc)

age=resultspatstruc$summary.random$agemonths
```

```
plot(resultspatstruc$summary.random$agemonths$ID,resultspatstruc$summary.r
andom$agemonths[,5],xlab="Age in Months",ylab="effect", main="Spatial
Structured Random Effect model")


#4) MODEL 4: SPATIAL UNSTRUCTURED AND STRUCTURED##
spat_unstru_struc<-
mal~f(agemonths,model="rw2")+as.factor(region)+as.factor(residence)+wealth
_mid_prop+wealth_rich_prop+
educ_prim_prop+educ_sectert_prop+heard_yes_prop+altitude2+INNET_noITN_prop
+INNET_yesITN_prop+f(ID_2,model="bym",graph="Malawi.graph", scale.model =
TRUE, hyper = list(prec.unstruct=list(prior = "loggamma",param = c(0.0011,
0.001)), prec.spatial=list(prior = "loggamma",param = c(0.0011, 0.001))))
resultspat_unstru_struc<-
inla(spat_unstru_struc,family="nbinomial",data=data,control.compute=list(d
ic=TRUE,cpo=TRUE), offset = log(numchildtested))
summary(resultspat_unstru_struc)
resultspat_unstru_struc$dic$dic;resultspat_unstru_struc$dic$p.eff
phi= resultspat_unstru_struc$summary.fixed
plot(resultspat_unstru_struc)
exp(phi)
age=resultspat_unstru_struc$summary.random$agemonths
plot(resultspat_unstru_struc$summary.random$agemonths$ID,resultspat_unstru
_struc$summary.random$agemonths[,5],xlab="Age in Months",ylab="effect",
main="Unstructured,Structured Random Effects model")


#5) MODEL 5: SPATIOTEMPORAL I.I.D
#
spattemporal_IID<-
mal~f(agemonths,model="rw2")+as.factor(region)+as.factor(residence)+wealth
_mid_prop+wealth_rich_prop+
educ_prim_prop+educ_sectert_prop+heard_yes_prop+altitude2+INNET_noITN_prop
+INNET_yesITN_prop+f(ID_2,model="bym",graph="Malawi.graph", scale.model =
TRUE, hyper = list(prec.unstruct=list(prior = "loggamma",param = c(0.0011,
0.001)), prec.spatial=list(prior = "loggamma",param = c(0.0011,
0.001))))+f(year,model="iid")
resultspattemp_IDD<-
inla(spattemporal_IID,family="nbinomial",data=data,control.results=list(re
turn.marginals.random=TRUE,return.marginals.predictor=TRUE),control.comput
e=list(dic=TRUE,cpo=TRUE,graph=TRUE), offset = log(numchildtested))
summary(resultspattemp_IDD)
resultspattemp_IDD$dic$p.eff

rand<-resultspattemp_IDD$summary.random
thetas = resultspattemp_IDD$summary.fixed
exp(thetas)
plot(resultspattemp_IDD)

age=resultspattemp_IDD$summary.random$agemonths
plot(resultspattemp_IDD$summary.random$agemonths$ID,resultspattemp_IDD$sum
mary.random$agemonths[,5],xlab="Age in Months",ylab="effect",
main="Spatiotemporal IID model")


#7) MODEL 6: SPATIOTEMPORAL RW1
#
spattemporal_RW1<-
mal~f(agemonths,model="rw2")+as.factor(region)+as.factor(residence)+wealth
_mid_prop+wealth_rich_prop+
educ_prim_prop+educ_sectert_prop+heard_yes_prop+altitude2+INNET_noITN_prop
+INNET_yesITN_prop+f(ID_2,model="bym",graph="Malawi.graph", scale.model =
TRUE, hyper = list(prec.unstruct=list(prior = "loggamma",param = c(0.0011,
0.001)), prec.spatial=list(prior = "loggamma",param = c(0.0011,
0.001))))+f(year,model="ar1")
resultspattemp_RW1<-
inla(spattemporal_RW1,family="nbinomial",data=data,control.results=list(re
turn.marginals.random=TRUE,return.marginals.predictor=TRUE),control.comput
e=list(dic=TRUE,cpo=TRUE,graph=TRUE), offset = log(numchildtested))
```

```
summary(resultspattemp_RW1)
resultspattemp_RW1$dic$p.eff

rand<-resultspattemp_RW1$summary.random
omegas = resultspattemp_RW1$summary.fixed
exp(omegas)
plot(resultspattemp_RW1)

age=resultspattemp_RW1$summary.random$agemonths
plot(resultspattemp_RW1$summary.random$agemonths$ID,resultspattemp_RW1$sum
mary.random$agemonths[,5],xlab="Age in Months",ylab="effect",
main="Spatiotemporal RW1 model")

###PLOTTING MAPS
vitu <- resultspattemp_RW1$marginals.random$ID_2[1:256]
zeta <- lapply(vitu, function(x) inla.emarginal(exp,x))
#zeta.cutoff <- c(0.83, 0.9, 0.95, 0.999, 1, 1.01, 1.05,1.1, 1.2)
zeta.cutoff <- c(0.1, 0.5, 0.75, 1.0, 1.5, 2.5)


cat.zeta <- cut(unlist(zeta), breaks = zeta.cutoff,include.lowest = TRUE)
maps.cat.zeta <- data.frame(anyDuplicated(data$ID_2), cat.zeta = cat.zeta)
data.malawi <- attr(Malawi, "data")
attr(Malawi, "data") <- merge(data.malawi, maps.cat.zeta, by.x = "ID_2",
by.y = "anyDuplicated.data.ID_2.")

spplot(obj = Malawi, zcol = "cat.zeta", col.regions =
diverge_hsv(8),scales = list(draw = TRUE), asp = 1)

#7) MODEL 7: SPATIOTEMPORAL IID interaction
year2<-year
ID.area.int <- data$ID_2
ID.year.int <- data$year
MODEL7 <-
mal~f(agemonths,model="rw2")+as.factor(region)+as.factor(residence)+wealth
_mid_prop+wealth_rich_prop+
educ_prim_prop+educ_sectert_prop+heard_yes_prop+altitude2+INNET_noITN_prop
+INNET_yesITN_prop +
  f(ID_2, model = "bym"  ,graph="Malawi.graph" , scale.model=TRUE,
    hyper=list(prec.unstruct=list(prior="loggamma",param=c(0.001,0.001)),
               prec.spatial=list(prior="loggamma",param=c(0.1,0.01)))) +
  f(year, model = "iid") + f(ID.year.int,model="iid", group=ID.area.int,
                             control.group=list(model="besag",
                                                graph="Malawi.graph"))
resultsMODEL7<-
inla(MODEL7,family="nbinomial",data=data,control.results=list(return.margi
nals.random=TRUE,return.marginals.predictor=TRUE),control.compute=list(dic
=TRUE,cpo=TRUE,graph=TRUE), offset = log(numchildtested))

summary(resultsMODEL7)
resultsMODEL7$dic$p.eff
rand<-resultsMODEL7$summary.random
kappas = resultsMODEL7$summary.fixed
exp(kappas)
plot(resultsMODEL7)

age=resultsMODEL7$summary.random$agemonths
plot(resultsMODEL7$summary.random$agemonths$ID,resultsMODEL7$summary.rando
m$agemonths[,5],xlab="Age in Months",ylab="effect")

#8) MODEL 8: SPATIOTEMPORAL AR1 interaction
ID.area.int <- data$ID_2
ID.year.int <- data$year
MODEL8 <-
mal~f(agemonths,model="rw2")+as.factor(region)+as.factor(residence)+wealth
_mid_prop+wealth_rich_prop+
educ_prim_prop+educ_sectert_prop+heard_yes_prop+altitude2+INNET_noITN_prop
+INNET_yesITN_prop +
  f(ID_2, model = "bym"  ,graph="Malawi.graph" , scale.model=TRUE,
```

```
    hyper=list(prec.unstruct=list(prior="loggamma",param=c(0.001,0.001)),
               prec.spatial=list(prior="loggamma",param=c(0.1,0.01)))) +
  f(year, model = "ar1") + f(ID.year.int,model="iid", group=ID.area.int,
                             control.group=list(model="besag",
                                                graph="Malawi.graph"))
resultsMODEL8<-
inla(MODEL8,family="nbinomial",data=data,control.results=list(return.margi
nals.random=TRUE,return.marginals.predictor=TRUE),control.compute=list(dic
=TRUE,cpo=TRUE,graph=TRUE), offset = log(numchildtested))

summary(resultsMODEL8)
resultsMODEL8$dic$p.eff
rand<-resultsMODEL8$summary.random
khi = resultsMODEL8$summary.fixed
exp(khi)
plot(resultsMODEL8)

age=resultsMODEL8$summary.random$agemonths
plot(resultsMODEL8$summary.random$agemonths$ID,resultsMODEL8$summary.rando
m$agemonths[,5],xlab="Age in Months",ylab="effect", main="Spatiotemporal
Interaction model")


plot(resultsMODEL8, plot.fixed.effects = TRUE, constant = FALSE,
     plot.lincomb = TRUE, plot.random.effects = TRUE,
     plot.hyperparameters = TRUE, plot.predictor = TRUE,
     plot.q = TRUE, plot.cpo = TRUE, single = TRUE)

plot(resultsMODEL8, plot.fixed.effects = TRUE, constant = FALSE,
     plot.cpo = F, single = F)

###PLOTTING OVERALL RISK MAP
vitu <- resultsMODEL8$marginals.random$ID_2[1:256]
zeta <- lapply(vitu, function(x) inla.emarginal(exp,x))
#zeta.cutoff <- c(0.83, 0.9, 0.95, 0.999, 1, 1.01, 1.05,1.1, 1.2)
zeta.cutoff <- c(0.1, 0.5, 0.75, 1.0, 1.5, 2.5)


cat.zeta <- cut(unlist(zeta), breaks = zeta.cutoff,include.lowest = TRUE)
maps.cat.zeta <- data.frame(anyDuplicated(data$ID_2), cat.zeta = cat.zeta)
data.malawi <- attr(Malawi, "data")
attr(Malawi, "data") <- merge(data.malawi, maps.cat.zeta, by.x = "ID_2",
by.y = "anyDuplicated.data.ID_2.")

spplot(obj = Malawi, zcol = "cat.zeta", col.regions =
diverge_hsv(8),scales = list(draw = TRUE), asp = 1)

#fillmap(Malawi,"Spatial Pattern for Malawi Malaria Prevalence Risk",
UH,n.col=4,leg.loc = "bottomright", leg.cex = 0.9, main.cex = 1.4)

##############################################
###TEMPORAL EFFECTS MAP FROM INTERACTION MODEL
##############################################
delta.intIII <-
data.frame(delta=exp(resultsMODEL8$summary.random$ID.year.int[,2]),tempC=r
ep(1:4, each = 256),

ID.area=resultsMODEL8$summary.random$ID.year.int[,1])
delta.intIII.matrix <- matrix(delta.intIII[,1], 256,4,byrow=FALSE)
rownames(delta.intIII.matrix)<- delta.intIII[1:256,3]

save.image("st_model3.RDA")


cutoff.interaction <- c(0.1,0.5,1,1.5,2)
data.mlw <- attr(Malawi, "data")
delta.intIII.factor <- data.frame(NAME=data.mlw$ID_2)
for(i in 1:4){
```

```
  delta.factor.temp <-
cut(delta.intIII.matrix[,i],breaks=cutoff.interaction,include.lowest=TRUE
)
  delta.intIII.factor <- cbind(delta.intIII.factor,delta.factor.temp)
}
colnames(delta.intIII.factor)<- c("NAME",seq(1,4))

# *** Code for Figure 7.6
attr(Malawi, "data") <- data.frame(data.mlw,  intIII=delta.intIII.factor)
trellis.par.set(axis.line=list(col=NA))

png(filename=paste0("temp_","img.png") , width = 25.47 , height = 27.57 ,
units = "cm" , res=300)

spplot(obj=Malawi, zcol=c("intIII.1","intIII.2","intIII.3","intIII.4"),
col.regions=diverge_hsv(8), names.attr=seq(1,4),main="")
dev.off()
```