STATISTICAL MODELLING ON CHILDHOOD ANAEMIA, MALARIA AND STUNTING IN MALAWI, LESOTHO, AND BURUNDI

By

Rugiranka Tony Gaston

(Student number: 210511400)

A dissertation in fulfillment of the requirement of the degree of

Doctoral of Philosophy (PhD)

in

Applied Statistics

in the School of Mathematics, Statistics, and Computer Science

University of KwaZulu-Natal, Pietermaritzburg Campus, South Africa

Supervisors: Prof Shaun Ramroop

Co-supervisor: Dr Faustin Habyarimana

March 2023

DECLARATION

I, **Rugiranka Tony Gaston**, hereby confirm that this thesis titled 'Statistical modelling of childhood anaemia, malaria, and stunting in Malawi, Burundi and Lesotho' is my original work. Apart from the recognised support of my supervisors and co-supervisor, my contribution, and where people's work was incorporated, it was fully acknowledged by the citations. The thesis has not been previously submitted by me to another institution to get a qualification award. Precisely, this work was carried out in the School of Mathematics, Statistics, and Computer Science Pietermaritzburg Campus, Department of Agriculture, Engineering and Science at the University of KwaZulu-Natal, South Africa.

…… *3 March 2023*

Candidate: Rugiranka Tony Gaston *Date*

…… *3 March 2023*

Supervisor: Prof Shaun Ramroop *Date*

…… *3 March 2023*

Co-Supervisor: Dr Faustin Habyarimana *Date*

DEDICATION

This thesis is dedicated to all of my family members, alive or dead. The dedication also goes to my beautiful wife Ayanda Mazomba-Gaston and our lovely daughter Anelisa Uwase Gaston for their unconditional love, being my pillars of strength and support.

ACKNOWLEDGEMENTS

I want to thank God for the gift of life and protection during this journey and everyone who provided a valuable contribution towards the achievement of the goal of this study: Many thanks go to my supervisors Prof Shaun Ramroop and Dr Faustin Habyarimana for providing supervision, encouragement, and moral support.

I would like to acknowledge the Demographic and Health Survey (DHS) and Malaria Indicator survey programs (MIS), for providing free access to the data used for this study. Many thanks go to Health Economics and HIV/AIDS Research Division (HEAD) for awarding me the scholarship during this study and all the staff members for their support and encouragement. I am thankful to the University of KwaZulu-Natal (UKZN), Pietermaritzburg Campus, School of Mathematics, Statistics and Computer Science, all staff members, and researchers for hosting the study, and for the opportunity to undertake this PhD study using their facilities.

Many thanks go to the Tshwane University of Technology (TUT), for offering me a job during my studies, and thanks to all staff members for the warm welcome and support when I joined the University, especially, Dr Annemarie Breytenbach, Mr Manqoba Jeffrey Mthethwa, and my mentor Mrs Sarie Mouton for her support and trust in me in my job.

I would like to thank my beautiful wife Ayanda Mazomba-Gaston and our daughter Uwase Anelisa Gaston for their love and support during the period of doing my PhD.

Thanks go to all my family members, especially Prof Paul Francis Muzindutsi for his endless support ever since I entered the UKZN. Furthermore, I would like to thank Dr Moise Majyambere for his financial and moral support that contributed positively to the completion of this study. I also want to thank Dr Rosemary Kuhn for financial support and proofreading of my PhD thesis.

I am extremely grateful to my friends Prof Ferdinand Niyimbanira, Dr JB Twayigira, Dr Egide Nshimiyimana, Dr Thomas Habanabakize, Dr Chrysostome Kanamugira, Dr Olivier Niyitegeka, Dr Abubakar Hakizimana, Dr Alexis Munyengabe, Mr Lambert Nizeyimana, Mr Seth Hakizimana, Mr Celesi Dukuziyaremye, Mr Eric Bariraje, Mr Jean Claude Mugisha, Mr Emmanuel Nshimiyimana, and many others for their constant support and encouragement throughout the study. Last but not least, I would like to thank my guardians Michael Bower and Michele Bower for their advice, support and encouragement ever since I came into their life.

LIST OF ABBREVIATIONS

95% CI	95% confidence interval
AIC	Akaike information criterion
AIDS	Acquired Immunodeficiency Syndrome
BIC	Bayesian information criterion
BDHS	Burundi Demographic and Health Survey.
CV	Cross-validation
DHS	Demographic and Health Survey
EAs	Enumeration areas
GAM	Generalised additive model
GAMM	Generalised additive mixed model
GCV	Generalised cross-validation
GLM	Generalised linear model
GLMM	Generalised linear mixed model
HIV	Human Immunodeficiency Virus
LDHS	Lesotho Demography and Health Survey
MIS	Malaria Indicator Survey
MLE	Maximum likelihood estimation
MOHSW	Ministry of Health and Family Welfare
NMCP	National Malaria Control Program
OR	Adjusted odds ratio
PLME	Pseudo maximum likelihood
PSU	Primary sampling unit
RDT	Rapid diagnostic test
SDGs3	Sustainable Development Goals
SEM	Structural equation modelling
SLRM	Survey logistic regression model
UNICEF	United Nations International Children's Emergency Fund
USAID	United States Agency for International Development
WHO	World Health Organization

LIST OF PUBLICATIONS

This thesis is based on the following publications:

Gaston, R.T., Ramroop, S. and Habyarimana, F., 2018. Determinants of factors associated with anaemia among children under five years in Lesotho. *African Population Studies*, *32*(1), pp.3893–902.

Gaston, R.T. and Ramroop, S., 2020. Prevalence of and factors associated with malaria in children under five years of age in Malawi, using malaria indicator survey data. *Heliyon*, *6*(5), p.e03946.

Gaston, R.T., Ramroop, S. and Habyarimana, F., 2021. Joint modelling of malaria and anaemia in children less than five years of age in Malawi. *Heliyon*, *7*(5), p.e06899.

Gaston, R.T., Habyarimana, F. and Ramroop, S., 2022. Joint modelling of anaemia and stunting in children less than five years of age in Lesotho: a cross-sectional case study. *BMC Public Health*, *22*(1), pp.1-11.

Gaston, R.T., Ramroop, S. and Habyarimana, F., 2022. Modelling the interrelationships between potential risk factors and childhood co-morbidity of malaria, anaemia, and stunting in children less than five years in Burundi (**Under review**).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The current research aimed to produce and expand statistical models in the discipline of biostatistics with a focus on childhood anaemia, malaria, and stunting. Malaria, anaemia, and stunting together continue to be public health issues worldwide in both industrialised and underdeveloped countries, particularly in children younger than 5 years (Osazuwa and Ayo, 2010; Kanchana et al., 2018). Malaria, anaemia, and stunting are dangerous, mostly in children from underdeveloped nations and they still remain the biggest contributor to morbidity and mortality. In addition, anaemia, malaria, and stunting are associated, and if not treated on time can damage children's emotional, physical, mental status and poor performance at school (Gaston et al., 2022). The current study evaluates the link between anaemia, stunting, and malaria simultaneously. Furthermore, the study assessed whether socioeconomic, geographical, environmental, and child demographic variables have a significant effect on childhood malaria, anaemia, and stunting. This study used a national secondary cross-sectional data from Malawi Malaria Indicator Survey (MMIS); Lesotho Demographic Health Survey (LDHS); and Burundi Demographic Health Survey (BDHS). The data was collected based on multi-stage sampling, stratified, and cluster sampling with an unequal chance of sampling. It is for this reason we first used the survey logistic regression model in Chapter 3, which accounted for the complexity of sampling design and heterogeneity between observations from the same cluster. However, this model includes only the fixed effect and does not have the option of adding the random effect to model the correlation between observations. We extend the model in Chapter 4, to a generalised mixed additive model (GAMM) to include the random effect. The GAMM is also an extension of the generalised linear mixed model (GLMM) and enables the parametric fixed effects from GLMM to be modelled as a non-parametric model using the additive smooth function. These models were applied to single response variables, and we wanted to evaluate the relationship which might exist between anaemia, stunting, and malaria. We then explore the multivariate joint model under GLMM in Chapter 5 to simultaneously joint either malaria and anaemia or anaemia and stunting. Finally, we introduce a structural equation model (SEM) in Chapter 6, to evaluate the complex interrelationships between socioeconomics, demographics, and environmental factors, as well as their direct or indirect relationship with childhood malaria, anaemia and stunting co-morbidity. The previous chapters could not address these interrelationships among the variables of interest. Each model used in this study has its weaknesses and strengths which can depend on the goal of the

researcher. However, the multivariate model under GLMM and the structural equation model were found to be more adaptive and attractive to researchers interested in innovative scientific research.

The findings from this study revealed that the child's nutrition status, age, the child with fever, diarrhoea, altitude, place of residence, toilet facility, access to electricity, children who slept under a mosquito bed net the night before the survey, mother's education level, and mother's body mass index have a significant effect on both childhood anaemia and malaria. The age of a child, the mother's educational status, place of residence, wealth index, and child weight at birth were the determinants of stunting or malnutrition. The findings also indicated that the geographical, geophysical, environmental, household and child demographic factors were statistically significant and have either a direct or an indirect effect on childhood co-morbidity factors. The geographical factors were statistically significant and had a positive direct effect on childhood malaria, anaemia, and stunting. The estimated indirect path for the impact of geophysical factors on childhood co-morbidity factors, as mediated by household factors was statistically significant and positive. However, the estimated indirect paths for the effect of geophysical factors on childhood co-morbidity factors, as mediated by environmental factors were statistically significant but negative.

The child demographic factors revealed a direct statistically significant impact on childhood co-morbidity factors. Furthermore, the estimated indirect path effect on childhood co-morbidity as mediated effect on household factors was statistically significant and negative. Moreover, household and environmental factors indicate a positive direct effect on childhood co-morbidity anaemia, malaria, and stunting. Finally, the results of this study revealed a positive relationship between stunting, anaemia, and malaria. This means that malaria, anaemia, and stunting increase or decrease in the same direction. Hence, controlling one or two between malaria, anaemia, and stunting can reduce the effect of other(s), which can assist the policymakers and government in the allocation of financial resources to fight against childhood comorbidity anaemia, malaria, and stunting. Furthermore, understanding the link between anaemia, malaria, and stunting other factors associated with them will assist in focusing on those areas and go a long way toward achieving the United Nations Sustainable Development Goals (SDGs3), known as the complete elimination of under-5 mortality by 2030.

Keywords: Malaria; Stunting; Anaemia; Children under five years; Multivariate joint model; Structural equation model.

CHAPTER ONE:

INTRODUCTION

In this chapter, we introduced the background on malaria, anaemia, and malanutrition in children younger than five years. The chapter also highlight the problem statement, and the objectives of the study. Furthermore, this chapter highlight the significance, and outline of the chapters of this study.

1.1. BACKGROUND

The discipline of biostatistics has attracted many researchers and investigators who are interested in under-five mortality rates for several reasons. These reasons include general assessments of a child's welfare in a country and the prevalence and distribution of different common diseases amongst children (e.g., malnutrition, anaemia, malaria, intestinal parasites, schistosomiasis, fever, diarrhoea, kwashiorkor, HIV/AIDS) because biostatistics can provide insights to inform interventions to prevent morbidities and mortalities (Nkoka et al., 2019).
A limitation in this field is that biostatistical modelling often focuses on one health disease, which either ignores or does not adequately cover the common phenomena of co-infection of children by two or more diseases and, in co-infected instances, the contribution of one or more present disease to morbidity or death from another present diseases (Kabaghe et al., 2017; McGann et al., 2018; Yimgang et al., 2021). This is particularly evident in the case of biostatistical studies on anaemia, stunting (malnutrition), and malaria amongst children younger than five years from Malawi, Lesotho and Burundi. Each one of malaria, anaemia, and stunting is independently, a significant cause of high morbidity and mortality rates amongst under-fives in sub-Saharan Africa but, critically, high morbidity and mortality rates within the same age cohort are also known to be due to co-infection with two and, frequently, both malaria, anaemia, and stunting.

Child morbidity and mortality rates are frequently used as a measure of a nation's overall rate of children's health and nutritional status (WHO, 2007; Gaston et al., 2022). Child morbidity rates, in particular, are a focus of health agencies because they indicate potential long-term health problems amongst affected children. Notably, the chief cause of morbidity rates is the lack of adequate nutrition among many of the affected children, which, in the long term can affect their mental, cognitive, and physical development (Kotecha, 2011; Gaston et al., 2018).

Malnutrition is the most common cause of mortality and comorbidities in young children under the age of five and may influence both susceptibilities to, and manifestations of malaria and anaemia (Osterbauer et al., 2012).

Malnutrition is described as a shortage of adequate nutrition, as a result of a lack of food or not eating food with sufficient vitamins, minerals, and other nutrients to maintain a body in sound health. The nutrient adequacy of children under the age of five is usually measured based on three anthropometric measurements (WHO, 1995, Gaston et al., 2022). The first one is height-for-age (HA) or stunting which measures body height in relation to age to determine within known parameters whether an infant or child potentially has a low height which can be due to genetic factors as well as stunting (impaired physical growth). The second is weight-for-height (WH) or wasting which measures body weight in relation to body height to determine within known parameters whether an infant or child is overweight or underweight. Last, is weight-for-age (WA) or underweight which is a composite index, which can reflect wasting or indicate stunting? An underweight person is a person whose body weight is considered too low to be healthy and whose weight is 15% to 20% below that normal for their age and height group (WHO, 1995).

Anaemia is also defined as a disease that can affect the mental, cognitive and physical development of a child. It is caused by the reduction of red blood cells, which can be due to a lack of sufficient nutrients in the body as well as certain parasitic infections, like diarrhoea and malaria (Benoist et al., 2008; Kotecha 2011; Gaston et al., 2018). Hemoglobin, which is a protein molecule found in red blood cells, transports oxygen to cells throughout the body and returns carbon dioxide to the lungs. A reduction of the red blood cells in the human body reduces the flow and exchange of oxygen and carbon dioxide (and nitric oxide) to and from body tissues, thereby threatening the normal metabolism of a body and thus, the reduction can lead to illness and death (Simbauranga et al., 2015, Gaston et al., 2022).

Malaria is defined as a deadly illness caused by the *Plasmodium* species, and *P. falciparum* which is mainly found in Africa, is considered the most threatening. Other less dangerous parasites include *P. ovale, P. vivax, P. knowlesi*, and *P. malariae* (WHO, 2015, Gaston and Ramroop, 2020). *Plasmodium falciparum* is Africa's most prevalent parasite, with 99% of malaria cases in 2016 from Sub-Saharan Africa, while *P. vivax* is the most widespread vector in other continents (WHO, 2017; Gaston and Ramroop, 2020). The individuals' contract malaria by the bite of *Anopheles* mosquito, and symptoms of the illness appear 10 to 15 days

after infection (Perkins et al., 2011; Gaston and Ramroop, 2020). Malaria parasites invade the red blood cells and decrease the number of blood cells in the person's body, which can result in severe anaemia (Noland et al., 2012; Seyoum, 2018). Malaria is not contagious, yet it is possible to contract the disease via blood transfusions or organ transplants (WHO, 2016; Gaston et al., 2021). The season of rain, humidity, and high temperature with low altitudes favor the breeding and growth of malaria vectors, and malaria is thus common in that season (Chirombo et al., 2014; Gaston and Ramroop, 2020). In light of the aforementioned, we note that malaria, anaemia, and stunting can occur together or in sequence in a body and thus they represent a considerable health risk generally and amongst children in particular. A child can be malnourished, anaemic, and have malaria. A child can suffer from malaria, which in turn may lead to anaemia, and together they can accentuate the health challenges of malnutrition. Malaria can be the overt cause of death in a child, but the risk of mortality is accentuated by the presence of malnutrition and/or anaemia (Shankar, 2000; Gaston et al. 2021, Gaston et al., 2022).

Despite the efforts and resources devoted to combat malaria, anaemia, and malnutrition, remain a significant global health concern, mostly in emergent nations (Leal et al., 2011; Yang et al., 2012; Aheto et al., 2015; Kanchana et al., 2018; Gaston et al., 2022). Malaria, anaemia, and malnutrition are thought to be interrelated and associated with morbidity and mortality around the world, notably in pregnant women and children (Black et al., 2013; Kavosi et al., 2014; Aheto et al., 2017; Wanzira et al., 2017).

In 2017, there were 151 million stunted children under the age of five in the world. The majority of these children came from Africa and South-East Asia, which contribute 75% of the total number. Globally, approximately 43% of children under the age of five had anaemia in 2011, with the highest rates of incidence found in South Asian and African areas (Milman, 2011; WHO, 2011; Kejo et al., 2018). Approximately 219 million occurrences of malaria have been confirmed in 2017, with 200 million cases arising in Africa (WHO, 2018). Furthermore, the WHO stated that there were 43000 malaria-related deaths worldwide, with 93% of those cases occurring in Africa. Children under the age of five accounted for 61% of all deaths worldwide and 93% of cases were from the African region (WHO, 2018). In 2017, the prevalence of malaria, anaemia and stunting in children younger than five years from Burundi was 27%, 61%, and 56% respectively (ISTEEBU, MSPLS, and ICF, 2017), while the prevalence of malaria, anaemia, and stunting in children younger than five years from Malawi in 2017 was 24%, 63% and 37% respectively (NSO and ICF, 2017). These percentages are high and show that

anaemia, malaria, and the nutrition status of child remain a health problems. However, a challenge for understanding the interaction between malaria, anaemia, and stunting is that there exists no study where malaria and anaemia, or anaemia and stunting, or all three together have been considered simultaneously in Malawi, Lesotho, and Burundi.

Consequently, it is difficult for health programme planners to predict the relative contribution of malaria, anemia, and stunting to morbidity and mortality amongst children younger than five years in a national population. Some models accommodate a combination between those two health problem, but their limitations for predicting population health threats have long been established (Caulfield et al., 2004; Deribew et al, 2010; Gahutu et al., 2011; Kandala et al., 2011; Zhao et al., 2012, Throne et al., 2013; Smithson et al., 2015; Shikur et al., 2016; Zgambo et al., 2017; Gari et al., 2018). There have been a few attempts to model malnutrition, anaemia, and malaria simultaneously (Ehrhardt et al., 2006; Olney et al., 2009; Osazuwa et al., 2010; Osterbauer et al., 2012; Kateera et al., 2015; Maketa et al., 2015).

Most of these studies used ordinary logistic regression model, they did not consider multivariate joint models. The logistic regression model is used only for predicting categorical or multinomial outcomes and assumes that all the variables have fixed effects. This model cannot predict continuous outcomes and there is no capacity in the model to include random effects for binary response (Dey and Raheem, 2016).

This study confronts the challenge of devising a biostatistical model, which can accommodate simultaneously both malaria, anaemia, and stunting. Ideally, the exercise will yield a generalisable model for accommodating two or both malaria, anaemia, and stunting simultaneously. At the very least, it will provide the statistical science foundations for modelling the interaction of anaemia, stunting (malnourished), and malaria co-infection amongst children younger than five years, which would assist health programme practitioners in predicting and responding to morbidities and mortalities amongst children with one or both malaria, anaemia, and stunting.

1.2. PROBLEM STATEMENT

Biostatistical modelling has become a powerful tool for understanding epidemiological patterns of diseases in human populations. However, a limitation of the science of biostatistics is the lack of models which can accommodate simultaneously, stunting, anaemia and malaria.

Malaria, anaemia, and stunting are known to be strongly correlated with persistently high morbidity and mortality rates amongst children younger than five years in developing countries. Development of such a model would greatly enhance the capabilities of health agencies to predict and, thereby mitigate more effectively than is possible at present, morbidities and mortalities amongst children younger than five years in a national population. The goal of this research is to devise such a model, and because of the above reason, the study seeks a model that can simultaneously evaluate the association between anaemia, malaria, and stunting in children younger than five years. This research will also consider other risk factors that may be linked to malaria, anaemia, and stunting.

1.3. AIM AND OBJECTIVES OF THE STUDY

The aim of this research was to generate an effective and flexible model that accommodates anaemia, malaria, and stunting simultaneously among children younger than five years and address the association between malaria, anaemia, and stunting.
The specifics objectives are:

- To take full responsibility for the difficulty of the sampling design by fitting the Survey Logistic Model (SLM) on anaemia among children younger than five years.
- To deal with nonlinear effects of continuous covariates by fitting the semiparametric generalised additive mixed model to the data and distinguish the key determinants of malaria.
- To develop a model that can combine either malaria and anaemia or stunting and anaemia concurrently in children under the age of five.
- To develop a model that accounts for the complex interrelationships between explanatory factors, as well as their direct or indirect relationship with childhood malaria, anaemia, and stunting co-morbidity.

1.4 SIGNIFICANCE OF THE STUDY

Anaemia, malaria, and stunting are public health problems globally, especially in children younger than five years from developing countries (Leal et al., 2011; Aheto et al., 2015). Malaria, anaemia, and stunting mostly coexist and affect children under five years in developing countries, which include Malawi, Burundi, and Lesotho (McCuskee et al., 2014; Ajakaye and Ibukunoluwa, 2020). Research has shown that both malaria, anaemia, and stunting jointly contribute to a higher percentage of mortality and morbidity in children younger than

five years in these three countries (Osterbauer et al., 2012; Kateera et al., 2015). Furthermore, anaemia, stunting, and malaria occur simultaneously due to sharing the risk factors. The geographical, demographic, environmental, and socio-economic risk factors of anaemia are more likely to be the same as malaria and stunting (McCuskee et al., 2014; Kabaghe et al., 2017). Hence, there is a need to use a reliable statistical models and methods to explore and reveal the extent of various risk factors of anaemia, malaria, and stunting, which are crucial for determining and forecasting factors affecting the prevalence of Malaria, anaemia, and stunting. In addition, the joint modeling of the two, or both malaria, anaemia, and stunting outcomes will offers a variety of possibilities for investigating the causes of malaria, anaemia, and stunting, the common contributing factors that will promote, and boost efficiency control strategies.

As a result, the findings of this research will significantly provide useful insights to the governments and policymakers in planning, controlling and the elimination of both malaria, anaemia, and stunting. Furthermore, the statistical models used in this study will help other researchers to compare findings and referencing.

1.5. OUTLINE OF THE THESIS

This dissertation is divided into seven chapters, which are arranged as follows. Chapter 1 gives an introduction to childhood comorbidity and mortality of anaemia, malaria, and stunting. It also indicates the study's objectives and research problems. Chapter 2 provides the underlying characteristics of the data and the data exploratory analysis. It also explains the data collection, and area of study and gives an overview of the prevalence of anaemia, stunting, and malaria in each country of interest. Chapter 3 introduces the survey logistic model (SLM) and fits the model to childhood anaemia in Lesotho. Furthermore, the chapter explains the model in detail and the interpretation of the results. Chapter 4 extends the SLM to a nonparametric generalized additive mixed model (GAMM) to investigate childhood malaria as well as other risk factors for malaria in Malawi and estimate the nonlinear influence of some explanatory variables. Chapter 5 introduces a multivariate joint model under the generalised linear mixed model (GLMM) to determine the relationship between anaemia and malaria, and anaemia with stunting. Chapter 6 introduces a structural equation model (SEM) to gain a better understanding of the complex interrelationship among multifactorial indicators and their direct or indirect influence on childhood malaria, stunting and anaemia co-morbidity in Burundi. In addition, we assess the association between anaemia, malaria, and stunting. Lastly, Chapter 7 provides a discussion and conclusions of the current study.

CHAPTER TWO:

DATA AND EXPLORATORY ANALYSIS

2.1. STUDY AREA

In this chapter, we describe the datasets used from three Sub-Saharan countries, which are Lesotho, Malawi, and Burundi. The selection of the countries was based on the availability of information regarding childhood malaria, anaemia, and stunting. Lesotho is one of the world's smallest countries, bordered by South Africa. The country covers an area of 30355 square kilometres (Moteetee, 2005; MOHSW and ICF, 2016). Lesotho has a dual legal system based on customary and common law, with the Prime Minister as the government leader, and the King as the head of state. The country is divided into ten politico-administrative regions, and the capital city is Maseru. The country is experiencing a serious HIV/AIDS epidemic, widespread poverty, high unemployment, malnutrition, and the burden of various illnesses (MOHSW and ICF, 2016; Gaston et al., 2018). In addition, the country is susceptible to climate change and natural calamities like droughts, torrential rains, and flooding (Renzaho, 2006; Letsie and Grab, 2015).

The second country is Malawi, which is south of the equator and is bordered by Tanzania in the north and northeast; Mozambique to the east and southwest; and Zambia to the west and northwest (NMCP and ICF, 2018). Approximately 118484 square kilometers make up the total area of Malawi, of which 9 4276 square kilometers are land and the rest is Lake Malawi. There are 28 districts and three regions in the entire country. There are three districts in the Northern region, nine in the Central region, and 13 in the Southern region (NMCP and ICF, 2018).

Malawi has a tropical continental climate with impacts from the sea, including altitude and distance from Lake Malawi influencing variations in rainfall and temperature. The anopheles mosquitoes thrive in the tropical climate, and their reproductive rate rises during the rainy season, which lasts from November until April. The country experiences cold, dry weather from May to August, and malaria transmission is lower than during the rainy season (Kazembe, 2007; Gaston and Ramroop, 2020). Malawi is one of the world's poorest countries, its economy is built on agriculture, and its healthcare is inadequate compared to other African nations (WHO, 2018; Gaston et al., 2021).

The last country considered in the current study is Burundi, with the capital city being Bujumbura, and the country covers a total area of 27834 square kilometres, (ISTEEBU, MSPLS, and ICF, 2017; Sinzinkayo et al., 2021). The country has a daily temperature variation throughout the region and a tropical highland climate. More than 70% of people live in poverty in Burundi, and this makes the country to be among the poorest in the world. The neighbourhood countries of Burundi are Tanzania to the south and east, in the northern region Rwanda, Lake Tanganyika in the southwest, and to the west the Democratic Republic of the Congo (ISTEEBU, MSPLS, and ICF, 2017; Nimpagaritse et al., 2020; Sinzinkayo et al., 2021). The morbidity and mortality rates are caused by malaria, anaemia, and stunting in the country, especially, among children under the age of five. The majority of the population works in the agriculture industry, which is the foundation of the nation's economy (Moise et al., 2016; Sinzinkayo et al., 2021).

2.2. DATA SOURCES

The current scientific setting used a secondary cross-sectional dataset from the 2014 Lesotho Demography and Health Survey (LDHS), the 2017 Malawi Malaria Indicator Survey (MMIS), and the 2017 Burundi Demographic and Health Survey (BDHS). The 2014 LDHS dataset usage ethical consent was assessed and approved through the Ethics Committee from Lesotho and the Ministry of Health Research, with assistance from ICF International's Institutional Review Board (MOHSW and ICF, 2016). The ethical approval for the 2017 MMIS dataset was reviewed and offered by the Ethics Committee and the Malawian Ministry of Health Research, with the endorsement of ICF International's Institutional Review Board (NMCP and ICF, 2018). Finally, the 2017 BDHS ethical clearance was approved by the Ministry of Public Health and the Fight Against AIDS. This was supported by the Institute of Statistics and Economic Studies of Burundi (ISTEEBU), the United States Agency for International Development (USAID), the United Fund for Childhood (UNICEF), the World Health Organization (WHO), The Swiss Agency for Development and Cooperation, and the Belgian Cooperation (ISTEEBU, MSPLS, and ICF, 2017).

2.3. SAMPLING DESIGN AND DATA COLLECTION

The current study used a nationally representative population' dataset, which was collected using multi-stage sampling, stratified and cluster sampling with the unequal likelihood of sampling. The first process of sampling involved the selection of clusters from the enumeration

areas (EAs) and denote the primary sampling units (PSUs). The second process of sampling involved the systematic selection of the households in each cluster or enumeration (MSPLS, and ICF, 2017; Gaston et al., 2018; Gaston et al., 2020).

The study used the weighted sample to gain insights that were illustrative of the nation and to account for the complex sample design from the data set (Heeringa et al., 2010; MOHSW and ICF, 2016; ISTEEBU, MSPLS, and ICF, 2017; NMCP and ICF, 2018). As part of the sampling process, each person selected in each area should equally provide equitably the size of the entire sample in the area. Most of the time, the regions may have low population density compared to others, and this unadjusted representation does not fairly reflect the exact population. Consequently, to avoid these problems, the area with a low population is oversampled, as a result, the current study used a weighted sample (MOHSW and ICF, 2016; NMCP and ICF, 2018; Gaston and Ramroop, 2020).

The calculation of sample size was based on the following formula:

$n = \frac{z^2 q(1-q)}{c^2}$, where n is the sample, z shows the number at 90% confidence interval which is 1.96, q indicates the occurrence of anaemia, and c is the significance level at a 5%. Considering the 2014 LDHS dataset, out of 3112 children a weighted sample of 1297 children for stunting and 1138 children for anaemia was used. The remaining children were regarded as incomplete or missing data and were rejected (MOHSW and ICF, 2016; Gaston et al., 2021).

The 2017 MMIS used considered a weighted sample of 2724 children aged between 6 to 59 months for both malaria and anaemia out of 2688 children (NMCP and ICF, 2018; Gaston et al., 2021).

Finally, the 2017 BDHS, used a weighted sample of 6493 children between 6 to59 months of age from13611 children who were surveyed, and the rest were considered as missing data (ISTEEBU, MSPLS, and ICF, 2017).

In both countries, the interview involved women between 15 to 49 years of age and children aged 6 to59 months who remained in or visited the designated households the night before the survey. Under the guidance of their parents or caregivers, all children in the survey had their height measured and tested for anaemia, and malaria. The children who were less than 6 months of age were not included and the system missing value and were consequently ignored (MOHSW and ICF, 2016; Gaston et al., 2022). Following national regulations, the children who tested positive for anemia, stunting, or malaria received immediate treatment.

The results from both anaemia, malaria, and stunting testing were recorded on the Biomarker Questionnaire as well as on the handbook left in the household with data on the causes and prevention of both malaria, anaemia, and stunting (NMCP and ICF, 2018; Gaston et al., 2021).

Anaemia blood is collected by puncturing a child's finger or heel with a spring-loaded sterile lancet. A blood sample was placed in a microcuvette, and a HemoCue analyser was used to determine the hemoglobin (Hb) level. For hygiene purposes, the lancet, gloves, alcohol swabs, and microcuvette were only used once. The children's results were received in less than 10 minutes and children with Hb levels under 7 g/dl were immediately taken to the closest hospital for further testing (NMCP and ICF, 2018; Gaston et al., 2021).

Malaria blood samples were taken from children's fingertips or heel-prick using the SD Biolne Malaria Ag P.f/P, a rapid diagnostic test (RDT). However, microscopy can be used to test for malaria. In this study, the RDT was taken into consideration for its effectiveness in diagnosing malaria, being easy to learn and determining the types of plasmodia compared to microscopy. (MOHSW and ICF, 201; NMCP and ICF, 2018; Gaston et al., 2021).

Nutrition status assessment was done by determining the child's weight using an electronic scale, and the child's height using a tape measure provided by UNICEF. The weight of children was determined using a Seca gauging scale that was set to zero. Children were either undressed or dressed in light clothing by their parents or guardians. The weight of a child who was unable to stand was calculated by deducting the parent's weight from the parent's weight while carrying the child. A small board that was flat on either the ground or standing upright was used to measure the height of children who did not wear shoes (MOHSW and ICF, 2016; Gaston et al., 2022). Each child who is less than 87cm was assessed in the prone position, and his/her nutritional status was calculated from the child's weight, height, and age (i.e., weight-for-age, height-for-age, and weight-for-height). The weight-for-age (underweight), height-for-age (stunting), and weight-for-height (wasting) are measured in z-scores, and children with a z-scores below -2 standard deviation (SD) are considered malnourished otherwise normal (WHO, 2007; Gaston et al., 2022).

2.4. VARIABLES OF INTEREST

2.4.1. DEPENDENT (RESPONSE) VARIABLES

The outcome variables of interest in this study were anaemia, malaria, and stunting status among children under five years. The response (outcome) variables were recorded as binary

exposure since the purpose of the research was to determine whether the child is anaemic or not, tested positive for malaria or negative, and stunted or normal. The key factor used to categorize children's anaemia is their blood's hemoglobin concentration, which is expressed in grams per deciliter (g/dl). A child is regarded as anaemic when his/her hemoglobin concentration level adjusted for altitude is lower than 11.0 g/dl, otherwise, the child is not (WHO, 2015; Gaston et al., 2018). Stunting was used to assess the child's nutritional status, and all children with z-scores below -2 SD were considered malnourished (stunted), otherwise nourished (normal) (Kazembe, 2013; Gaston et al., 2022). To determine if the child has malaria (positive) or (negative), the RDT test was used (NMCP and ICF, 2018; Gaston and Ramroop, 2020).

2.4.2. INDEPENDENT (EXPLANATORY) VARIABLES

The explanatory variables considered included a variety of demographic, socio-economic, and environmental characteristics and were chosen based on those found in the literature to have some association with anaemia, malaria, and/or stunting as well as those anticipated to be determinants of each outcome (Kotecha, 2011; Bennett et al., 2013; Alegana et al., 2014; Caminade et al., 2014; Habyarimana et al., 2017; Kabaghe et al., 2017), among others. As a result, this serves as the theoretical foundation for the current research.

The socio-economic variables included the type of place of residence (rural or urban); wealth quantile; mother's highest education level; source of drinking water; type of toilet facility; the main material of the walls, floor, and roof of the rooms; the household share of toilet facility; children under 5 who slept under a mosquito bed net the night before the survey; access to information through television, the household access to electricity; whether the child had a fever or not; cough or diarrhoea in the two weeks before the survey or not; whether the child had received drugs for intestinal worms or vitamin A supplementation in the six months before the survey or not. All these variables were collected at the household level (Buchwald et al., 2016). The demographic variables included child age; gender; and birth order of the child and were gathered at individual level (Ayele et al., 2014a; Zgambo et al., 2017). Finally, the environmental variables comprised of the minimum temperature, maximum temperature, rainfall, proximity to water, land surface temperature, enhanced vegetation index (EVI), aridity, wet days, and travel times.

2.5. DESCRIPTIVE STATISTICS

In this chapter, we used cross-tabulation techniques to assess the association between potential explanatory variables and childhood anaemia, malaria, and stunting. The analysis was done using the Statistical Package for Social Sciences (SPSS) version 24.0. The Chi-square test was used to check whether the independent variables are statistically significantly associated with the response variables (anaemia, malaria, or stunting) or not. To account for any possible multi-collinearity and confounding between the covariates, all variables with a *p*-value less than 0.2 are included in the analysis of multivariate models (Schneider et al., 2008, Gari et al., 2017). We also used the percentages to determine the category with low or high effect as the number of surveyed in each category are not the same and can only be compared using the percentages.

The prevalence of anaemia and stunting in children younger than five years from Lesotho was 51% and 43% respectively, with 35.2 % of children having both malaria, anaemia, and stunting.

The frequency distribution, *p*-values, and percentages of childhood anaemia and stunting with their associated factors are indicated in Table 2.1. Childhood stunting and anaemia were both related to child age but in opposing directions (i.e., stunting increased with age, while anaemia decreased with age). However, stunting itself was further associated with sex, having visited healthcare facilities, maternal education, wealth index, access to electricity, drinking water, and dwelling characteristics (wall and floor material). Furthermore, anaemia was more specifically related to fever in the previous two weeks, recent diarrhoea, and roof characteristics of the dwelling. The prevalence of anaemia was higher in children aged between 0-19 months (62.7%), and then decreased in children aged 20-39 months (55.0%) and then again in those aged 40-59 months (44.4%). The anaemia prevalence was higher in children who experienced fever in the last two weeks (62.8%) compared to children who did not (37.2%).

The prevalence of stunting was lower in younger children (age group 0-19 months, 17.8%), but increased in children aged (20-39 months, 33.3%; and 40-59 months, 31.2%). The prevalence of stunting was higher in children from mothers with low education (72.7%) and reduced as the level of mother's education increased by primary (31.3%) and post-primary (21.7%), respectively.

Table 2.1: Childhood anaemia and stunting by categorical variables in Lesotho.

Variables	Categories	Anaemic	Stunting	*p*-value
Sex of the child	Male	56.2% (301)	30.7% (184)	0.185[1]
	Female	52.2% (315)	23.0% (160)	0.002[2]
Child's age in months	0-19	62.7% (227)	17.8% (92)	<0.001[1]
	20-39	55.0% (224)	33.3% (136)	<0.001[2]
	40-59	44.4% (164)	31.2% (115)	
Child's birth weight	<2500g	58.5% (48)	50.6% (45)	0.359[1]
	≥2500g	53.2% (467)	22.4% (227)	<0.001[2]
Had fever in last two weeks	Yes	62.8% (120)	21.2% (44)	0.008[1]
	No	52.3% (493)	27.6% (299)	0.054[2]
Had diarrhoea recently	Yes	62.7% (94)	27.2% (44)	0.024[1]
	No	52.8% (520)	26.5% (299)	0.855[2]
Had cough in last two weeks	Yes	54.4% (184)	26.1% (97)	0.883[1]
	No	54.0% (429)	26.7% (246)	0.807[2]
Received vitamin A in last 6 months	Yes	54.3% (357)	26.0% (176)	0.858[1]
	No	53.7% (253)	27.2% (166)	0.633[2]
Birth order	1st	53.2% (231)	24.5% (126)	0.728[1]
	2-3	56.0% (261)	27.0% (145)	0.529[2]
	4-5	52.1 % (85)	29.8 % (51)	
	≥6	51.3% (39)	28.4% (21)	
Mother's BMI	<18.5	56.0% (14)	29.6% (8)	0.830[1]
	≥18.5	53.8% (596)	26.3% (333)	0.702[2]
Mother's education level	No education	53.8% (7)	72.7% (8)	0.193[1]
	Primary	51.1% (267)	31.3% (186)	<0.001[2]
	Post primary	56.6% (341)	21.7% (150)	
Visited health facility	Yes	52.7% (479)	24.4% (257)	0.069[1]
	No	59.4% (136)	35.7% (87)	<0.001[2]
Wealth Index	Poor	58.0% (280)	33.8% (184)	0.059[1]
	Middle	53.1% (127)	30.1% (80)	<0.001[2]
	Rich	50.1% (209)	16.3% (79)	
Place of residence	Rural	55.5% (457)	27.3% (257)	0.104[1]
	Urban	50.2% (158)	24.2% (86)	0246[2]
Household with electricity	Yes	50.0% (150)	18.2% (64)	0.102[1]
	No	55.5% (465)	29.6% (279)	<0.001[2]
Toilet facility	Toilet with flush	53.7% (306)	26.4% (170)	0.148[1]
	Pit latrine	48.3% (87)	19.8% (41)	0.027[2]
	No facility	57.1% (222)	29.8% (133)	
Type of drinking water	Tap water	54.0% (274)	29.9% (173)	0.166[1]
	Protected water	49.1% (109)	19.2% (48)	0.006[2]
	Unprotected	57.0% (233)	26.1% (122)	
Main roof material	Thatch/Palm leaf	58.8% (248)	34.4% (165)	0.014[1]
	Corrugated metal	52.3% (335)	23.2% (169)	<0.001[2]
	Stick &mud	42.7% (32)	11.1% (10)	
Main wall material	Wood/Mud	56.6% (233)	33.0% (194)	0.270[1]
	Bricks	51.5% (202)	20.1% (89)	<0.001[2]
	Cement /Block	52.6% (120)	22.6% (60)	
Main floor material	Earth/Sand	52.8% (152)	20.9% (72)	0.441[1]
	Mud block/Wood	56.6% (233)	33.8% (158)	<0.001[2]
	Cement/ Block	52.5% (229)	23.5% (114)	

***value*[1]** *= p-value for anaemia status;* ***value*[2]** *= p-value for stunting status*

The prevalence of anaemia and malaria in children younger than five years from Malawi was 56.9 % and 37.2%, respectively with 61.5 % of children having both malaria, anaemia, and stunting. Table 2.2 indicates the frequency distribution, *p*-values, and percentages of childhood anaemia and malaria respectively with their associated factors. The results indicated that all independent variables were significantly associated with childhood anaemia with the *p*-value less than 0.05, except for the sex of the child, and households who share a toilet. The results also showed that all independent variables were significantly associated with childhood malaria with the *p*-value less than 0.05, except for the sex of the child.

Table 2.2 shows that the prevalence of anaemia was higher in children from mothers with no education (68.9%) and lower in mothers with primary (55.8%) and post-primary (52.2%) respectively. The same results indicated a decrease in the prevalence of anaemia in children from wealthier classes (55.2%), and an increase in the middle (59.2%), and poorer classes (67.7%). The prevalence of anaemia in children from a rural area (63.0%) was higher compared to those from an urban area (54.0%). The results also indicated a decrease in the prevalence of anaemia as the age of a child increased. The prevalence of anaemia was 78.6%, 58.0%, and 51.0% among children aged between 6-23, 24-41, and 42-59 months, respectively.

The same results revealed that the prevalence of malaria was higher in children from mothers with no education (46.3%) and reduced where mothers had primary (39.6%) and post-primary education (16.6%), respectively. The same results indicated that the prevalence of malaria was lower in children from the wealthier (21.3%) and increased in the middle (39.6%), and poorer (48.0%) classes, respectively. The prevalence of malaria in children from rural areas was higher (41.9%); while for those from the urban area it was lower (7.3%). **Table 2.2** also shows that the prevalence of malaria increased as the child's age increased, which was 25.6%, 40.6%, and 43.4% among children aged between 6-23, 24-41, and 42-59 months, respectively.

Table 2.2: Childhood anaemia and malaria by categorical variables in Malawi.

Variables	Category	Anaemic	Malaria+	*P*-value
Child's age in months	6-23 months	609 (78.6%)	200(25.6%)	
	24-31 months	466 (58.0%)	411(40.6%)	<0.001[1]
	32-59 months	474 (51.0%)	403(43.4%)	<0.001[2]
Sex of the child	Male	771 (61.0%)	463(36.5%)	0.432[1]
	Female	779 (62.6%)	447(35.8%)	0.716[2]
Sleep under mosquito bed net	All children	956 (58.0%)	553(33.6%)	
	Some children	136 (52.7%)	88 (34.1%)	0.002[1]
	None	458 (55.9%)	372(45.4%)	<0.001[2]
Region	North	145 (53.5%)	62 (20.8%)	
	Central	686 (64.4%)	452(39.8%)	0.004[1]
	South	718 (61.3%)	500(38.7%)	<0.001[2]
Place of residence	Rural	1366(63.0%)	987 (41.9%)	0.001[1]
	Urban	184 (54.0%)	27 (7.3%)	<0.001[2]
Wealth index	Poorer	793 (67.7%)	611(48.0%)	
	Middle	284 (59.2%)	204(39.6%)	<0.001[1]
	Richer	473 (55.2%)	199(21.3%)	<0.001[2]
Mother's education level	No education	226 (68.9%)	152(46.3%)	
	Primary	1125(55.8%)	799(39.6%)	< 0.001[1]
	Post-primary	198(52.2%)	63 (16.6%)	<0.001[2]
Altitude	0-500 metres	240 (72.9%)	119(34.3%)	
	501-1000 metres	604 (60.0%)	492(43.8%)	< 0.001[1]
	>1000 metres	706 (60.2%)	403(32.2%)	<0.001[2]
Toilet facility	Toilet with flush	198 (51.3%)	88 (21.1%)	
	Pit latrine	1235(63.3%)	832(39.3%)	<0.001[1]
	No facility	117 (68.8%)	95 (50.8%)	<0.001[2]
Type of drinking water	Tap water	339 (53.6%)	164(25.9%)	
	Protected water	290 (54.3%)	122(31.9%)	0.024[1]
	Unprotected	920 (59.1%)	728(42.6%)	<0.001[2]
Main roof material	Thatch/Palm leaf	920 (61.4%)	686(45.8%)	
	Corrugated metal	557 (50.7%)	283(25.8%)	<0.001[1]
	Stick & mud	73 (57.5%)	45 (35.7%)	<0.001[2]
Main wall material	Wood/Mud	475 (61.9%)	331(43.2%)	
	bricks	624 (53.6%)	467(40.1%)	0.001[1]
	Cement /Block	451 (56.9%)	216(27.2%)	<0.001[2]
Main floor material	Earth/Sand	1143(59.2%)	840(43.5%)	
	Mud block/Wood	110 (58.5%)	74 (39.4%)	<0.001[1]
	Cement/ Block	297 (49.2%)	100(16.6%)	<0.001[2]
Electricity	Yes	122 (42.5%)	21 (7.3%)	<0.001[1]
	No	1427(58.6%)	992(40.7%)	<0.001[2]
Household share toilet	Yes	566 (55.4%)	329(32.2%)	0.380[1]
	No	872 (57.1%)	600(39.3%)	<0.001[2]

value[1] *= p-value for anaemia status;* ***value***[2] *= p-value for malaria status, and* ***malaria*** *+=positive malaria*

The prevalence of anaemia, stunting, and malaria in children younger than five years in Burundi was 59.1 %, 47.5%, and 35.7 %, respectively. The prevalence of both anaemia and malaria was 48.6%, anaemia and stunting 55.9%, and malaria and stunting 60.4%. **Table 2.3** represents the frequency distribution, *p*-values, and percentages of childhood anaemia, malaria, and stunting with their associated variables. The results from **Table 2.3** point out that all explanatory variables associated with childhood malaria, anaemia, and stunting were statistically significant with a *p*-value less than 0.05, except for the sex of the child, and children from households sharing the toilet. The prevalence of malaria in children aged between 6-23, 24-41, and 42-59 months was 29.5%, 34.6%, and 38.8% respectively. The prevalence of stunting in children aged between 6-23, 24-41, and 42-59 months was 34.7%, 50.5%, and 57.6% respectively. Finally, the prevalence of anaemia in children aged between 6-23, 24-41, and 42-59 months was 71.9%, 54.5%, and 54.1% respectively. Based on these results, malaria and stunting increase as the child gets older, while anaemia decrease as the child gets older. According to Table 2.3**,** the prevalence of anaemia was higher in children born to mothers with no education (63.8%) and lower in children born to mothers with primary (58.5%) and post-primary (44.0%) education respectively. The prevalence of malaria was higher in children whose mothers had no education (40.7%) and reduced in children whose mothers had primary (33.9%) and post-primary (13.5%) education respectively. Lastly, the prevalence of stunting was higher in children from mothers with no education (53.6%) and lower in mothers with primary (47.9%) and post-primary (22.0%) education respectively.

The same results showed a decrease in the prevalence of anaemia in children from wealthier classes (49.2%), and an increase in the middle (61.1%), and lower (67.8%) classes. The prevalence of malaria was lower in children from the wealthier (19.3%) and increased to the middle (36.0%) and lower (51.6%) classes, respectively. The results also point out the increase in the prevalence of stunting in children from lower classes (58.5%) and reduced in children from the middle (54.3%) and wealthier (33.1%), classes, respectively.

Table 2.3: Childhood anaemia, malaria, and stunting by categorical variables in Burundi.

Variables	Category	Anaemia status	Malaria status	Nutrition status	*P*-values
		Anaemic	Malaria+	Stunting	
Child's age in months	6-23 months	1136 (71.9%)	466 (29.5%)	753 (34.7%)	<0.001[1]
	24-31months	1096 (54.5%)	695 (34.6%)	1016 (50.5%)	<0.001[2]
	32-59 months	1178 (54.1%)	843 (38.8%)	1251 (57.6%)	<0.001[3]
Sex of the child	Male	1761 (60.7%)	1007 (34.8%)	1590 (49.8%)	0.016[1]
	Female	1649 (57.6%)	997 (34.8%)	1430 (45.2%)	0.956[2]
					<0.001[3]
Sleep under mosquito bed net	All children	1298 (55.5%)	801 (29.4%)	1075 (41.2%)	<0.001[1]
	Some children	326 (57.5%)	203 (32.3%)	314 (47.9%)	<0.001[2]
	None	1766 (62.4%)	1347 (41.2%)	1615 (52.9%)	<0.001[3]
Place of residence	Rural	408 (45.5%)	2320 (39.8%)	2770 (51.8%)	<0.001[1]
	Urban	3002 (61.6%)	147 (13.6%)	250 (24.8%)	<0.001[2]
					<0.001[3]
Wealth index	Poorer	1568 (67.8%)	1433 (51.6%)	1480 (58.5%)	<0.001[1]
	Middle	725 (61.1%)	506 (36.0%)	703 (54.3%)	<0.001[2]
	Richer	1117 (49.2%)	528 (19.3%)	837 (33.1%)	<0.001[3]
Mother's education level	No education	1620 (63.8%)	1033 (40.7%)	1491 (53.6%)	<0.001[1]
	Primary	1333 (58.5%)	772 (33.9%)	1209 (47.9%)	<0.001[2]
	Post-primary	302 (44.0%)	92 (13.5%)	174 (22.0%)	<0.001[3]
Altitude in metre (m)	Less than 1000m	432 (54.4%)	198 (20.7%)	277 (31.1%)	<0.001[1]
	Between1001-2000m	2799 (60.4%)	2239 (40.3%)	2568 (50.5%)	<0.001[2]
	More than 2000m	161 (52.1%)	12 (03.3%)	158 (46.1%)	<0.001[3]
Toilet facility	Toilet with flush	72 (35.0%)	15 (05.9%)	26 (11.1%)	<0.001[1]
	Pit latrine	1708 (55.5%)	1043 (28.4%)	1514 (44.4%)	<0.001[2]
	No facility	1630 (65.6%)	1409 (47.2%)	1480 (54.7%)	<0.001[3]
Type of drinking water	Tap water	1036 (53.1%)	547 (23.0%)	782 (35.7%)	<0.001[1]
	Protected water	1764 (61.9%)	1416 (41.9%)	1669 (53.8%)	<0.001[2]
	Unprotected	610 (63.3%)	504 (43.7%)	569 (53.5%)	<0.001[3]
Main roof material	Thatch/Palm leaf	617 (64.9%)	567 (50.9%)	622 (59.2%)	<0.001[1]
	Corrugated metal	1658 (54.7%)	978 (26.7%)	1374 (40.9%)	<0.001[2]
	Stick & mud	1135 (63.5%)	922 (43.2%)	1024 (52.5%)	<0.001[3]
Main wall material	Wood/Mud	2159 (60.7%)	1681 (39.4%)	1946 (49.8%)	0.006[1]
	bricks	8 (47.1%)	3 (15.0%)	2 (11.1%)	<0.001[2]
	Cement /Block	1243 (56.7%)	783 (29.8%)	1072 (44.1%)	<0.001[3]
Main floor material	Earth/Sand	2966 (62.5%)	2320 (40.7%)	2732 (52.5%)	<0.001[1]
	Mud block/Wood	42 (48.8%)	17 (15.9%)	40 (40.0%)	<0.001[2]
	Cement/ Block	402 (43.1%)	130 (11.7%)	248 (23.6%)	<0.001[3]
Electricity	Yes	212 (39.8%)	46 (7.1%)	108 (17.8%)	<0.001[1]
	No	3198 (61.1%)	2421 (38.7%)	2912 (50.7%)	<0.001[2]
					<0.001[3]
Household share toilet	Yes	565 (59.2%)	330 (28.9%)	435 (41.0%)	0.773[1]
	No	2764 (58.7%)	2061 (36.5%)	2516 (48.6%)	<0.001[2]
					<0.001[3]

value[1] *= p-value for anaemia;* ***value[2]*** *= p-value for malaria,* ***value[3]*** *= p-value for stunting,* ***malaria+*** *= positive malaria*

2.6. SUMMARY

In this chapter, we introduced the explanatory analysis of Malaria Indicator Survey (MIS) and Demographic Health Survey (DHS) data sets in detail. The chapter summarised the frequency

distribution and percentages of childhood anaemia, malaria, and stunting respectively with their associated factors. The cross-tabulation techniques were used to analyse the data and summarise the results. Pearson's Chi-square test and *p*-values were used to investigate whether the independent (explanatory) variables were statistically significantly associated with each one of the responses (dependent) variables or not and this was done using SPSS version 24.0. The prevalence of anaemia and stunting in children younger than five years in Lesotho was 51% and 43%, respectively with 35.2 % of children having both malaria, anaemia, and stunting. The child's age indicated a significant effect on childhood stunting and anaemia.

The prevalence of anaemia and malaria in children younger than five years from Malawi was 56.9 % and 37.2% respectively, with 61.5 % of children having both malaria, anaemia, and stunting. The prevalence of anaemia, stunting, and malaria in children younger than five years from Burundi was 59.1 % and 47.5%, and 35.7 %, respectively. The prevalence of both anaemia and malaria was 48.6%, anaemia and stunting 55.9%, and malaria and stunting 60.4%. Based on these results in each country of interest, it can clearly be seen that anaemia, malaria, and stunting are still health problems in children younger than five years. These results also indicate an association between anaemia, malaria, and stunting. In the next chapters, we will fit different models and include all the variables with a *p*-value less than 0.2 to account for any possible multi-collinearity and confounding between the covariate (Schneider et al., 2008; Gari et al., 2017; Gaston et al., 2018).

CHAPTER THREE:

SURVEY LOGISTIC REGRESSION MODEL

3.1. INTRODUCTION

In this chapter, we introduce a survey logistic regression model (SLRM), which is a member of the generalised linear models to analyse our data. The generalised linear models (GLMs) are an adaptable generalisation of ordinary linear regression developed by Nelder and Wedderburn (1972). The GLMs generalise linear regression by accepting the linear model to be related to the dependent (response) variable through a link function and enable the magnitude of each estimation's variance to be a function of its predicted value (Jain et al., 2017).

The generalised linear model can be extended to accommodate models of different numbers of classes such as marginal, random effects, and conditional models, in case of correlated observations (Diggle et al., 2002). Intending to understand each type of model, different distributions are applied, and we can first describe the exponential family distribution (Diggle et al., 2002; Aerts et al., 2002).

The Exponential Family: Assume a random variable $\boldsymbol{Y}$ has an exponential distribution as well as its probability density function to be expressed as follows:

$$f(y/\delta,\theta) = exp\left\{\frac{y\delta-\theta(\delta)}{b} - d(y,b)\right\}. \tag{3.1}$$

The functions δ and d are known, while the parameters δ and b are unknown, with δ called the natural or canonical parameter, and b the scale or dispersion parameter (Diggle et al., 2002; Aerts et al., 2002). In order to obtain the mean and variance of $\boldsymbol{Y}$, the property of $\int f(y/\delta,b)dy = 1$, can be applied, and if we take the first and second derivatives with respect to δ to both sides of equation (3.1) we get the following:

$\int[y-\theta'(\delta)]f(y/\delta,b)dy$ and $\int\left[b^{-1}\left(y-\theta'(\delta)\right)-\theta''(\delta)\right]f(y/\delta,b)dy = 0$. Then we have $E(Y) = \mu = \theta'(\delta)$ and $Var(Y) = \theta''(\delta)b$, with θ' and θ'' first and second derivatives of $\theta(\delta)$ with respect to θ. Thus, the mean and variance are linked by the relation, and can be shown as follow $\sigma^2 = b\theta''[\theta'^{-1}(\mu)] = b\omega(\mu)$, where $\omega(\mu) = \theta''[\theta'^{-1}(\mu)]$ is referred to as the variance function.

Binomial random variable: The binomial distribution has a variable $\boldsymbol{Y}$ representing the number of successes in n independent trials with a probability of success π for every trial, and the probability distribution is given by:

$$f(y) = \binom{n}{y}\pi^y(1-\pi)^{n-y} = exp\left[y\ log\left(\frac{\pi}{1-\pi}\right) + n\ log(1-\pi) + log\binom{n}{y}\right]. \quad (3.2)$$

Based on equation (3.1), the canonical (natural) parameter is $\delta = \log\left(\frac{\pi}{1-\pi}\right)$ and is noted as the logit(π), and on the other hand, $\delta = \log\left(\frac{\mu}{n-\mu}\right)$, where $\mu = n\pi$ (McCullagh and Nelder, 1989; Diggle et al., 2002). Regarding δ, the probability of success and failure can be expressed as follow: $\pi = \frac{exp(\delta)}{1+exp(\delta)}$ and $1-\pi = \frac{1}{1+exp(\delta)}$ respectively. Considering the structure of an exponential probability density function (p.d.f), $\theta(\delta) = -n\ log(1-\pi) = n\ log[1 - exp(\delta)]; b = 1$ and $d(y,b) = log\binom{n}{y}$. Moreover, the expected value can be defined as $E(Y) = \theta'(\delta) = n\frac{exp(\delta)}{1+exp(\delta)} = n\pi$, and the variance as $Var(Y) = \theta''(\delta)b = \frac{n\ exp(\delta)[1+exp(\delta)]-n\ exp(\delta)exp(\delta)}{[1+exp(\delta)]^2} = n\pi(1-\pi)$.

As a result, in this instance $\omega(\mu) = \mu\left(1-\frac{\mu}{n}\right)$ as $\mu\pi$ (Diggle et al., 2002; Molenberghs and Verbeke, 2005). When the response variable is binary and the number of trials $n = 1$, the binomial distribution becomes a special case called the Bernoulli distribution (Molenberghs and Verbeke, 2005).

Logistic and Probit regression for Binary data: The natural link is the logit and if $Y_j \sim Bernoulli\ (\pi_j)$ we have the following equation for the linear model

$ln\left(\frac{\pi_j}{1-\pi_j}\right) = z_j'\alpha$, and in covariates context becomes $\pi_j = \frac{exp\left(z_j'\alpha\right)}{\left[1+exp\left(z_j'\alpha\right)\right]}$. In this case, the natural parameter depends on the covariate z_j, and besides the probit link, we can also use the model of $b^{-1}\left(\pi_j = z_j'\alpha\right)$, and $\pi_j = b\left(z_j\alpha\right)$, with b known as the distribution of a basic normal random variable. Regarding a binomial distribution variable, the $Y_j \sim B\left(n_j\pi_j\right)$ and logit $\left(\pi_j\right) = z_j'\alpha$ (Molenberghs and Verbeke, 2005).

Poisson regression for counts: The classic Poisson regression model $Y_j \sim Poisson\left(\mu_j\right)$ with $ln\left(\mu_j\right) = z_j'\alpha$. By applying the exponent on both sides we get a quantity that is always

negative and is given by $\mu_j = exp(z_j'\alpha)$. The μ_j is the average occurrence rate, and the logarithm is the natural link function (Diggle et al., 2002; Molenberghs and Verbeke, 2005).

Generalised linear models have been used in a wide range of academic subjects, more details on GLMs can be found in the studies by McCullagh and Nelder (1989); McCulloch and Searle (2004); Molenberghs and Verbeke (2005); and Dobson and Barnett (2018), among others.

The GLMs are intended to simulate the relationship between binary data and a group of response variables. Nevertheless, the standard logistic regression model is unsuitable for analysing survey data with clustering and stratification, which may be observable in survey designs. The dataset used in this study is from Health Demographic Survey and data are collected based on multi-stage sampling, stratified, and cluster sampling with an unequal probability of selection. Heeringa (2010) emphasised the importance of accounting for the complexity of sampling design in order to avoid underestimating variance and making incorrect inferences. Thus, for this reason in this chapter, we used a survey logistic regression model to account for the complexity of sampling design, clustering, as well as the possible association between observations from the same cluster (Heeringa, 2010; Gaston et al., 2018).

3.2. MODEL OVERVIEW

The survey logistic regression model (SLRM) has the same properties as the ordinary logistic regression models (Heeringa, 2010). However, the SLRM is the best model to analyse the data from a complex sampling design to avoid underestimating variance and drawing the wrong conclusions (Heeringa, 2010; Agresti, A., 2015). In the survey logistic regression model, the first stage from each stratum surveyed is modelled as the primary sampling units (PSUs). Let us assume y_{jhk} be the response variable such that $j = 1,2,3, \ldots m_j$; $h = 1,2,3, \ldots m_{jh}$; and $k = 1,2,3, \ldots K$, where k is the stratum, h is the cluster, and j is the household and means the sampling weight for the jhk^{th} observation as V_{jhk} and Z_{jhk} the row vector of design matrix related to the j^{th} household in h^{th} PSU, nested in k^{th} stratum.

Suppose that Z_{jhk} is the row vector of the design matrix related to the j^{th} household in h^{th} PSU, nested in k^{th} stratum. In addition, we assume that the row Y denotes $m = \sum_{j=1}^{J} \sum_{h=1}^{mh} mjh$

Observation of the response variable Y, and the rows of the Z are m observation of t explanatory variables $Z_1, Z_2, Z_3, \ldots, Z_t$. We then let Y_{hjk} relate to the exponential family of distributions, with the following sampling distribution:

$$f(y_{jhk}, \delta_{jhk}, \theta) = exp\left(\frac{y_{jhk}\delta_{jhk} - c(\delta_{jhk})}{b(\theta)} + d(y_{jhk}, \theta)\right), \quad (3.3)$$

where $f(.)$ is the density function of y_{jhk}, and δ_{jhk} is the natural parameter, while θ is the dispersion parameter. Moreover, we assume a model for the mean vector μ_{jhk}=E$[y_{jhk}]$, which can be expressed as $\mu_{jhk} = n(Z_{jhk}, \alpha)$, where $n(.)$ is a vector-valued function of Z_{jhk} and $q \times 1$ vector α of unknown parameters. Then the model can be converted to a linear model by employing a suitable link function and the mean of the distribution, which gives the association between the linear predictors. The linear model of GLM can be written as:

$$\varphi_{jhk} = g(\mu_{jhk)} = Z_{jhk}\alpha, \quad (3.4)$$

where Z_{jhk} is a covariate vector and $g: \mathcal{R} \to \mathcal{R}$ indicates the link function for a binary outcome, and the link function of a survey logistic regression is $\varphi_{jhk} = logitg(\mu_{jhk)}$, and the generalised logit model becomes:

$$logit\left(\pi_{jhk}\right) = Z'_{jhk}\alpha. \quad (3.5)$$

In the estimation methods for survey logistic regression parameters and standard error for complex survey data, we use two main approaches. The first approach is based on weighted least square estimation and was developed by Grizzle et al. (1969).

The second approach was developed by Binder (1983) and is based on pseudo maximum likelihood (PLME) for fitting logistic regression and other generalised linear models to complex sample survey data. For parameter estimates, the PLME approach was combined with a linearised estimator of the variance-covariance matrix while considering complex sample design and was used in this chapter for that reason. The PLME approach can be expressed as follows:

$$\mathrm{l} = \log(\alpha/y) = \sum_{j=1}^{J}\sum_{h=1}^{m_j}\sum_{k=1}^{m_{jh}}\rho_{jhk}\phi_{hjk}\ln f(y_{jhk,}\delta_{jhk}, \theta), \quad (3.6)$$

where ρ_{jhk}, ϕ_{hjk}, and $f(y_{jhk,}\delta_{jhk})$ are the weight, frequency, and the probability density function for the k^{th} individual in the h^{th} household nested within the j^{th}stratum, respectively (Binder, 1983; Breslow and Clayton, 1993). To calculate the parameters from equation (3.6), we have to differentiate the log-likelihood function indicated in equation (3.6) with respect to α. This will give us the gradient function that can be used to calculate the intended results, and the new equation will be:

$$D(\alpha) = \frac{\partial l}{\partial l}\sum_{j=1}^{J}\sum_{h=1}^{m_j}\sum_{k=1}^{m_{jh}}\rho_{jhk}\phi_{hjk}D'_{jhk}\Sigma'[y_{jhk} - \mu_{jhk}] = 0, \quad (3.7)$$

where, $D_{jhk=}\left[\frac{\partial \mu_{jhk}}{\partial \tau_{jhk}}\right] A_{jhk}$, A_{jhk} is the covariance matrix of explanatory variables, and Σ is the covariance matrix of y_{jhk}.

Mostly, there is no closed form of a fast remedy to equation (3.7), and an iterative framework is needed to get the maximum likelihood estimates of the unknown parameters α (Binder, 1983; Wolter, 2007). Generally, no closed version of a fast remedy to equation exists.

3.3. PARAMETERS ESTIMATION

The maximum likelihood estimator is used for survey logistic regression, and we assumed the outcome variable of y_{jhk} to follow a Bernoulli distribution with a density function and is shown in the following equation

$$f(y_{jhk}) = \pi_{jhk}^{y_{jhk}}(1-\pi_{jhk})^{1-y_{jhk}}. \tag{3.8}$$

The mean of y_{jhk} is presented as follow

$\mu_{jhk} = \frac{exp(Z'_{jhk}\alpha)}{1+exp(Z'_{jhk}\alpha)}$, where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_t)'$ is the vector of parameters.

The variance of $y_{jhk}$is given by

$\sigma^2 = \mu_{jhk}(1-\mu_{jhk})$ (Molenberghs and Verbeke, 2005). We can now write the log-likelihood function, which is the base for maximum likelihood estimation as

$$l = \sum_{j=1}^{J}\sum_{h=1}^{m_j}\sum_{k=1}^{m_{jh}} \rho_{jhk}\phi_{jhk}\left[y_{jhk}log(\mu_{jhk}) + (1-y_{jhk})log(1-\mu_{jhk})\right]. \tag{3.9}$$

Then substitute the value of μ_{jhk} into equation (3.9), to get the following:

$$l = \sum_{j=1}^{J}\sum_{h=1}^{m_j}\sum_{k=1}^{m_{jh}} \rho_{jhk}\phi_{jhk}\left[y_{jhk}log\left(\frac{e^{z'_{jhk}\alpha}}{1+e^{z'_{jhk}\alpha}}\right) + (1-y_{jhk})log\left(1-\frac{e^{z'_{jhk}\alpha}}{1+e^{z'_{jhk}}}\right)\right]. \tag{3.10}$$

To obtain the unknown parameters, we should first differentiate the log-likelihood from equation (3.10), with respect to α to achieve the following equation

$$\frac{\partial l}{\partial \alpha} = \sum_{j=1}^{J}\sum_{h=1}^{m_j}\sum_{k=1}^{m_{jh}} \phi_{jhk}\, D'_{jhk}\frac{e^{z'_{jhk}\alpha}}{\left(1+e^{z'_{jhk}\alpha}\right)}\left[\frac{y_{jhk}}{1-\left(1+e^{z'_{jhk}\alpha}\right)} - \frac{1-y_{jhk}}{1+e^{z'_{jhk}\alpha}}\right] z'_{jhk},$$

$$\frac{\partial l}{\partial \alpha} = \sum_{j=1}^{J}\sum_{h=1}^{m_j}\sum_{k=1}^{m_{jh}} \phi_{jhk}\, D'_{jhk}\left[\sigma^2(y_{jhk})\right]^{-1}\left[y_{jhk}-\mu_{jhk}\right], \tag{3.11}$$

where $D'_{jhk} = \mu_{jhk}(1 - \mu_{jhk})z'_{jhk}$.

When the fitted model is adequate, the estimators are asymptotically normally distributed with the exact value as the mean and the inverse Fisher information matrix as the covariance matrix (Binder, 1983; Molenberghs and Verbeke, 2005). Then the Fisher information matrix for the parameters of the Bernoulli model is deduced from the second derivative of log-likelihood with respect to α and can be expressed as follows:

$$\mathcal{F} = -\mathrm{E}\left[\frac{\partial^2 l}{\partial \alpha \alpha'}\right] = T z'^2_{jhk} e f^{-2}\left[y_{jhk}(1 - 2f^{-1}) - \left(1 - y_{jhk}\right)\left((1+e)^{-1} - 3(1+e)^{-2}e\right)\right],$$

where $T = \rho_{jhk}\phi_{jhk}$, $e = exp\left(z'_{jhk}\alpha\right)$, and $f = \left[1 + exp\left(z'_{jhk}\alpha\right)\right]$.

The final Fisher information is obtained by simplifying the above equation and becomes

$$\mathcal{F} = \sum_{j=1}^{J}\sum_{h=1}^{m_j}\sum_{k=1}^{m_{jh}} \rho_{jhk}\phi_{jhk}\, D'_{jhk}\left[\sigma^2\left(y_{jhk}\right)\right]^{-1} D_{jhk}\ . \qquad (3.12)$$

3.4. APPROXIMATE COVARIANCE MATRIX

The parameter estimation of the survey logistic regression model is obtained using the maximum likelihood method. Nevertheless, evaluating the standard errors of parameter estimates is complicated for data derived from complex designs. The complex survey design and weighting technique introduced in sampling contributes to the challenges in variability assessment. However, sampling information must be included to accurately assess a statistic's variance (Park, 2008). The process of weighting and sample designs are primarily useful in improving the efficiency of a statistic. Hence, their inclusion in the variance estimation methodology is critical and should be explored in detail (Schaefer et al., 2003; Lehtonen and Pahkinen, 2004). Therefore, the covariance-covariance matrix can be estimated using different techniques, such as Taylor linearisation, Jackknife, bootstrap, balanced replication, and random groups (McCarthy, 1969; Miller, 1974; Davison and Hinkley, 1997; Wolter, 2007).

Taylor approximation procedure

The Taylor series approximation procedure proposed by Binder (1983) is the most common procedure used to evaluate the variance of a linear statistic and covariance matrix of complex survey data. The use of the Taylor linearisation procedure on nonlinear statistics is simulated by the linear structure of observation through the application of the first-order terms in a suitable Taylor series. Improving the Taylor series could result in the generation of second-

order approximations, with an exception of biases in populations estimation. Nevertheless, in practice the first-order approximation generally produces a good result (Wolter, 2007). The Taylor series expansion is applied to evaluate the variance of the general estimator, by considering the population sample N. Then we assume q to be the dimensional parameter vector represented by $Z = (Z_1, Z_2, \dots, Z_q)'$, with Z_q indicating total population or means for q various survey properties, while $\hat{Z}_j$ represents the standard estimators of Z_j. According to the sample size s of $n(s)$, the associated estimator vector of Z is defined by $\hat{Z} = (\hat{Z}_1, \hat{Z}_2, \dots, \hat{Z}_q)'$ (Lehtonen and Pahkinen, 2004; Wolter, 2007). As a result, the estimators Z_j with $j = 1,2, \dots, q$ rely on the sampling design used to produce the sample s. Assume that a nonlinear parameter $\Phi = g(Z)$ with its reliable estimator is defined by $\hat{\Phi} = g(\hat{Z})$. Consequently, the main goal is to obtain an estimated expression for a designed variance of $\hat{\Phi}$ and to build an adequate variance estimator of $\hat{\Phi}$ (Wolter, 2007). Based on a Taylor series expression's linear terms, and assume that the function $g(Z)$ has a continuous second-order derivative, we get the following approximate linearised expression:

$$\hat{\Phi} - \Phi = \sum_{j=1}^{s} \frac{\partial g(Z)}{\partial Z_j} (\hat{Z}_j - Z_j), \tag{3.13}$$

where $\frac{\partial g(Z)}{\partial Z_j}$ terms are the partial derivative of $g(Z)$ with respect to Z_j.

Considering equation (3.13), we can get the variance approximation of $\hat{\Phi}$ as follow:

$$V(\hat{\Phi}) = V\left[\sum_{j=1}^{s} \frac{\partial g(Z)}{\partial Z_j} (\hat{Z}_j - Z_j)\right] = \frac{\partial g(Z)}{\partial Z_j} \times \frac{\partial g(Z)}{\partial Z_h} \times V(\hat{Z}_j, \hat{Z}_h), \tag{3.14}$$

where $V(\hat{Z}_j, \hat{Z}_h)$ represents the variance and covariance of the estimators $\hat{Z}_j$ and $\hat{Z}_h$ respectively. As a result, a nonlinear estimator's variance is simplified to a function of variances and covariance of s linear estimators Z_j (Fuller, 1975; Wolter, 2007). Moreover, substituting the variance and covariance estimators $\hat{V}(\hat{Z}_j, \hat{Z}_h)$ for the corresponding parameters $V(\hat{Z}_j, \hat{Z}_h)$ from equation (3.14), we get the variance estimator $V(\hat{\Phi})$ (Skinner et al., 1989; Wolter, 2007).

The achieved variance is defined as the first-order approximation, and by extending the Taylor series, second and even higher-order approximations may be obtained. In the practice, the first order approximation tends to produce consistent results, except in the case of extremely skewed populations (Wolter, 2007). The linearised statistic can be subjected to standard variance estimation methods. This indicates that the Taylor linearisation method is commonly utilised

and simple in any situation for which an estimator already exists for totals. The Taylor linearisation variance estimator, on the other hand, is a biased estimator. Its bias arises from its proclivity to underestimate the actual value, which is affected by the sample size and the complexity of the estimated statistic (Särndal et al., 2003; Wolter, 2007). However, when the statistic is straightforward, such as the weighted sample mean, the bias is statistically insignificant for small samples, and becomes zero for large samples. For a complex estimator such as the variance, large samples are required before the bias becomes negligible. However, it is a reliable estimator in every particular instance (Särndal et al., 2003).

Jackknife estimator

The jackknife method was devised by Quenouille (1949) and Quenouille (1956) to split the sample into different and distinct sections. Then one portion must be dropped, and the remaining portion must be used to recalculate the statistic of interest using the partial sample. The removed section is brought back into the sample, and the procedure is carried out gradually until every part has been taken out from the original sample. Then to evaluate the proportional variance, these duplicated statistics are applied. In basic random sampling, the previously described distinct portions could be a single observation, whereas, in multistage cluster sampling systems, they could be groups of units. There are several alternative representations of jackknife variance depending on how sampling units are input and re-entered into the sample. It should also be noted that the jackknife technique for variance estimation is more applicable in replacement designs, though it can be used in other situations as well such as surveys without replacement when the sampling fraction is low (Wolter, 2007). However, for a business survey, this is not always the case. The study by Shao and Tu (1995) revealed that the implementation of the jackknife estimator requires an adjustment to account for the sampling fractions when the first phase of sampling is done without replacement.

Regardless of their structure, Jackknife variance calculations appear to be more suitable for single or multistage cluster designs, where one cluster is excluded from the calculation in each replicate. In the case of large-scale surveys, the calculation of replicate estimates is time-consuming if the number of discontinuous portions is high, and makes the entire process time-consuming (Yung and Rao, 2000). Consequently, various jackknife approaches have been developed, where the adjustment of the common jackknife estimator relying on linearisation becomes the jackknife linearised variance estimation (Efron, 1982). The fundamental idea behind it is that analytical differentiation should be used instead of repeatedly recalculating a

statistic. For instance, the adjusted bias of variance formula for a stratified cluster sample with replacement becomes

$\widehat{\boldsymbol{V}} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}}(\mathbf{1} - \boldsymbol{g_k})\frac{\mathbf{1}}{\boldsymbol{n_{k.}(n_k-1)}}\sum_{\boldsymbol{h=1}}^{\boldsymbol{n_k}}\boldsymbol{t_{kh.}^2}$, where t_{kh} is the observed value for the h^{th} cluster in k stratum (Canty and Devison, 1999). The estimation of t_{kh} depends on the complexity of the statistic, and a linear estimator $\widehat{\Phi}$ with stratified cluster sampling gives the following formula: $\widehat{\Phi} = \sum_{hk} Z'_{hk}$, where $Z'_{hk} = \sum_k \psi_{jhk}$ is the sum of Z'_s in every cluster h in each stratum k, and ψ_{jhk} is the design weight. Then $t_{kh} = n_k Z'_{kh}$. Hence, when two calibrated estimators are compared, we get the following equation $t_{kh} = \frac{t_{kh}^{y} - \widehat{\Phi} t_{kh}^{z}}{t^T W_z}$ where $\widehat{\Phi} = \frac{t^T W_y}{t^T W_z}$ with y and z as the vectors of the dataset's observations, while t_{kh}^{y} , t_{kh}^{z} and W are computed and analysed from the data (Canty and Devison, 1999). The fundamental benefit of the jackknife estimator is that it requires less calculation while usually maintaining the positive characteristics of the original jackknife approach. On the other hand, for a non-linear statistic, it is necessary for the derivation of independent equations. As a result, its applicability for complex sample designs or complex analyses of survey data is sometimes restricted (Rao, 1997; Canty and Devison, 1999; Holmes and Skinner, 2000).

Bootstrap estimator

The bootstrap method was developed to apply to different data sets. It was established beyond the domain of the survey sampling concept (Efron, 1979; Efron, 1981; Efron, 1982). This method was developed for the samples of independent and similar distribution observations. However, non-independence between observations when sampling without replacement along with other complexities are still concerned to be investigated (Efron, 1982).

The bootstrap estimator has been the subject of extensive theoretical research, and has gained popularity as a method for typical data analysis (Särndal et al., 1992; Shao and Tu, 1995). The bootstrap effectively takes a number of separate samples from the observed data. Thereafter, employ the same sampling procedure as the one used to select the original sample from the entire population and estimate the size of each bootstrap sample (Rao and Wu, 1988).

Balanced repeated replication method

The balanced repeated replication (BRR), pseudo-replication method was developed for the survey design with a huge number of strata. The method is only restricted to many strata with

two parameters per strata and a cluster design with two final stage components. The purpose of this method is to select a cluster sample from $2k$ stratum population and assess each independently, then use those approximate values to draw a biased sample for the variance estimator (Judkins, 1990; Särndal et al., 1992).

The structure of pseudo samples in the BRR procedure begins with k strata and a sample cluster or primary sampling units (PSUs) per stratum $k = 2$. Such procedures become replicated when there are no cluster or primary sampling units (PSUs) per stratum. The total sample can be divided into 2^H intersecting half-samples each with H as sample clusters. As result, one can build an estimate $\widehat{\Phi}_j$ for every half sample and apply it to estimate the $V(\widehat{\Phi})$ (Judkins, 1990).

However, assessing those 2^H possible of $\widehat{\Phi}_j$ at times seems to be difficult to solve. Thus, a consistent set of half-samples with a minimum multiple of 4, and larger than H needs to be selected.

Therefore, the variance estimator can be shown as follows:

$$V(\widehat{\Phi}) = \sum_{j=1}^{h} \frac{(\Phi_j - \widehat{\Phi})^2}{h}. \tag{3.15}$$

The estimator from equation (3.15) and the estimator obtained from all 2^H half-samples have similar asymptotic accuracy. The additional computation required must be weighed against the improvement in variance estimate accuracy over simple replication. The accuracy of estimated variance over the simple replication can be adjusted to the additional estimation needed (Rao and Wu, 1985). The study by Rao and Shao (1996) revealed that the exact asymptotic estimator could be attained through repeated division. As a result, using BRR in business surveys is problematic since stratification is frequently used, and making the modification with both data and software is challenging. The advantage of the BRR method induces strong asymptomatic assumptions for smooth and non-smooth functions. Nevertheless, the BRR method is not as highly adaptable for a random sample size n_h as the Bootstrap and Jackknife methods (Rao, 1997).

Random groups method

The random groups method was initially implemented to minimise the variance estimation for a complex survey design. The parameter estimation is done by selecting a sub-sample from the population. While the estimation of variance is based on the variations from the

combination of all sub-samples (Wolter, 2007). However, the survey's structure should include s independent replicates of the same design to evaluate the variance (Skinner et al., 1989).

Assume $\hat{\Phi}$ to be an estimator of Φ from the entire sample, and any estimator s for the entire sample can be determined at each of the s replicates yield $\hat{\Phi}_1, \ldots, \hat{\Phi}_s$. The estimator $\hat{\Phi}_j$ is achieved from the s^{th} random cluster with $\hat{\Phi}' = \sum_{j=1}^{s} \frac{\hat{\Phi}_j}{s}$. Thus, the variance estimator can be calculated as $V(\hat{\Phi}') = \frac{1}{s(s-1)} \sum_{j=1}^{s} (\hat{\Phi}_j - \hat{\Phi}')^2$. Then, $\hat{\Phi}$ can be calculated based on $V(\hat{\Phi}')$ where

$V(\hat{\Phi}) = \frac{1}{s(s-1)} \sum_{j=1}^{s} (\hat{\Phi}_j - \hat{\Phi})^2$ (Wolter, 2007). The random group's procedure can be categorised into two groups considering the sub-samples are independent or otherwise. When the sub-sample is independent, the random group's technique gives unbiased linear estimators, even if bias may arise in the evaluation of non-linear survey data. However, with dependent random groups, the results will be biased and such bias is negligible in surveys conducted with a low sampling percentage (Hansen et al., 1953; Wolter, 2007). In such cases, the homogeneity of the fundamental sampling strategy of each sub-sample is needed to guarantee the characteristics of variance estimators for a random groups method (Wolter, 2007).

3.5. MODEL SELECTION

The model selection can be done in different ways; however, the most common methods used to select the variables to be included in the model are forward, backward, and stepwise. In the forward selection process, we start with the intercept coefficient and add one explanatory variable at a time. The backward procedure is often used when the predictor variables are few in the model and the process begins with all explanatory variables and subtracts one explanatory variable at a time (Hosmer et al., 2000). The only difference between the stepwise selection procedure and forward selection is that in stepwise all variables included in the model are observed for exclusion when a new variable enters the model. When the study involves large data, the stepwise procedure is more advisable as the process reduces the chances of keeping unnecessary variables. When all variables in the model reach the criterion to remain and no variables outside the model satisfy the criterion to enter, the stepwise selection of variables ends. After that one excludes a variable with a nonsignificant effect and assesses the contribution of the remaining variables. This process will be repeated until the model has only significant effects (Hosmer et al., 2000; Hosmer et al., 2013).

3.6. MODEL DIAGNOSTICS

It is always good to check the adequacy of the model after fitting the model to the data. The adequacy of the model is known as goodness-of fit, and after checking how well the model fits the data, it can be accepted or required to be revised. The most commonly used method to assess the goodness-of-fit are the log-likelihood ratio (deviance) and Pearson's Chi-square (Fahrmeir et al., 1994; Hosmer et al., 2000; Kutner et al., 2005). To assess the difference between the maximum log-likelihood achievable and achieved log-likelihood, the deviance is used and can be calculated as follow:

$D\ (Y, \hat{\mu}) = 2\{l(y, y) - l(\hat{\mu}, y)\}$, where (y, y) is the log-likelihood in the maximum achieved model or saturated model, while $(\hat{\mu}, y)$ is the log-likelihood in the present model. The target is to minimize $(Y, \hat{\mu})$ by maximising $(\hat{\mu}, y)$, and we test the goodness-of-fit by setting the null hypothesis $H_0 = model\ is\ adequate$ vs the alternative $H_0 = model\ is\ inadequate$. The null hypothesis is rejected when $> \chi^2_{n-p,\alpha}$, with n as the observed number, p the number of parameters in the model, and α the level of significance (Hosmer et al., 2000; Jiang, 2001). The deviance test is not reliable in measuring the goodness-of-fit when one has ungrouped data. In this case, the Hosmer-Lemeshow goodness-of-fit test will be relevant. Then the predicted probabilities $(\hat{\mu}_i's, i = 1,2,3, \dots, n)$ generated from the current model are used to produce g groups with an estimated g/n subjects (Hosmer et al., 2000). The grouping techniques are based on the percentile strategy developed by Hosmer et al. (2000) as follows.

The first group subjects are roughly n/g subjects, with $\hat{\mu}_i's$ less or equal to the $100/g^{th}$ percentile of all $\hat{\mu}_i's$. In the second group, 10 subjects are roughly n/g subjects n/g with $\hat{\mu}_i's$ more than $\left(1 - \frac{1}{g}\right) \times 100^{th}$ percentile of all $\hat{\mu}_i's$. The last group k subject, for $k = 1,2,3, \dots, g-1$ are approximately n/g, with $\hat{\mu}_i's$ greater than the $\frac{k-1}{g} \times 100/g^{th}$ percentile and less than or equal to the $\frac{k}{g} \times 100/g^{th}$ of all $\hat{\mu}_i's$ (Hosmer et al., 2000). When n becomes larger, the value of $g = 10$ assists in making a consistent conclusion and evaluating the observed and expected frequencies of outcome $y = 0\ and\ y = 1$ for each group. Therefore, the Hosmer-Lemeshow goodness-of-fit χ^2 statistic test is achieved by calculating the Pearson Chi-square statistic from $2 \times g$ tables of observed and expected frequencies. Then, Hosmer-Lemeshow goodness-of-fit statistic can be expressed as

$\chi^2 = \frac{(O_k - N_k\hat{\pi}_k)}{N_k\hat{\pi}_k(1-\hat{\pi}_k)}$, where N_k is the total frequency of the subject in the k^{th} group, O_k is the total frequency of the outcome in the k^{th}, and $\hat{\pi}_k$ is the estimated average probability for the outcome in k^{th}, with $\hat{\pi}_k = \sum_{i=1}\left(\frac{n_i\hat{\pi}_i}{N_k}\right)$ and n_i is the number of the subject of z_i and $O_k = \sum_{i=1}^{c_i} y_k$ as a response between the c_i covariate patterns. Then, the Hosmer-Lemeshow statistic can be compared with the critical value of Chi-square distribution (χ^2) with $(g-n)$ degree of freedom, with a specified n. When the χ^2 is non-statistically significant, indicating the model goodness-of-fit otherwise, the model is not well fitted (Hosmer et al., 2000; Jiang, 2001). Besides the significances tests, the Akaike Information Criterion (AIC), and Schwartz Criterion (SC), or Bayesian Information Criterion (BIC) can be used to evaluate goodness-of-model-fit. The smaller the AIC and BIC of the full model compared to the corresponding AIC and BIC of the reduced model, the better the full model is. More details of standard criteria for model selection such as AIC, BIC among others can be found in studies such as those by Akaike (1974); Schwarz (1978); Gameroff (2005); Lumley and Scott (2015) just to name a few.

3.7. APPLICATION OF THE SURVEY LOGISTIC REGRESSION MODEL TO CHILDHOOD ANAEMIA DATA

Assume y_{jhk} to be the response variable, anaemia status of child *j* from the h^{th} cluster and k^{th} stratum. The outcome variable is defined as a dichotomous variable such that $y_{jhk} = 1$ if the child *j* is anaemic, and $y_{jhk} = 0$ if the child *j* is not anaemic. In the present study, we have assumed that outcome variable y_{jhk} is Bernoulli distributed as $y_{jhk}|\mu_{jhk} \sim Bernoulli\ (\mu_{jhk})$, with μ_{jhk} known as the mean and is defined as $E(y_{jhk}) = \mu_{jhk}$. It is linked to the independent variables :
$g(\mu_{jhk}) = z'_{jhk}\alpha$, where g (.) is the logit link function, and α is an *m*-dimensional vector of categorical explanatory variables.

3.8. DATA ANALYSIS

The present study used both 2009 and 2014 Lesotho Demography and Health Survey (LDHS) data to assess anaemia in children younger than five years. The survey logistic regression model was used to account for the complexity of the sampling design and heterogeneity between observations from the same cluster, to avoid the underestimation of the variance and wrong inference (Schneider et al., 2008; Gari et al., 2017). The analysis was done using proc survey logistic from SAS software version 9.4 and did not support a forward, backward, or stepwise

variable selection approach. Consequently, one variable was selected from the model at a time and the impact of each variable was checked. The model fit was assessed based on the smallest Akaike Information Criteria (AIC), and -2 log-likelihood (-2LogL). Any variable that was significant at a 5% level of significance was considered otherwise and it was dropped in the analysis in both data sets of 2009 and 2014.

3.9. RESULTS AND INTERPRETATIONS

The results from the 2014 data showed that the full model is the best fit compared to the model with Intercept since it is the one with the lowest AIC. In addition, we checked the possible interactions and none were statistically significant (Habyarimana et al., 2016; Gaston et al., 2018).

Table 3.1 presents the parameter estimate, standard deviation (STD), *p*-value, odds ratio (OR), and confidence interval (CI). The results revealed that the child's age group, his/her nutritional status (stunting), fever in the last two weeks and the mother's body mass index were significantly associated with anaemia among children under five years in Lesotho. A child whose age is less than 19 months was found to be 0.471 (OR: 0.471 (0.323; 0.687)), *p*-value <0.001) times less likely to be anaemic as compared to those in the age group of 40-50 months while a child between 20-39 months of age was 0.687 (OR: 0.687 (0.482; 0.977), *p*-value=0.037) times less likely to be anaemic than those who were in the age group of 40-59 months. A child who had a fever in the two weeks before the survey was 1.674 (OR: 1.674 (1.103; 2.540), *p*-value= 0.016) times more likely to be anaemic than a child who did not have a fever in the two weeks before the survey. It was noted that a stunted child was 1.787 (OR: 1.787 (1.219; 2.619)), *p*-value=0.003) times more likely to be anaemic than a child who was not stunted. It was also noted that the mother's BMI was significantly associated with childhood anaemia. A child born to an underweight mother was 1.542 (OR: 1.542 (1.024; 2.321), *p*-value=0.038) times more likely to be anaemic compared to a child born to normal weight or obese mother.

Table 3.1: Survey logistic regression analysis in Lesotho.

Variable	Estimate	STD	*p*-value	OR	95% CI
Intercept	-0.474	0.609	0.437	-	-
Stunting(ref=no)	-	-	-	-	-
yes	0.581	0.194	0.003	1.787	1.219;2.619
Fever (ref=no)	-	-	-	-	-
Yes	0.515	0.212	0.016	1.674	1.103;2.540
BMI (ref=less than18.5)	-	-	-	-	-
More or equal 18.5	0.433	0.208	0.038	1.542	1.024;2.321
Child age (ref=40-59)	-	-	-	-	-
0-19	-0.752	0.191	<0.001	0.471	0.323;0.687
20-39	-0.376	0.179	0.037	0.687	0.482;0.977

***OR**: odd ratio; **95% CI**: Confidence Interval, **STD**: standard deviation*

3.10. DISCUSSION AND SUMMARY

The prevalence of anaemia among children under five years in Lesotho was 48%, 47% and 51% in 2004, 2009 and 2014 respectively. This shows an increase of 4% in five years. The main objective of the current study was to assess the risk factors of anaemia among children under five years based on the 2014 LDHS.

The findings from this study revealed that childhood anaemia in Lesotho decreases with the increasing age of the child. These findings agree with many previous studies such as those by Kotecha (2011); Zhao et al. (2012), Pita et al. (2014); Dey and Raheem (2016); Gaston et al. (2018). This means that older children have a lower likelihood of getting anaemia than younger children.

The findings from this study also highlighted the association of height for age with anaemia among children under five years. This finding has been identified elsewhere (Ngnie-Teta et al., 2007; Kotecha, 2011; Zhao et al., 2012; Leite et al., 2013; Khan et al., 2016; Gaston et al., 2018). This is an indication of chronic malnutrition which might cause iron deficiency (Woldie et al., 2015; Gaston et al., 2018).

The findings revealed a significant association between a mother's body mass index and childhood anaemia. The results showed that children born to an underweight mother had a higher likelihood of being anaemic. A similar result was found in studies by Fonseca et al. (2016); Habyarimana et al. (2017); and Gaston et al.(2018). This may be because underweight people are more likely to have other associated co-morbidity illnesses that may be associated with anaemia.

In addition, the findings from this study demonstrated that the recent incidence of fever had a positive impact on childhood anaemia. The results showed that a child who had a fever in the two weeks before the survey had a higher likelihood of being anaemic than a child who did not have a fever. This is in line with results found in studies by Santos et al. (2011); Konstantyner et al. (2012); Gayawan et al. (2014); Gaston et al. (2018) among others. This may be due to the fact that fever is commonly accompanied by a number of diseases and morbidities that are known to positively affect anaemia such as diarrhoea, cough and malaria, among others (Konstantyner et al., 2012; Gaston et al., 2018).

The findings from this chapter highlighted that the child's age and stunting were common risk factors for childhood anaemia. It was also found that fever and mother's body mass index were significant factors associated with childhood anaemia in 2014. In addition, the mother's anaemia status was not statistically significant.

The findings also revealed that the child's malnutrition status, child's age, fever, and mother's body mass index were determinants of childhood anaemia in Lesotho. The findings from this study may assist public health institutions in Lesotho and policy makers to formulate preventative measures and design intervention strategies that target children under five years.

The survey logistic regression models used in this chapter are powerful tools and fit our data well. The model accounted for the complexity of sampling design and heterogeneity between observations from the same cluster, to avoid the underestimation of the variance and wrong inference. However, the model assumes that all variables have a fixed effect and does not allow for the inclusion of random effects. The study used DHS data and had the primary sampling units (clusters) variable, which is considered as a random effect. In addition, survey logistic is a parametric model. In some cases, the parametric models may struggle with their rigidity when simulating complex relationships between the outcome variable and the predictor factors.

Therefore, in the next chapter we will introduce the generalised additive mixed model (GAMM) which will take care of the mentioned issues.

CHAPTER FOUR:

GENERALISED ADDITIVE MIXED MODEL

4.1. INTRODUCTION

In Chapter 3, we used the survey logistic regression model, which is a parametric model. The parametric models produce an effective tool for modeling the association between response variables and predictor variables. Nevertheless, these parametric models cannot change the parametric fixed effects to be modelled as nonparametric (Hastie and Tibshirani, 1990; Gaston and Ramroop, 2020). Hence, the nonparametric models were developed to minimise possible modelling biases in parametric models (Hastie and Tibshirani, 1990; Lin and Carroll, 2000). The literature on nonparametric models and their applications are reviewed in different studies such as those of Hastie and Tibshirani (1990); Davis (1991); Akritas (1996); Holmes et al. (1996); Wood (2006); Enjuanes et al. (2016); Silverman (2018) among others. Nonparametric approaches aim to adjust a higher set of predictors to minimise the bias in the model. These models also enable statisticians to find nonlinear types of models that adequately explain the available information. Furthermore, they are useful for parametric nonlinear modelling and modelling diagnostics tests (Hastie and Tibshirani, 1990). The models have a large number of regression and smoothing techniques. These techniques include kernel-smoothing, spline fitting or smoothing, L-smoothing, R-smoothing, M-smoothing, and locally weighted scatterplot smoothing and are linked to one another, although each method has unique properties that are beneficial in various studies (Härdle, 1990; Wu and Zhang, 2006). However, there are some issues with the additive (nonparametric) models such as model selection, overfitting, and multicollinearity.

Numerous studies have investigated potential solutions to nonparametric challenges. Different methods have been improved with considerable effort to decrease the complexity of estimators with high dimensions and enable slightly parametric modelling (Härdle, 1990; Hastie and Tibshirani, 1990). The use of both parametric and nonparametric techniques is complementary rather than competitive. Nonparametric methods can sometimes be used to support or recommend a parametric model (Wu and Zhang, 2006). In many applications, combining nonparametric and parametric models is more powerful than using a single method (Härdle et al., 2004; Hastie and Tibshirani, 1990; Ruppert et al., 2003; Wu and Zhang, 2006).

Since the relationships between the outcome and covariate variables might have an unknown functional form, it has to be indicated in several applications. This inspires the researchers to investigate the semiparametric models called generalised additive models (GAM) and its extension generalised additive mixed model (GAMM). This was achieved by including a nonlinear parametric factor into the additive indicator to the link scale (Hastie and Tibshirani, 1990; Wood, 2006).

As a result, this chapter aims to introduce a generalised additive mixed model (GAMM) to evaluate childhood malaria and other risk factors associated with malaria using the 2017 Malawi Malaria Indicator Survey.

Generalised additive mixed model (GAMM) is an extension of the generalised additive model (GAM) that adds the random effect in the model to assess the association among the observations (Wang, 1998; Gaston and Ramroop, 2020). The choice of GAMM over GAM is because GAM includes only the covariance effect in models without the random effect (Wang, 1998; Lin and Zhang, 1999; Gaston and Ramroop, 2020). In addition, the generalised additive mixed model is an extension of the generalised linear mixed model (GLMM) introduced by Breslow and Clayton (1993). By employing the additive smooth function and the GAMM, it is possible to express the parametric fixed effects from the GLMM as a non-parametric model (Hastie and Tibshirani, 1990; Gaston and Ramroop, 2020).

4.2. MODEL FORMULATION

The generalised additive model (GAM) is the same as the semiparametric additive model, which was developed by Hastie and Tibshirani (1986). The GAM is applied to the data, to identify the relationship between the response and covariates variables. The parametric models also have powerful tools in modelling the relationship between the response and predictors variables when their assumptions for a linear, or some parametric form of the covariate effects are not violated. The parametric models in applications such as determining the relationship between the response and covariates variables, may have unknown functional form and are complicated (Ayele et al., 2014a). The unknown functions may lead to applications of semiparametric additive models which are flexible to allow non-normal error distributions. Furthermore, semiparametric additive models relax the assumption of normality and linearity in linear regression (Lin and Zhang, 1999; Ayele et al., 2014a).

The use of a semiparametric additive model may allow the response variable to be modelled with Poisson and binomial distribution. Moreover, the nonparametric models are flexible for modelling the continuous predictor variables. The GAM extends the generalised linear model (GLM) by allowing the predictor function to include the unspecified nonlinear function for some or all of the covariate variables (Hastie and Tibshirani, 1990). The linear form was replaced with the additive form and hence the general equation for GLM as $\sum_{i=1}^{n} x_i \beta_i$ becomes $\sum_{i=1}^{n} f(x_i)$ in GAM. Thus, the equation of GAM is written as follows:

$$g(\mu_i) = X_i + f_i(x_{1i}) + f_i(x_{2i}) + f_i(x_{1i}) + \cdots + f_k(x_{ki}). \quad (4.1)$$

From Equation (4.1), X_i is the designed matrix, f_i are the smooth functions of covariates, while g (.) is the monotonic differentiable function, with $\mu_i = E(y_i/b)$ (Wood, 2017). If there is no linear component in equation (4.1), the model is known as nonparametric, whilst the models whose predictors have both linear and unspecified nonlinear function are semi-parametric. To estimate the parameters, the standardised condition of the smooth functions f_i should be satisfied such that E $[f_i X_i]$ =0, apart from that, each function will have free constants (Hastie and Tibshirani, 1990).

When the data has repeated measurement or correlations, the model includes a random variable, and this leads to the extension of GAM. Hence, GAM becomes the generalised additive mixed model (GAMM) in the same way as the generalised linear mixed models (GLMM) are an extension of GLM (Hastie and Tibshirani, 1990). The GAMM was introduced by Breslow and Clayton (1993) to include the random effect in the GAM and model the correlation between the observations.

The equation of GAMM can be expressed as follows:

$$g(\mu_i) = \beta_i + f_i(x_{1i}) + f_i(x_{2i}) + \cdots + f_k(x_{ki}) + Z_i b, \quad (4.2)$$

where g (.) is monotonic differentiable link function, $y_i, i = 1, \ldots,$ n is outcome variable, k covariates $X_i = (1, x_{1i}, \ldots, x_{1k})'$ associated with fixed effects and $q \times 1$ vector of covariates Z_i associated with random effects. Thus, the given $q \times 1$ vector of random effect b, the observations y_i are assumed to be conditionally independent with means, $E(y_i/b) = \mu_i$ and $(y_i/b) = \psi v(\mu_i)$, where v (.) is specified variance function and ψ is a scale parameter. Moreover, $f_i(.)$ is a centred twice differentiable smooth function and the random effects b is assumed to be distributed as $N\{0, G(\rho)\}$ and ρ is a $c \times 1$ vector of variance components. In

addition, when f_i is a linear function, the GAMM reduces to GLMM (Lin and Zhang, 1999; Gaston and Ramroop, 2020).

For a specified variance component, θ the log-likelihood function of (β, f_i, θ) is expressed by Lin and Zhang (1999) in the following equation:

$$\exp\left[l\{y; \beta_0 f_1(.), \dots, f_k(.), \theta\} \propto\right] |G|^{-1/2} \int \exp\{-\frac{1}{2\tau}\sum_{i=1}^{n} di(y; \mu_i) - \frac{1}{2} b' G^{-1} b\} db, \qquad (4.3)$$

where $y_i = (y_1, \dots, y_n)'$ and $d_i(y, \mu) \propto -2\int_{y_i}^{\mu_i} \frac{y_i - \mu}{w(u) d_u}$, define the conditional deviance the function of (β, f_i, θ) given b. The statistical inference for GAMM on nonparametric function f_i requires the estimate of smoothing parameter τ and the inference on variance component θ.

The smoother spline estimators and the linear mixed models are closely related (Wang, 1998; Green and Silverman, 1993; Lin and Zhang, 1999). Moreover, the natural cubic smoothing spline estimators of function f_i maximise the penalised log-likelihood for the same given τ and θ and give the following equation:

$$\exp[l\{y;\ \beta_0, f_1(.), \dots, f_k(.), \theta\}] - \frac{1}{2}\sum_{i=1}^{k} \tau_i \int_{s_i}^{t_i} f_i'' x^2 dx$$

$$= l\left[y;\ \beta_0 f_1(.), \dots, f_k, \theta\right] - \frac{-1}{2} \sum_{i}^{k} \tau_i\, f_i' S_i f_i\ , \qquad (4.4)$$

where the s_i and t_i indicate the range of i^{th} covariate and τ_i are the smoothing parameters that manage the trade-off between goodness-of-fit and the smoothness of the estimated functions (Lin and Zhang, 1999; Ayele et al., 2014a; Gaston and Ramroop, 2020). Moreover, $f_i(.)$ is an $r_j \times 1$ unknown vector of the values of $f_i(.)$, estimated at r_j ordered values of the x_{kj}, where k= (1,…, n) and s_j is the smoothing matrix (Green and Silverman, 1993). By using the matrix form, the GAMM given in equation (4.2), can be written as:

$$g(\mu_i) = 1\beta_1 + N_1 f_1 + \cdots + N_p f_p + Z_i b, \qquad (4.5)$$

where $g(\mu_i) = [g(\mu_1), g(\mu_2), \dots, g(\mu_n)]$, $n \times 1$ the vector of ones, $N_i = m \times r$ matrix, such that, the k^{th} component of $N_i f_i\ is\ f_i x_{ki}$ and $Z = Z_1, Z_{2,\dots,} Z_n$. To evaluate the equation (4.6), the numerical integration is required. Additionally, calculating the natural cubic smoothing spline estimators of f_i by maximising equation (4.5) is sometimes complicated. Consequently, Li and Zhang (1999) resolved this problem by suggesting the double penalised quasi-likelihood (DPQL) model as an alternative approach. Hence, the estimation of nonparametric function f_i

can be obtained by applying double quasi-likelihood. The function f_i is re-parameterised in terms of β_i and a_k in one-to-one transformation as:

$$f_k = X_k^* \beta_k + \beta_k a_k, \tag{4.6}$$

where X_k^* is $r_k \times 1$ vector with the r_k centred and ordered distinct values of the x_{ki} $(k = 1,2,\dots,n)$ and $\beta_{k=}L(L_k L_k)^{-1}$ and L_k is an $r_k(r_k - 2)$ full rank matrix satisfying $S_k = LL'$ and $L_k' x_k^* = 0$. Thus, the double penalised quasi-likelihood with respect to $(\beta_0 f_i)$ and b becomes:

$$-\frac{1}{2\tau}\sum_{i=1}^{n} di(y;\mu_i) - \frac{1}{2} b' G^{-1} b - \frac{1}{2} a' D^{-1} a, \tag{4.7}$$

where $f_k' S_k f_k = a_k' a_k$, $a = (a_1', a_2', \dots, a_k')$ and $D = diag(\rho_1 I, \rho_2 I, \dots, \rho_k I)$ with $\rho_k = \frac{1}{\tau_k}$. Note that the small values of $\rho_k = (\rho_{1,}\rho_{2,\dots,}\rho_k)$ correspond to over smoothing (Breslow and Clayton, 1993; Lin and Zhang, 1999; Gaston and Ramroop, 2020).

4.3. MODEL SELECTION AND DIAGNOSTICS

The smoothing parameter value has a significant impact on model fit, and for that reason, the selection and goodness fit of an appropriate smoothing parameter is crucial. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be used in the section and goodness of model fit (Akaike, 1998; Schwarz, 1978). However, the most common used are the cross-validation-criterion (CV), and generalised cross-validation criterion (GCV) (Hastie and Tibshirani, 1990).

The AIC procedure proposed first by (Akaike, 1998) is good to use when the model has more predictors, as it can prevent data prediction errors by penalizing high model challenges, and the small value of AIC the better the model fit. Besides applying a penalty for the addition of variables, the AIC considers both the goodness of fit and the number of variables that need to be estimated to obtain the degree of fit needed (Akaike, 1998; Everitt, 1998). Assume that a parameter estimate is defined as ϑ, then the AIC can be written as $AIC = -2\left[l(\hat{\vartheta}) + q\right]$. The $l(\hat{\vartheta}) = log[f(y/\vartheta)]$ is the maximum log-likelihood value of the model, and q is the number of estimated parameters used in the model.

The BIC approach was introduced by Schwarz (1978)**,** and works as AIC, where the small value of BIC the better the model fit. However, this approach is good to use when the model

has few covariates, while AIC works better with many covariates in the model. The Bayesian Information Criterion can be expressed as follow

$BIC = -2\, l(\hat{\vartheta}) + q\ log(n)$, where $l(\hat{\vartheta})$ is the maximized value of the log-likelihood function, with q indicates the number of parameters, and n the sample size. The penalty term is determined by the sample size, with larger sample size resulting in a larger penalty, and reduced as the sample size decreases (Schwarz, 1978).

The smoothing parameter selected has a significant impact on model fit, and the selection criteria can also use cross validation (CV) and generalised cross-validation criterion (GCV).

The two approaches select the smoothing parameter in such way the smooth function estimate $\hat{f}$ could be close to the actual f. Assume that the cross-validation criterion chooses the value τ that minimizes the following expression $CV(\tau) = \sum_i^n \left[y_i - \hat{f}_i(x_i; y_i)\right]^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{(1 - S_\tau, ii)^2}$.

The τ is the smoothing parameter, and $\hat{f}$ the regression estimator, while the $(x_i; y_i)$ and $S_\tau,\ ii$ are the i^{th} diagonal elements. Then, the generalized cross validation criteria select the smoothing parameter τ that minimises $CV(\tau) = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\left(1 - \frac{tr(S_\tau)}{n}\right)^2}$. The cross-validation and generalised are very similar, however, the GCV has a good attribute property (Wood, 2006).

4.4. MODEL FITTING

The study used R software to analyse the 2017 Malawi Malaria Indicator Survey (MMIS) data with the application of "mgcv" packages. The GAMM was used to model the effect of age and altitude non-parametrically, while other covariates were used as parametric. These factors have a continuous effect and might have non-linear relationships with malaria (Ayele et al., 2014a). The R software has packages with numerous choices for controlling the smoothness in the GAMM using splines. Various splines can be used such as cubic smoothing splines, bin smoothers, shrinkage smoothers, locally-weighted running line smoothers and kernel smoothers among others (Hastie and Tibshirani, 1990; Ruppert et al., 2003). However, this study used shrinkage smoothers (splines) to fit the GAM model, due to its advantages such as assisting to control the knot placement. Furthermore, the shrinkage smoother is constructed in such a way that the smooth terms are rebuffed away all around (Wood, 2006; Gaston and Ramroop, 2020). The study also considered the fundamental impact and possible two-way interaction effect. The *p*-value of the individual smooth term and the AIC of each model,

together with the inference of smooth were analysed. The selection of the model was based on the smallest AIC, the higher value of degree of freedom and high statistical significance. Hence, the final model for this study is given in equation (4.8) as follows:

$$g(\mu_{ij}) = \beta_0 + \beta_1 Anemia_j + \beta_2 Region_j + \beta_3 Residence_j + \beta_4 Toilet\_facility_j + \beta_5 Wealt\_Index_j + \beta_6 Electricity_j + \beta_7 Mothers'education_j + f_1(Age_j) + f_2(Altitude_j) + b_{oj}, \quad (4.8)$$

where, $g(\mu_{ij})$ is the logit link function, $\beta's$ are the parametric regression coefficients, $f'_j s$ are centred smooth functions, while b_{oj} is the random effects, which can be written as $b_{oj} \sim N(0, G(\theta))$.

4.5. INTERPRETATION OF RESULTS

Table 4.1 illustrates the parameter estimates for the model, standard error, z-value, odds ratio and *p*-values. The study reported the variables with significant impact on the malaria RDT such as anaemia, electricity, region, residence, wealth index, toilet facilities and mother's education status. The study checked all possible interactions. However, the two-way interaction effect was not included since it did not add any significant effect to the model with non-significant p-values.

Table 4.1 shows that the children with no anaemia were 0.233 times less likely to test positive for malaria using an RDT, compared with anaemic children. The results also revealed that the odds of positive malaria results in an RDT for children living in the Central region of Malawi were 1.936 times more likely than for those who lived in the North region. Similarly, the odds of positive malaria in an RDT for children living in the South region were 1.179 times more likely than for those who lived in the North region. The children living in rural areas were 4.318 times more likely to test positive for malaria in RDT results compared to those living in urban areas. The study also showed that the odds of positive malaria results in an RDT test for children from a household with no toilet facilities were 2.938 times more likely than those with flush toilets. Furthermore, the children from households with pit latrines were 1.389 times more likely to test positive for malaria in an RDT, compared to those with flush toilets. The results indicated that the children from the middle classes were 0.743 times less likely to test positive for malaria using an RDT, compared to those from the poorer classes. In addition, the children from the richer classes were 0.571 times less likely to test positive for malaria in an RDT than those from the poorer classes. Lastly, the results indicated that the odds of positive malaria

resulting in an RDT for children from households with access to electricity was 0.435 times less likely than those from households with no access to electricity.

Table 4.1: The parameter estimates and odds ratio of GAMM model of the main parametric models in Malawi.

Variable	Estimate	Standard Error	z-value	Odds ratio	*p*-value
Intercept	-2.086	0.4620	-4.514	0.124	<0.001
Anaemia (Ref=Yes)	-	-	-	-	-
No	-1.455	0.152	-9.577	0.233	<0.001
Region (Ref=North)	-	-	-	-	-
Central	0.668	0.279	2.395	1.936	0.017
South	0.143	0.317	0.450	1.179	0.653
Residence (Ref= Urban)	-	-	-	-	-
Rural	1.463	0.293	4.994	4.318	<0.001
Toilet facility (Ref=Flush toilet)	-	-	-	-	-
Pit Latrine	0.329	0.226	1.456	1.389	0.145
No facility	1.078	0.343	3.143	2.938	0.002
Wealth Index (Ref=Poorer)	-	-	-	-	-
Middle	-0.298	0.182	-1.631	0.743	0.103
Richer	-0.560	0.190	-2.956	0.571	0.003
Electricity (Ref=No)	-	-	-	-	-
Yes	-0.832	0.317	-2.623	0.435	0.009
Mothers' education (Ref=No education)	-	-	-	-	-
Primary	-0.102	0.201	-0.508	0.903	0.611
Secondary	-0.541	0.273	-1.982	0.582	0.047
Tertiary	-2.274	1.251	-1.817	0.103	0.069

Table 4.2 below shows that the age of the child and the altitude of their region of residence has a significant impact on malaria RDT results.

The letter S in **Table 4.2** represents the smooth term and the number in parentheses shows the estimated degree of freedom (EDF). The test statistics for child age and altitude of region of residence (22.340; 90.420 respectively) together with their *p*-value (<0.001; <0.001) shows that there is no linear trend associated either for child age or for altitude. This is confirmed in **Figure 4.1**, where the trend shows that the effect of malaria RDT results increases with age to approximately 35 months and afterward remains constant with no sign of decreasing. The same results indicate that the effect of malaria RDT results increases with the altitude of their region of residence approximately 750 metres and afterward starts decreasing.

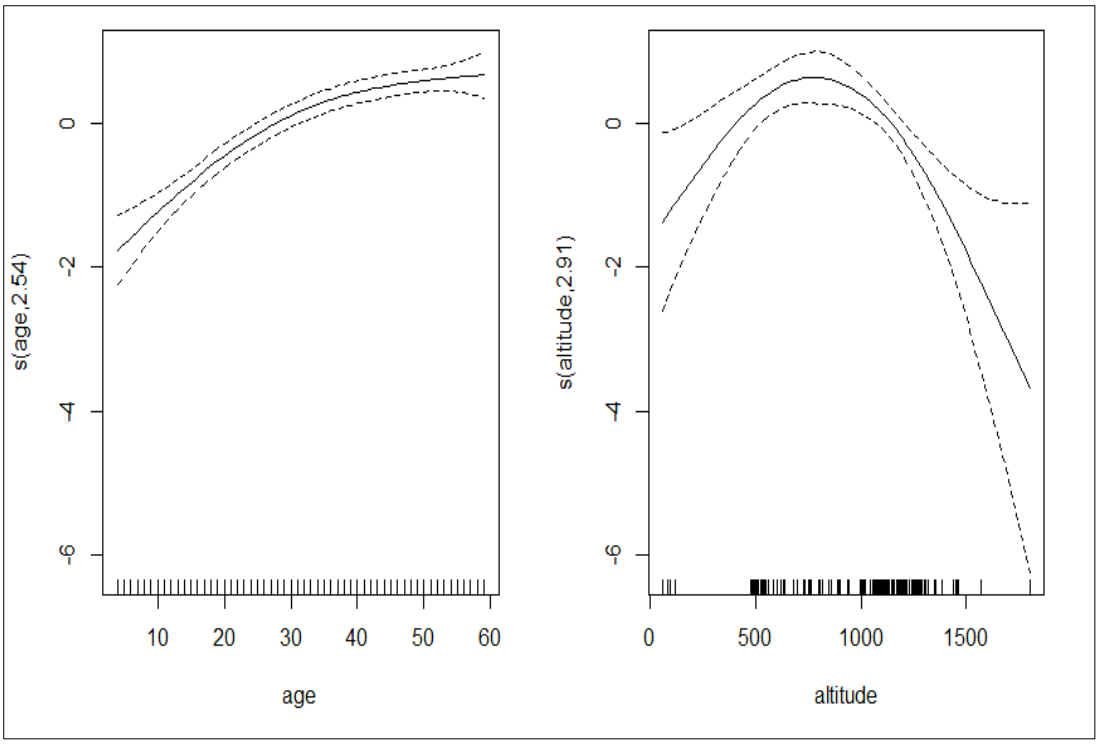

Figure 4.1: Smoothing components of malaria RDT test with age and altitude.

Table 4.2: Approximate significance of the smooth terms .

Source	Degree of freedom (df)	Chi squared	*p*-value
S (age)	2.423	90.420	<0.001
S (altitude)	2.875	22.340	<0.001

4.5. DISCUSSION AND SUMMARY

The present study utilised the generalised additive mixed model (GAMM) to investigate the risk factors associated with malaria using the 2017 Malawi Malaria Indicator Survey nationwide. The previous studies used the parametric model such as the generalised linear mixed model (GLMM) to analyse the malaria RDT data (Ayele et al., 2013; Gaston and Ramroop, 2020). The parametric models are useful to model the relationship between a response variable and covariance. However, non-parametric models are flexible to allow non-normal error distributions, modelling continuous predictor variables, and relax the assumption of normality and linearity in linear regression (Lin and Zhang, 1999; Gaston and Ramroop, 2020). The parametric and non-parametric models should complement each other and for this

reason, the combination of the two methods is more useful (Wu and Zhang, 2006). Thus, the study used the parametric model first to create geographical and socio-economic status such as type of place of residence, region, wealth quantile, mother's highest education level, type of toilet facility and availability of electricity.

The effect of age and altitude was modelled as non-parametric and was statistically significant. The interaction effect was not included in the model as it was not statistically significant to improve the original model. The results from the parametric part revealed that the probability of increasing a positive malaria RDT was lower in the richer and the middle classes compared to the poorer classes. These results confirmed that the prevalence of malaria is linked to socio-economic factors, where the poorer people are more vulnerable (Hay et al., 2004; Chitunhu and Musenge, 2016; Gaston and Ramroop, 2020). This is due to the limited access to health care and the affordability of treatment (Worral et al., 2003). The study revealed that households with access to electricity are less likely to increase the positive malaria RDT rates. Moreover, households with no toilet facilities are more likely to increase the positive malaria RDT rates. This shows that the households with access to electricity and toilet facilities are not poor. Hence, these factors are indicators of socio-economic status, and this is in line with the study by Ayele et al. (2014a) and Gaston and Ramroop (2020).

The results from the study also showed that the probability of increasing a positive malaria RDT in a mother's education reduces as the mother's children further their education. This means that the more the mother becomes educated and the more the children know, the more they become aware of malaria and its prevention. Furthermore, this might be linked to socio-economic status, as educated people are more likely to have a better life. The study was consistent with previous studies such as those by Zgambo et al. (2017) and Sultana et al. (2017); Gaston and Ramroop (2020).

The study revealed that households from rural areas are more likely to increase the positive malaria RDT rates than those from urban areas. This might be explained by not having access to so many things such as good houses, drinking water, access to health care, few educated people, and so forth (Jenkins et al., 2015; Kazembe and Mathanga, 2016; Sultana et al., 2017; Gaston and Ramroop, 2020).

The study indicated that the probability of reducing a positive malaria RDT in children with anaemia status is very much higher compared to non-anaemic children. This might be

explained by the link between anaemia and malaria, as has been shown by previous studies by Biemba et al. (2000); Sultana et al. (2017); Gaston and Ramroop (2020).

The study indicates a large variation among the three regions, where the households from the Central region are more likely to show an increase in the probability of having a positive malaria RDT. This is because the region is covered by large plains and the low-lying zone along the lake. Moreover, the lake might be an area conducive to the breeding of malaria vectors (Minakawa et al., 2012; Zgambo et al., 2017; Gaston and Ramroop, 2020).

The results from the non-parametric model indicate that the probability of a positive malaria RDT increases as the child's age increases. This could be the impact of maternal immunity in the child before one year of age. In addition, children younger than one-year-old are more protected and well taken care of and this helps to fight any kind of disease. This reduces as the children get older. These results are consistent with the studies by Ayele et al. (2014a); Chirombo et al. (2014); and Gaston and Ramroop (2020).

The study revealed that the risk of having a positive malaria RDT result increases as the altitude increases up to 750 metres and starts showing a decrease going higher. This may be explained by the very high temperatures at lower altitude as the mosquitos develop in hot areas. In addition, as the altitude increases, the temperatures reduce and this reduces the risk of having a positive malaria RDT result (Lindsay and Martens, 1998; Chirombo et al., 2014; Gaston and Ramroop, 2020).

The aim of this chapter was to assess the prevalence and factors associated with malaria in under five-year-old children in Malawi using GAMM. The findings from this chapter show that the government should consider other factors associated with malaria especially in children under five years of age such as anaemia, region, residence, toilet facilities, wealth index, use of electricity, mothers' education, children's age and altitude of the region of residence. The findings from this study revealed that malaria is still a major problem and is linked with socio-economics and geographical location. The government should focus on poorer communities from rural and high-altitude areas, especially in the Central region. In addition, children with anaemia should be the priority to get all the proper health care and support needed.

The key findings also show that there is a need to educate the population through workshops, mobile clinics and various social media platforms on how to prevent malaria in children under five years of age. Moreover, the education of mothers should be considered and supported so that they can take care of and protect their children, especially, in their first six months, when

they are more likely to be exposed to malaria vectors. The findings from this chapter will help the government and donors to control and eliminate malaria in children under five years old. The key priority should be in children with anaemia, with consideration of the factors of mother's education level, wealth index, child's age, altitude of the place of residence, region, place of residence, toilet facility and electricity facilities.

The GAMM fitted our data well but could not join either malaria and anaemia or anaemia and stunting simultaneously as there is an association between malaria, anaemia, and stunting (Ayele et al., 2014b; Adebayo et al., 2016; Gaston et al., 2021). Therefore, the next chapter will resolve this issue by employing a multivariate joint model to assess the correlation between malaria, anaemia, and stunting and their predictor's factors.

CHAPTER FIVE:

MULTIVARIATE JOINT MODELLING

5.1. INTRODUCTION

In this chapter, we used a multivariate joint model within the ambit of the generalised linear mixed model (GLMM) to assess the association between anaemia and malaria using the 2017 Malawi Malaria Indicator Survey (MMIS). The model was also applied to the 2014 Lesotho Demographic Health Survey to determine the link between anaemia and stunting (LDHS). In previous chapters, we reported on the use of a separated model for childhood anaemia or malaria. Although, the model has its advantages but it cannot take into account a potential link between malaria and anaemia concurrently. The joint model is required to effectively simulate anaemia and malaria or anaemia and stunting to explore the link between them along with determining related variables. When compared to single models, the multivariate joint model under a GLMM provides reference points, such as the efficient performance of type I errors in numerous tests. Apart from that, the multivariate joint model is more effective in terms of parameter estimation capabilities and the ability to answer some multivariate issues. Moreover, in order to show the association between two or more variables, the GLMM incorporates the random effect into the model (Gueorguieva, 2001; Hedeker, 2005; Agresti, 2015; Habyarimana et al., 2016; Gaston and Ramroop, 2020).

Thus, for the above-mentioned reason, in this chapter we aim to concurrently simulate the relationship between malaria and anaemia or anaemia and stunting. In addition, we assess the factors that might affect childhood malaria, anaemia, and stunting by utilising the joint model for a multivariate under generalised linear mixed model (GLMM).

5.2. MODEL OVERVIEW

Assume that you have bivariate response variables, which can be extended to two response variables. There are numerous techniques for joint modelling such as Plackett-Dale and the Probit normal generalised linear mixed model approaches among others. However, in this chapter we decided to use a generalised linear mixed model approach and account for both random effect and sequential correlations (McCullagh and Nelder, 1989; Neuhaus et al., 1991).

The study by Molenberghs and Verbeke (2005) indicated that a generalised linear mixed model can be expressed as follows:

$$y_j = \gamma_j + \epsilon_j, \tag{5.1}$$

with

$$\gamma_j = \gamma_j(\theta_j) = k(\alpha x_j + c_j z_j), \tag{5.2}$$

where $c_j \sim N(0, D)$ are the P-dimensional random effects. The component of inverse link functions k may differ depending on the type of the various explanatory variables in y_j. Moreover, the variance of ϵ_j is determined by the mean-variance links of the various dependent variable; it also includes a covariance matrix $\mathcal{R}_j(\beta)$ and the over-dispersion variable ψ. Whenever the equation (5.2) has no random variables, it decreases to a marginal model known as the marginal generalised linear model (MGLM). Although, when the matrix $\mathcal{R}_j(\beta)$ has no residual correlations, it reduces to a mainly random effects model or a conditional independence model, and both are generalised linear mixed models (Molenberghs and Verbeke, 2005).

The y_j variance-covariance matrix is derived from a general first-order estimated equation given by Molenberghs and Verbeke (2005) as follows:

$$V_j = Var(y_j \approx \pi_j z_j D z_j' \pi_j' + \varepsilon_j), \tag{5.3}$$

where $\pi_j = \left(\frac{\partial \gamma_j}{\partial \theta_j}\right) / c_j = 0$ and $\varepsilon_j = \psi_j^{1/2} B_j^{1/2} \mathcal{R}_j(\beta) B_j^{1/2} \psi_j^{1/2}$ with B_j considered as a diagonal matrix comprising a variance from a generalised linear classification of y_{jh} with $h = 1,2$, for a specified random effect, and $c_j = 0$. In addition, ψ_j is also considered as a diagonal matrix along with the overdispersion variables across diagonally.

The variance-covariance in residual error ϵ_j is expressed by ε_j and the first term in the right hand of equation (5.3) stands for the random effects structure of $k(x_j \alpha) + c_j z_j$. While $\mathcal{R}_j(\beta)$ is the correlation matrix (Neuhaus et al., 1991; Molenberghs and Verbeke, 2005). Additionally, when the dependent variable is normally distributed, the overdispersion variable is σ_j^2 and the variance coefficient is 1. Moreover, when the dependent variable is binary with logit link, we deduce the following expression:

$$\gamma_{jh}(c_j = 0)[1 - \gamma_{jh}(c_j = 0)], \tag{5.4}$$

with $c_j = 0$ estimated from a Taylor series extension of the mean value of $c_j = 0$.

If the exponential family requirements for all elements are observed with the canonical link, $\pi_j = B_j$, then the variance matrix of y_j can be expressed by

$$Var(y_j \approx \pi_j z_j D z_j' \pi_j' + \psi_j^{1/2} \pi_j^{1/2} \mathcal{R}_j(\beta) \pi_j^{1/2} \psi_j^{1/2}), \tag{5.5}$$

and because of conditional independence $\mathcal{R}_j$ fades away, and equation (5.5) becomes

$$Var(y_j) = z_j D z_j' \pi_j' + \psi_j^{1/2} \pi_j^{1/2} + \psi_j^{1/2} \pi_j^{1/2} \psi_j^{1/2} . \tag{5.6}$$

When the model has no random variables for the residual, generalised linear model (MGLM) can be expressed in the following equation

$$\begin{pmatrix} y_{j1} \\ y_{j2} \end{pmatrix} = \begin{pmatrix} \gamma_1 + \tau c_j + \beta x \\ \frac{exp(\gamma_2 + c_j + \alpha x_j)}{1 + exp(\gamma_2 + c_j + \alpha x_j)} \end{pmatrix} + \begin{pmatrix} \epsilon_{j1} \\ \epsilon_{j2} \end{pmatrix} . \tag{5.7}$$

Given this, the continuous and categorical responses are assessed on various levels, and τ is the scale parameter added in the continuous random intercept model (McCullagh and Nelder, 1989; Molenberghs and Verbeke, 2005). Suppose that $z_j = \begin{pmatrix} \tau \\ 1 \end{pmatrix}, \pi_j \begin{pmatrix} 1 & 0 \\ 0 & \omega_{j2} \end{pmatrix}, \psi = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 1 \end{pmatrix}$, where $\omega_{j2} = \gamma_{j2}(c_j = 0)[1 - \gamma_{12}(c_j = 0)]$. Furthermore, assuming that Σ is the correlation between ϵ_{j1} and ϵ_{j2}, and z_j on the other hand, it is not a design matrix since it contains unknown values. Thus, the variance-covariance matrix of y_j given in equation (5.1) becomes

$$V_j = \begin{pmatrix} \tau^2 & \omega_{j2}\tau \\ \omega_{j2}\tau & \omega_{j2}^2 \end{pmatrix} \eta^2 + \begin{pmatrix} \sigma^2 & \Sigma\sigma\sqrt{\omega_{j2}} \\ \Sigma\sigma\sqrt{\omega_{j2}} & \omega_{j2} \end{pmatrix} =$$
$$\begin{pmatrix} \tau^2\eta^2 + \sigma^2 & \omega_{j2}\tau^2\eta^2 + \Sigma\sigma\sqrt{\omega_{j2}} \\ \omega_{j2}\tau\eta^2 + \Sigma\sigma\sqrt{\omega_{j2}} & \omega_{j2}^2\eta^2 + \omega_{j2} \end{pmatrix}. \tag{5.8}$$

As a result, the derived estimated marginal correlation function is expressed as follows:

$$\Sigma(\alpha) = \frac{\omega_{j2}\tau\eta^2 + \Sigma\sigma\sqrt{\omega_{j2}}}{\sqrt{\tau^2\eta^2 + \sigma^2} \times \sqrt{\omega_{j2}^2\eta^2 + \omega_{j2}}} . \tag{5.9}$$

This equation depends on the fixed effects via ω_{j2}, and the model with no random effects becomes

$$\begin{pmatrix} y_{j1} \\ y_{j2} \end{pmatrix} = \begin{pmatrix} \gamma_2 + \alpha x_j \\ \frac{exp(\gamma_1 + c_j + \alpha x_j)}{1 + exp(\gamma_1 + \alpha x_j)} \end{pmatrix} + \begin{pmatrix} \epsilon_{j1} \\ \epsilon_{j2} \end{pmatrix}. \quad (5.10)$$

Based on the fully marginal classification, equation (5.8) reduces to Σ with an assumption of independence, and equation (5.9) reduces to the following function of fixed effects:

$$\Sigma(\alpha) = \frac{\omega_{j2}\tau\eta^2}{\sqrt{\tau^2\eta^2 + \sigma^2} \times \sqrt{\omega_{j2}^2\eta^2 + \omega_{j2}}}. \quad (5.11)$$

If both equations (5.9) and (5.11) are binary with a constant correlation Σ but without a random effect and no residual correlation, the equation (5.11) becomes

$$\Sigma(\alpha) = \frac{\omega_{j2}\omega_{j2}}{\sqrt{\omega_{j1}\eta^2 + \omega_{j1}} \times \sqrt{\omega_{j2}^2\eta^2 + \omega_{j2}}}. \quad (5.12)$$

The equation (5.12) can be solved using a random effects design matrices z_i and for more than two factors that are not simply continuous and binary (Neuhaus et al., 1991; Molenberghs and Verbeke, 2005).

In the case where the response variables are both binary, then a generalised linear mixed model (GLMM) with correlated random effects can be expressed as follows:

$$\begin{pmatrix} y_{j1} \\ y_{j2} \end{pmatrix} = \begin{pmatrix} \frac{exp(\beta_1 + \alpha_1 x_j + c_{j1})}{1 + exp(\beta_0 + \alpha_1 x_j + c_{j1})} \\ \frac{exp(\beta_2 + \alpha_2 x_j + c_{j2})}{1 + exp(\beta_0 + \alpha_2 x_j + c_{j2})} \end{pmatrix} + \begin{pmatrix} \epsilon_{j1h} \\ \epsilon_{j2h} \end{pmatrix}, \quad (5.13)$$

with the random effects c_{j1} and c_{j2} considered as normally distributed; while the terms ϵ_{j1h} and ϵ_{j12} are independent (Faes et al., 2008). The variance of the two independent variables is supposed to be $Var(c_{j1}) = \omega_{j1} = \gamma_{j1h}(c_{j1=}0)[1 - \gamma_{j1h}(c_{j1} = 0)]$, and $Var(c_{j2}) = \omega_{j2} = \gamma_{j2h}(c_{j2=}0)[1 - \gamma_{j2h}(c_{j2} = 0)]$. For subject j and h, the estimated variance-covariance matrix of two binary response variables is given by

$V_{j1} = \begin{pmatrix} \omega_{j1h}^2\eta_1^2 + \omega_{j1h}\Sigma\eta_2\omega_{j1h}\omega_{j2h} \\ \omega_{j1h}\Sigma\eta_2\omega_{j2h} + \omega_{j2h}^2\eta_2^2 + \omega_{j2h} \end{pmatrix} + \begin{pmatrix} \epsilon_{j1h} \\ \epsilon_{j2h} \end{pmatrix}$ (Faes et al., 2008; Habyarimana et al., 2016).

Moreover, the correlation between the two response variables can be expressed as

$$\Sigma y_1 y_2 = \frac{\Sigma\eta_1\eta_2\omega_{j1h}\omega_{j2h}}{\sqrt{\omega_{j1h}^2\eta_1^2 + \omega_{j1h}} \times \sqrt{\omega_{j2h}^2\eta_2^2 + \omega_{j2h}}}. \quad (5.14)$$

In case both response variables are continuous, and linear mixed models have a random effect, the correlation between the two response variables can be written as follows (Habanabakize, 2021):

$$\Sigma y_1 y_2 = \frac{\Sigma \eta_1 \eta_2}{\sqrt{\eta_1^2+\sigma_1^2} \times \sqrt{\eta_2^2+\sigma_2^2}}. \quad (5.15)$$

In general, the specification of a full model is not required, when assuming the first outcome to be continuous and the second one to be Bernoulli distributed. By using marginal correlation, we can still keep the joint moments measurement to the second, yet the conditional on the random effects, as well as normality criteria concerning the random effects are important (Molenberghs and Verbeke, 2005; Faes et al., 2008; Habyarimana et al., 2016).

5.3. MAXIMUM LIKELIHOOD ESTIMATION

In GLMM, the maximum likelihood is obtained by integrating the random variables and getting the following equation

In GLMM, the maximum likelihood is obtained by integrating the random variables and getting the following equation

$$\prod_{j=1}^{m} \iint \left[\prod_{h=1}^{m_{j1}} g_1(y_{j1}/c_{1j}; \alpha_1, \psi_1) \prod_{h=1}^{m_{j2}} g_2(y_{j2}/c_{2j}; \alpha_2, \psi_2)\right] g(c_{1j}, c_{2j}; \varpi) dc_{1j} dc_{2j}. \quad (5.16)$$

Most of the time, the integral (5.16) is unsolvable; hence an analytical, stochastic, or statistical approach must be applied (Breslow and Clayton, 1993; Gueorguieva, 2001).

There are several methods to fit GLMM such as marginal maximisation using Gaussian quadrature or Monte Carlo approximation, penalised quasi-likelihood, Monte Carlo EM algorithm, Monte Carlo Newton-Raphson algorithm and simulated maximum likelihood (Breslow and Clayton, 1993; Fahrmeir et al., 1994; Wolfinger et al., 1994; McCulloch, 1997; Booth and Hobert, 1999). In multivariate, all these methods can be used, except in the maximum likelihood when the number of response variables is more than three and becomes difficult for estimation (Molenberghs and Verbeke, 2005; Habyarimana et al., 2016). However, this problem can be resolved by using the extension to higher-dimension order.

5.4. EXTENSION TO HIGHER-DIMENSIONAL DATA

Suppose that n is the number of response variables to be modelled jointly, and is expressed as $y_{jh} = y_{jh1}, y_{jh2}, \dots, y_{jhn_j}$, with $h = 1, 2, \dots, n$. The y_{jh} is a vector of n_{jh} observations performed on subject j, for outcome h and y_{jh} is restricted to any response variables, can be either continuous or binary or mixed (Faes et al., 2008).

As a result, all m response variable outcomes can be modelled simultaneously by describing a joint distribution for the random effects, the same as binary outcomes, but using $p \times n$ dimensional random effects vector c_j (Breslow and Clayton, 1993; Habyarimana et al., 2016). In application, the conditionally on the random effects $c_{1j}, c_{2j}, \dots, c_{nj;}\, y_{1j}, y_{2j}, \dots, y_{nj}$ are assumed to be independent. Then, the model can be evaluated by considering the variable c_j of all random effects for item j to be multivariate normal having a mean of zero and covariance Φ as follows:

$$c_j = \begin{bmatrix} C_{1j} \\ c_{2j} \\ \vdots \\ c_{nj} \end{bmatrix} \sim i.i.d.\, MVN(0, \phi) = MVN\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1n} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{n1} & \phi_{n2} & \cdots & \phi_{nn} \end{bmatrix} \right). \tag{5.17}$$

The matrices ϕ_{kl} illustrate covariance between c_{kj} and c_{lj}, with $and\ l = 1,2, \dots, n$. The matrix ϕ is represented by ϕ_{kl} as the input blocks, and the calculation with conclusion relied on the vector y_j as the marginal model of all individual j observations (Faes et al., 2008). Given that the response variables are expected to be independent with the vector c_j of random variables, the likelihood effect on subject j can be expressed as follows:

$$L_j(\Omega / y_{j1}, y_{j2}, \dots, y_{jn}) = \int_{\mathcal{R}^{np}}^{n} \prod_{h=1}^{m_j} g_{jh}\left(y_{j1h}, y_{j2h}, \dots, y_{jnh} / c_j; \Omega\right) g\left(c_j / \phi\right) dc_j \tag{5.18}$$

where $\Omega = (\alpha, \beta, \phi)$. When the response variables are of a different type, and the n increases to the $n \times p$-dimensional integral, the calculation of likelihood becomes difficult. Thus, the pseudo-likelihood method is introduced to solve this problem, and works like the pairwise modelling method introduced by Fieuws and Verbeke (2006). Hence, the pseudo-likelihood function substitutes the full likelihood relating to j observations, and produces the following equation:

$$PL_j = \prod_{r=1}^{n-1} \prod_{q=r+1}^{n} L_{jrq}\left(\frac{\Omega}{y_{jr}}, y_{jq}\right)$$

$$= \prod_{h=1}^{n-1} \prod_{l=h+1}^{n} \int_{\mathcal{R}^{2p}}^{n} \prod_{h=1}^{m_j} g_{jh}\left(y_{jrh}, y_{jqh}/c_j^{rq}; \Omega\right) g\left(c_j^{rq}/\phi\right) dc_j^{rq}\,, \tag{5.19}$$

in which the contribution L_{jrq} is similar to the multivariate likelihood function for outcomes r and q. Then, the $n \times p$-dimensional integration issue decreases to $2 \times p$-dimensional integrations. To achieve this the data must be divided into different groups of outcomes and assume that conditional m random effects of all associations of r, q groups and subject j are independent (Molenberghs and Verbeke, 2005; Fieuws and Verbeke, 2006). The assumption of Ω focuses on the pseudo-likelihood concept and subject to a sandwich-type robust variance predictor (Arnold and Strauss, 1991; Habyarimana et al., 2016). The equation of asymptotic multivariate normal distribution for $\widehat{\Omega}$ can be written as

$$\sqrt{M}\left(\widehat{\Omega} - \Omega\right) \sim MVN[0, H(\Omega)^{-1} R(\Omega) H(\Omega)^{-1}], \tag{5.20}$$

with $H = H(\Omega)$ is the matrix defined as: $-\sum_{r=1}^{n-1} \sum_{q=r+1}^{n} E\left(\frac{\partial ln L_{jrq}(\Omega/y_{jr}, y_{jq})}{\partial \theta_s \partial \theta_t}\right)$ and $R = R(\Omega)$ is a symmetric matrix given by $-\sum_{r=1}^{n-1} \sum_{q=r+1}^{n} E\left(\frac{\partial ln L_{jrq}(\Omega/y_{jr}, y_{jq}) \partial ln L_{jrq}(\Omega/y_{jr}, y_{jq})}{\partial \theta_s \qquad \partial \theta_t}\right)$.

The pseudo-likelihood ratio test is good when a variable has a huge impact on the model compared to Wald tests (Geys et al., 1997; Habyarimana et al., 2016). If we wish we can conduct a test and assume that the null hypothesis $H_0: \lambda = \lambda_0$, and λ is an k-dimensional sub-vector of the t-dimensional vector of regression coefficient α, where α is expressed as (λ^T, ν^T).

Thus, the pseudo-likelihood ratio test is provided as follows

$$F^{*2} = \frac{2}{\tau}\{PL(\hat{\alpha}_N) - PL[\lambda_0, \hat{\nu}(\lambda_0)]\}, \tag{5.21}$$

where F^{*2} is a Chi (χ_k^2) distribution, and $\hat{\alpha}_N$ is a pseudo-likelihood parameter estimate of α, while $\hat{\nu}(\lambda_0)$ is the maximum pseudo-likelihood estimator in the subspace when $\lambda = \lambda_0$.

Furthermore, $\hat{\tau}$ represents the mean of eigenvalue of $\left(H^{\lambda\lambda}\right)^{-1} \phi_{\lambda\lambda}$ and $H^{\lambda\lambda}$ is the $k \times k$ submatrix inverse of H, while $\phi_{\lambda\lambda}$ describes the submatrix of $\phi = H^{-1} R H^{-1}$ (Geys et al., 1997; Molenberghs and Verbeke, 2005; Habyarimana et al., 2016).

5.5. DATA ANALYSIS

A number of methods may be employed to calculate the variables in joint models, such as numerical approximation, Gaussian quadrature, adaptive Gaussian quadrature, or Laplace approximation, or by approximating the data using the pseudo-likelihood method. The pseudo approach in a dataset is generated using mean linearisation and is used when evaluating variables in residual models and random effects with or without correlation. The quadrature or Laplace approximations however can only calculate the variables of conditionally independent random effects models (Molenberghs and Verbeke, 2005; Gaston et al., 2021). To determine whether the two response variables had identical distribution and link functions, the SAS 9.4 PROC GLIMMIX technique was used. The PROC GLIMMIX procedure can also use different link functions for the two response variables. However, the NLMIXED procedure can also be applied, when estimating the parameter using Laplace approximation or Gaussian quadrature (Habyarimana et al., 2016; Gaston et al., 2021; Gaston et al., 2022).

Furthermore, based on the lowest value of the Akaike information criteria (AIC), numerous covariance structures were considered, and unstructured (UN) were deemed to be appropriate for our assessment. Besides that, we examined the potential interactions and found none to be statistically significant (Gaston et al., 2021; Gaston et al., 2022).

5.6. MODEL FORMULATION FOR TWO OUTCOMES

In this chapter, malaria and anaemia status of a child were the two response variables of interest. Assume that y_{j1} represents the RDT status for malaria, where a value of one (1) indicates a positive result and a value of zero (0) indicates a negative result. The second response variable y_{j2}, represents anaemia status, with one (1) being an anaemic child and zero (0) representing a non-anaemic child.

The distinguished outcomes are derived using a bivariate Bernoulli distribution, with p_{j1} representing the risk of malaria appearing in a child j and p_{j2} representing the probability of anaemia occurring in a child j. Thus, the binary generalised linear model may be expressed as:

$$y_1(\mu_{j1}) = X_{j1}\alpha_1 + Z_{j1}c_1 \tag{5.22}$$

$$y(\mu_{j1}) = X_{j2}\alpha_2 + Z_{j2}\,c_2 \tag{5.23}$$

where α_1 and α_2 are considered as the variables of fixed effects, c_1 and c_2 are the variables of the random effects, while X_{j1}, X_{j2}, Z_{j1} and Z_{i2} are the intended matrices for fixed and random effects, respectively. As a result, the following equation for the variance-covariance matrices model is written as:

$$c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \sim i.i.d.MVN\ (0, \phi\ ) = MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}\right), \quad (5.24)$$

where the ϕ_{11} and ϕ_{22} are the variance determinants of childhood malaria and anaemia respectively, while ϕ_{12} and ϕ_{21} are identical correlation determinants between malaria and anaemia. In case the correlation coefficients from equation (5.24), ϕ_{12}= ϕ_{21}=0, the multivariate joint under generalised linear mixed model becomes a single model (Molenberghs and Verbeke, 2005; Habyarimana et al., 2016; Gaston et al., 2021).

5.6. RESULTS AND INTERPRETATION FOR JOINING MODEL TO ANAEMIA AND MALARIA

The results shown in **Table 5.1** indicate the parameter estimates, standard error (SE), odds ratio (OR), and *p*-values. The study reported only the exploratory variables with statistically significant impact on malaria and anaemia (p <0.05). The variables with a significant effect on both malaria and anaemia were the child's age, mother's education level, availability of electricity, toilet facilities, and children under five who slept under a mosquito bed net the night before the survey. The residence of the household and altitude of residence had only a statistically significant effect on malaria.

The results in **Table 5.1** indicate that children aged 6-23 months were 0.367 (OR: 0.367, 95% CI: 0.274; 0.490) times less likely to test positive for malaria using an RDT when compared with those in the reference group (42-59 months). In contrast, the children aged 6-23 months were 4.289 (OR: 4.289, 95% CI: 3.418; 5.382) times more likely to have anaemia compared with those in the age group 42-59 months.

The same results showed that children from mothers with post-primary levels were 0.505 (OR: 0.505, 95% CI: 0.305; 0.835) times less likely to have malaria compared to those from the mother with no education. The children from mothers with primary school education were 0.710 (OR: 0.710, 95% CI: 0.506; 0.997) times less likely to have anaemia compared with those in the reference category group. The results revealed that the odds of testing positive for

malaria in an RDT for children from households with no access to electricity were 2.296 (OR: 2.296, 95% CI: 1.415; 3.745) times more likely than those from households with access to electricity. The same results indicated that children from households with no access to electricity were 1.279 (OR: 1.279, 95% CI: 1.005; 1.732) times more likely to have anaemia compared to those from households with access to electricity.

The study also revealed that children from a household with pit latrines were 0.625 (OR: 0.625, 95% CI: 0.401; 0.975) times less likely to test positive for malaria than those with no toilet facilities, while those from households with flush toilets were 0.470 (OR: 0.470, 95% CI: 0.271; 0.815) times less likely to have malaria compared to those with no toilet facilities. Furthermore, children from households with flush toilets were 0.580 (OR: 0.580, 95% CI: 0.369; 0.913) times less likely to have anaemia, compared to those with no toilet facilities.

The results indicated that the odds of testing positive for malaria in an RDT was 1.586 (OR: 1.586, 95% CI: 1.045;2.406) times more likely in households where children did not sleep under a mosquito bed net the night before the survey, compared to households where all children slept under a mosquito bed net. However, the odds of anaemia in some children who slept under a mosquito bed net the night before the survey, were 1.439 (OR: 1.439, 95% CI: 1.064; 1.946) times more likely than all children who slept under a mosquito bed net.

Table 5.1 shows that children living in rural areas were 3.611 (OR: 3.611, 95% CI: 2.111; 6.178) times more likely to test positive for malaria compared to those living in urban areas. The same results indicate that children who live in a residence at an altitude of more than 1000 metres were 0.421 (OR: 0.421, 95% CI: 0.244; 0.725) times less likely to have malaria than those living in a residence at an altitude between 501 and 1000 metres. However, the children living in residences in rural and at-altitude areas were not statistically significant to anaemia; hence, we did not interpret the results.

Table 5.1: Parameter estimates for a joint marginal model for malaria and anaemia in Malawi.

	Malaria				Anemia			
Indicator variables	**Estimate**	**OR**	**95% CI**	***P*-value**	**Estimate**	**OR**	**95% CI**	***P*-value**
Child's age	-	-	-	-	-	-	-	-
Ref: 42-59								
6-23 months	-1.003	0.367	[0.274;0.490]	0.000	1.456	4.289	[3.418;5.382]	<0.001
24-41 months	-0.069	0.933	[0.728;1.197]	0.589	-0.008	0.992	[0.815;1.207]	0.939
Residence	-	-	-	-	-	-	-	-
Ref: Urban								
Rural	1.284	3.611	[2.111;6.178]	<0.00	-0.149	0.856	[0.649;1.145]	0.303
Altitude	-	-	-	-	-	-	-	-
(Ref:501-1000 m)								
>1000 metres	-0.866	0.421	[0.244;0.725]	0.002	-0.090	0.914	[0.698;1.197]	0.512
0-500 metres	-0.497	0.608	[0.295;1.256]	0.179	0.280	1.323	[0.919;1.906]	0.131
Education	-	-	-	-	-	-	-	-
Ref:No education								
Primary	-0.062	0.940	[0.649;1.361]	0.743	-0.342	0.710	[0.506;0.997]	0.048
Post-primary	-0.684	0.505	[0.305;0.835]	0.008	-0.325	0.723	[0.486;1.075]	0.109
Electricity	-	-	-	-	-	-	-	-
Ref: Yes								
No	0.831	2.296	[1.415;3.745]	0.001	0.277	1.279	[1.005;1.732]	0.046
Toilet facilities	-	-	-	-	-	-	-	-
Ref: No facility								
Pit latrine	-0.470	0.625	[0.401;0.975]	0.038	-0.222	0.801	[0.537;5.930]	0.277
Flush Toilet	-0.755	0.470	[0.271;0.815]	0.007	-0.544	0.580	[0.369;0.913]	0.019
Child sleeping under net	-	-	-	-	-	-	-	-
Ref: All children								
Some children	0.024	1.024	[0.691;1.516]	0.906	0.364	1.439	[1.064;1.946]	0.018
No children	0.461	1.586	[1.045;2.406]	0.030	0.055	1.057	[0.763;1.462]	0.742

***OR:** odds ratio; **95% CI** confidence interval*

Table 5.2 below shows the variance components and covariance between anaemia and malaria. The coefficient estimate of 0.700 indicates the relationship between anaemia and malaria. This means that there is a statistically significant positive correlation between malaria and anaemia.

The overall fitted model was highly significant as the coefficient of covariance parameter indicated the p-value<0.001. Hence, including the random cluster effect in the model was necessary (Molenberghs and Verbeke, 2005; Zhang and Lin, 2008; Gaston et al., 2021).

The pseudo-likelihood test rejected the null hypothesis of zero correlation with Pearson chi-squared test =4591.22 and the ratio of Pearson and degree of freedom of 0.91 with p-value<0.001. This revealed that there is a good variability in the dataset and there was no residual over dispersion (Molenberghs and Verbeke, 2005; Gaston et al., 2021). Furthermore,

this indicated that the association between the prevalence of malaria and anaemia was highly significant. In addition, the odds ratio (2.014) also confirmed that anaemia and malaria are highly associated. The results from the present study are in line with the study by Seyoum (2018); and Gaston et al. (2021).

Table 5.2: Variance components and covariance between anaemia and malaria in Malawi.

Variables	Estimates	Odds Ratio (OR)	95% CI (Lower, Upper)	*P*-value
Variance (Malaria)	1.014	2.757	[0.610;1.418]	<0.001
Variance (Anaemia)	0.118	1.125	[0.020;0.216]	0.009
Correlation between Anaemia and Malaria	0.700	2.014	[0.333;1.067]	<0.001

5.7. DISCUSSION AND SUMMARY FOR JOINING MODEL TO ANAEMIA AND MALARIA

In this chapter, we used the joint model for a multivariate generalised linear mixed model (GLMM) to simultaneously model the association between malaria and anaemia and identify factors associated with malaria and anaemia. The study indicated that anaemia and malaria are highly associated. This means that malaria and anaemia move in the same direction, where any increase in malaria in children, will also result in an increase in anaemia. This finding is consistent with existing literature (Noland et al., 2012; Zgambo et al., 2017; Gaston et al., 2021). The same change can be in an opposite direction; where the number of children with malaria reduces, so does anaemia. Therefore, the change between both malaria and anaemia can be interpreted to mean that controlling malaria can result in effectively reducing anaemia (Reithinger et al., 2013; Hershey et al., 2017; Yimgang et al., 2021; Gaston et al., 2021). In addition, controlling anaemia in the area with a high prevalence of malaria can result in a reduction in deaths related to malaria (Korenromp et al., 2004; Seyoum, 2018; Gaston et al., 2021).

The findings from this study revealed that children from mothers with primary and post-primary education levels were less likely to have both malaria and anaemia compared to those from mothers with no education. This shows that the risk of having malaria or anaemia reduces as the education levels of their mothers increases. This might be connected to socio-economic position, since educated individuals live better lives than uneducated ones. Additionally, educated individuals have a better understanding of health-related concerns and can readily

buy malaria prevention tools. The findings from this study are aligned with the studies by Adebayo et al. (2016); Seyoum (2018); and Gaston and Ramroop (2021).

The results indicated that children from households with no access to electricity are at higher risk of having malaria and anaemia. Furthermore, households with no toilet facilities are more likely to see increased rates of positive malaria RDT and anaemia. Access to electricity and toilet facilities are indicators of socio-economic factors. Therefore, households with good access to facilities such as the above are more likely to be able to access healthcare, eat healthy food and can easily afford medical treatment (Ayele et al., 2014b; Gaston and Ramroop, 2021).

The findings from the present study also indicated that children who did not sleep under a mosquito bed net the night before the survey were at greater risk of having malaria and anaemia. This might be due to the fact that children who sleep under mosquito bed nets are more protected from being bitten by *Anopheles* mosquitoes which is the cause of malaria (Gaston and Ramroop, 2020; Gaston et al., 2021). The same results were found in previous studies, such as those by Ayele et al. (2014b); Zgambo et al. (2017); and Gaston et al. (2021).

The research shows that children living in high-altitude residences are less likely to have malaria. This is due to the markedly higher temperatures in lower altitude residence areas, which favour the growth of mosquitoes (Chirombo et al., 2014; Teh et al., 2018; Gaston and Ramroop, 2020; Gaston et al., 2021). The results also revealed that children from rural areas are more likely to test positive for malaria in an RDT. The findings from this study are in line with the studies by Adebayo et al. (2016); Gaston and Ramroop (2020).

The findings from this study revealed that the probability of being positive for malaria increased as the child's age increased. The children aged 6-23 months were less likely to test positive for malaria. These results are in contrast with the results found in the study by Seyoum (2018). However, the results are in line with the findings in studies by Zgambo et al. (2017); Gaston and Ramroop (2020); Gaston et al. (2021); Yimgang et al. (2021); and Gaston et al. (2021).

In contrast, the probability of having anaemia reduced as the child's age increased. Children aged 6 -23 months were more likely to have anaemia. This might be because anaemia is also associated with other factors such as nutritional deficiencies, disease infections such as HIV, intestinal worms, intake of iron, folate, vitamin B12 deficiency, and other parasitic infections (Ayoya et al., 2013; Gaston et al., 2018; Gaston et al., 2021). Furthermore, the immune system of young children is not strong enough to fight against different diseases but becomes stronger

as they grow older. Thus, for this self-same reason young children are at high risk of getting anaemia, and this reduces as they grow older. The results from this study are consistent with the studies by Hershey et al. (2017); Kuziga et al. (2017); Gaston et al. (2018) ; Gaston et al. (2021); and Yimgang et al. (2021).

The main objective of this cross-sectional study was to examine the association between malaria and anaemia using a multivariate joint model under the GLMM in children 6-59 months of age in Malawi. The current scientific setting also checked other factors which might be associated with both malaria and anaemia. Finally, we examined all possible interaction effects between the exploratory variables, and these were not included in the results since none were statistically significant.

The findings from this study indicate that there is an association between anaemia and malaria and any change in one disease has a similar effect on the other disease. This means that as malaria increases so does anaemia and vice versa. Therefore, the study suggests that any policy change to malaria will impact on anaemia. Furthermore, malaria and anaemia are associated with socio-economic, demographic, and geographical factors, which makes malaria and anaemia a persistent and a current problem.

Based on the findings from this study, there is a need for educating the population, particularly those from rural areas, on how to prevent malaria and anaemia in children under five years of age. The policy makers and Malawian government should focus on improving toilet facilities, access to electricity, and providing more mosquito bed nets, mostly for the individuals who live in rural areas and at low altitudes. In addition, the education of the mothers should be prioritised so that they can treat and take care of their children, especially those in the age group 6-23 months, as they are more vulnerable. Understanding the relationship between anaemia and malaria together with other factors associated with malaria and anaemia can provide useful insights to the government and policymakers in planning, controlling, and eliminating both malaria and anaemia. In addition, the statistical model used in this study will help other researchers to compare findings and references.

5.8. RESULTS AND INTERPRETATION FOR JOINING MODEL TO ANAEMIA AND STUNTING

Table 5.3 below presents the estimation results for the fixed effects of the joint GLMM. Several socio-economic and demographics factors showed significant relationships with both stunting and anaemia. A child's age had a significant effect on both anaemia and stunting. Children aged less than 20 months were less likely to be stunted (OR: 0.44, 95% CI: 0.298; 0.638) but did not differ in their risk of anaemia compared to the reference group (40-59 months). However, children aged 20-39 months did not differ in their risk of stunting but were more likely to be anaemic (OR: 1.7, 95% CI: 1.207; 2.396) compared to the reference group. Fever and diarrhoea were not linked to stunting but were both associated with a higher risk of anaemia (OR: 0.491, 95% CI: 0.341; 0.707 and OR: 0.609, 95% CI: 0.410; 0.905, respectively). High birth weight (≥2500g) and living in a rural area were protective against stunting (OR: 0.24, 95% CI: 0.182; 0.452 and OR: 0.52, 95% CI: 0.333; 0.814, respectively) compared to lower birth weight children; while the odds of being stunted increased with higher levels of poverty (OR: 3.5, 95% CI: 2.149; 5.703) compared to those from wealthier homes. Children from the middle-wealth tertile were 2.9 times more likely to be stunted compared to those from the top-wealth tertile. Lastly, lower maternal education was also associated with stunting.

Table 5.3: Parameter estimates for a joint marginal model for stunting and anaemia in Lesotho.

	Stunting				**Anaemia**			
	Estimate	**OR**	**95% CI**	***P*-value**	**Estimate**	**OR**	**95% CI**	***P*-value**
Child's age								
Ref: >39 months	-	-	-	-	-	-	-	-
20-39 months	0.226	1.254	[0.870;1.806]	0.225	0.531	1.701	[1.207;2.396]	0.003
< 20 months	0.829	0.436	[0.298;0.638]	<0.001	0.096	0.908	[0.655;1.260]	0.565
Child had fever								
Ref:Yes	-	-	-	-	-	-	-	-
No	0.153	1.165	[0.762;1.782]	0.482	-0.712	0.491	[0.341;0.707]	<0.001
Child had Diarrhoea								
Ref:Yes	-	-	-	-	-	-	-	-
No	-0.227	0.797	[0.507;1.254]	0.326	-0.496	0.609	[0.410;0.905]	0.014
Child's birth weight								
< 2500g	-	-	-	-	-	-	-	-
≥ 2500g	-1.248	0.240	[0.182;0.452]	<0.001	0.038	1.039	[0.662;1.631]	0.868
Residence								
Ref: Urban	-	-	-	-	-	-	-	-
Rural	-0.653	0.520	[0.333;0.814]	0.004	0.102	1.107	[0.753;1.629]	0.603
Wealth Index								
Ref: Richer	-	-	-	-	-	-	-	-
Middle	1.065	2.901	[1.788;4.707]	<0.001	0.224	1.251	[0.847;1.848]	0.262
Poorer	1.253	3.501	[2.149;5.703]	<0.001	0.282	1.326	[0.896;1.962]	0.158

Education level								
Ref:No education	-	-	-	-	-	-	-	-
Primary	-1.473	0.229	[0.049;1.064]	0.060	0.772	0.462	[0.094;2.270]	0.342
Post Primary	-1.841	0.159	[0.121;2.933]	0.020	-0.519	0.595	[0.121;2.933]	0.524

***OR**= odd ratio; **CI**=confidence interval*

The variance components and covariance between anaemia and stunting are presented in **Table 5.4** below. The covariance coefficient estimate of 1.000 indicated a positive relationship between anaemia and stunting, meaning that changes in either nutrition or anaemia in a child impacts the likelihood of both anaemia and stunting. In addition, the odds ratio of 2.718 confirmed that anaemia and stunting are highly associated. The overall fitted model was highly significant as the coefficient of covariance parameter indicated the *p*-value <0.001. Hence, including the random effect in the model was shown to be very important (Molenberghs and Verbeke, 2005; Zhang and Lin, 2008; Gaston et al., 2021; Gaston et al., 2022).

The test covariance parameters based on pseudo-likelihood rejected the null hypothesis of zero correlation with Pearson chi-squared test =2644.470 and *p*-value <0.001. This revealed that the association between anaemia and stunting was significant and not zero (Tuerlinckx et al., 2006). Furthermore, the results from the fitted statistics for conditional distribution indicated the Pearson chi-squared =2123.070 with 0.94 degrees of freedom. This is an indication of a good variability in the dataset and residual over-dispersion was not present (Molenberghs and Verbeke, 2005; Gaston et al., 2021; Gaston et al., 2022).

Table 5.4: Variance components and covariance between anaemia and stunting in Lesotho.

Variables	Estimate; SE	OR	95% CI	*P*-value
Variance (stunting)	0.104; 0.010	1.110	1.088; 1.132	0.149
Variance (anaemia)	0.314; 0.170	1.369	0.981; 1.910	0.033
Covariance between anaemia and stunting	1.000; 0.141	2.718	2.063; 3.582	0.001

5.9. DISCUSSION AND SUMMARY FOR JOINING MODEL TO ANAEMIA AND STUNTING

This cross-sectional study used secondary data from the 2014 LDHS. To our knowledge, this was the first study to simultaneously model the association between anaemia and stunting in children less than five years of age in Lesotho. The study utilised a multivariate joint model under the scope of GLMM to associate both anaemia and stunting and explore their associated socio-economic and demographic factors. Anaemia and stunting show a significant positive association confirming that anaemia and stunting should be considered interrelated health

problems in children where anaemia and stunting are more likely to coexist and inter-influence their manifestations. Thus, coordinated interventions aiming to improve both stunting and anaemia are likely to produce synergetic effects on child health. The association between anaemia and stunting can be interpreted as an indication of chronic malnutrition which might cause iron deficiency. Similar results were described in studies by Yang et al. (2012); Gari et al. (2017); Mohammed et al. (2019); Rahman et al. (2019); Rivadeneira et al. (2020); Gaston et al. (2022). Our findings also indicate that child age has a significant effect on both anaemia and malnutrition but impacts different age groups. The chance of having anaemia or stunting reduced as the children grew older.

This may be explained by the fact that the immune systems of children are still weak and need more nutrients to support rapid body growth. In addition, many children at an early age are not breastfed which makes them more susceptible to exposure to various illnesses. Some of these illnesses reduce the hemoglobin level in the blood which may lead to anaemia and stunting. Furthermore, when older children are introduced to foods and eat a greater variety of foods, this reduces the risk of them being anaemic or stunting. Similar results were found in previous studies (Anticona and Sebastian, 2014; Gari et al., 2017; Gaston et al., 2018; Kejo et al., 2018; Adhikari et al., 2019; Rahman et al., 2019 Gaston et al.,2022). However, the studies by Anticona and Sebastian (2014); and Oliveira et al. (2015) showed that stunting increased as the children grew older.

The findings from this study revealed that the risk of anaemia was related to having experienced recent fever and diarrhoea. This may be because fever and diarrhoea are commonly accompanied by a number of diseases and morbidities which are associated with anaemia. This has also been previously described (Habyarimana et al., 2016; Gaston et al., 2018; Rivadeneira et al., 2020; Gaston et al., 2022). The probability of being stunted is reduced with increasing levels of maternal education. This might be linked to socio-economic status, where educated individuals are more likely to have a better standard of living, and knowledge of balancing food. In addition, educated individuals can easily access and improve their nutritional status as most of them have a monthly income. This is also in line with findings from previous studies such as those by Kavosi et al. (2014); Aheto et al. (2015); Adebayo et al. (2016); Aheto et al. (2017); Adhikari et al. (2019); Gaston et al. (2022). Child birth weight significantly impacts stunting in children, with a lower risk in children born with a higher weight (≥2500g). This connection can be explained by the fact that children with low birth weight are more likely to have other co-morbidity illnesses that might be associated with stunting. Similar findings were

found in studies by Yang et al. (2012); Habyarimana et al. (2016); Kejo et al. (2018); Gaston et al. (2022).

We found that children living in rural areas have a lower risk of stunting, an effect that is debated in the field. This might be explained by the fact that some individuals in rural areas are educated and they eat fresh food and fruits with more nutrients. Also, individuals from rural areas may be breastfed for long periods, which can contribute to fighting stunting at an early age. Some studies have described similar results, such as the study by Kavosi et al. (2014); Gaston et al. (2022), while others have described contrasting results Yang et al. (2012); El Kishawi et al. (2015).

Lastly, children living within families from the middle and top tertile wealth index have a lower risk of malnutrition. This confirms that malnutrition is linked to socio-economic factors, where the children from the lower end of the wealth index cannot afford the proper food, maintain hygiene and access to health care services. Similar results were found in previous studies such as those by Gari et al. (2017); Mohammed et al. (2019); Rivadeneira et al. (2020); Gaston et al. (2022).

This study aimed to determine the association between anaemia and stunting in children less than five years of age in Lesotho using the multivariate joint model under GLMM. The study also assessed the association of socio-economic and demographic factors with anaemia and stunting. Lastly, we evaluated possible interaction effects between independent variables, but none passed the significance threshold. We found a significant positive association between anaemia and stunting which indicates that when malnutrition increases in children less than five years, anaemia also increases and vice-versa. Thus, a change in childhood stunting can have a significant impact on anaemia status. In addition, several socio-economic and demographic factors impact both malnutrition and anaemia such as family wealth, maternal education, urban vs. rural living environments. In addition, children that were of low birthweight or who have recently experienced fever, or diarrhoea should be prioritised for intervention.

Knowledge on the relationship between anaemia and malnutrition together with other determinants can provide useful insights to policymakers, donors, and government in planning and fighting to improve childhood anaemia and stunting through tailored public health messages and interventions.

This chapter revealed the association between anaemia and malaria, also anaemia and stunting. Anaemia, malaria, and stunting are correlated and are the most dominant health problem in children younger than five years. However, we could not fit anaemia, malaria, and stunting simultaneously. In addition, in this chapter, we could not address the complex interrelationships between explanatory factors, as well as their direct or indirect relationship with childhood malaria, anaemia, and stunting co-morbidity. Thus, for that reason in the next chapter, we employ a structural equation modelling to resolve the issues.

CHAPTER SIX:

STRUCTURAL EQUATION MODEL

6.1. INTRODUCTION

In this chapter, we introduce a structural equation modelling (SEM) to evaluate the complex interrelationships between socio-economics, demographics, and environmental factors, as well as their direct or indirect relationship with childhood malaria, anaemia, and stunting co-morbidity in Burundi. The previous chapters could not address these interrelationships among the variables of interest.

Structural equation modelling has become a general method in science for analysing and understanding multivariate relationships among the variables of interest. The analysis of covariance using structural equations, also known as latent variable analysis, is a new area of statistics. However, this method has been applied in econometrics and psychometrics for a long time (Gould and Golob, 1997; Golob, 2003; Lowry and Gaskin, 2014). The SEM usually is a multivariate model that links an attribute and unmeasured latent variables (Bollen, 1989, Fan et al., 2016; Caraka et al., 2021). The structural equation model can assess complex interrelationships between different variables and related unobserved and observed variables. The assessment can be done by calculating the sample covariance matrix of the observed variables and the population covariance matrix produced by the SEM framework (Bollen, 1989; Austin and Wolfle, 1991; Kaplan, 2008; Byrne, 2013; Fan et al., 2016). The SEM is very important in its extension for calculating measurement errors via latent variables. Structural equation modelling allows the evaluation of numerous sets of observed variables to define the non-measurable variable (latent or construct variables) and allows these latent variables to be related to each other (Bollen and Hoyle, 2012; Schumacker and Lomax, 2016). The SEM integrates several variables, which are not measured directly but along their effects or indicators. Furthermore, the SEM methodology involves multivariate data analysis tools that merge features of multiple regression and factor analysis to simultaneously calculate the series of interrelationships between variables and relations of dependency that permit the methodology of directly including measurement error in the model (Ditlevsen et al., 2005; Byrne, 2013; Ainur et al., 2017). The theoretical model of SEM can be demonstrated by applying mathematical equations and graphs (path diagrams), to summarise a set of hypotheses (Bollen and Hoyle, 2012; Caraka et al., 2021).

Moreover, SEM methodology can accommodate both observed and unobserved (latent or constructed) variables, which is one of the most important distinctions between structural equation modelling and other statistical modelling tools (Bentler, 1990; Byrne, 2013). In addition, the SEM can define the reciprocal effect, when the two variables are affecting one another through the feedback loop (Bollen, 1989; Byrne, 2013; Schumacker and Lomax, 2016).

6.2. MODEL FORMULATION

The application of structural equation modelling includes different steps such as the development of the theoretical model; conceptual model; specification of the mathematical model; determination of the model's evidence; and determining the model fit and evaluation of the goodness-of-fit the model (Byrne, 2013; Fan et al., 2016).

6.3. THEORETICAL CONCEPTUAL AND PATH MODELS

The latent (constructs) variables cannot be measured directly and are known as theoretical concepts. The latent variable is measured by observed (indicator) variables, and they assist the expansion and estimation of casual relationships in SEM (Chavance et al., 2010; Fan et al., 2016). The latent variable related to each other in the model must be indicated first, and the impact that these variables apply to each other is categorised as exogenous and endogenous (Kaplan, 2008; Byrne, 2013).

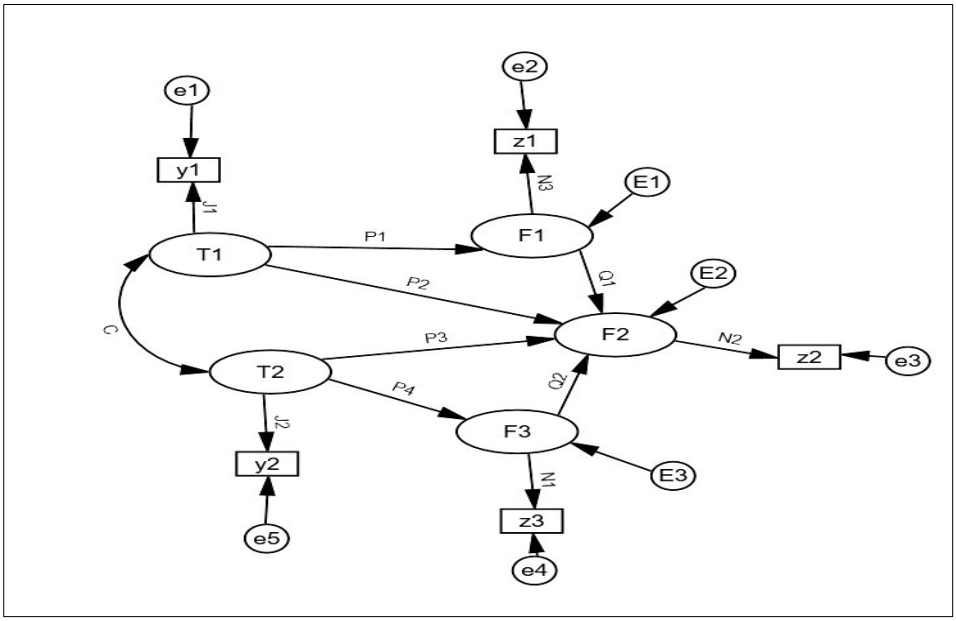

Figure 6.1: Graphical path of a structural model.

In the model F_i are the latent endogenous variables; T_i the latent exogenous variables; z_i are the observed endogenous variables; y_i the observed exogenous variables; e_i is the measurement errors, E_i are structural errors; c is the coefficient correlation between the latent exogenous variable, while P_i, Q_i, J_i, and N_i are the coefficients.

The exogenous (predictive) variables are not manipulated by the effect of other variables in the model and are measured without error. However, the endogenous (dependent) variables are influenced by the effect of other variables included in the model (Schumacker and Lomax, 2016).

Structural equation modelling offers the unique capacity to describe latent or unobserved variables in a linear model, in contrast to other statistical approaches where observed variables in a specific data set are employed for statistical analysis (Bollen, 1989; Fan et al., 2016; Grace et al., 2010; Ainur et al., 2017).

The SEM includes the measurement and structural models which can be written as follow:

$$\varphi = \alpha\varphi + \tau\delta + \epsilon, \tag{6.1}$$

where φ denotes a vector $m \times 1$ of latent endogenous variables, δ denotes a vector of $k \times 1$ latent endogenous variables, α represents an $m \times m$ matrix of coefficients connecting the latent endogenous variables to each order, τ denotes an $m \times k$ coefficients of the matrix that links the endogenous variables to the exogenous variables, and ϵ denotes an $m \times 1$ vector of the structural disturbances or errors. The main diagonal components of α are often zeros with $\delta \; and \; \epsilon$ viewed as mutually independent and normally distributed (Jöreskog and Sörbom, 1982; Kaplan, 2008; Schumacker and Lomax, 2016). The measurement model, which is theoretically described independently for the endogenous and the exogenous variables, links the observable and latent variables together and is expressed by:

$$Y = \pi_y\varphi + \mu \text{ , and } X = \pi_x\delta + \theta, \tag{6.2}$$

where $\pi_y(p\times m)$ and $\pi_x(q\times k)$ are coefficient matrices illustrating the relationship between latent endogenous and exogenous factors and the observable variables, respectively. Consequently, $\mu \; and \; \theta$ are $p\times 1$ and $q\times 1$ vectors of measurement errors in Y and X respectively.

To generate a scale for the related latent variables, each column of the π matrices normally has a value that is set to one. As an alternative, this can also be accomplished by setting the

variances of the latent exogenous variables in a matrix defined as γ, the matrix corresponding to the exogenous variable's covariance matrix, to zero. Furthermore, the measurement model's fit to the confirmatory factor analysis (CFA) application, which identifies the latent components must be addressed (Kaplan, 2008; Byrne, 2013; Fan et al., 2016).

The CFA examines the theoretical measurement model and determines whether the hypothesised measurement model produces a variance-covariance matrix that is identical to the sample variance-covariance matrix. Based on equation (6.2), the measurement errors μ and θ, each having a multivariate normal distribution, are considered to have zero expectations.

The errors presume independent of each other, independent of latent endogenous variables φ, latent exogenous variables δ, and independent of the disturbances ϵ. Additionally, it is assumed that the latent exogenous variables have a multivariate normal distribution and that the observations are independently sampled. However, for the endogenous variables that are accurately measured, this assumption is irrelevant (Kaplan, 2008; Byrne, 2013).

The structural errors ϵ, on the other hand, are unaffected by the latent exogenous variables δ and have zero expectation with a multivariate normal distribution. The observed indicators X and Y in this case exhibit a multivariate normal distribution and can be expressed as follows:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \tilde{} N_{p+q}(0, \rho), \tag{6.3}$$

where ρ, is the indicators' population covariance matrix, which is a function of the model's parameter $\phi = \alpha, \tau, \pi_x, \pi_y, \eta, \lambda_\delta, \pi\lambda_\epsilon, \beta, \xi$ and can be calculated as:

$$\rho = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix} = \begin{pmatrix} \pi_y(1-\alpha)^{-1}(\tau\gamma\tau')[(1-\alpha)^{-1}]'\pi'_y + \lambda_\mu & \pi_y(1-\alpha)^{-1}\tau\gamma\pi'_x \\ \pi_y\gamma\tau'[(1-\alpha)^{-1}]'\pi'_y & \pi_x\gamma\pi'_x + \lambda_\theta \end{pmatrix}, \tag{6.4}$$

where γ, indicates a $k \times k$ covariance matrix of the latent exogenous variables δ, τ denotes the $m \times m$ covariance matrix of the disturbance term, $\lambda_\mu$and λ_θ represent the covariance matrices of the measurement errors μ and θ.

6.4. MODEL ESTIMATION

To obtain the matrix ρ, related to the confirmatory factor analysis (CFA) of equation (6.4), it should be expected that $\alpha = 0$; $\tau = 0$; $\gamma = 0$; $\pi_y = 0$; and $\lambda_\mu = 0$ (Ditlevsen et al., 2005). In any given model, the restriction is essential in certain components of the matrix ρ, which

incorporates setting a few parameters to zero. With adequate limitation, the maximum likelihood estimates (MLE) might be achieved for the parameters of the model, and the log-likelihood related to the model can be portrayed as a function of the model's parameters of ρ and T, as the sampling covariance matrix between the observable variables. In the structural equation model, this method focuses on estimating the parameters ϕ, in order to minimise the inconsistency of function $F(T, \rho)$. The inconsistency function $F(T, \rho)$ is a scalar that estimates the distance between the examining covariance T and the adjusted covariance matrix $\hat{\rho}$ (Marsh and Hocevar, 1985; West et al., 2012; Ainur et al., 2017). In structural equation modelling, the MLE and generalised least squares (GLS) are the most commonly used estimation methods. A brief discussion and the required criteria for using MLE and GLS techniques can be found in the study by Bollen (1989); Raykov and Marcoulides (2012); Fan et al. (2016); Ainur et al. (2017). The MLE methods are characterised for parameters so that the two matrices T and $\hat{\rho}$ are pretty much as close as could really be expected, for the likelihood logarithm estimates the vicinity between the two matrices. The asymptotic standard errors are calculated from the square root of the matrix diagonals, in the parameter estimates. It is also, expected that the structural relationship between the latent endogenous variables φ and the latent exogenous variables δ are linear, as they are the interrelationships exuding between the indicator variables and the latent constructs.

6.5. MODEL IDENTIFIABILITY

A principal stage in model evaluation is to check the model's identifiability of latent variables, which is a complicated task in SEM without a straightforward answer (Golob, 2003; West et al., 2012). The model can be reported as nonidentifiable when the system of equations cannot be solved. The counting rules for identifiability is generally the number of free parameters in the model, and must not be more than the number of variance and covariance between the observable variables, and can be shown as

$$\frac{(k+l)(k+l+1)}{2}, \tag{6.5}$$

where, k represents the number of endogenous variables and l represents the number of exogenous variables in the model (Kline, 2015; Schumacker and Lomax, 2016).

However, the counting rule is not generally enough condition, as the condition can be quickly achieved while still resulting in a non-identifiable model. Hence, more conditions that allow a

model to be identified, such as when measurement errors are not correlated when at least two exclusive indicators exist for each of the latent variables when a single indicator for latent variables equally assumed without error is possible, and when the structural model includes only observed variables (Byrne, 2013; Fan et al., 2016). The SEM is not stable with a small sample size, and the minimum sample size depends on the involvement of the model, the degree of freedom, and the size effect (Byrne, 2013; Schumacker and Lomax; Fan et al., 2016).

6.6. MODEL DIAGNOSTICS

In the structural equation model, the verification of the model fit is based on various goodness-of-fit-model criteria and is created to assess the model under several assumptions. In addition, the verification of the goodness-of-fit is not a direct process as accessible in other multivariate techniques (Kaplan, 2008; Schumacker and Lomax, 2016). The chi-square (χ^2) test is widely regarded as the only statistical test of significance that is commonly used to evaluate the theoretical model in SEM. The insignificant results indicate a resemblance between the original sample variance-covariance matrix and the variance-covariance matrix estimated by the model. Although, the application of χ^2 is very difficult when the sample size is big, and rejecting the null hypothesis becomes hard (Kline, 2015; Ainur et al., 2017). When the chi-square has a zero value, means there is a good fit or no difference between the values in the sample covariance matrix and the model-suggested covariance matrix (ρ) generated based on a theoretical model. However, in structural equation modelling, it is recommended that various goodness-of-fit criteria be used in conjunction with overall fit measurements (Bentler, 1990; Kock and Lynn, 2012; Schumacker and Lomax, 2016). As a result, the measurement indices range from poor fit to perfect fit, and various structural equation modeling programmes report a range of the most common model fit as follows:

Goodness of fit index (GIF) =1-$\frac{\chi^2_{model}}{\chi^2_{null}}$, where 1 indicates perfect fit.

Root mean-square error of approximation (RMSEA) =$\sqrt{\frac{\chi^2_{model}-df_{model}}{\langle N-1\rangle df_{model}}}$, as a value less than 0.05 indicates the model's good fit.

Comparative fit index (CFI) =1-$\frac{\chi^2_{model}-df_{model}}{\chi^2_{null}-df_{null}}$, where the good fit the value greater than 0.9 is expected.

Tucker-Lewis index (TLI)=$\left[\left(\frac{\chi^2_{model}}{\chi^2_{null}}\right)-\left(\frac{\chi^2_{model}-df_{model}}{\chi^2_{null}-df_{null}}\right)-1\right]$,

Incremental fit index (IFI)=$\frac{\chi^2_{null}-\chi^2_{model}}{\chi^2_{null}-df_{model}}$, and

Normal fit index (NFI) = $\frac{\chi^2_{null}-\chi^2_{model}}{\chi^2_{null}}$.

These mentioned above goodness-of-fit criteria are based on differences in variance-covariance matrices between observed (original, T) and model-implied (replicate, ρ) (West et al., 2012; Kock and Lynn, 2012; Schumacker and Lomax, 2016; Ainur et al., 2017).

6.7. APPLICATION OF MODEL

Initially, we evaluated a theoretical model for individual variables (household, environmental, child demography, and geophysical factors) to guarantee that the theoretical relationships between the observable variables and their corresponding factors were upheld by the data.
The CFA was used to evaluate whether the measurement model and relationship between all the latent, and manifest variables are relevant. Numerous model fit indices were used in the analysis of this study, however, the common techniques for assessing model fit are χ^2, and should not be significant for a good model (Bagozzi and Foxall, 1996; Schumacker and Lomax, 2016). The model fit indices and their conditions incorporates the use of CFI, GFI, IFI, TLI and NFI which should be greater than 0.90 for good model (Kock and Lynn, 2012; West et al., 2012; Ainur et al., 2017).

The RMSEA $\leq$ 0.05 indicates a best-fit model, however, the values between 0.05 and 0.08 show a reasonable fit model (Bentler, 1990; Kock and Lynn, 2012; Kline, 2015). The validity of the structural model used a cross-validation method, which includes categorising the data into two different sample sizes (Byrne, 2013; Kline, 2015). The first sample was considered an adjustment sample, while the second one was the validation sample. We first tested the SEM on the adjustment sample and analysed the goodness-of-fit model, once the model reaches a good fit for the adjustment sample, the model can be assessed on the validation sample (Bentler, 1990; Byrne, 2013). The validity of the model is obtained, when the covariance structure of the model reaches the best fit in both the adjustment and validation sample. The maximum likelihood estimation techniques are used for full structural equation modelling in the calibration sample. In general, the less the value of χ^2, the better goodness of the fit to the data.

However, we cannot depend only on χ^2, as the test statistics are sensitive when the sample size is big, and tend to reject the model (Bentler, 1990; Carpentier et al., 2012; Schumacker and Lomax, 2016). Therefore, other indices which are not dependent on the sample size, are included. These indices and their cut-off indicating a good fit are NFI $\geq$ 0.95, CFI $\geq$ 0.95, AGFI $\geq$ 0.90, and RMSEA $\leq$0.07 (Hooper et al., 2008; Fan et al., 2016; Byrne, 2013; Ainur et al., 2017). In addition, to check the validity and reliability of the internal consistency between various items, Cronbach's alpha (coefficient alpha) method was used in this study. The Cronbach's alpha coefficient vary between 0 and 1, and the acceptable coefficient is $\geq$ 0.7 (Bell and Bryman, 2011; Bonett and Wright, 2015; Bujang et al., 2018).

In the theoretical model, we used the household factors (residence, wealth index, source of drinking water, type of toilet facility, the household share of toilet facility, mother's educational attainment, mother's access to information through television, household access to electricity, household's main roof, floor, and wall material) are directly or indirectly related to childhood co-morbidity of malaria, anaemia, and stunting. The geophysical factors including geographical regions of the children, travel times, and nightlight composites directly and indirectly influence childhood susceptibility to the co-morbidity of malaria, anaemia, and stunting through the mediating effects of household factors.

The environmental factors (rainfall, proximity to water, land surface temperature, enhanced vegetation index (EVI), aridity, wet days, and cluster) directly or indirectly influence childhood co-morbidity with respect to malaria, anaemia, and stunting. The child demographic factors (child's age, gender, and whether child slept under a mosquito net) directly and indirectly influence childhood co-morbidity to malaria, anaemia, and stunting.

Relying upon those assumptions, the conceptual framework can be defined by latent variables, which are deduced through observable variables because they cannot be directly measured, to assess their impact on childhood malaria, anaemia, and stunting co-morbidity. In a full SEM, the latent factors are then normalised on other factors based on the theory, empirical research, and suitable observed indicators (Schumacker and Lomax, 2016).

In order to assess the model fit and evaluate the variables of interest (child demography, household-level, environmental, and geophysical factors), we used the two approaches of equations (6.1) and (6.2) by conducting a confirmatory factor analysis (CFA) to test the hypothesised associations among these variables. We then created a conceptual path model diagram (Figure 6.1) that comprised the endogenous and exogenous variables to represent the

causal structure shown in the figure and define all of the conceptual relationships between these factors with respect to childhood co-morbidity. Thereafter, we evaluate the estimation model with all latent and observed factors included in the full model. The association between the direct and indirect variables, together with the childhood co-morbidity of malaria, anaemia, and stunting was assessed using structural techniques. The model to be accepted or rejected we perform a goodness of fit test using IBM SPSS AMOS 27 version software (Bentler and Wu, 2005; Carpentier et al., 2012).

6.8. RESULTS AND INTERPRETATIONS

Figure 6.2 below indicates the results of the full model, after deleting some items in the model for the best fit. The household constructs had an 11 factor higher order construct comprising residence, wealth index, source of drinking water, type of toilet facility, the household share of toilet facility, mother's educational attainment, and mother's access to information through television, household access to electricity, household's main roof, floor, and wall material. The Cronbach's α score was acceptable after deleting one variable from the model, and 10 items were left as indicated in **Figure 6.2**.

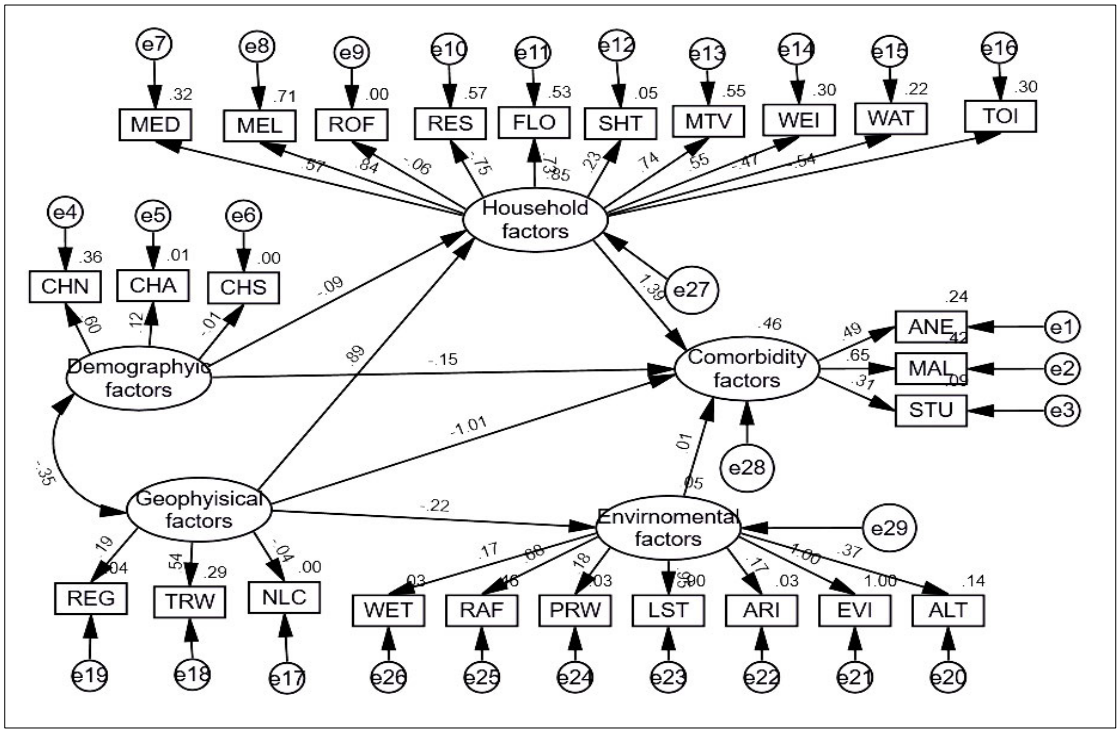

Figure 6.2: Graph for full structural model results.

The environmental factors include rainfall, proximity to water, minimum and maximum temperature, land surface temperature, enhanced vegetation index (EVI), aridity, wet days, and cluster. The Cronbach's α score was acceptable after deleting two variables from the model, leaving seven items. The geophysical factors resulted in a higher order of three factors in the model comprising geographical regions of the children, travel times, and nightlight composites and the Cronbach's α score was acceptable.

The environmental factors (rainfall, proximity to water, land surface temperature, enhanced vegetation index (EVI), aridity, wet days, and cluster) are directly or indirectly influencing childhood co-morbidity of malaria, anaemia, and stunting. The child demographic factors (child's age, gender, and whether the child slept under a mosquito net), also have an acceptable Cronbach's α score (Bentler and Wu, 2005; Hooper et al., 2008).

The non-significant Bollen-Stine p statistic, together with the underlying model statistics, demonstrated that the model was a good fit for the data in each testing model case, illustrating factor validity (Bentler and Wu, 2005; Kline, 2015). The Cronbach's α score in every case was above the recommended cut-off of 0.75, which is an indication of good scale reliability (Kline, 2015). Then we improved the CFA model in order to get a better model with the highest goodness-of-fit as indicated in **Table 6.1**. The results presented in **Table 6.1** also indicate that GFI=.969, CFI=.976, IFI=.926, TLI=.912, and RMSEA=.004 in a full model favoured the fit of the model (Kock and Lynn, 2012; Ainur et al., 2017).

The calibration sample model fit results showed that there is no point in changing the model; standardised regression weights for this model were all significant at 95% confidence intervals (CI).

Table 6.1: The goodness-of-fit indices in the two models (CFA and SEM).

Model	χ^2/df	GFI	CFI	IFI	TLI	NFI	RMSEA
Conceptual CFA model	732.025	0.651	0.645	0.640	0.545	0.648	0.046
Full structural model	435.020	0.968	0.976	0.926	0.912	0.913	0.04

The full path results of SEM are indicated in **Figure 6.2**, where both the direct and indirect effects of geophysical, and child demography factors are statistically significant on childhood co-morbidity (p<0.001).

The direct and indirect interrelationships between contextual factors and their impact on co-morbidity of childhood malaria, anaemia, and stunting are summarised in **Tables 6.2** and **6.3**.

In this chapter, the testing for the partial mediation effect assumed the estimation of two phases. The direct effect was used in phase one to estimate the effect of predictors (household, child demographic, environmental and geophysical factors) on childhood co-morbidity of malaria, anaemia, and stunting. The direct path coefficient from both geophysical and child demography factors on malaria, anaemia, and stunting co-morbidity was statistically significant and revealed a negative direct effect on childhood co-morbidity as indicated in **Table 6.2**. The direct path coefficient from both household and environmental factors on malaria, anaemia, and stunting co-morbidity was statistically significant and revealed a positive direct effect on childhood co-morbidity. There was a positive direct association between geophysical and household factors (β=0.889, *p*=0.034), and a negative direct association between geophysical and environmental factors (β=-0.217, *p*=0.008). Finally, we observed a negative direct association between child demographic and household factors (β=-0.087, *p*=0.004).

In the second phase, we involve the testing of the indirect relationship between geophysical factors on co-morbidity, and child demography on co-morbidity. The geophysical factors had a positive indirect association with childhood co-morbidity via the mediating effect of the household. However, the indirect association between geophysical factors and childhood co-morbidity via the mediating effect of environmental factors was negative. Lastly, the child demography factors had a negative indirect association with childhood co-morbidity factors via the mediating effect of household factors.

Table 6.2: Standardised direct and indirect effects of factors on childhood co-morbidity of malaria, anaemia, and stunting.

Factors	Total effect	Direct effect	Indirect effect
Geophysical factors → Household	0.889*	0.889*	-
Geophysical factors → Environmental	-0.217**	-0.217**	-
Geophysical factors →Co-morbidity	0.231*	-1.006*	1.237*
Household factor→Co-morbidity	1.394*	1.394*	-
Child demography factor→Co-morbidity	-0.271*	-0.150*	-0.121*
Child demography factor→Household	-0.087***	-0.087***	-
Environmental → Co-morbidity	0.014**	0.014**	-

*The * on the numbers indicates the P values. $p^{*} \leq 0.05$; $p^{**} \leq 0.01$, and $p^{***} \leq 0.001$*

Table 6.3: Standardised regression coefficients for the full structured model.

	Parameters	Estimate	95% CI Lower	95% CI Upper	*P*-value
Environmental factors	Geophysical factors	-0.217	0.802	0.908	<0.001
Household factors	Demographic factors	-0.087	0.881	0.956	0.004
Household factors	Geophysical factors	0.889	2.305	2.573	<0.001
Comorbidity factors	Environmental factors	0.014	0.967	1.062	0.038
Comorbidity factors	Household factors	1.394	3.193	5.150	<0.001
Comorbidity factors	Demographic factors	-0.150	0.820	0.904	0.010
Comorbidity factors	Geophysical factors	1.006	0.286	0.461	<0.001
ANE	Comorbidity factors	0.493	1.481	1.809	<0.001
MAL	Comorbidity factors	0.650	1.691	2.164	<0.001
STU	Comorbidity factors	0.306	1.256	1.468	<0.001
FLO	Household factors	0.731	1.958	2.203	<0.001
WAT	Household factors	-0.465	0.608	0.649	<0.001
MEL	Household factors	0.842	2.259	2.385	<0.001
ROF	Household factors	-0.062	0.929	0.951	<0.001
TOI	Household factors	-0.544	0.563	0.599	<0.001
WEI	Household factors	0.549	1.649	1.818	<0.001
MTV	Household factors	0.739	2.052	2.136	<0.001
SHT	Household factors	0.230	1.246	1.271	<0.001
RES	Household factors	-0.752	0.458	0.485	<0.001
MED	Household factors	0.567	1.659	1.874	<0.001
ALT	Environmental factors	0.371	0.916	2.293	<0.001
EVI	Environmental factors	0.999	1.442	5.114	<0.001
ARI	Environmental factors	0.171	0.709	1.986	<0.001
LST	Environmental factors	0.951	1.598	4.191	<0.001
PRW	Environmental factors	0.178	0.685	2.085	<0.001
RAF	Environmental factors	0.680	1.239	3.146	<0.001
WET	Environmental factors	0.170	0.725	1.939	<0.001
REG	Geophysical factors	0.194	1.096	1.344	<0.001
TRW	Geophysical factors	0.535	1.099	2.654	<0.001
NLC	Geophysical factors	-0.036	0.603	1.543	<0.001
CHN	Demographic factors	0.598	1.702	1.962	<0.001
CHS	Demographic factors	-0.007	0.963	1.024	0.043
CHA	Demographic factors	0.120	1.033	1.231	<0.001

***ANE:** Anaemia, **MAL:** Malaria, **STU:** Stunting, **FLO:** Type of floor material, **WAT:** Type of drinking water, **MEL:** Mother's access to electricity, **ROF:** Type of roof material, **TOI:** Household has toilet facility, **WEI:** Wealth index, **MTV:** Mother's access to television, **SHT:** Whether household share toilet facility, **RES:** Type of place of residence, **MED:** Mother's education, **ALT:** Cluster altitude, **EVI:** Enhanced vegetation index, **ARI:** Aridity, **LST:** Land surface temperature, **PRW:** Proximity to water, **RAF:** Rainfall, **WET:** Wet days, **REG** Region, **TRW:** Travel times, **NLC:** Night light composite, **CHN:** Child slept under mosquito net, **CHS:** Child's sex, **CHA:** Child's age*

6.9. SUMMARY

In this chapter, we used the structural equation model, to better understand the complex interrelationships between socioeconomics, demographics, and environmental factors, as well as their direct or indirect relationship with childhood malaria, anaemia, and stunting co-morbidity in Burundi. We assumed that the observed variables are dependent on the latent variables. The model fit our data well and explained the complex interrelationships between variables in a dataset. The findings from this chapter revealed that the geographical factor (geographical region where a child resides, travel times, and nightlight composites), were

statistically significant determinants of childhood malaria, anaemia, and stunting, and have a direct and indirect effect on childhood co-morbidity factors. The estimated indirect path for the effect of geophysical factors on childhood co-morbidity factors, as mediated by household factors were statistically significant and positive.

However, the estimated indirect paths for the effect of geophysical factors on childhood co-morbidity factors, as mediated by environmental factors were statistically significant but negative. The child demography factors such as child's age, child's gender and child sleeping under a mosquito net, were statistically significant predictors of childhood co-morbidity factors. The estimated indirect path effect on childhood co-morbidity factors via the mediating effect of household factors was statistically significant and negative.

The household factors comprising residence, wealth index, source of drinking water, type of toilet facility, the household share of toilet facility, mother's educational attainment, mother's access to information through television, household access to electricity, household's main roof, floor, and wall material were also statistically significant predictors in childhood co-morbidity factors. The study also indicates that environmental factors such as rainfall, proximity to water, land surface temperature, enhanced vegetation index, aridity, wet days, and cluster were statistically significant predictors of childhood co-morbidity factors.

The SEM model was good to fit the data and indicated the complex interrelationships between geographical, child demography, household, and environmental factors, as well as their direct or indirect relationship with childhood co-morbidity factors in Burundi. In addition, the SEM indicates a positive association between anaemia, malaria, and stunting.

CHAPTER SEVEN:
DISCUSSION AND CONCLUSION

The main purpose of this study was to investigate the association between stunting, anaemia, and malaria, along with other factors related to them in children younger than five years. Despite the efforts and resources committed to overcoming malaria, anaemia, and stunting continue to be a major public health concern around the world, particularly in low-income countries. Malaria, anaemia and stunting coexist and are associated with both mortality and morbidity, notably in children younger than five years. In Malawi, the incidence of anaemia and malaria in children younger than five years was 56.9% and 37.2%, respectively, with 61.5% of children having both anaemia and malaria. In Lesotho, the prevalence of both anaemia and stunting was 51% and 43%, respectively, with 35.2% of children suffering from both anaemia and stunting. Furthermore, in Burundi, the rate of anaemia, stunting, and malaria in children under the age of five was 59.1%, 47.5%, and 35.7%, respectively. Whereas anaemia and malaria were both prevalent at 48.6%, anaemia and stunting were prevalent at 55.9%, with malaria and stunting prevalent at 60.4%. Based on the above percentages anaemia was more prevalent than malaria and stunting in all three countries. Lesotho has a lower prevalence of anaemia and the country is considered malaria-free. This might be due to a higher prevalence of anaemia among malaria-infected children (Brabin et al., 2004; Gaston et al., 2021). As a result, controlling malaria with caution can lead to a decrease in anaemia.

In this study, we employed a national secondary data set from Malawi Malaria Indicator Survey (MMIS), Lesotho Demographic Health Survey (LDHS), and Burundi Demographic Health Survey (BDHS). In our analysis, we took into account a wide range of covariates such as socioeconomic, geographical, environmental, and demographic factors. These factors have been previously proposed in the literature depending on either theoretical analysis or statistical findings. The exploratory data analysis in Chapter 2 facilitated an understanding of the trends between the covariates factors and childhood anaemia, malaria, and stunting. Furthermore, exploratory data analysis provides evidence for integrating more reliable and innovative statistical techniques. The cross-tabulation method was used to summarise the data and the chi-squared test was used to investigate whether the exploratory variables were statistically significantly associated with each response variable or not. The results indicate that the socioeconomic, geographical, environmental, and demographic factors were statistically associated with childhood anaemia, malaria, and stunting. The findings from this study are in

line with the previous studies (Alegana et al., 2014; Caminade et al., 2014; Habyarimana et al., 2017; Kabaghe et al., 2017; Zgambo et al., 2017). The data for this study was gathered through stratified, and cluster sampling with an uneven chance of being chosen. Heering (2010) emphasised the importance of accounting for the complexity of sampling design to prevent misrepresenting variance and making incorrect inferences. As a result, we first employed the Survey Logistic Regression Model (SLRM), in Chapter 3, which took into consideration the intricacy of the sampling process as well as variability between data taken from the same cluster. The SLRM was applied to assess the relevant factors linked with anaemia in children under the age of five from Lesotho.

The findings from Chapter 3 revealed that the age of a child has a substantial influence on childhood anaemia and childhood anaemia decreases as the child gets older. Similar results were found in studies by Dey and Raheem (2016); Kuziga et al. (2017); Yimgang et al. (2021); among others. The findings also indicate a statistically significant effect between stunting or chronic malnutrition and anaemia, and this indicates the association between stunting and anaemia. This is in line with existing studies (Yang et al., 2012; Gari et al., 2017; Rahman et al., 2019). The findings showed a significant association between a mother's body mass index and childhood anaemia. This means that a child delivered to a mother who is either underweight or overweight has a greater chance of being anaemic and has a higher likelihood of being anaemic. The results from this study are consistent with those from other studies (Fonseca et al., 2016; Habyarimana et al., 2017). Furthermore, the findings showed that anaemia is more prevalent in children with a fever. This might be attributed to the fact that fever is often coupled with a variety of illness and morbidity that are known to positively affect anaemia such as diarrhoea, cough, and malaria among others. The findings from this study are in line with these from the studies by Santos et al. (2011); Konstantyner et al. (2012); Gayawan et al. (2014). In addition, the findings highlight that there is a higher chance of anaemia in children born to an anaemic mother. Same results we found in previous studies (Ngnie-Teta et al., 2007; Yang et al., 2012; Pita et al., 2014).

The survey logistic regression model used in Chapter 3 is robust and well-suited to our data. To avoid underestimation of variance and incorrect inference, the model accounted for the difficulties of sampling techniques along with the variation between the data from the same cluster. However, this model assumes that all variables have a fixed effect and does not enable random effects to be included. The DHS data used in our study has the primary sampling units (clusters) variable, which is considered as a random effect.

In addition, SLRM is parametric, and cannot be modified to a model with non-parametric fixed effects. As a result, in Chapter 4, we addressed these issues using the generalised additive mixed model (GAMM).

The generalised additive mixed model (GAMM) was used to examine potential factors linked with malaria using Malawian national malaria survey data. The GAMM is a generalised linear mixed model (GLMM) extension that allows the GLMM's parametric fixed effects to be modelled as a non-parametric model using the smooth additive function. To model the relationship between a response variable and covariance, parametric models are useful. However, non-parametric models are more adaptable, allowing for non-normal error distributions, modeling continuous predictor variables, and relaxing the assumption of normality and linearity in linear regression. Hence, combining the parametric and non-parametric models is more effective since they should enhance one another. In this Chapter 4, we first modelled the type of place of residence, region, wealth quantile, mother's highest education level, type of toilet facility, and availability of electricity factors as parametric. Then, age and altitude were modelled as non-parametric. We also checked the interaction effect, which might exist between the covariates factors and was statistically insignificant.

The findings from the parametric part revealed that the likelihood of increasing a positive malaria RDT was lower in children from the middle and top tertile wealth index. This confirmed that the prevalence of malaria is related to socioeconomic status, where children from within the lower levels of the wealth index are more susceptible. This is due to limited access to health care and the generally high cost of treatment. Similar results were found in previous studies (Hay et al., 2004; Chitunhu and Musenge, 2016; Zgambo et al., 2017).

The results also indicate that households with access to electricity and toilet facilities are less likely to experience increases in positive malaria rates. This shows that the households with access to electricity and toilet facilities factors are the indicators of socio-economic status. The findings from this study are in line with the studies by Ayele et al. (2014b); Chitunhu and Musenge (2016); Zgambo et al. (2017). The findings also highlighted a decrease in the positive malaria rate among the children whose mothers have a higher level of education. This suggests that the children of educated mothers, have a better life and can access health care easily which can help in reducing malaria. Similar results were found in the studies by Adebayo et al. (2016); Sultana et al. (2017); Zgambo et al. (2017). Children from rural areas indicated a high chance of having positive malaria. This might be explained by the lack of access to decent houses,

clean water, medical care, and other necessities. These results are consistent with previous studies (Jenkins et al. 2015; Kazembe and Mathanga, 2016; Sultana et al., 2017).

In addition, the parametric results indicate a significant difference between the three regions, with households in the Central region more likely to have positive rates of malaria. This is because the region is dominated by large plains and a low-lying zone along the lake, which are breeding grounds for malaria vectors. Similar results were found in previous studies (Minakawa et al., 2012; Zgambo et al., 2017). The non-parametric results indicate that the probability of positive malaria in children increases as a child gets older. This could be the result of children younger than one year being more protected, their maternal immunity aids in the fight against the disease. These results are in contrast with the study by Seyoum (2018). However, the same results are in line with the studies by Chirombo et al. (2014); Zgambo et al. (2017); and Yimgang et al. (2021). The findings also revealed children in lower altitude areas are more likely to have malaria, which reduces as the altitude increases. In higher altitude, the temperature reduces and the risk of getting malaria decreases in low temperatures. These findings are consistent with previous studies (Chirombo et al., 2014; The et al., 2018). The GAMM fitted our data well; however, the model applied to a single disease, and could not join two or both malaria, anaemia, and stunting simultaneously. Hence, in Chapter 5, we employed a multivariate joint model under GLMM to assess the correlation between either anaemia, malaria, or anaemia and stunting and their predictor factors.

In this Chapter 5, we used a multivariate joint model within the ambit of the generalised linear mixed model (GLMM) to assess the link between malaria and anaemia using the 2017 Malawi Malaria Indicator Survey (MMIS). The model also was used to examine the relationship between stunting and anaemia, and was applied to the 2014 Lesotho Demographic Health Survey (LDHS). The joint model is required to concurrently model two or both anaemia, malaria, and stunting to address their association and identify associated factors. The multivariate joint model under a GLMM has key elements compared to a single model, as an example, enhanced control of type I error degrees in the different tests and increased the ability of parameter estimate. In addition, to model the association between two or more variables, the GLMM introduces the random effect into the model.

The findings from this Chapter indicated a positive relationship between anaemia and stunting. The same results were found in previous studies (Zhao et al., 2012; Khan et al., 2016; Gari et al., 2018; Mohammed et al., 2019). In addition, the findings revealed a positive relationship between anaemia and malaria, and this shows that anaemia, stinting, and malaria are related, which was the key factor of our study. This means that malaria, anaemia, and stunting progress

in the same manner: where malaria and stunting increase or decrease so does anaemia. The findings from this study are in line with the studies by Adebayo et al. (2016); McCuskee et al. (2014); Ajakaye and Ibukunoluwa (2020). This implies that any change in policy regarding stunting or malaria will affect anemia.

The findings also indicate that children from mothers with advanced levels of education, especially those with access to electricity, toilet facilities, and sleeping under mosquito bed nets are less likely to have malaria, anaemia, and stunting. This means that educated mothers, with access to electricity and toilet facilities, can easily access healthcare, eat healthy food and afford treatment for their children. Similar results were found in study by Aheto et al. (2015); Adebayo et al. (2016); Ajakaye and Ibukunoluwa (2020); Yimgang et al. (202).

The results also show that children living in low altitudes and from rural areas are more exposed to malaria and least prone to stunting. This is a result of the noticeably higher temperatures in the residential area at lower altitudes, where the mosquitoes breed. The results from this study are consistent with the previous studies (Kavosi et al., 2014; Teh et al., 2018;).

The findings also highlight that fever, diarrhoea, and birth weight of a child were statistically significant on childhood anaemia and stunting. Similar results were found in studies by Konstantyner et al. (2012); Gayawan et al. (2014); Kejo et al. (2018); Rivadeneira et al. (2020).

Moreover, the findings from Chapter 5 indicated that the chances of developing malaria increase with age, while anaemia and stunting decrease. The same results were found in previous studies (Kuziga et al., 2017; Zgambo et al., 2017; Yimgang et al., 2021).

This Chapter revealed a positive relationship between anaemia, stunting, and malaria, which implies that they are the most significant health concern in children younger than five years. However, we could not address the complex interrelationships between explanatory factors, as well as their direct or indirect relationship with childhood malaria, anaemia, and stunting co-morbidity. Thus, for that reason in Chapter 6, we employed a structural equation modelling (SEM) to investigate the complex interrelationships between dependent variables as well as their direct or indirect relationship with childhood malaria, anaemia, and stunting co-morbidity. The structural equation model can assess complex interrelationships between different variables and related unobserved and observed variables. The assessment can be done by calculating the sample covariance matrix of the estimated parameters, as well as the population covariance matrix produced by the SEM framework. The finding from Chapter 6 highlights a positive association between anaemia, malaria, and stunting which confirmed the results from Chapter 5. The geographical factors were statistically significant and had a positive direct effect on childhood malaria, anaemia, and stunting. These results are in line with the previous studies (McCuskee et al. 2014; Kateera et al., 2015; Ajakaye and Ibukunoluwa, 2020).

The estimated indirect path for the effect of geophysical factors on childhood co-morbidity factors, as mediated by household factors was statistically significant and positive. However, the estimated indirect paths for the effect of geophysical factors on childhood co-morbidity factors, as mediated by environmental factors were statistically significant but negative. The child demographic factors revealed a direct statistically significant impact on childhood co-morbidity factors. Furthermore, the estimated indirect path effect on childhood co-morbidity as a mediated effect on household factors was statistically significant and negative. Moreover, household and environmental factors indicate a positive direct effect on childhood co-morbidity anaemia, malaria, and stunting. The current study revealed that anaemia, malaria and malnutrition are related.

Furthermore, the study indicated the factors associated with childhood anaemia, malaria, and stunting. Hence, understanding the link between anaemia, malaria, and stunting, and associated factors will assist the donors and policymakers to support and focus on those areas. Furthermore, this will contribute to achieving the United Nations Sustainable Development Goals (SDGs3), known as the complete elimination of under-5 mortality by 2030. To reduce anaemia, malaria, and stunting in children younger than five years, effective prevention measures must be established. Hence, we recommend that the policymakers and the state work together to educate each household about malaria, anaemia, and stunting. We further recommend enhancing medical care, sanitation amenities, distribution of mosquito bed nets, and nutrition status mainly in children from rural settings with diarrhoea, fever, and low birth weight. Moreover, children from uneducated mothers within the low quartile index, and high altitude should be considered and supported. This may be accomplished through a variety of channels, including social media, television, radio, promotional events, and even rural seminars.

This research contributes to the literature by modelling anaemia, stunting, and malaria in children younger than five years using a multivariate joint model within the framework of a generalised linear mixed model (GLMM). To our knowledge, no study assessed the connection between anaemia, malaria, and malnutrition in children younger than five years. Few studies attempted to model either malaria and anaemia, or anaemia, and stunting or both malaria, anaemia, and stunting simultaneously using logistic regression which does not account for the random variable in the regression model. Furthermore, a further novel advantage is the use of the structural equation model (SEM) to determine the factors which may directly or indirectly have an impact on childhood co-morbidity of anaemia, malaria, and stunting.

There is no study without limitation; hence, the present study could not use a multivariate spatial model to account for the spatial variability for anaemia, malaria, and malnutrition between the regions. This variation of health problems in districts could have helped in identifying the districts at high risk and informed allocation of resources and other support that can go directly to those districts at risk.

The study also used stunting as a measure of nutrition status, while wasting and underweight can be used as the measure of nutrition. Furthermore, the data set used in this research was cross-sectional survey data, and this means that a causal relationship cannot be addressed.

In addition, the current research considered the anaemia status of a child as a binary response variable (anaemic or non-anaemic child). However, the status of anaemia in children can be summarised as severe, moderate, mild, or non-anaemic.

Further research can use longitudinal data to address the causal association between the response variables, and account for the spatial variability of multiple diseases between the regions.

BIBLIOGRAPHY

Aerts, M., Molenberghs, G., Ryan, L.M. and Geys, H., 2002. *Topics in modelling of clustered data*. (1st Edition), New York, Chapman and Hall/CRC.

Adebayo, S.B., Gayawan, E., Heumann, C. and Seiler, C., 2016. Joint modelling of anaemia and malaria in children under five in Nigeria. *Spatial and spatio-temporal epidemiology*, *17*, pp.105-115.

Adhikari, R.P., Shrestha, M.L., Acharya, A. and Upadhaya, N., 2019. Determinants of stunting among children aged 0–59 months in Nepal: findings from Nepal Demographic and Health Survey, 2006, 2011, and 2016. *BMC Nutrition*, *5*(1), pp.1-10.

Agresti, A., 2010. *Analysis of ordinal categorical data.* (2nd Edition), Hoboken, New Jersey, John Wiley & Sons.

Agresti, A., 2015. *Foundations of linear and generalised linear models*. Hoboken, New Jersey, John Wiley & Sons, Inc.

Aheto, J.M.K., Keegan, T.J., Taylor, B.M. and Diggle, P.J., 2015. Childhood malnutrition and its determinants among under-five children in Ghana. *Paediatric and perinatal epidemiology*, *29*(6), pp.552-561.

Aheto, J.M.K., Taylor, B.M., Keegan, T.J. and Diggle, P.J., 2017. Modelling and forecasting spatio-temporal variation in the risk of chronic malnutrition among under-five children in Ghana. *Spatial and spatio-temporal epidemiology*, *21*, pp.37-46.

Ajakaye, O.G. and Ibukunoluwa, M.R., 2020. Prevalence and risk of malaria, anemia and malnutrition among children in IDPs camp in Edo State, Nigeria. *Parasite epidemiology and control*, *8*, p.e00127.

Ainur, A.K., Sayang, M.D., Jannoo, Z. and Yap, B.W., 2017. Sample size and non-normality effects on goodness of fit measures in structural equation models. *Pertanika journal of science & technology*, *25*(2), pp.575-586.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), pp.716-723.

Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199-213). New York, Springer.

Akritas, M.G., 1996. On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*, pp.1342-1362.

Alegana, V.A., Wright, J.A., Nahzat, S.M., Butt, W., Sediqi, A.W., Habib, N., Snow, R.W., Atkinson, P.M. and Noor, A.M., 2014. Modelling the incidence of Plasmodium vivax and Plasmodium falciparum malaria in Afghanistan 2006–2009. *PloS one*, *9*(7), pp. e102304.

Anticona, C. and San Sebastian, M., 2014. Anemia and malnutrition in indigenous children and adolescents of the Peruvian Amazon in a context of lead exposure: a cross-sectional study. *Global health action*, *7*(1), p.22888.

Arnold, B.C. and Strauss, D., 1991. Pseudolikelihood estimation: some examples. *Sankhyā: the Indian journal of statistics, Series B*, *53*(2), pp.233-243.

Austin, J.T. and Wolfle, L.M., 1991. Annotated bibliography of structural equation modelling: Technical work. *British journal of mathematical and statistical psychology*, *44*(1), pp.93-152.

Ayele, D.G., Zewotir, T.T. and Mwambi, H.G., 2013. The risk factor indicators of malaria in Ethiopia. *International journal of medicine and medical sciences*, *5*(7), pp.335-347.

Ayele, D.G., Zewotir, T.T. and Mwambi, H.G., 2014a. Semiparametric models for malaria rapid diagnosis test result. *BMC Public Health*, *14*(1), pp.31.

Ayele, D.G., Zewotir, T.T. and Mwambi, H.G., 2014b. Modelling the joint determinants of a positive malaria Rapid Diagnosis Test result, use of mosquito nets and indoor residual spraying with insecticide. *Occupational health Southern Africa*, *20*(4), pp.20-27.

Ayoya, M.A., Ngnie-Teta, I., Séraphin, M.N., Mamadoultaibou, A., Boldon, E., Saint-Fleur, J.E., Koo, L. and Bernard, S., 2013. Prevalence and risk factors of anemia among children 6–59 months old in Haiti. *Anemia*, *2013*: 2013, p. 502968.

Bagozzi, R.P. and Foxall, G.R., 1996. Construct validation of a measure of adaptive-innovative cognitive styles in consumption. *International journal of research in marketing*, *13*(3), pp.201-213.

Bell, E., Bryman, A. and Harley, B., 2022. *Business research methods*. (6th Edition), New York, Oxford university press.

Bennett, A., Kazembe, L., Mathanga, D.P., Kinyoki, D., Ali, D., Snow, R.W. and Noor, A.M., 2013. Mapping malaria transmission intensity in Malawi, 2000–2010. *The American journal of tropical medicine and hygiene*, *89*(5), pp.840-849.

Benoist, B.D., McLean, E., Egll, I., Cogswell, M. and World Health Organization, 2021. *Worldwide prevalence of anaemia 1993–2005: WHO global database on anaemia.* Geneva, World Health Organization.

Bentler, P.M., 1990. Comparative fit indexes in structural models. *Psychological bulletin*, *107*(2), p.238.

Bentler, P.M. and Wu, E.J., 2005. *EQS 6.1 for Windows. Structural equations program manual.* Encino, CA: Multivariate Software.

Biemba, G., Dolmans, D., Thuma, P.E., Weiss, G. and Gordeuk, V.R., 2000. Severe anaemia in Zambian children with *Plasmodium falciparum* malaria. *Tropical medicine and international health*, *5*(1), pp.9-16.

Binder, D.A., 1983. On the variances of asymptotically normal estimators from complex surveys. *International statistical review* 51 (3), pp.279-292.

Black, R.E., Victora, C.G., Walker, S.P., Bhutta, Z.A., Christian, P., De Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R. and Uauy, R., 2013. Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, *382*(9890), pp.427-451.

Bollen, K.A., 1989. *Structural equations with latent variables* (Vol. 210). New York, John Wiley & Sons.

Bollen, K. A., & Hoyle, R. H. (2012). Latent variables in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*. New York, the Guilford Press, pp. 56–67.

Bonett, D.G. and Wright, T.A., 2015. Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of organizational behavior*, *36*(1), pp.3-15.

Booth, J.G. and Hobert, J.P., 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B*, *61*(1), pp.265-285.

Brabin, B.J., Kalanda, B.F., Verhoeff, F.H., Chimsuku, L.H. and Broadhead, R.L., 2004. Risk factors for fetal anaemia in a malarious area of Malawi. *Annals of tropical paediatrics*, *24*(4), pp.311-321.

Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), pp.9-25.

Bujang, M.A., Omar, E.D. and Baharum, N.A., 2018. A review on sample size determination for Cronbach's alpha test: a simple guide for researchers. *The Malaysian journal of medical sciences: MJMS*, *25*(6), p.85.

Byrne, B.M., 2013. *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. City, Routledge.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A.M., Morse, A.P., Colón-González, F.J., Stenlund, H., Martens, P. and Lloyd, S.J., 2014. Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences*, *111*(9), pp.3286-3291.

Canty, A.J. and Davison, A.C., 1999. Resampling-based variance estimation for labour force surveys. *Journal of the Royal Statistical Society: Series D (The Statistician), 48*(3), pp.379-391.

Caraka, R.E., Noh, M., Chen, R.C., Lee, Y., Gio, P.U. and Pardamean, B., 2021. Connecting climate and communicable disease to penta helix using hierarchical likelihood structural equation modelling. *Symmetry*, *13*(4), p.657.

Carpentier, J., Mageau, G.A. and Vallerand, R.J., 2012. Ruminations and flow: Why do people with a more harmonious passion experience higher well-being? *Journal of happiness studies*, *13*(3), pp.501-518.

Caulfield, L.E., Richard, S.A. and Black, R.E., 2004. Undernutrition as an underlying cause of malaria morbidity and mortality in children less than five years old. *American journal of tropical medicine and hygiene*, *71*(2), pp.55-63.

Chavance, M., Escolano, S., Romon, M., Basdevant, A., de Lauzon-Guillain, B. and Charles, M.A., 2010. Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology. *BMC medical research methodology*, *10*(1), pp.1-10.

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*, *9*(7), pp. e101116.

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.

Davis, C.S., 1991. Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in medicine*, *10*(12), pp.1959-1980.

Davison, A.C. and Hinkley, D.V., 1997. *Bootstrap methods and their application*. United Kingdom, Cambridge, Cambridge University Press.

Deribew, A., Alemseged, F., Tessema, F., Sena, L., Birhanu, Z., Zeynudin, A., Sudhakar, M., Abdo, N., Deribe, K. and Biadgilign, S., 2010. Malaria and under-nutrition: a community-based study among under-five children at risk of malaria, south-west Ethiopia. *PLos one*, *5*(5), p.e10775.

Dey, S. and Raheem, E., 2016. Multilevel multinomial logistic regression model for identifying factors associated with anemia in children 6–59 months in North-Eastern states of India. *Cogent mathematics*, *3*(1), p.1159798.

Diggle, P., Diggle, P.J., Heagerty, P., Liang, K.Y. and Zeger, S., 2002. *Analysis of longitudinal data*. (2nd Edition), United Kingdom, Oxford University Press.

Ditlevsen, S., Christensen, U., Lynch, J., Damsgaard, M.T. and Keiding, N., 2005. The mediation proportion: a structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. *Epidemiology*, pp.114-120.

Dobson, A.J. and Barnett, A.G., 2018. *An introduction to generalized linear models*. (4th Edition), New York, Chapman and Hall/CRC.

Efron, B., 1979. (1979a). Bootstrap methods: Another look at the jackknife. *Annals of statistics*, *7*(1), pp.1-26.

Efron, B., 1981. Nonparametric standard errors and confidence intervals. *Canadian journal of statistics*, *9*(2), pp.139-158.

Efron, B., 1982. *The jackknife, the bootstrap and other resampling plans*. Philadelphia, Society for Industrial and Applied Mathematics.

Ehrhardt, S., Burchard, G.D., Mantel, C., Cramer, J.P., Kaiser, S., Kubo, M., Otchwemah, R.N., Bienzle, U. and Mockenhaupt, F.P., 2006. Malaria, anemia, and malnutrition in African children-defining intervention priorities. *The journal of infectious diseases*, *194*(1), pp.108-114.

Enjuanes, C., Bruguera, J., Grau, M., Cladellas, M., Gonzalez, G., Meroño, O., Moliner-Borja, P., Verdú, J.M., Farré, N. and Comín-Colet, J., 2016. Iron status in chronic heart failure: impact on symptoms, functional class and submaximal exercise capacity. *Revista española de cardiología*, *69*(3), pp.247-255.

Everitt, B., 1998. *The Cambridge dictionary of statistics*. United Kingdom, Cambridge, Cambridge University Press.

Faes, C., Aerts, M., Molenberghs, G., Geys, H., Teuns, G. and Bijnens, L., 2008. A high-dimensional joint model for longitudinal outcomes of different nature. *Statistics in medicine*, *27*(22), pp.4408-4427.

Fahrmeir, L., Tutz, G., Hennevogl, W. and Salem, E., 1994. *Multivariate statistical modelling based on generalized linear models*. New York, Springer-Verlag.

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S.R., Park, H. and Shao, C., 2016. Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecological processes*, *5*(1), pp.1-12.

Fieuws, S. and Verbeke, G., 2006. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, *62*(2), pp.424-431.

Fonseca, C., Marques, F., Robalo Nunes, A., Belo, A., Brilhante, D. and Cortez, J., 2016. Prevalence of anaemia and iron deficiency in Portugal: the Empire study. *Internal medicine journal*, *46*(4), pp.470-478.

Fuller, W.A., 1975. Regression analysis for sample survey. *Sankhya series C*, *37*(3), pp.117-132.

Gahutu, J.B., Steininger, C., Shyirambere, C., Zeile, I., Cwinya-Ay, N., Danquah, I., Larsen, C.H., Eggelte, T.A., Uwimana, A., Karema, C. and Musemakweri, A., 2011. Prevalence and risk factors of malaria among children in southern highland Rwanda. *Malaria journal*, *10*(1), pp.1-12.

Gameroff, M.J., 2005. Using the proportional odds model for health-related outcomes: Why, when, and how with various SAS procedures. *Proceedings of the thirtieth annual SAS users group international conference*, 205–230.

Gari, T., Loha, E., Deressa, W., Solomon, T., Atsbeha, H., Assegid, M., Hailu, A. and Lindtjørn, B., 2017. Anaemia among children in a drought affected community in south-central Ethiopia. *PloS one*, *12*(3), p.e0170898.

Gari, T., Loha, E., Deressa, W., Solomon, T. and Lindtjørn, B., 2018. Malaria increased the risk of stunting and wasting among young children in Ethiopia: results of a cohort study. *PloS one*, *13*(1), p.e0190983.

Gaston, R.T., Ramroop, S. and Habyarimana, F., 2018. Determinants of factors associated with anaemia among children under five years in Lesotho. *African population studies*, *32*(1), pp.3893-3902.

Gaston, R.T. and Ramroop, S., 2020. Prevalence of and factors associated with malaria in children under five years of age in Malawi, using malaria indicator survey data. *Heliyon*, *6*(5), p.e03946.

Gaston, R.T., Ramroop, S. and Habyarimana, F., 2021. Joint modelling of malaria and anaemia in children less than five years of age in Malawi. *Heliyon*, *7*(5), p.e06899.

Gaston, R.T., Habyarimana, F. and Ramroop, S., 2022. Joint modelling of anaemia and stunting in children less than five years of age in Lesotho: a cross-sectional case study. *BMC public health*, *22*(1), pp.1-11.

Gayawan, E., Arogundade, E.D. and Adebayo, S.B., 2014. Possible determinants and spatial patterns of anaemia among young children in Nigeria: a Bayesian semi-parametric modelling. *International health*, *6*(1), pp.35-45.

Geys, H., Molenberghs, G. and Ryan, L.M., 1997. Pseudo-likelihood inference for clustered binary data. *Communications in statistics-theory and methods*, *26*(11), pp.2743-2767.

Golob, T.F., 2003. Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological*, *37*(1), pp.1-25.

Gould, J. and Golob, T.F., 1997. Shopping without travel or travel without shopping, an investigation of electronic home shopping. *Transport reviews*, *17*(4), pp.355-376.

Grace, J.B., Anderson, T.M., Olff, H. and Scheiner, S.M., 2010. On the specification of structural equation models for ecological systems. *Ecological monographs*, *80*(1), pp.67-87.

Green, P.J. and Silverman, B.W., 1993. *Nonparametric regression and generalized linear models: a roughness penalty approach*. London, CRC Press.

Grizzle, J.E., Starmer, C.F. and Koch, G.G., 1969. Analysis of categorical data by linear models. *Biometrics*, *25*(3), pp.489-504.

Gueorguieva, R., 2001. A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical modelling*, *1*(3), pp.177-193.

Habanabakize, T. (2021). Determining the household consumption expenditure's resilience towards petrol price, disposable income and exchange rate volatilities. *Economies*, *9*(2), 87.

Habyarimana, F., Zewotir, T. and Ramroop, S., 2016. Joint modeling of poverty of households and malnutrition of children under five years from demographic and health survey data: case of Rwanda. *Journal of economics and behavioral studies*, *8*(2), pp.108-114.

Habyarimana, F., Zewotir, T. and Ramroop, S., 2017. Structured additive quantile regression for assessing the determinants of childhood anemia in Rwanda. *International journal of environmental research and public health*, *14*(6), p.652.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G., 1953. Sample survey methods and theory. New York, Wiley.

Härdle, W., 1990. *Applied nonparametric regression*. Cambridge UK: Cambridge University Press.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A., 2004. *Nonparametric and semiparametric models*. Berlin, Springer.

Hastie, T.J. and Tibshirani, R., 1986. Generalized additive models. *Statistical science*, 1, pp.297-318.

Hastie, T. J. and Tibshirani, R., 1990. *Generalized additive models*. London, CRC Press.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M. and Snow, R.W., 2004. The global distribution and population at risk of malaria: past, present, and future. *The lancet infectious diseases*, *4*(6), pp.327-336.

Hedeker, D., 2005. Generalized linear mixed models. *Encyclopedia of statistics in behavioral science*. City, John Wiley and Sons, Ltd.

Heeringa, S., West, B.T., Berglund, P.A., 2010. *Applied survey data analysis: Statistical in the social and behavioral sciences series*. New York, Chapman & Hall/CRC.

Hershey, C.L., Florey, L.S., Ali, D., Bennett, A., Luhanga, M., Mathanga, D.P., Salgado, S.R., Nielsen, C.F., Troell, P., Jenda, G. and Ye, Y., 2017. Malaria control interventions contributed to declines in malaria parasitemia, severe anaemia, and all-cause mortality in children less than 5 years of age in Malawi, 2000–2010. *The American journal of tropical medicine and hygiene*, *97*(3_Suppl), pp.76-88.

Holmes, A.P., Blair, R.C., Watson, J.D.G. and Ford, I., 1996. Nonparametric analysis of statistic images from functional mapping experiments. *Journal of cerebral blood flow & metabolism*, *16*(1), pp.7-22.

Holmes, D.J. and Skinner, C.J., 2000. Variance estimation for labour force survey estimates of level and change. *Government statistical service methodology series 21*.

Hooper, D., Coughlan, J. and Mullen, M.R., 2008. Structural equation modelling: Guidelines for determining model fit. *Electronic journal of business research methods*, *6*(1), pp.53-60.

Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X., 2000. *Introduction to the logistic regression model*. (2nd edition), New York, John Wiley & Sons.

Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013. *Applied logistic regression*. (3rd edition), New York: John Wiley & Sons.

Jain, N., Gupta, R. and Gupta, R., 2017. Generalized additive and generalized linear modeling for children diseases. *Quest Journals-journal of research in applied mathematics*, *3*(5), pp.0-10.

Jenkins, R., Omollo, R., Ongecha, M., Sifuna, P., Othieno, C., Ongeri, L., Kingora, J. and Ogutu, B., 2015. Prevalence of malaria parasites in adults and its determinants in malaria endemic area of Kisumu County, Kenya. *Malaria journal*, *14*(1), pp.263.

Jiang, J. (2001). A non-standard χ 2 -test with application to generalized linear model diagnostics. *Statistical and probability letters*, *53*(1), pp.101–109.

Jöreskog, K.G. and Sörbom, D., 1982. Recent developments in structural equation modeling. *Journal of marketing research*, *19*(4), pp.404-416.

Judkins, D.R., 1990. Fay's method for variance estimation. *Journal of official statistics*, *6*(3), pp.223-239.

Kabaghe, A.N., Phiri, M.D., Phiri, K.S. and van Vugt, M., 2017. Challenges in implementing uncomplicated malaria treatment in children: a health facility survey in rural Malawi. *Malaria journal*, *16*(1), pp.419.

Kanchana, M.S., Ahuja, S. and Nagaraj, N., 2018. Prevalence and risk factors of anemia in under five-year-old children in children's hospital. *International journal of contemporary pediatrics*, *5*(2), pp.499-502.

Kandala, N.B., Madungu, T.P., Emina, J.B., Nzita, K.P. and Cappuccio, F.P., 2011. Malnutrition among children under the age of five in the Democratic Republic of Congo (DRC): does geographic location matter? *BMC public health*, *11*(1), pp.1-15.

Kaplan, D., 2008. *Structural equation modeling: Foundations and extensions* (Vol. 10). City, SAGE publications.

Kateera, F., Ingabire, C.M., Hakizimana, E., Kalinda, P., Mens, P.F., Grobusch, M.P., Mutesa, L. and van Vugt, M., 2015. Malaria, anaemia and under-nutrition: three frequently co-existing conditions among preschool children in rural Rwanda. *Malaria journal*, *14*(1), pp.1-11.

Kavosi, E., Rostami, Z.H., Kavosi, Z., Nasihatkon, A., Moghadami, M. and Heidari, M., 2014. Prevalence and determinants of under-nutrition among children under six: a cross-sectional survey in Fars province, Iran. *International journal of health policy and management*, *3*(2), p.71.

Kazembe, L.N., 2007. Spatial modelling and risk factors of malaria incidence in northern Malawi. *Acta Tropica*, *102*(2), pp.126-137.

Kazembe, L.N., 2013. An additive regression model for investigating the relationship between childhood health and socio-economic status. *Spatial and spatio-temporal epidemiology*, *6*, pp.71-84.

Kazembe, L.N. and Mathanga, D.P., 2016. Estimating risk factors of urban malaria in Blantyre, Malawi: A spatial regression analysis. *Asian Pacific journal of tropical biomedicine*, *6*(5), pp.376-381.

Kejo, D., Petrucka, P.M., Martin, H., Kimanya, M.E. and Mosha, T.C., 2018. Prevalence and predictors of anemia among children under 5 years of age in Arusha District, Tanzania. *Pediatric health, medicine and therapeutics*, *9*, p.9.

Khan, J.R., Awan, N. and Misu, F., 2016. Determinants of anemia among 6–59 months aged children in Bangladesh: evidence from nationally representative data. *BMC Pediatrics*, *16*(1), pp.1-12.

Kishawi, E., Rafiq, R., Soo, K.L., Abed, Y.A. and Wan Muda, W.A.M., 2015. Anemia among children aged 2–5 years in the Gaza Strip-Palestinian: a cross sectional study. *BMC public health*, *15*(1), pp.1-8.

Kline, R. B. 2015. *Principles and practice of structural equation modeling*, New York, NY, Guilford publications.

Kock, N. and Lynn, G., 2012. Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the association for information systems*, *13*(7).

Konstantyner, T., Roma Oliveira, T.C. and de Aguiar Carrazedo Taddei, J.A., 2012. Risk factors for anemia among Brazilian infants from the 2006 National Demographic Health Survey. *Anemia*, *2012*, 1012, 850681.

Korenromp, E.L., Armstrong-Schellenberg, J.R., Williams, B.G., Nahlen, B.L. and Snow, R.W., 2004. Impact of malaria control on childhood anaemia in Africa–a quantitative review. *Tropical medicine & international health*, *9*(10), pp.1050-1065.

Kotecha, P.V., 2011. Nutritional anemia in young children with focus on Asia and India. *Indian journal of community medicine: Official Publication of Indian Association of Preventive & Social Medicine*, *36*(1), p.8.

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W., 2005. *Applied linear statistical models.* New York, McGraw Hill Irwin.

Kuziga, F., Adoke, Y. and Wanyenze, R.K., 2017. Prevalence and factors associated with anaemia among children aged 6 to 59 months in Namutumba district, Uganda: a cross-sectional study. *BMC Pediatrics*, *17*(1), pp.1-9.

Leal, L.P., Batista Filho, M., Lira, P.I.C.D., Figueiroa, J.N. and Osório, M.M., 2011. Prevalence of anaemia and associated factors in children aged 6-59 months in Pernambuco, North-Eastern Brazil. *Revista de Saude Publica*, *45*, pp.457-466.

Lehtonen, R. and Pahkinen, E., 2004. *Practical methods for design and analysis of complex surveys*. (2nd edition), Chichester, John Wiley & Sons.

Leite, M.S., Cardoso, A.M., Coimbra, C.E., Welch, J.R., Gugelmin, S.A., Lira, P.C.I., Horta, B.L., Santos, R.V. and Escobar, A.L., 2013. Prevalence of anemia and associated factors among indigenous children in Brazil: results from the First National Survey of Indigenous People's Health and Nutrition. *Nutrition journal*, *12*(1), pp.1-11.

Letsie, M.M. and Grab, S.W., 2015. Assessment of social vulnerability to natural hazards in the Mountain Kingdom of Lesotho. *Mountain Research and Development*, *35*(2), pp.115-125.

Lin, X. and Zhang, D., 1999. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series b (Statistical Methodology)*, *61*(2), pp.381-400.

Lin, X. and Carroll, R.J., 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, *95*(450), pp.520-534.

Lindsay, S.W. and Martens, W.J., 1998. Malaria in the African highlands: past, present and future. *Bulletin of the World Health Organization*, *76*(1), pp.33.

Lowry, P.B. and Gaskin, J., 2014. Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE transactions on professional communication*, *57*(2), pp.123-146.

Lumley, T. and Scott, A., 2015. AIC and BIC for modeling with complex survey data. *Journal of survey statistics and methodology*, *3*(1), pp.1-18.

Maketa, V., Mavoko, H.M., da Luz, R.I., Zanga, J., Lubiba, J., Kalonji, A. and Lutumba, P., 2015. The relationship between Plasmodium infection, anaemia and nutritional status in asymptomatic children aged under five years living in stable transmission zones in Kinshasa, Democratic Republic of Congo. *Malaria journal*, *14*(1), pp.1-9.

Marsh, H.W. and Hocevar, D., 1985. Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological bulletin*, *97*(3), p.562.

McCarthy, P.J., 1969. Pseudo-replication: Half samples. *Review of the international statistical institute* *37*(3), pp.239-264.

McCullagh, P. and Nelder, J.A., 1989. *Generalized linear models*. (2nd edition), New York, Chapman & Hall.

McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, *92*(437), pp.162-170.

McCulloch, C.E. and Searle, S.R., 2004. *Generalized, linear, and mixed models*. New York, John Wiley & Sons.

McCuskee, S., Brickley, E.B., Wood, A. and Mossialos, E., 2014. Malaria and macronutrient deficiency as correlates of anemia in young children: a systematic review of observational studies. *Annals of global health*, *80*(6), pp.458-465.

McGann, P.T., Williams, A.M., Ellis, G., McElhinney, K.E., Romano, L., Woodall, J., Howard, T.A., Tegha, G., Krysiak, R., Lark, R.M. and Ander, E.L., 2018. Prevalence of inherited blood disorders and associations with malaria and anemia in Malawian children. *Blood advances*, 2(21), pp.3035-3044.

Miller, R.G., 1974. The jackknife-a review. *Biometrika*, *61*(1), pp.1-15.

Milman, N., 2011. Anemia still a major health problem in many parts of the world. *Annals of hematology*, *90*(4), pp.369-377.

Minakawa, N., Dida, G.O., Sonye, G.O., Futami, K. and Njenga, S.M., 2012. Malaria vectors in Lake Victoria and adjacent habitats in western Kenya. *PloS one*, *7*(3), pp. e32725.

Ministry of Health [Lesotho] and ICF International (2016). *Lesotho Demographic and Health Survey 2014*. Maseru, Lesotho, Ministry of Health and ICF International.

Ministry at the Presidency in charge of good governance and planning [Burundi] (MPBGP), Ministry of health and the fight against AID [Burundi] (MPLS), Institute of Statistics and Economic studies of Burundi (ISTEEBU), and ICF. 2017. *Third Demographic and Health Survey*. Bujumbura, Burundi, ISTEEBU, MSPLS, and ICF.

Mohammed, S.H., Larijani, B. and Esmaillzadeh, A., 2019. Concurrent anemia and stunting in young children: prevalence, dietary and non-dietary associated factors. *Nutrition journal*, *18*(1), pp.1-10.

Moise, I.K., Roy, S.S., Nkengurutse, D. and Ndikubagenzi, J., 2016. Seasonal and geographic variation of pediatric malaria in Burundi: 2011 to 2012. *International journal of environmental research and public health*, *13*(4), p.425.

Moise, I.K., 2018. Causes of morbidity and mortality among neonates and children in post-conflict Burundi: a cross-sectional retrospective study. *Children*, *5*(9), p.125.

Molenberghs, G. and Verbeke, G., 2005. *Models for discrete longitudinal data*. New York, Springer-Verlag.

Moteetee, M.M., 2005. Report on the quality and responsiveness of health services to meet the needs of women in crisis settings. *WHO, June*, pp. l–8.

National Statistical Office (NSO) [Malawi] and ICF. 2017. *Malawi Demographic and Health Survey 2015-16.* Zomba, Malawi, and Rockville, Maryland, USA, NSO and ICF.

National Malaria Control Programme (NMCP) [Malawi] and ICF. 2018. *Malawi Malaria Indicator Survey 2017*. Lilongwe, Malawi, and Rockville, Maryland, USA: NMCP and ICF.

Nelder, J.A. and Wedderburn, R.W., 1972. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), pp.370-384.

Neuhaus, J.M., Kalbfleisch, J.D. and Hauck, W.W., 1991. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International statistical review*, *59*(1), pp.25-35.

Ngnie-Teta, I., Receveur, O. and Kuate-Defo, B., 2007. Risk factors for moderate to severe anemia among children in Benin and Mali: insights from a multilevel analysis. *Food and nutrition bulletin*, *28*(1), pp.76-89.

Nimpagaritse, M., Korachais, C. and Meessen, B., 2020. Effects in spite of tough constraints-A theory of change-based investigation of contextual and implementation factors affecting the results of a performance-based financing scheme extended to malnutrition in Burundi. *PLoS One*, *15*(1), p.e0226376. City, Springer-Verlag.

Nkoka, O., Chipeta, M.S., Chuang, Y.C., Fergus, D. and Chuang, K.Y., 2019. A comparative study of the prevalence of and factors associated with insecticide-treated nets usage among children under 5 years of age in households that already own nets in Malawi. *Malaria journal*, *18*(1), p.43.

Noland, G.S., Ayodo, G., Abuya, J., Hodges, J.S., Rolfes, M.A. and John, C.C., 2012. Decreased prevalence of anaemia in highland areas of low malaria transmission after a 1-year interruption of transmission. *Clinical infectious diseases*, *54*(2), pp.178-184.

Oliveira, D., Ferreira, F.S., Atouguia, J., Fortes, F., Guerra, A. and Centeno-Lima, S., 2015. Infection by intestinal parasites, stunting and anemia in school-aged children from southern Angola. *PloS one*, *10*(9), p. e0137327.

Olney, D.K., Kariger, P.K., Stoltzfus, R.J., Khalfan, S.S., Ali, N.S., Tielsch, J.M., Sazawal, S., Black, R., Allen, L.H. and Pollitt, E., 2009. Development of nutritionally at-risk young children is predicted by malaria, anaemia, and stunting in Pemba, Zanzibar. *The journal of nutrition*, *139*(4), pp.763-772.

Osazuwa, F. and Ayo, O.M., 2010. Contribution of malnutrition and malaria to anemia in children in rural communities of Edo state, Nigeria. *North American journal of medical sciences*, *2*(11), p.532.

Osterbauer, B., Kapisi, J., Bigira, V., Mwangwa, F., Kinara, S., Kamya, M.R. and Dorsey, G., 2012. Factors associated with malaria parasitaemia, malnutrition, and anaemia among HIV-exposed and unexposed Ugandan infants: a cross-sectional survey. *Malaria journal*, *11*(1), pp.1-6.

Park, I., 2008. Primary sampling unit (PSU) masking and variance estimation in complex surveys. *Survey methodology*, *34*(2), p.183-194.

Perkins, D.J., Were, T., Davenport, G.C., Kempaiah, P., Hittner, J.B. and Ong'echa, J.M., 2011. Severe malarial anemia: innate immunity and pathogenesis. *International journal of biological sciences*, *7*(9), p.1427.

Pita, G.M., Jiménez, S., Basabe, B., García, R.G., Macías, C., Selva, L., Hernández, C., Cruz, M., Herrera, R., O'Farrill, R. and Calderius, I., 2014. Anemia in children under five years old in Eastern Cuba, 2005-2011. *Medicc review*, *16*(1), pp.16-23.

Quenouille, M.H., 1949. Problems in plane sampling. *The annals of mathematical statistics*, *20*(3), pp.355-375.

Quenouille, M.H., 1956. Notes on bias in estimation. *Biometrika*, *43*(3/4), pp.353-360.

Rahman, M.S., Mushfiquee, M., Masud, M.S. and Howlader, T., 2019. Association between malnutrition and anaemia in under-five children and women of reproductive age: Evidence from Bangladesh Demographic and Health Survey 2011. *PloS one*, *14*(7), p. e0219170.

Raykov, T. and Marcoulides, G.A., 2012. *A first course in structural equation modeling*. New York, Routledge.

Rao, J.N.K. and Wu, C.J., 1985. Inference from stratified samples: Second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association*, *80*(391), pp.620-630.

Rao, J.N. and Wu, C.F.J., 1988. Resampling inference with complex survey data. *Journal of the American Statistical Association*, *83*(401), pp.231-241.

Rao, J.N.K. and Shao, J., 1996. On balanced half-sample variance estimation in stratified random sampling. *Journal of the American Statistical Association*, *91*(433), pp.343-348.

Rao, J.N.K., 1997. Developments in sample survey theory: an appraisal. *Canadian journal of statistics*, *25*(1), pp.1-21.

Reithinger, R., Ngondi, J.M., Graves, P.M., Hwang, J., Getachew, A., Jima, D., Ethiopia Malaria Indicator Survey Working Group, Amena, M., Bergeron, L., Bilak, H. and Chirwa, B., 2013. Risk factors for anaemia in children under 6 years of age in Ethiopia: analysis of the data from the cross-sectional Malaria Indicator Survey, 2007. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *107*(12), pp.769-776.

Renzaho, A., 2006. Mortality, malnutrition and the humanitarian response to the food crises in Lesotho. *Journal of emergency primary health care*, *4*(4), pp.1447-4999.

Rivadeneira, M.F., Moncayo, A.L., Tello, B., Torres, A.L., Buitrón, G.J., Astudillo, F., Fredricks, T.R. and Grijalva, M.J., 2020. A Multi-Causal Model for Chronic Malnutrition and Anaemia in a Population of Rural Coastal Children in Ecuador. *Maternal and child health Journal*, *24*(4), pp.472-482.

Ruppert, D., Wand, M.P. and Carroll, R.J., 2003. *Semiparametric regression*. Cambridge, Cambridge University Press.

Santos, R.F.D., Gonzalez, E.S.C., Albuquerque, E.C.D., Arruda, I.K.G.D., Diniz, A.D.S., Figueroa, J.N. and Pereira, A.P.C., 2011. Prevalence of anemia in under five-year-old children in a children's hospital in Recife, Brazil. *Revista brasileira de hematologia e hemoterapia*, *33*, *33*(2), pp.100-104.

Särndal, C. E., Swensson, B., and Wretman J., 1992. *Model assisted survey sampling.* (Springer series in statistics). New York,, Springer-Verlag.

Särndal, C.E., Swensson, B. and Wretman, J., 2003. *Model assisted survey sampling*. New York, Springer-Verlang. Inc.

Schaefer, E., Potter, F., Williams, S., Diaz-Tena, N., Reschovsky, J.D. and Moore, G., 2003. Community Tracking Study: Comparison of Selected Statistical Software Packages for

Variance Estimation in the CTS Surveys. *Technical publication*, (40). Center for Studying Health System Change 600 Maryland Ave., SW, Suite 550, Washington, DC.

Schneider, J.M., Fujii, M.L., Lamp, C.L., Lönnerdal, B.O., Dewey, K.G. and Zidenberg-Cherr, S., 2008. The use of multiple logistic regression to identify risk factors associated with anemia and iron deficiency in a convenience sample of 12–36-mo-old children from low-income families. *The American journal of clinical nutrition*, *87*(3), pp.614-620.

Schumacker, R.E. and Lomax, R.G., 2016. *A beginner's guide to structural equation modeling*. (4th edition), New York, N,: Routledge.

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The annals of statisti*cs, 6(2), 461–464.

Seyoum, S., 2018. Analysis of prevalence of malaria and anaemia using bivariate probit model. *Annals of data science*, *5*(2), pp.301-312.

Shankar, A.H., 2000. Nutritional modulation of malaria morbidity and mortality. *The journal of infectious diseases*, *182*(1), pp. S37-S53.

Shao, J. and Tu, D., 1995. *Jackknife and bootstrap*. New York, Springer-Verlag.

Shikur, B., Deressa, W. and Lindtjørn, B., 2016. Association between malaria and malnutrition among children aged under-five years in Adami Tulu District, south-central Ethiopia: a case–control study. *BMC public health*, *16*(1), pp.1-8.

Silverman, B.W., 2018. *Density estimation for statistics and data analysis*. (1st Edition), New York, Routledge.

Simbauranga, R.H., Kamugisha, E., Hokororo, A., Kidenya, B.R. and Makani, J., 2015. Prevalence and factors associated with severe anaemia amongst under-five children hospitalized at Bugando Medical Centre, Mwanza, Tanzania. *BMC hematology*, *15*(1), pp.1-9.

Sinzinkayo, D., Baza, D., Gnanguenon, V. and Koepfli, C., 2021. The lead-up to epidemic transmission: malaria trends and control interventions in Burundi 2000 to 2019. *Malaria journal*, *20*(1), pp.1-7.

Skinner, C.J., Holt, D. and Smith, T.M.F., 1989. *Analysis of complex surveys*. New York, John Wiley & Sons. Inc.

Smithson, P., Florey, L., Salgado, S.R., Hershey, C.L., Masanja, H., Bhattarai, A., Mwita, A., McElroy, P.D. and Tanzania Malaria Impact Evaluation Research Group, 2015. Impact of malaria control on mortality and anemia among Tanzanian children less than five years of age, 1999–2010. *PLoS one*, *10*(11), p.e0141112.

Sultana, M., Sheikh, N., Mahumud, R.A., Jahir, T., Islam, Z. and Sarker, A.R., 2017. Prevalence and associated determinants of malaria parasites among Kenyan children. *Tropical medicine and health*, *45*(1), pp.25.

Teh, R.N., Sumbele, I.U.N., Meduke, D.N., Ojong, S.T. and Kimbi, H.K., 2018. Malaria parasitaemia, anaemia and malnutrition in children less than 15 years residing in different altitudes along the slope of Mount Cameroon: prevalence, intensity and risk factors. *Malaria journal*, *17*(1), pp.1-13.

Thorne, C.J., Roberts, L.M., Edwards, D.R., Haque, M.S., Cumbassa, A. and Last, A.R., 2013. Anaemia and malnutrition in children aged 0–59 months on the Bijagos Archipelago, Guinea-Bissau, West Africa: a cross-sectional, population-based study. *Paediatrics and international child health*, *33*(3), pp.151-160.

Tuerlinckx, F., Rijmen, F., Verbeke, G. and De Boeck, P., 2006. Statistical inference in generalized linear mixed models: A review. *British journal of mathematical and statistical psychology*, *59*(2), pp.225-255.

UNICEF. (2017). *State of the World's children: Children in a digital World*. New York, NY, UNICEF.

Wang, Y., 1998. Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society: Series b (Statistical Methodology)*, *60*(1), pp.159-174.

Wanzira, H., Katamba, H., Okullo, A.E., Agaba, B., Kasule, M. and Rubahika, D., 2017. Factors associated with malaria parasitaemia among children under 5 years in Uganda: a secondary data analysis of the 2014 Malaria Indicator Survey dataset. *Malaria journal*, *16*(1), pp.1-9.

West, S.G., Taylor, A.B. and Wu, W., 2012. Model fit and model selection in structural equation modeling. *Handbook of structural equation modeling*, New York, Guilford Presspp.209-231.

Woldie, H., Kebede, Y. and Tariku, A., 2015. Factors associated with anemia among children aged 6–23 months attending growth monitoring at Tsitsika Health Center, Wag-Himra Zone, Northeast Ethiopia. *Journal of nutrition and metabolism*, *2015*, 2015, 928632.

Wolfinger, R., Tobias, R. and Sall, J., 1994. Computing Gaussian likelihoods and their derivatives for general linear mixed models. *SIAM journal on scientific computing*, *15*(6), pp.1294-1310.

Wolter, K.M., 2007. *Introduction to variance estimation*. (2nd Edition), New York, Springer science & business media.

Wood, S.N., 2006. *Generalized additive models: an introduction with R*. (1st Edition), New York, Chapman and hall/CRC.

Wood, S.N., 2017. *Generalized additive models: an introduction with R*. (2nd Edition), London, UK, Chapman and Hall/CRC.

World Health Organization, 1995. Physical status: The use and interpretation of anthropometry. Report of the expert committee.

World Health Organization, 2007. *WHO child growth standards: head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: methods and development*. Geneva, World Health Organization, 2007.

World Health Organization, 2011. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity (No. WHO/NMH/NHD/MNM/11.1). Geneva, World Health Organization.

World Health Organization, 2015. *Investing to overcome the global impact of neglected tropical diseases: third WHO report on neglected tropical diseases 2015* (Vol. 3). World Health Organization.

World Health Organization, 2016. *World Malaria Report 2016*. Geneva, World Health Organization.

World Health Organization, 2017. *Nutritional Anaemias: Tools for Effective Prevention and Control.* Geneva, World Health Organization.

World Malaria Report 2018. Geneva, World Health Organization. Licence: CC BY-NC-SA 3.0 IGO. https://apps.who.int/iris/bitstream/handle/10665/275867/9789241565653-eng.pdf. Accessed on 15 December 2021.

Worrall, E., Basu, S. and Hanson, K., 2003. The relationship between socio-economic status and malaria: a review of the literature. *Background paper for ensuring that malaria control interventions reach the poor,* London*,* London School of Hygiene and Tropical Medicine.

Wu, H. and Zhang, J.T., 2006. *Nonparametric regression methods for longitudinal data analysis.* John Wiley & Sons, New York.

Yang, W., Li, X., Li, Y., Zhang, S., Liu, L., Wang, X. and Li, W., 2012. Anemia, malnutrition and their correlations with socio-demographic characteristics and feeding practices among infants aged 0–18 months in rural areas of Shaanxi province in North-Western China: a cross-sectional study. *BMC public health*, *12*(1), pp.1-7.

Yimgang, D.P., Buchwald, A.G., Coalson, J.E., Walldorf, J.A., Bauleni, A., Kapito-Tembo, A., Mathanga, D.P., Taylor, T.E., Laufer, M.K. and Cohee, L.M., 2021. Population attributable fraction of anemia associated with Plasmodium falciparum infection in children in southern Malawi. *The American journal of tropical medicine and hygiene*, *104*(3), pp.1013-1017.

Yung, W. and Rao, J.N.K., 2000. Jackknife variance estimation under imputation for estimators using poststratification information. *Journal of the American Statistical Association*, *95*(451), pp.903-915.

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), pp. e0175537.

Zhang, D. and Lin, X., 2008. Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. In *Random effect and latent variable model selection* (pp. 19-36). Springer, New York, NY.

Zhao, A., Zhang, Y., Peng, Y., Li, J., Yang, T., Liu, Z., Lv, Y. and Wang, P., 2012. Prevalence of anemia and its risk factors among children 6–36 months old in Burma. *The American journal of tropical medicine and hygiene*, *87*(2), p.306.

APPENDIX

B. CODES

Survey Logistic Regression Codes Using SAS Software (Chapter 3)

```
Proc surveylogistic data=T;
class stunting (ref=’no’)  Fever (ref=’no’)  BMI (ref=’less than 18.5’) age (ref= ’40-59’)/
param = reference;
model anaemia (descending)= stunting  Fever  BMI   age;
strata Strata;
cluster Cluster;
weight SamplingWeight;
run;
```

==========

Generalized Additive Mixed Model (GAMM) codes Using R software (Chapter 4)

```
library (foreign)
library (lattice)
install.packages ("R2jags")
library (R2jags)
library (nlme)
library (lme4)
library (mgcv)
library (gamm4)
library (gam)
library (Hmisc)
library (ggplot2)
library (splines2)
setwd ("C:/Users/T/Desktop/Gamma")
data= read.spss("malaria.sav",to.data.frame = TRUE)
```

```
data
region = data$HV024
residence = data$HV025
malaria = data$fHML35
gender = data$fHC27
education = data$education
floor = data$floor_material
wall = data$fwall_material
roof = data$froof_material
toilet = data$fToilet
wealth = data$fWealth_Index
anemia = data$fAnemia
water = data$drinking_water
sleep_under_net = data$Child_sleeping_under_net
ITN= data$HML10
LLIN= data$HML20
TV = data$HV208
Radio = data$HV207
electricity = data$HV206
family_size = data$HV009
total_rooms = data$HV216
age = data$HC1
altitude = data$HV040
sleep_under_net = data$HML2
number_bed_net = data$HML1
family_size = data$HV009
gender = data$gender
number_of_rooms = data$HV216
number_of_rooms = data$fHV216
Wealth_Index = data$Wealth_Index
```

```
gam7< gamm4
(malaria~anemia+gender+region+water+electricity+LLIN+education+Wealth_Index
toilet+wall+residence+radio+wall+roof+sleep_under_net+s (age, bs = "ps”) +s (altitude, bs =
"ps”), random =~ (1|HV001), family = "binomial", data = data)

summary (gam7$gam) # working, keep adding and see ones which are significance

plot (gam7$gam,pages=1)
```

Joint model under GLMM codes Using SAS software (Chapter 5)

```
data   T;

length dist $11;

set FG;

response= (Malaria=1);

dist = &amp;quot;binary1&amp;quot;

output;

response = (Anemia=2);

dist = &amp;quot;binary2&amp;quot;

output;

keep Wealth_Index Toilet floor_material wall_material roof_material

Residence Child_age Region Education Sleeping Altitude share_toilet

Electricity Water response HV021 Gender dist;

run;

ods html;

proc freq;

tables response dist;
```

```
run;
ods html close;
ods html;
ods graphics on;
proc glimmix data= T asycov asycorr odds ratio ORDER=DATA MAXOPT=200
METHOD= quad (qpoints=25)/* plots=all*/;
class HV021 dist Residence Gender Wealth_Index Toilet floor_material
wall_material roof_material Child_age Region Education Sleeping Altitude
share_toilet Electricity Water;
model response= dist dist*Residence dist*Gender dist*Wealth_Index
dist*Toilet dist*floor_material dist*wall_material dist*roof_material
dist*Child_age dist*Region dist*Education dist*Sleeping dist*Altitude
dist*share_toilet dist*Electricity dist*Water / dist = byobs(dist) solution;
random dist/ subject =HV021 type = unr;
covtest/cl wald classical est;
run;
ods graphics off;
ods html close;
```

==========

B. PUBLICATIONS

Determinants of factors associated with anemia among children under five years in Lesotho

Rugiranka Tony Gaston, Shaun Ramroop & Faustin Habyarimana

School of Mathematics, Statistics and Computer Sciences,

University of KwaZulu-Natal, Pietermaritzburg Campus

South Africa

rugiratony@gmail.com

Abstract

Context: Anaemia is a global public health problem which occurs mostly in developing countries. The objective of this study is to assess the prevalence and risk factors associated with anaemia among children under five years of age in Lesotho.

Data and method: The logistic regression model was used to analyse the Lesotho Demographic and Health Survey data for 2009 and 2014.

Findings: The results from the 2009 data set revealed that the nutritional (stunting) status of child, child's age and mother's anaemia status were the risk factors associated to childhood anaemia, whereas the findings from the 2014 data set showed that the nutritional status of child, whether the child had a fever in the last two weeks prior to the survey, child's age and mother's body mass index were risk factors associated with anaemia among children under five years.

Conclusion: There is a need to improve the child health at an early age and nutritional status.

Keywords: Anaemia, Children, Logistic regression; LDHS.

Introduction

Anaemia remains a public health problem in many countries. It is most prevalent among children under five years and pregnant women, especially in developing countries (Gorospe et al. 2014). According to the World Health Organisation (WHO 2015), anaemia is defined as low blood hemoglobin concentration. The hemoglobin concentration level, when it is below 7g/dl, can lead to death if there is no treatment provided on time (Ngesa and Mwambi 2014). Anaemia in early childhood can negatively affect mental development, performance in school, and physical and behavioural growth (McCann and Ames 2007; Kotecha 2011). Half of the micronutrients that contribute to the global burden of anaemia is iron deficiency (Benoist et al. 2008). However, there are small numbers of other micronutrients such as folate, vitamins A and B12 that also contribute. Parasitic infections, such as malaria, filariasis and chronic diarrhea, can also result in anaemia (Benoist et al. 2008). The estimates suggested that 47.4% of children globally under five years of age are anaemic (McLean et al. 2009). In Lesotho, the 2009 and 2014 nationally representative survey showed that 47% and 51% of the children under five years are anaemic; while in men 12% and 14% are anaemic; and in women 26% and 27% are anaemic (Population and ICF 2011; MOHSW and ICF 2010; MOHSW and ICF 2016). This shows that anaemia has been increasing in Lesotho, especially in children, and more focus is needed on improvement in children's health as their health is the future of the country (Isiugo-Abanihe and Oke, 2011). More research about anaemia under five years has been done in different countries (Kuziga et al. 2017; Kisiangani et al. 2015). It has been shown that the prevalence of anaemia might not be necessarily the same in each country, as there is variability in accessing health care and food with enough micronutrients across different countries. Currently, there is not much research on anaemia in Lesotho; only few studies such as (Makonnen et al. 2003; Oguntibeju 2003; Yasutake et al. 2013; Mugomeri et al. 2016) among others have been carried out. However, of all of these studies none was done on children under five years nationwide.

In addition, most of the studies done on anemia (Adegoke et al. 2012; Dey and Raheem 2016; Gorospe et al. 2014) used logistic regression and this model is very powerful when for instance the assumption of independence of observation is not violated. But when the data comes from a complex survey design, the measurements from the same cluster may be correlated and thus violate the assumption of independence. Therefore, the present study addressed this problem via survey logistic regression that accounts for the complexity of

sampling design, and clustering and possible correlation between observations from the same cluster. According to our understanding there was no study in the literature using the survey logistic regression to identify the risk factors associated with childhood anaemia in Lesotho nationwide.

Material and methods

Study area

Lesotho is a hilly landlocked country which is completely surrounded by South Africa and is situated between 28° and 30° south, and between 27° and 30° east (Moteetee 2005). The country covers an area of around 30,355 square kilometers with an estimated population of 2.203821 million. It has ten politico-administrative districts where the Maseru district is the capital city. Lesotho is an established government with the King as the Head of State and the Prime Minster as Head of Government and a double law framework comprising of customary law and common law. The country faces a severe HIV and AIDS problem, widespread poverty, high unemployment, food insecurity and other diseases which affect maternal health and nutritional status adversely (MOHSW and ICF 2016).

Data source

The study used the 2009 and 2014 Lesotho Demography and Health Survey (LDHS). The survey used a sampling frame from the 2006 population and housing census supplied by Lesotho Bureau of Statistics. The data was collected based on multistage sampling technique with stratification. In the first stage 400 enumeration areas were selected (282 rural and 118 urban). In the second stage, systematic sampling was used among the selected households, where 25 households were selected in each selected enumeration area or cluster. More details on sampling techniques on the 2009 and 2014 data collection can be found in MOHSW and ICF (2010); and MOHSW and ICF (2016). The 2009 and 2014 LDHS provided the children's data set and this was used in the current study. A weighted total of 1295 children were obtained from 2009 LDHS and 1138 from 2014 LDHS.

We have used 2009 and 2014 LDHS in order to assess whether the determinants of anemia in Lesotho remained the same or differed in these two periods. In addition, we wanted to know which of the explanatory variables was significant in both periods, so that more attention can be given to policy making and underwriting.

Data analysis

Dependent variable

The outcome variable of interest in this study was anaemia status among children under five years. The anaemia status among children is classified mainly based on hemoglobin concentration level in blood measured in grams per deciliter (g/dl). A child is considered as anaemic if his/her hemoglobin concentration level adjusted for altitude is less than 11.0 g/dl otherwise it is not considered anaemic (WHO 2015).

Independent variables

The independent variables used in this study were also used in various similar studies on childhood anaemia in other areas (Benoist et al. 2008; McLean et al. 2009; Kotecha 2011; Habyarimana et al. 2016; Habyarimana et al. 2017) among others. Consequently, this forms the theoretical framework that will underpin the current research. These independent variables are grouped as socio-economic and demographic factors which include: sex of child (female, male); whether the child had fever, or not; cough or diarrhea in the two weeks prior to the survey or not; the birth order of the child; the child's birth weight; the child's age in months; whether the child had received drugs for intestinal worms or vitamin A supplementation in the six months prior to the survey or not; child nutritional status (underweight or not, stunted or not and wasted or not); mother's anaemia status; mother's education level (no education, primary, secondary and more); mother's age at birth; mother's body mass index; place of residence; and wealth index of the household.

Statistical analysis

The data used in this study was collected based on multi-stage sampling, stratified and cluster sampling with unequal probability of selection. Heering (2010) advised the necessity of accounting for the complexity of sampling design in order to avoid underestimation of the variance and also avoid wrong inferences.

Therefore, the current study used the survey logistic regression that accounted for complexity of sampling design and heterogeneity between observations from the same cluster inherent in DHS data.

Model formulation

Let y_{ikn} denote the anaemia status of child i from k^{th} stratum and n^{th} cluster. The outcome variable is defined as a dichotomous variable such that $y_{ikn} = 1$ if the child *i* is anaemic and $y_{ikn} = 0$ if the child *i* is not anaemics. In the present study, we have assumed

that outcome variable y_{ikn} is Bernoulli distributed as $y_{ikn}|\mu_{ikn} \sim Bernoulli\ (\mu_{ikn})$, with μ_{ikn} known as the mean and is defined as $E(y_{ikn}) = \mu_{ikn}$. It is linked to the independent variables as :

$$g(\mu_{ikn}) = X'_{ikn}\beta$$

where g (.) is the logit link function, β is a m-dimensional vector of categorical explanatory variables.

Data analysis

The present study used bivariate techniques to identify the association between potential explanatory variables and childhood anaemia. In the bivariate analysis, cross-tabulation technique was used for both 2009 and 2014 LDHS data. The p-values and Pearson's chi-square test were used to check whether the independent variables are significantly associated with anaemia or not. This was done using SPSS version 24.0. Any variables from bivariate results with p-value less than 0.2 were included in the analysis of multivariate survey logistic regression. This was done in order to account for any possible multicollinearity and confounding between the covariate (Schneider et al. 2008; Gari et al. 2017). Thereafter, any variable that was significant at 5% level of significance was reported in multivariate analysis in both 2009 and 2014 data sets. The analysis was done using Proc Survey logistic from SAS software version 9.4 and the model fit statistics was assessed based on Akaike information criteria (AIC) and -2 Log-Likelihood (-2LogL).

Results

The current study used the survey weights provided by the Lesotho Demographic and Health Survey data set in order to ascertain a national level representation. The prevalence of anaemia among children under five years in Lesotho was 47% and 51% in 2009 and 2014 respectively. In bivariate analysis in 2009, the results from cross-tabulation analysis are summarized in Table 1. The Pearson Chi-squared test was used to assess the association between the explanatory variables and childhood anaemia. It was found that the child's age, the incidence of cough in the last two weeks prior to the survey, stunting, children who received vitamin A and the mother anaemia status were significantly associated with childhood anaemia. The child's age group was significantly associated with childhood anemia (p-value < 0.001). The prevalence of anaemia was 62.8%, 45.4% and 38.6% among children aged between 0-19; 20-39 and 40-59 months, respectively. It was also found that the incidence of cough was significantly associated with childhood anaemia (p-value=0.018). The prevalence of anaemia was 53.7% among children who had a cough and 46.6% among children who did not coughing. It was observed that the mother's anaemia level was significantly associated with childhood anaemia (p-value=0.008). The prevalence of anaemia was 46.8% among children from anaemic mothers and 55.4% among children from none anaemic mothers. It was also observed that the stunting children were significantly associated with childhood anaemia (p-value=0.047). The prevalence of anaemia was 46.6% among stunted children and 51.9% among non-stunted children. The prevalence of anaemia among children who received vitamin A was 46.5% and 54.9% among children who did not receive vitamin A.

Table 1: Childhood anaemia by categorical variables 2009

Variable	Category	Anaemic	Not anaemic	p-value
Sex of the child	male	316 (48.8%)	331 (51.2%)	
	female	316 (48.8%)	332 (51.2%)	0.978
Birth order	1st	248 (51.5%)	234(48.5%)	
	2-3	253 (47.6%)	278(52.4%)	
	4-5	92 (46.9%)	104(53.1%)	
	6	39 (45.3%)	47 (54.7%)	0.508
Child's birth weight	< (2500g)	42 (48.8%)	44 (52.2%)	
	≥ (2500g)	403 (49.4%)	412 (50.6%)	0.914
Child's age in months	0-19	263 (62.8%)	156 (37.2%)	
	20-39	207 (45.4%)	249 (54.6%)	
	40-59	162 (38.6%)	258 (61.4%)	0.000
Had fever in the last two weeks	yes	113 (46.5%)	130 (53.5%)	
	no	517(49.2%)	533 (50.8%)	0.442
Had coughing in the last two weeks	yes	209 (53.7%)	180 (46.3%)	
	no	421 (46.6%)	483 (53.4%)	0.018

Had diarrhea in the last two weeks	yes	91 (55.2%)	74 (44.8%)	
	no	539 (47.9%)	587 (52.1%)	0.080
Had received drugs for intestinal worms	yes	61 (47.7%)	67 (52.3%)	
	no	561 (49.4%)	575 (50.6%)	0.711
Received vitamin A	yes	185 (46.5%)	213 (53.5%)	
	no	335 (54.9%)	275 (45.1%)	0.009
Stunted	yes	386 (46.6%)	443 (53.4%)	
	no	216 (51.9%)	200 (48.1%)	0.047
Wasting	yes	22(53.7%)	19 (46.3%)	
	no	579 (48.1%)	624 (51.9%)	0.486
Underweight	yes	107 (47.6%)	118 (52.4%)	
	no	494 (48.5%)	525(51.5%)	0.802
Mother's BMI	<18.5	25 (45.5%)	30 (54.5%)	
	≥18.5	600 (48.9%)	628 (51.1%)	0.621
Mother's education level	no education	14 (43.8%)	18 (56.2%)	
	primary	356 (48.4%)	379 (51.6%)	
	secondary & higher	261 (49.4%)	267 (50.6%)	0.800
Mother's anaemia level	anaemic	461 (46.8%)	525 (53.2%)	
	not anaemic	168 (55.4%)	135 (44.6%)	0.008
Mother's age at birth	<25 years	261 (51.5%)	246 (48.5%)	
	≥25 years	371(47.1%)	417 (52.9%)	0.122
Wealth index	poor	302 (50.7%)	294(49.3%)	
	middle	124 (47.3%)	138 (52.7%)	
	rich	206 (47.2%)	230 (52.8%)	0.477
Place of residence	urban	509 (47.9%)	553 (52.1%)	
	rural	123 (52.8%)	110 (47.2%)	0.179

Table 2 showed that the prevalence of anaemia among children was 51% in 2014. It was observed that the child age was significantly associated with childhood anaemia (p-value<.0001). The prevalence of anaemia was 62.7%, 55.0% and 43.6% among children aged from 0-19, 20-39 and 40-59 months respectively. It was also found that the incidence of fever was significantly associated with childhood anaemia (p-value=0.008). The prevalence of anaemia was 52.3% among children who had fever and 62.8% among children who did not have fever. It was observed that the incidence of diarrhea was significantly associated with childhood anaemia (p-value=0.024). The prevalence of anaemia was 62.7% among children who had diarrhea in the two weeks prior to the survey and 52.8% among children who did not have diarrhea in the two weeks prior to the survey. It was also observed that the stunting children were significantly associated with childhood anaemia (p-value=0.002). The prevalence of anaemia was 49.7% among stunted children and 64.5% among non-stunted children. The prevalence of anaemia was 62.0% among children who received drugs for intestinal worms and 51.5% among children who did not received the drugs for intestinal worms. Mother's age at birth was significantly associated with childhood anaemia (p-value= 0.011). The prevalence of anaemia was 58.9% among mothers who were less than 25 years old and 51.2% among those who were greater or equal to 25 years old. The mother's body mass index was significantly associated with childhood anaemia (p-value =0.001). The prevalence of anaemia was 51.7% among mothers whose body mass index was less than 18.5 and 64.9% among mothers whose body mass index was greater or equal to 18.5.

Table 2. Childhood anaemia by categorical variables 2014

Variable	Category	Anaemic	Not anaemic	p-value
Sex of the child	male	301 (56.2%)	235(43.8%)	
	female	315 (52.2%)	288 (47.8%)	0.185
Birth order	1st	230 (52.6%)	207 (47.4%)	
	2-3	260 (55.7%)	207(44.3%)	
	4-5	81 (51.3%)	77(48.7%)	
	6	46 (51.1%)	44 (48.9%)	0.673
Child's birth weight	< (2500g)	45 (58.4%)	32 (41.6%)	
	≥(2500g)	568 (53.7%)	489 (46.3%)	0.424
Child's age in months	0-19	227 (62.7%)	135 (37.3%)	
	20-39	224 (55.0%)	183 (45.0%)	0.000
	40-59	160 (43.6%)	207 (56.4%)	
Had fever in the last two	yes	493 (52.3%)	449 (47.7%)	
weeks	no	120(62.8%)	71 (37.2%)	0.008
Had coughing in the last two	yes	184 (54.4%)	154 (45.6%)	
weeks	no	429 (54.0%)	366 (46.0%)	0.883
Had diarrhea in the last two	yes	94 (62.7%)	56 (37.3%)	
weeks	no	520 (52.8%)	465 (47.2%)	0.024
Had received drugs for	yes	152 (62.0%)	93 (38.0%)	
intestinal worms	no	439 (51.5%)	413 (48.5%)	0.004
Received Vitamin A	yes	363 (54.3%)	306 (45.7%)	
	no	131 (53.7%)	113 (46.3%)	0.878
Stunted	yes	392 (49.7%)	397 (50.3%)	
	no	213 (64.5%)	117 (35.5%)	0.000
Wasting	yes	592 (54.3%)	499 (45.7%)	
	no	14 (48.3%)	15 (51.7%)	0.523

Underweight	yes	499 (53.0%)	443 (47.0%)	
	no	107 (60.5%)	70 (39.5%)	0.067
Mother's BMI	<18.5	467 (51.7%)	437 (48.3%)	
	≥18.5	111 (64.9%)	60 (35.1%)	0.001
Mother's education level	no education	7 (53.8%)	6 (46.2%)	
	primary	508 (53.2%)	447 (46.8%)	
	secondary & higher	99 (58.6%)	70 (41.4%)	0.432
Mother's anaemia level	anaemic	173 (57.3%)	129 (42.7%)	
	not anaemic	439 (52.8%)	393 (47.2%)	0.177
Mother's age at birth	<25 years	247 (58.9%)	172 (41.1%)	
	≥25 years	368 (51.2%)	351(48.8%)	0.011
Wealth index	poor	147 (58.3%)	105 (41.7%)	
	middle	259 (52.2%)	210 (44.8%)	
	rich	209 (50.1%)	208 (49.9%)	0.095
Place of residence	urban	158 (50.2%)	157 (49.8%)	
	rural	457 (55.5%)	366 (44.5%)	0.104

Multivariate analysis

The results from survey logistic regression in both 2009 and 2014 showed that the model with intercept only has AIC (1.71175E9) and (1.45661E9) respectively, while the full model has AIC (1.64201E9) and (1.39001E9). Therefore, the full model was considered as the best fit model since it is the one with smallest AIC. Table 3 presents the parameter estimate, standard deviation (STD), P-value, odds ratio (OR) and confidence interval (CI). The results from the 2009 data set showed that stunting child, the mother's anaemia status and the child's age were significantly associated with anaemia at 5% level of significance. A child whose age was less than 19 months was 2.994 (OR: 2.994 (2.161; 4.147), p-value<.0001) times more likely to be anaemic as compared to those in age group 40-59 months while a child between 20-39 months of age was 1.366 (OR: 1.366 (1.016; 1.836), p-value=0.039) times more likely to be anaemic than those who were in the age group of 40-59 months.

The nutritional status of the child was significantly associated with childhood anaemia. A stunted child was 1.416 (OR: 1.416 (1.081; 1.854), p-value=0.012) times more likely to be anaemic than a non-stunted child. It is observed from the results that mother's anaemia status had a significant impact of childhood anaemia. A child born to a non-anaemic mother was 0.688(OR: 0.688 (0.514; 0.922), p-value=0.012) times less likely to be anaemic than a child born to an anaemic mother. The 2014 results revealed that the child's age group, his/her nutritional status (stunting), fever in the last two weeks prior to the survey and the mother's body mass index were significantly associated with anaemia among children under five years in Lesotho. A child whose age is less than 19 months was found to be 0.471 (OR: 0.471 (0.323; 0.687)), p-value=0.000) times less likely to be anaemic as compared to those in the age group of

40-50 months while a child between 20-39 months of age was 0.687 (OR: 0.687 (0.482; 0.977), p-value=0.037) times less likely to be anaemic than those who were in the age group of 40-59 months. A child who had fever in the last two weeks prior to the survey was 1.674 (OR: 1.674 (1.103; 2.540), p-value= 0.016) times more likely to be anaemic than a child who did not have fever in the last two weeks prior to the survey. It was noted that a stunted child was 1.787 (OR: 1.787 (1.219; 2.619)), p-value=0.003) times more likely to be anaemic than a child who was not stunted. It was also noted that mother's BMI was significantly associated with childhood anaemia. A child born to an underweight mother was 1.542 (OR: 1.542 (1.024; 2.321), p-value=0.038) times more likely to be anaemic compared to a child born to a normal weight or obese mother.

Table 3: Multivariate survey logistic regression analysis.

Variable	Estimate	STD	p-value	OR	95% CI
Year: 2009					
Intercept	-0.363	0.160	0.024	--	--
Stunting(ref=no)					
yes	0.348	0.137	0.012	1.146	1.081;1.854
Mother with anaemia (ref=no)					
yes	-0.373	0.148	0.012	0.688	0.514;0.922
Child age (ref=40-59)					
0-19	1.097	0.166	0.000	2.994	2.161;4.417
20-39	0.312	0.152	0.039	1.366	1.016;1.836
Year:2014					
Intercept	-0.474	0.609	0.437	--	--
Stunting (ref=no)					
yes	0.581	0.194	0.003	1.787	1.219;2.619
Fever (ref=no)					
Yes	0.515	0.212	0.016	1.674	1.103;2.540
BMI (ref=less than18.5) more or equal 18.5	0.433	0.208	0.038	1.542	1.024;2.321
Child age (ref=40-59)					
0-19	-0.752	0.191	0.000	0.471	0.323;0.687
20-39	-0.376	0.179	0.037	0.687	0.482;0.977

Discussion

The prevalence of anaemia among children under five years in Lesotho was 47% and 51% in 2009 and 2014 respectively. This shows an increase of 4% in the period of five years. The main objective of the current study was to assess the risk factors of anaemia among children under five years based on the 2009 and 2014 LDHS data sets. The findings from this study revealed that childhood anaemia in Lesotho decreases with increasing the age of the child in both the 2009 and 2014 analyses. These findings are consistent with many previous studies such as Kotecha (2011); Zhao et al. (2012); and Dey and Raheem (2016). This means that older children have a lower likelihood of being anaemic than younger children.

The findings from this study also highlighted the association of stunting with anaemia among children under five years in 2009 and 2014. This finding was found elsewhere (Kotecha 2011; Zhao et al. 2012; Leite et al. 2013; Khan et al. 2016). This is an indication of chronic malnutrition which might cause iron deficiency (Woldie et al. 2015).

The findings from the present study pointed out a significant association between mother's body mass index and childhood anaemia only in the 2014 data set. The results showed that children born to an underweight mother had a higher likelihood of being anaemic whilst normal, overweight and obese mothers are less likely to have children susceptible to anaemia. A similar result was found in studies by Habyarimana et al. (2017) and Fonseca et al. (2016). This may be due to the fact that underweight people are more likely to have other associated co-morbidity illnesses that may be associated with anaemia.

It was also found from the 2009 results that mother's anaemia status was significantly associated with childhood anaemia. The results showed that a

child born to anaemic mothers had a higher likelihood of being anaemic than a child born to a non-anaemic mother. This finding was also found in other similar studies (Ngnie-Teta et al. 2007; Yang et al. 2012; Pita et al. 2014).

In addition, the findings from the analysis of the 2014 data set demonstrated that the recent incidence of fever had a positive impact on childhood anaemia. The results showed that a child who had fever in the last two weeks prior to the survey had a higher likelihood of being anaemic than a child who did not have fever. This is in line with results found in studies by Santos et al. (2011); Konstantyner et al. (2012) and Gayawan et al. (2014), among others. This may be due to the fact that fever is commonly accompanied by a number of diseases and morbidity that are known to positively affect anaemia such as diarrhea, cough and malaria among others (Konstantyner et al. 2012). The findings from this study highlighted that the child's age and stunting were common risk factors of childhood anaemia in both studies (2009 and 2014). It was also found that fever and mother's body mass index were significant factors associated to childhood anaemia in 2014. In addition, mother's anaemia status was not statistically significant in 2014. A possible reason could be that in a period of 5 years the mothers have begun to practice a more healthy diet that reduces their risk of anaemia.

In the current bivariate study, a statistical association between childhood anaemia and place of residence, cough and diarrhea in the last two weeks prior to the survey, wasting and vitamin A supplementation was found but the study did not find any statistical significance in multivariate survey logistic regression in both data sets.

Conclusion

The current study identified the risk factors associated with anaemia among children under five years in Lesotho from 2009 to 2014. The findings from the study revealed that stunting and child's age were significant determinants of childhood anemia in both surveys (2009 and 2014) and this suggests that much attention on child health at an early age and the improvement of nutritional status especially stunting of children under five years must still be addressed with diligence. The findings from the current study also revealed that the determinants of childhood anemia in Lesotho increased from 2009 data set to 2014 and the prevalence of anemia also increased by 4%. This could possibly be due to a shortage of educating the public of proper diet and eating habits. Consequently, more workshops, roadshows, paraphernalia and even social media could be used to further educate the adult population with respect to diet and eating habits. The findings from this study also suggest an improvement of of the incidence of fever. The findings of this study may assist public health institutions in Lesotho and policy makers to formulate preventative measures and design intervention strategies that target children under five years.

Limitation

This study used cross-sectional data from LDHS and this data may not be able to address the causality but rather association; therefore, a longitudinal study is suggested to solve this problem.

Acknowledgments

The authors acknowledge Ministry of Health [Lesotho] and ICF International for the data and HEARD for funding this research.

References

Adegoke, S., Ayansanwo, A., Oluwayemi, I. and Okeniyi, J., 2012. Determinants of mortality in Nigerian children with severe anaemia. *SAMJ: South African Medical Journal*, *102*(10), pp.807-810.

Benoist, B.D, McLean, E., Egll, I. and Cogswell, M (2008). "Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia." Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia.

Dey, S. and Raheem, E (2016). "Multilevel multinomial logistic regression model for identifying factors associated with anemia in children 6–59 months in northeastern states of India." Cogent Mathematics **3**(1): 1159798.

Fonseca, C., Marques, F., Robalo Nunes, A., Belo, A., Brilhante, D. and Cortez, J (2016). "Prevalence of anaemia and iron deficiency in Portugal: the EMPIRE study." Internal medicine journal **46**(4): 470-478.

Gari, T., Loha. E., Deressa, W., Solomon, T., Atsbeha, H., Assegid, M., Hailu, A. and Lindtjørn, B (2017). "Anaemia among children in a drought affected community in south-central Ethiopia." PloS one **12**(3): e0170898.

Gayawan, E., Arogundade, E.D. and Adebayo, S.B (2014). "Possible determinants and spatial patterns of anaemia among young children in Nigeria: a Bayesian semi-parametric modelling." International health **6**(1): 35-45.

Gorospe, J., Bismonte, S.A., Arcilla, R.G. and Gironella, G (2014). "Ordinal Logistic Regression Analyses on Anemia for Children Aged 6 Months to 5 Years Old in the Philippines."

Habyarimana, F., 2016. Key determinants of malnutrition of children under five years of age in

Rwanda: Simultaneous measurement of three anthropometric indices. *African Population Studies*, *30*(2).

Habyarimana, F., Zewotir, T. and Ramroop, S (2017). "Structured Additive Quantile Regression for Assessing the Determinants of Childhood Anemia in Rwanda." International journal of environmental research and public health **14**(6): 652.

Heering, S.G., West, B.T. and Berglund, P.A. Applied survey data analysis: Statistical in the social and behavioral sciences series. New York: Chapman & Hall/CRC 2010.

Isiugo-Abanihe, U.C. and Oke, O.A., 2011. Maternal and environmental factors influencing infant birth weight in Ibadan, Nigeria. *African Population Studies*, *25*(2).

Khan, J.R., Awan, N. and Misu, F (2016). "Determinants of anemia among 6–59 months aged children in Bangladesh: evidence from nationally representative data." BMC pediatrics **16**(1): 3.

Kisiangani, I., Mbakaya, C. and Makokha, A., 2015. Prevalence of anaemia and associated factors among preschool children (6-59 months) in Western Province, Kenya. *Public Health and Preventive Medicine*, *1*(1), pp.28-32.

Konstantyner, T., Roma Oliveira, T.C. and de Aguiar Carrazedo Taddei, J.A (2012). "Risk factors for anemia among Brazilian infants from the 2006 National Demographic Health Survey." **2012**.

Kotecha, P.V (2011). "Nutritional anemia in young children with focus on Asia and India." Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine **36**(1): 8.

Kuziga, F., Adoke, Y. and Wanyenze, R.K., 2017. Prevalence and factors associated with anaemia among children aged 6 to 59 months in Namutumba district, Uganda: a cross-sectional study.

Leite, M.S., Cardoso, A.M., Coimbra, C.E., Welch, J.R., Gugelmin, S.A., Lira, P.C.I., Horta, B.L., Santos, R.V. and Escobar, A.L (2013). "Prevalence of anemia and associated factors among indigenous children in Brazil: results from the First National Survey of Indigenous People's Health and Nutrition." Nutrition journal **12**(1): 69.

Makonnen, B., Venter, A. and Joubert, G (2003). "A randomised controlled study of the impact of dietary zinc supplementation in the management of children with protein–energy malnutrition in Lesotho. I: Mortality and Morbidity." Journal of tropical pediatrics **49**(6): 340-352.

McCann, J.C. and Ames, B.N (2007). "An overview of evidence for a causal relation between iron deficiency during development and deficits in cognitive or behavioral function–." The American journal of clinical nutrition **85**(4): 931-945.

McLean, E., Cogswell, M., Egli, I., Wojdyla, D. and De Benoist, B (2009). "Worldwide prevalence of anaemia, WHO vitamin and mineral nutrition information system, 1993–2005." Public health nutrition **12**(4): 444-454.

Ministry of Health and Social Welfare (MOHSW) [Lesotho] and ICF Macro (2010). *Lesotho Demographic and Health Survey 2009*. Maseru, Lesotho: MOHSW and ICF Macro.

Ministry of Health [Lesotho] and ICF International (2016). *Lesotho Demographic and Health Survey 2014*. Maseru, Lesotho: Ministry of Health and ICF International.

Moteetee, M.M., 2005. Report on the quality and responsiveness of health services to meet the needs of women in crisis settings. *WHO, June*, pp.1-8.

Mugomeri, E., Musa, N.K. and Chatanga, P (2016). "Heamoglobin response to routine iron and folate supplementation during pregnancy in an HIV-endemic rural area roma, Lesotho." Medical Technology SA **30**(1): 10-14.

Ngesa, O. and Mwambi, H (2014). "Prevalence and risk factors of anaemia among children aged between 6 months and 14 years in Kenya." PLoS One **9**(11): e113756.

Ngnie-Teta, I., Receveur, O. and Kuate-Defo, B (2007). "Risk factors for moderate to severe anemia among children in Benin and Mali: insights from a multilevel analysis." Food and Nutrition Bulletin **28**(1): 76-89.

Oguntibeju, O.O (2003). "Parasitic infestation and anaemia: The prevalence in a rural hospital setting." J Indian Acad Clin Med **4**: 210-212.

WHO (2015). "The global prevalence of anaemia in 2011."

Pita, G.M., Jiménez, S., Basabe, B., García, R.G., Macías, C., Selva, L., Hernández, C., Cruz, M., Herrera, R. and O'Farrill, R (2014). "Anemia in children under five years old in Eastern Cuba, 2005-2011." Medicc Review **16**(1): 16-23.

Population, C. and ICF, M (2011). "Lesotho 2009: results from the Demographic and Health Survey." Studies in family planning **42**(4): 305.

Santos, R.F.D., Gonzalez, E.S.C., Albuquerque, E.C.D., Arruda, I.K.G.D., Diniz, A.D.S., Figueroa, J.N. and Pereira, A.P.C (2011). "Prevalence of anemia in under five-year-old children in a children's hospital in Recife, Brazil." Revista brasileira de hematologia e hemoterapia **33**(2): 100-104.

Schneider, J.M., Fujii, M.L., Lamp, C.L., Lönnerdal, B., Dewey, K.G. and Zidenberg-Cherr, S (2008).

"The use of multiple logistic regression to identify risk factors associated with anemia and iron deficiency in a convenience sample of 12–36-mo-old children from low-income families–." The American journal of clinical nutrition **87**(3): 614-620.

Woldie, H., Kebede, Y. and Tariku, A (2015). "Factors associated with anemia among children aged 6–23 months attending growth monitoring at Tsitsika Health Center, Wag-Himra Zone, Northeast Ethiopia." Journal of nutrition and metabolism **2015**.

Yang, W., Li, X., Li, Y., Zhang, S., Liu, L., Wang, X. and Li, W (2012). "Anemia, malnutrition and their correlations with socio-demographic characteristics and feeding practices among infants aged 0–18 months in rural areas of Shaanxi province in northwestern China: a cross-sectional study." BMC Public Health **12**(1): 1127.

Yasutake, S., He, H., Decker, M.R., Sonenstein, F.L. and Astone, N.M (2013). "Anemia among adolescent and young women in low-and-middle-income countries." International Journal of Child Health and Nutrition **2**(2): 105-112.

Zhao, A., Zhang, Y., Peng, Y., Li, J., Yang, T., Liu, Z., Lv, Y. and Wang P (2012). "Prevalence of anemia and its risk factors among children 6–36 months old in Burma." The American journal of tropical medicine and hygiene **87**(2): 306-311.

Heliyon 6 (2020) e03946

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Prevalence of and factors associated with malaria in children under five years of age in Malawi, using malaria indicator survey data

Rugiranka Tony Gaston *, Shaun Ramroop

School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Pietermaritzburg Campus, Private Bag X01, Scottsville, 3209, South Africa

ARTICLE INFO

Keywords:
Malaria
Children under five years of age
GAMM
RDT
MMIS
Public heath
Infectious disease
Nutrition

ABSTRACT

Background: Malaria remains a public health problem in developing countries and Malawi is no different. Although there has been an improvement in reducing malaria in Malawi, it remains a problem, especially in children less than five years old. The primary objective of the study was to assess whether socio-economic, geographic and demographic factors are associated with malaria, using the generalized additive mixed model (GAMM).

Data and methodology: The study used a 2017 dataset from the Malawi Malaria Indicator Survey (MMI) with a total number of 2724 children under five years old. The study also utilized the GAMM to analyze data. The outcome was that either the child had malaria or did not, as detected using the malaria Rapid Diagnostic Test (RDT) (Ayele et al., 2014a).

Results: In this study, more than 37 % of the total number of children who were tested showed a positive malaria result. In addition, the results from this study using GAMM indicated that anaemia, mother's education level, wealth index, child's age, the altitude of the place of residence, region, place of residence, toilet facility and electricity were significantly associated with a positive malaria RDT.

Conclusion: The study revealed that socio-economic, geographical and demographic variables are the key factors in improving malaria vectors in children. Improving income levels and supporting the poorer rural community mostly from the Central Region would be a great achievement in reducing malaria vectors in Malawi. In addition, improving health care in rural areas, especially at higher altitudes, would contribute to controlling malaria and reducing anaemia.

1. Introduction

Malaria is a serious disease caused by a protozoan parasite called *Plasmodium* species and the most dangerous is *P. falciparum* which occurs mainly in Africa. There are four other different parasites which are not as dangerous as *P. falciparum*, these being *P. vivax, P. ovale, P. malariae* and *P. knowlesi* (WHO, 2015). *Plasmodium falciparum* is the most prevalent parasite in Africa especially in Sub Saharan Africa and is estimated to be the cause of 99% of malaria in 2016, while outside of Africa *P. vivax* is the most predominant vector (WHO, 2017).

The parasites that cause malaria are transmitted to humans through the bite of the female *Anopheles* mosquito and it takes 10 15 days to develop symptoms of the disease after being infected (Perkins et al., 2011). Malaria is more often transmitted during the rainy season and with higher temperatures. The disease mostly affects the poorest coun tries (Mhalu, 2005; Chirombo et al., 2014; Cibulskis et al., 2016). Malaria is not contagious; however, it is possible to contract the disease from another person through blood transfusions or organ transplants (WHO, 2016).

Despite the interventions and precautions taken against malaria, the disease remains a major health problem worldwide, especially in developing countries (WHO, 2016). Malaria is among the leading causes of morbidity and mortality especially in the Sub Saharan Africa countries and the most vulnerable are pregnant women and children (WHO, 2013; Semakula et al., 2016). The global estimate of malaria cases was 237 million, 211 million and 216 million in 2010, 2015 and 2016 respec tively, with Africa having the most cases each year (WHO, 2016). This shows that there has been a reduction in malaria cases generally, but with, a slight increase globally between 2015 and 2016.

The estimated number of deaths from malaria globally was 655 000, 446 000 and 445 000 in 2010, 2015 and 2016 respectively. It was esti mated that approximately 80% of these deaths were from Sub Saharan

* Corresponding author.
E-mail address: rugiratony@gmail.com (R.T. Gaston).

https://doi.org/10.1016/j.heliyon.2020.e03946
Received 15 November 2019; Received in revised form 10 March 2020; Accepted 5 May 2020

African countries in 2015, while in 2016 the figure was estimated as 91%. Moreover, 70% of all those who died in 2016 were children under five years old (WHO, 2015; WHO, 2016).

Malaria was the fourth highest cause of mortality in Sub Saharan Africa, accounting for 10% of children's deaths. This is the equivalent of approximately one child in Sub Saharan Africa dying of malaria every two minutes (WHO, 2015). In Malawi, as one of the Sub Saharan Africa countries, where malaria is endemic throughout the country, the disease remains a health problem. In this country, malaria most often affects individuals who live in the rural areas, which are hotter, wetter, more humid and tend to be low lying, rather than those who live in the dry, urban, highland areas (Kazembe et al., 2006; Dzinjalamala, 2009; Chi rombo et al., 2014; Kazembe and Mathanga, 2016).

In 2013, malaria in Malawi was the leading cause of hospital ad missions and death in children under five years of age and pregnant women. The disease accounted for 20% of all deaths of children under five (WHO, 2016). The prevalence of malaria in Malawi has decreased from 43% in 2010 to 33% in 2014 and 24% in 2018 (NMCP and ICF, 2010; NMCP and ICF, 2014; NMCP and ICF, 2018). This reduction was due to the efforts of the Malawian government and the international sponsors who put more resources into fighting the burden of malaria among children under five years of age (Mathanga et al., 2012). In 2010, the Malawian government introduced indoor residual spraying (IRS) in various districts around the country as one of the methods of reducing malaria. In 2012 to 2014, the government of Malawi and the sponsors distributed free long lasting insecticide treated nets (LLIN) to the chil dren and pregnant women in the whole country (Mathanga et al., 2012; Chanda et al., 2016). Although there has been a great reduction in ma laria in Malawi, the disease remains a health problem, especially in children under five (NMCP and ICF, 2012; NMCP and ICF, 2018).

The study by Zgambo et al. (2017) used the 2012 and 2014 Malawi Malaria Indicator Surveys (MMIS) to compare the prevalence of and factors associated with malaria parasitaemia in children under five years of age in Malawi. Their findings showed that the prevalence of malaria had increased from 28% in 2012 to 33% in 2014. This reveals that ma laria is still a health problem in the country and more research is needed using different methods to identify the risk factors associated with the disease, especially for children under five years of age in Malawi.

Post 2000 research about malaria in Malawi such as that by Buchwald et al. (2016); Chitunhu and Musenge (2016); Kazembe et al. (2006); Lazzerini et al. (2016); Hershey et al. (2017); Kabaghe et al. (2017); and Parvin et al. (2018) have used different parametric methods, such as logistic regression, generalized linear mixed models and other statistical models. These models are an amazing asset in modelling the relationship between the outcome variable and covariates. However, in numerous applications, this relationship between the outcome and some con founding covariates may have an unknown function form (Ayele et al., 2014a, b). Therefore, along these lines, with such kinds of parameters, it is very important to estimate non parametrically. This prompted an investigation into non parametric methods, which include semi parametric additive models. Thus, the current study used the generalized additive mixed model (GAMM) to overcome these challenges. The GAMM is an extension of the generalized additive model (GAM), which includes the random effect. The random effect is used to model the cor relation between observations. The GAM does not include random effect, only the model covariate effect (Wang, 1998; Lin and Zhang, 1999). The GAMM is also an extension of the generalized linear mixed model (GLMM) which is a parametric model introduced by Breslow and Clayton (1993). The GAMM enables the parametric fixed effects from GLMM to be modelled as a non parametric model using the additive smooth function (Hastie and Tibshirani, 1990).

According to our knowledge, researchers have not used the general ized additive mixed model (GAMM) nation wide to identify the factors associated with malaria in children under five years of age in Malawi. The study by Chirombo et al. (2014) used structured additive regression models, which include GAMM and the geo additive model in their research in Malawi for children under five years of age, using the 2010 MMIS data set. However, it must be stressed that the data set used is different from the 2017 MMIS data set used in the current study.

Therefore, this study aims to investigate the prevalence of and factors associated with malaria in children under five years old, using the MMIS for 2017 with the application of GAMM. The study also investigates the risk factors for malaria and whether or not they had remained the same after the study by Chirombo et al. (2014).

2. Methodology and material

2.1. Study area

Malawi is a Sub Saharan African country situated south of the equator and is bordered by Tanzania in the north and northeast; by Mozambique to the east and southwest; and Zambia to the west and northwest (NMCP and ICF, 2018). The total area is approximately 118 484 square kilo metres, in which 9 4276 square kilometres are land and the remaining area consists of Lake Malawi. The country is split into three regions and twenty eight districts. The Northern region has six districts, the Central region nine and the Southern region comprises thirteen districts (NMCP and ICF, 2018). Malawi has a tropical continental climate with sea in fluences, where the variation of rain and temperature depends on alti tude and proximity to Lake Malawi. The tropical climate is favourable for the breeding of *Anopheles Mosquitoes* and the breeding of *Anopheles* in creases in the rainy season from November to April. The weather in Malawi becomes cool and dry from May to August and the transmission of malaria is not as high as in the rainy season (Kazembe, 2007; NMCP and ICF, 2012). The economy of Malawi is based on agriculture and it is one of the poorest countries in the world. Healthcare is poor compared to other African countries (WHO, 2016; Team, 2018).

2.2. Data sources

The study used secondary data from the 2017 Malawi Malaria Indi cator Survey (MMIS) and was collected between 15 April and June 2017. The MMIS was implemented by the Malawi National Malaria Control Program (NMCP) through support from the President's Malaria Initiative (PMI). The United States Agency for International Development (USAID) provided financial support through the President's Malaria Initiative (PMI). They also funded the project by offering technical assistance in the implementation of population and health surveys as they do in countries worldwide (NMCP and ICF, 2018). The governing body of Malawi pro vided staff, office space and strategic help. Thereafter, the ICF provided technical support through the Demographic and Health Survey (DHS) program.

The 2017 MMIS data interviewed all residents or visitors who stayed in the selected households the night before the interview. The survey was population based on a household cluster survey. The data sampling followed the two stage sampling method. The first stage included a selection of 150 clusters from the enumeration areas (EAs) demarcated in the 2008 population and housing census. Out of the 150 clusters, 60 clusters were from urban and 90 from rural areas. The second stage sampling involved the systematic selection of a sample of 3 750 households. Out of the 3 750 households, 25 households were selected from each enumeration area (EA). All women aged 15 49 years who were living in or had visited the selected household the night before the survey were eligible to be interviewed. The children aged 6 59 months in these households were tested for malaria infection with the consent of their parents or guardians. The study used a total number of 2 724 children as a weighted sample in order to ascertain a national level representation (NMCP and ICF, 2018). The survey sample of this study is representative at the national and regional level, as well as for urban and rural areas.

In the sampling process, the number of women surveyed in each re gion should contribute to the size of the total sample in proportion to the

size of the region (NMCP and ICF, 2018). However, some regions may have small populations, and this unweighted distribution does not represent the exact population. To resolve this problem, regions with small populations are oversampled. Thus, the weighted sample used in the study to get statistics that are representative of the country and to account for the complex sample design from MMIS data set (NMCP and ICF, 2018). The SD Bioline Malaria Ag P.f/P, a rapid diagnostic test (RDT) was used to collect a blood sample from children's finger or heel prick. This test is appropriate in detecting the histidine rich protein II (HRP II), an antigen of Plasmodium falciparum and common Plasmodium lactate dehydrogenase (pLDH) of Plasmodium species in human blood (NMCP and ICF, 2018) The diagnostic test incorporates an expendable sample tool that comes in a standard package. A tiny volume of blood is caught on the instrument and placed in the well of a testing device. The RDTs for malaria offers the possibility to expand the arrangement of exact malaria diagnosis to the region where microscopy services are not accessible; for example in a remote area or after standard laboratory hours (Wongsrichanalai et al., 2007). Microscopic diagnosis does have some limitations such as insufficiently trained microscopists, lack of quality control; the chance of misdiagnosis because of low parasitaemia or blended diseases, and in some cases itis hard to determine the types of plasmodium (Ohrt et al., 2002). In the field, laboratory technicians were trained to use the RDT and results were available in 20 min. The children who tested positive were given medication by trained nurses according to the national guidelines (NMCP and ICF, 2018).

3. Data analysis

3.1. Dependent variable

The prevalence of malaria in children under the age of five years was detected using results from the RDT. Hence, the response variable (outcome of interest) was binary, where the child tested either positive (had malaria) or negative (did not have malaria).

3.2. Independent variables

The independent variables considered in this study include a number of socio economic, demographic, and environmental or geographic fac tors. The demographic factors associated with the malaria status in children were the age of the child, the gender of the child and family size. These variables of interest were collected at an individual level (Ayele et al., 2014a; Zgambo et al., 2017).

The socio economic variables were: type of place of residence; wealth quintile; mother's highest education level; source of drinking water; type of toilet facility; the main material of the walls, floors and roofs of the rooms; the total number of rooms inhabited; the total number of nets in dwellings; if mosquito nets were used for sleeping and if there had been antimalarial spraying. These were collected at the household level (Buchwald et al., 2016).

The environmental variables were the temperature; altitude; rainfall and humidity; the life cycle of the parasite in the mosquito, and the breeding and feeding habits of the vector (Bennett et al., 2013; Alegana et al., 2014).

4. Statistical analysis

The present study used bivariate procedures to show the association between independent variables and childhood malaria. The analysis for the bivariate method used cross tabulation techniques with an applica tion of SPSS version 2.50. The p value and Chi squared test were used to check whether the independent variables are significantly associated with childhood malaria or not. The variables from bivariate results with a p value less than 5% level of significance were included in multivariate GAMM analysis (Gaston et al., 2018).

4.1. Model formulation

The generalized additive model (GAM) is the same as the semi parametric additive model, which was developed by Hastie and Tib shirani (1986). The GAM is applied to the data to identify the relation ship between the response and covariates variables. The parametric models also have powerful tools for modelling the relationship between the response and predictors variables, when their assumptions are not violated. The parametric models in applications such as determining the relationship between the response and covariates variables, may have unknown functional form and are complicated (Ayele et al., 2014a). The unknown functions may lead to applications of semiparametric additive models, which are flexible to allow non normal error distributions. Furthermore, the semi parametric additive model relaxes the assumption of normality and linearity in linear regression (Lin and Zhang, 1999; Ayele et al., 2014a). The use of a semiparametric additive model may allow the response variable to be modelled with Poisson and binomial distribution. Moreover, the nonparametric models are flexible for modelling the continuous predictor variables. The GAM extends the generalized linear model (GLM) by allowing the predictor function to include the unspecified nonlinear function for some or all of the covariate variables (Hastie and Tibshirani, 1990). The linear form for the condi tional expectation expressed as follow:

$$(Y / X_1, x_2, \ldots, x_k) \quad \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k. \tag{1}$$

Eq. (1) was replaced with the additive form and hence, the general equation for GLM as $\sum_{i\ 1}^{n} x_i \beta_i$ becomes $\sum_{i\ 1}^{n} f(x_i)$ in GAM. Thus, the equation of GAM is written as follows:

$$g(\mu_i) \quad \theta X_i + f_i(x_{1i}) + f_i(x_{2i}) + f_i(x_{1i}) + \ldots + f_k(x_{ki}). \tag{2}$$

From Eq. (1), $\mu_i \quad E(Y_i)$ and X_β distributed some exponential family distribution, X_i is the designed matrix, while θ is the corresponding parameter vector and f_i are the smooth functions of covariates, while g (.) is the monotonic differentiable function (Wood, 2017). If there is no linear component in Eq. (2), the model is known as nonparametric, whilst the models whose predictors have both linear and unspecified nonlinear function are semiparametric. To estimate the parameters, the standard ized condition of the smooth functions f_i should be satisfied such that E $[f_i X_i] \quad 0$, apart from that, each function will have free constants (Hastie and Tibshirani, 1990).

When the data has repeated measurement or correlations, the model includes a random variable and this leads to the extension of GAM. Hence, GAM becomes the generalized additive mixed model (GAMM) in the same way as the generalized linear mixed models GLMM are an extension of GLM (Hastie and Tibshirani, 1990). The GAMM was intro duced by Breslow and Clayton (1993) to include the random effect in the GAM and model the correlation between the observations.

The equation of GAMM can be expressed as follows:

$$g(\mu_i) \quad \beta_i + f_i(x_{1i}) + f_i(x_{2i}) + \ldots + f_k(x_{ki}) + Z_i b, \tag{3}$$

where g (.) is monotonic differentiable link function, X_i $(1, x_{1i}, \ldots, x_{1k})'$ are n covariate associated with fixed effects and $q \times 1$ vector of covariates Z_i associated with random effects. Thus, the given $q \times 1$ vector of random effect b, the observations y_i are assumed to be conditionally independent with means $E(y_i / b) \quad \mu_i$ and variance, $(y_i / b) \quad \psi v(\mu_i)$, where v (.) is the specified variance function and ψ is a scale parameter. Moreover, $f_i(.)$ is a centred twice differentiable smooth function and the random effects b is assumed to be distributed as $N\{0, G(\rho)\}$ and ρ is a $c \times 1$ vector of variance components. In addition, when f_i is a linear function, the GAMM reduces to GLMM (Lin and Zhang, 1999).

For a specified variance component, θ the log likelihood function of (β, f_i, θ) is expressed by Lin and Zhang (1999) in the following equation:

$$expexp\,[l\{y;\beta_0,f_1(.),\ldots,f_k(.),\theta\}\propto]\,|G|^{1/2}\int exp\left\{\frac{1}{2\tau}exp\sum_{i\ 1}^{n}di(y;\mu_i)\right.$$

$$\left.\frac{1}{2}b'G^{\ 1}b\right\}db, \tag{4}$$

where $y_i \quad (y_1,\ldots,y_n)'$ and $d_i(y,\mu)\propto \quad 2\int_{y_i}^{\mu_i}\frac{y_i \quad \mu}{w(u)d_u}$ define the conditional deviance function of (β, f_i, θ) given $\boldsymbol{b}$. The statistical inference for GAMM on nonparametric function f_i requires the estimate of smoothing parameter τ and the inference on variance component θ. It is known that the smoothing spline estimators and linear mixed models have close connections (Green and Silverman, 1993; Wang, 1998; Lin and Zhang, 1999). Moreover, the natural cubic smoothing spline estimators of function f_i maximize the penalized log likelihood for the same given τ and θ and give the following equation:

$$expexp\,[l\{y;\ \beta_0,f_1(.),\ldots,f_k(.),\ \theta\}] \quad \frac{1}{2}\sum_{i\ 1}^{k}\tau_i\int_{s_i}^{t_i}f_i''x^2dx$$

$$l[y;\ \beta_0,f_1(.),\ldots,f_k,\ \theta] \quad \frac{1}{2}\sum_{i}^{k}\tau_i f_i' S_i f_i. \tag{5}$$

where the s_i and t_i indicate the range of i^{th} covariate and τ_i are the smoothing parameters that manage the trade off between goodness of fit and the smoothness of the estimated functions (Lin and Zhang, 1999; Ayele et al., 2014a). Moreover, $f_i(.)$ is an $r_j \times 1$ unknown vector of the values of $f_i(.)$, estimated at r_j ordered values of the x_{kj}, where $k \quad (1,\ldots,n)$ and s_j is the smoothing matrix (Green and Silverman, 1993). By using the matrix form, the GAMM given in Eq. (3), can be written as:

$$g(\mu_i) \quad 1\beta_1 + N_1f_1 + \ldots + N_pf_p + Z_ib, \tag{6}$$

where $g(\mu_i) \quad [g(\mu_1),\ g(\mu_2),\ldots,g(\mu_n)]$, $n\times 1$ the vector of ones, $N_i \quad m\times r$ matrix, such that, the k^{th} component of N_if_i is f_ix_{ki} and $Z \quad Z_1, Z_2,\ldots, Z_n$. In order to evaluate Eq. (7), the numerical integration is required. Additionally, to calculate the natural cubic smoothing spline estimators of f_i by maximizing Eq. (6) is sometimes complicated. Consequently, Lin and Zhang, (1999) resolved this problem by suggesting the double penalized quasi likelihood (DPQL) model as an alternative approach. Hence, the estimation of nonparametric function f_i can be obtained by applying double quasi likelihood. The function f_i is re parameterized in terms of β_i and a_k in one to one transformation as:

$$f_k \quad X_k^*\beta_k + \beta_k a_k, \tag{7}$$

where X_k^* is $r_k \times 1$ vector with the r_k centred and ordered distinct values of the x_{ki} $(k \quad 1,2,\ldots,n)$ and $\beta_k \ L(L_kL_k)^{-1}$ and L_k is an $r_k(r_k \quad 2)$ full rank

Table 1. Childhood malaria by categorical variables.

Variable	Category	Malaria (Positive)	Malaria (Negative)	p-value
Region	North	62 (20.8%)	236 (79.2%)	0.000
	Central	452 (39.8%)	683 (60.2%)	
	South	500 (38.7%)	791 (61.3%)	
Place of residence	Rural	987 (41.9%)	1369 (58.1%)	0.000
	Urban	27 (7.3%)	341 (92.7%)	
Wealth index	Poorer	611 (48.0%)	662 (52.0%)	0.000
	Middle	204 (39.6%)	311 (60.4%)	
	Richer	199 (21.3%)	737 (78.7%)	
Mother's education level	No education	152 (46.3%)	176 (53.7%)	0.000
	Primary	557 (37.5%)	930 (62.5%)	
	Secondary	61 (17.4%)	290 (82.6%)	
	Tertiary	1 (3.6%)	27 (96.4%)	
Child's age in months	6–23	200 (25.6%)	581 (74.4%)	0.000
	24–31	411 (40.6%)	602 (59.4%)	
	32–59	403 (43.4%)	526 (56.6%)	
Anemia level	Anemic	706 (45.6%)	843 (54.4%)	0.000
	Not anemic	307 (26.1%)	867 (73.9%)	
Sex of the child	Male	463 (36.5%)	804 (63.5%)	0.001
	Female	447 (35.8%)	800 (64.2%)	
Altitude	0–500metres	119 (34.3%)	228 (65.7%)	0.000
	501–1000m	492 (43.8%)	631 (56.2%)	
	>1000	403 (32.2%)	850 (67.8%)	
Toilet facility	Toilet with flush	88 (21.1%)	330 (78.9%)	0.000
	Pit latrine	832 (39.3%)	1287 (60.7%)	
	No facility	95 (50.8%)	92 (49.2%)	
Electricity	Yes	21 (7.3%)	265 (92.7%)	0.000
	No	992 (40.7%)	1445 (59.3%)	
Main roof material	Thatch/Palm leaf	686 (45.8%)	812 (54.2%)	0.000
	Corrugated &Metal	283 (25.8%)	816 (74.2%)	
	Stick & Mud	45 (35.7%)	81 (64.3%)	
Main wall material	Wood/Mud	331 (43.2%)	436 (56.8%)	0.000
	Bricks	467 (40.1%)	697 (59.9%)	
	Cement/block	216 (27.2%)	577 (72.8%)	
Main floor material	Earth/Sand	840 (43.5%)	1092 (56.5%)	0.000
	Mud block/wood	74 (39.4%)	114 (60.6%)	
	Cement/block	100 (16.6%)	504 (83.4%)	
Sleep under mosquito bed net	All children	553 (33.6%)	1093 (66.4%)	0.000
	Some	88 (34.1%)	170 (65.9%)	
	None	372 (45.4%)	447 (54.6%)	

matrix satisfying S_k LL' and $L'_k x^*_k$ 0. Thus, the double penalized quasi likelihood with respect to $(\beta_0 f_i)$ and b becomes:

$$\frac{1}{2\tau}\sum_{i\ 1}^{n} di(y;\mu_i) \quad \frac{1}{2}b' G^{\ 1} b \quad \frac{1}{2}a' D^{\ 1} a, \tag{8}$$

where $f'_k S_k f_k$ $a'_k a_k$, a $(a'_1, a'_2, \ldots, a'_k)$ and D $diag(\rho_1 I, \rho_2 I, \ldots, \rho_k I)$ with ρ_k $\frac{1}{\tau_k}$. Note that the small values of ρ_k $(\rho_{1,}\ \rho_{2,\ \ldots,}\rho_k)$ correspond to over smoothing (Breslow and Clayton, 1993; Lin and Zhang, 1999).

5. Results

The present study used survey weighted data in order to ascertain a national level presentation (NMCP and ICF, 2018). The results from cross tabulation analysis are summarized in Table 1. The results indi cated that all independent variables were significantly associated with childhood malaria (p value<0.05). The age of child and altitude were categorized, however, in multivariate GAMM are considered as contin uous. The results from Table 1 indicated that the prevalence of malaria was 25%, 40.6%, and 43.4% among children aged between 6 23, 24 41 and 42 59 months respectively. The prevalence of malaria in terms of the sex of the child was 36.5% for a male child and 35.8% for a female child.

It was observed that the prevalence of malaria was higher in the Central Region (39.8%), followed by the South Region (38.7) and lastly the North Region (20.8%). The results also showed that the prevalence of malaria was higher in children from poorer households (48.0%), middle income households (39.6%) and wealthy households (21.3%) respec tively. It was found that the prevalence of malaria among the children with anaemia (45.6%) was higher compared to those without anaemia (26.1%). The results also showed that the prevalence of malaria in chil dren living in rural areas is higher (41.9%) than in children living in urban areas (7.3%). The prevalence of malaria was highest in children from mothers with no education (46.3%), followed by those whose mothers had primary (37.5%), secondary (17.4) and tertiary education (3.6%) respectively.

5.1. Model fitting

The multivariate study used R software to analyze the data with the application of "mgcv" packages. The GAMM was used to model the effect of age and altitude non parametrically, while other covariates were used as parametric factors. These factors have a continuous effect and might have non linear relationships with malaria (Ayele et al., 2014a). The R software has packages with numerous choices for controlling the smoothness in the GAMM using splines. The various splines can be used such as the cubic smoothing splines, Bin smoothers, shrinkage smoothers, locally weighted running line smoothers, kernel smoothers, among others (Hastie and Tibshirani, 1990; Ruppert et al., 2003). However, this study used shrinkage smoothers (splines) to fit the GAM model, due to advantages such as assisting to control the knot placement. Furthermore, the shrinkage smoother is constructed in such a way that the smooth terms are rebuffed away all around (Wood, 2006). The study also considered the fundamental impact and possible two way interaction effect. The p value of the individual smooth term and the AIC of each model, together with the inference of smooth were analyzed. The se lection of the model was based on the smallest AIC, the higher value of degree of freedom and high statistical significance. Hence, the final model for this study is given in Eq. (9) as follows:

where, $g(\mu_{ij})$ is the logit link function, $\beta' s$ are the parametric regression coefficients, $f'_j s$ are centred smooth functions, while b_{oj} is the random effects, which can be written as $b_{oj} \sim N(0, G(\theta))$.

5.2. Interpretation of results

The results in model (9) are presented in Tables 2 and 3 and in Figure 1. Table 2 indicates the parameter estimates for the model, stan dard error, z value, odds ratio and p values. The study reported the variables with significant impact for malaria using a RDT such as anaemia, electricity, region, residence, wealth index, the toilet facilities and mother's education status. The study checked all possible in teractions. However, the two way interaction effect was not included, since it did not add any significant effect to the model with non significant p values.

Table 2 shows that the children with no anaemia were 0.233 times less likely to test positive for malaria using a RDT as compared with anaemic children. The results also revealed that the odds of positive malaria results in a RDT for children living in the Central Region were 1.936 times more likely than for those who lived in the North Region. Similarly, the odds of positive malaria in a RDT for children living in the South Region were 1.179 times more likely than for those who lived in the North Region. The children living in rural areas were 4.318 times more likely to test positive for malaria in terms of RDT results compared to those living in urban areas. The study also showed that the odds of positive malaria results in a RDT test for children from a household with no toilet facilities were 2.938 times more likely than those with flush toilets. Furthermore, the children from households with pit latrines were 1.389 times more likely to test positive for malaria in a RDT, compared to those with flush toilets. The results indicated that the children from the middle class were 0.743 times less likely to test positive for malaria using a RDT, compared to those from the poorer classes. In addition, the chil dren from the wealthier classes were 0.571 times less likely to test pos itive for malaria in a RDT than those from the poorer classes. Lastly, the results indicated that the odds of positive malaria in a RDT for children from households with access to electricity were 0.435 times less likely than those from households with no access to electricity. Table 3 shows that the age of a child and the altitude of their region of residence has a significant impact on malaria using a RDT.

The letter **S** in Table 3 represents the smooth term and the number in parentheses shows the estimated degree of freedom (**edf**). The test sta tistics for child age and altitude of the region of residence (22.340; 90.420 respectively) together with their p value (0.000; 0.000) shows that there is no linear trend associated either for child age or for altitude. This is confirmed in Figure 1, where the trend shows that the effect of malaria results in a RDT increases with age up to approximately 35 months and thereafter remains constant with no sign of decreasing. The same results indicate that the effect of malaria results in a RDT increases with the altitude of their region of residence up to approximately 750 m above sea level and above that starts decreasing.

6. Discussion

The present study utilized the generalized additive mixed model (GAMM) to investigate the risk factors associated with malaria using nationwide malaria survey data from Malawi. The previous studies had used the parametric models such as the generalized linear mixed model (GLMM) to analyze the malaria results using RDT (Ayele et al., 2013;

$$g(\mu_{ij}) \quad \beta_0 + \beta_1 Anemia_j + \beta_2 Wealt\_Index_j + \beta_3 Region_j + \beta_3 Toilet\_facility_j + \beta_5 Electricity_j + \beta_7 Education_J + \beta_8 Residence_j + f_1(Age_j) + f_2(Altitude_j) + b_{oj}, \tag{9}$$

Table 2. The parameter estimates and Odds ratio of GAM model of the main parametric models.

Variable	Estimate	Standard Error	z-value	Odds ratio	p-value
Intercept	-2.086	0.4620	-4.514	0.124	0.000
Anaemia (Ref Yes)	-	-	-	-	-
No	-1.455	0.152	-9.577	0.233	0.000
Region (Ref=North)	-	-	-	-	-
Central	0.66755	0.279	2.395	1.936	0.017
South	0.14260	0.317	0.450	1.179	0.653
Residence (Ref= Urban) Rural	1.463	0.293	4.994	4.318	0.000
Toilet facility (Ref=Flush toilet)	-	-	-	-	-
Pit Latrine	0.329	0.226	1.456	1.389	0.145
No facility	1.078	0.343	3.143	2.938	0.002
Wealth Index (Ref=Poorer)	-	-	-	-	-
Middle	-0.298	0.182	-1.631	0.743	0.103
Richer	-0.560	0.190	-2.956	0.571	0.003
Electricity (Ref=No) Yes	-0.832	0.317	-2.623	0.435	0.009
Mothers' education (Ref=No education)	-	-	-	-	-
Primary	-0.102	0.201	-0.508	0.903	0.611
Secondary	-0.541	0.273	-1.982	0.582	0.047
Tertiary	-2.274	1.251	-1.817	0.103	0.069

Table 3. Approximate significance of the smooth terms.

Source	Degree of freedom (Edf)	Chi squared	p-value
S (age)	2.423	90.420	0.000
S (altitude)	2.875	22.340	0.000

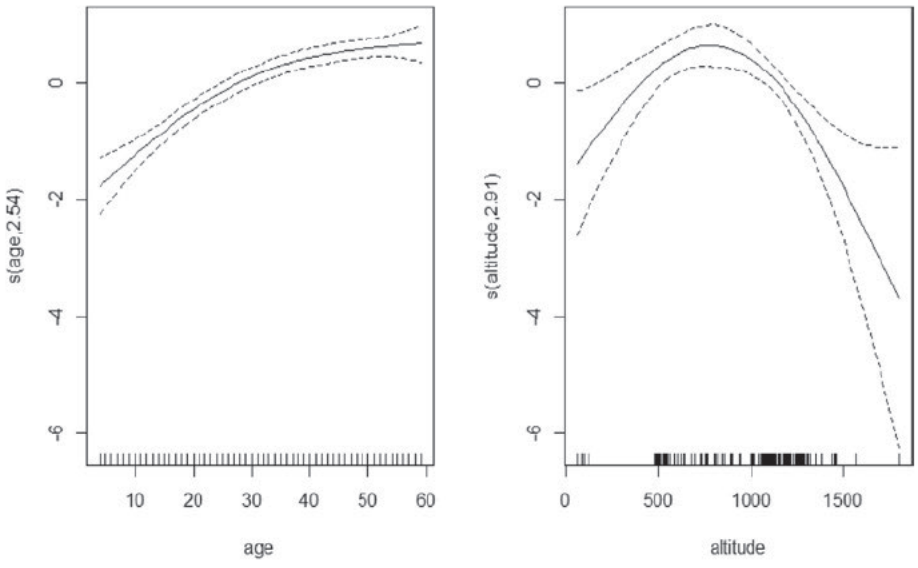

Figure 1. Smoothing components of malaria RDT test with age and altitude.

Roberts and Matthews, 2016). The parametric models are useful to model the relationship between a response variable and covariance. However, non parametric models are flexible enough to allow non normal error distributions, modelling continuous predictor variable, relaxes the assumption of normality and linearity in linear regression (Lin and Zhang, 1999). The parametric and non parametric models should com plement each other, and for this reason the combination of the two methods is more useful (Wu and Zhang, 2006). Thus, the study first used the parametric model to create geographical and social economic status variables such as type of place of residence, region, wealth quintile, mother's highest education level, type of toilet facility and availability of electricity.

The effect of age and altitude was modelled as non parametric and was statistically significant. The interaction effect was not included in the model as it was not statistically significant to improve the original model. The results from the parametric part revealed that the probability of increasing a positive malaria RDT was lower in the richer and the middle classes compared to the poorer class. These results confirmed that the prevalence of malaria is linked to socio economic conditions, where poorer people are more vulnerable (Hay et al., 2004; Chitunhu and

Musenge, 2016). This is due to the limited access to healthcare and the affordability of treatment (Worrall et al., 2002). The study revealed that the households with access to electricity are less likely to increase the positive malaria RDT rates. Moreover, the households with no toilet fa cilities are more likely to increase the positive malaria RDT rates. This shows that the households with access to electricity and toilet facilities can more easily access the healthcare and afford the treatment (Hay et al., 2004). Hence, these factors that are indicators of socio economic status are consistent with the study by Ayele et al. (2014a).

The results from the study also showed that the risk of malaria is lower in children from mothers with higher education. This might be linked to socio economic status, as educated individuals are more likely to have a better standard of living and better understand health related issues. Furthermore, the individuals with higher levels of education can more easily access healthcare and afford the use of mosquito nets, indoor residual spray and other preventive measures for malaria. These results were consistent with previous studies such as those by Snyman et al. (2015); Zgambo et al. (2017) and Sultana et al. (2017).

The study revealed that the households from rural areas have a higher prevalence of testing positive to malaria compared to those from urban areas. This may be explained by the individuals living in rural areas not having the same access to the many things that their urban counterparts have; such as proper formal houses, drinking water, education opportu nities, access to health care and so forth. The individuals living in rural areas often drink water from rivers which may attract mosquitoes, as river water is often dirty. Moreover, it takes longer to pass through the bush to reach these rivers, which may increase susceptibility to mosquito bites and therefore contracting malaria. In addition, most of the in dividuals living in rural areas live in poor housing conditions. A partic ular issue could be holes in the walls of the houses where the *Anopheles Mosquitoes* could enter the houses increasing the chance of malaria transmission through its bite (Lwetoijera et al., 2013; Jenkins et al., 2015; Kazembe and Mathanga, 2016; Sultana et al., 2017).

The study indicated that children without anaemia have a lower prevalence of testing positive to malaria compared to that of anaemic children. This might be explained by the link between anaemia and malaria, as has been shown by previous studies, such as those of Biemba et al. (2000) and Sultana et al. (2017).

The study indicates a large variation among the three regions, where the households from the Central Region being most likely to test positive for malaria. This is due to the fact that the region is covered by a large plain of land and the low lying zone along the lake. Moreover, the lake might be an area conducive to the breeding of malaria vectors (Mina kawa et al., 2012; Zgambo et al., 2017).

The results from the non parametric model indicate that the proba bility of a positive malaria RDT increases as the child's age increases. This could be due, in part, to the impact of maternal immunity in the child before one year of age. In addition, the children younger than one year old are more protected and well taken care of and this helps to fight any kind of disease. This reduces as the children get older. These results are consistent with the studies by Ayele et al. (2014a) and Chirombo et al. (2014).

The research reveals that the risk of having a positive malaria RDT result increases as the altitude increases, up to 750 m, and starts showing a decrease as altitude increases thereafter. This may be explained by the very high temperatures at lower altitudes as mosquitos develop in hotter areas. As the altitude increases, the temperatures decrease and this re duces the risk of having a positive malaria RDT result (Lindsay and Martens, 1998; Chirombo et al., 2014).

7. Conclusion

The aim of this study was to assess the prevalence of and factors associated with malaria in children under the age of five years in Malawi using GAMM. The current findings show that the government should consider other factors associated with malaria especially in children under five years of age; such as anaemia, region, residence type, toilet facilities, wealth index, the use of electricity, mothers' education, child ren's age and the altitude of the region of residence.

The findings from this study revealed that malaria is still a major problem and is linked to socio economic factors as well as geographical location. The government should focus on poorer communities from rural and low altitude areas, especially in the Central Region, as their target group of individuals to educate, support and help change mindsets. In addition, children with anaemia should take priority in receiving the necessary health care and support. The key findings also show that there is a need to educate the population through workshops, mobile clinics and various social media platforms on how to prevent malaria in children under five years of age.

Moreover, the education of mothers should be considered and sup ported so that they can take better care of and protect their children, especially after the child's first six months from birth, as they are more likely to be exposed to malaria vectors.

The study will help the government and donors to control and possibly eliminate malaria in children under five years of age. The main focus should be on children with anaemia, mother's education level, wealth index, children's age, the altitude of the place of residence, region, place of resi dence, toilet facility and electricity facilities. Furthermore, the model used in this study will help other researchers to compare findings.

Future research could use the joint model to model malaria and anaemia simultaneously in order to examine the possible correlation between the two diseases, as there is a link between them (Ayele et al., 2014b; Adebayo et al., 2016).

8. Limitation

The current data set was cross sectional and consequently cannot address causality. It would have been ideal to have a longitudinal data set to study the change in factors and prevalence over time.

Declarations

Author contribution statement

R.T. Gaston: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

S. Ramroop: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by the Health Economics and HIV/AIDS Research Division (HEARD).

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

The authors acknowledge the National Malaria Control Programme and the ICF through the DHS programme for providing the data.

References

Adebayo, S.B., Gayawan, E., Heumann, C., Seiler, C., 2016. Joint modelling of anaemia and malaria in children under five in Nigeria. Spatial Spatio Tempo Epidemiol. 17, 105–115.

Alegana, V.A., Wright, J.A., Nahzat, S.M., Butt, W., Sediqi, A.W., Habib, N., Snow, R.W., Atkinson, P.M., Noor, A.M., 2014. Modelling the incidence of Plasmodium vivax and Plasmodium falciparum malaria in Afghanistan 2006–2009. PloS One 9 (7), e102304.

Ayele, D.G., Zewotir, T.T., Mwambi, H.G., 2013. The risk factor indicators of malaria in Ethiopia. Int. J. Med. Med. Sci. 5 (7), 335–347.

Ayele, D.G., Zewotir, T.T., Mwambi, H.G., 2014. Semiparametric models for malaria rapid diagnosis test result. BMC Publ. Health 14 (1), 31.

Ayele, D.G., Zewotir, T.T., Mwambi, H.G., 2014. Modelling the joint determinants of a positive malaria Rapid Diagnosis Test result, use of mosquito nets and indoor residual spraying with insecticide. Occup. Health South Afr. 20 (4), 20–27.

Bennett, A., Kazembe, L., Mathanga, D.P., Kinyoki, D., Ali, D., Snow, R.W., Noor, A.M., 2013. Mapping malaria transmission intensity in Malawi, 2000–2010. Am. J. Trop. Med. Hyg. 89 (5), 840–849.

Biemba, G., Dolmans, D., Thuma, P.E., Weiss, G., Gordeuk, V.R., 2000. Severe anaemia in Zambian children with Plasmodium falciparum malaria. Trop. Med. Int. Health 5 (1), 9–16.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. 88 (421), 9–25.

Buchwald, A.G., Walldorf, J.A., Cohee, L.M., Coalson, J.E., Chimbiya, N., Bauleni, A., Nkanaunena, K., Ngwira, A., Kapito-Tembo, A., Mathanga, D.P., Taylor, T.E., 2016. Bed net use among school-aged children after a universal bed net campaign in Malawi. Malar. J. 15 (1), 127.

Chanda, E., Mzilahowa, T., Chipwanya, J., Ali, D., Troell, P., Dodoli, W., Mnzava, A.P., Ameneshewa, B., Gimnig, J., 2016. Scale-up of integrated malaria vector control: lessons from Malawi. Bull. World Health Organ. 94 (6), 475.

Cibulskis, R.E., Alonso, P., Aponte, J., Aregawi, M., Barrette, A., Bergeron, L., Fergus, C.A., Knox, T., Lynch, M., Patouillard, E., Schwarte, S., 2016. Malaria: global progress 2000–2015 and future challenges. Infect. Dis. Poverty 5 (1), 61.

Chirombo, J., Lowe, R., Kazembe, L., 2014. Using structured additive regression models to estimate risk factors of malaria: analysis of 2010 Malawi malaria indicator survey data. PloS One 9 (7), e101116.

Chitunhu, S., Musenge, E., 2016. Spatial and socio-economic effects on malaria morbidity in children under 5 years in Malawi in 2012. Spatial Spatio Temporal Epidemiol. 16, 21–33.

Dzinjalamala, F., 2009. Epidemiology of malaria in Malawi. Epidemiol. Malawi 203, 21.

Gaston, R.T., Ramroop, S., Habyarimana, F., 2018. Determinants of factors associated with anemia among children under five years in Lesotho. Afr. Popul. Stud. 32 (1), 3894–3895.

Green, P.J., Silverman, B.W., 1993. Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach. CRC Press, Landon.

Hastie, T.J., Tibshirani, R., 1986. Generalized additive models. Stat. Sci. 1, 297–318.

Hastie, T.J., Tibshirani, R., 1990. Generalized Additive Models. CRC press, London.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M., Snow, R.W., 2004. The global distribution and population at risk of malaria: past, present, and future. Lancet Infect. Dis. 4 (6), 327–336.

Hershey, C.L., Florey, L.S., Ali, D., Bennett, A., Luhanga, M., Mathanga, D.P., Salgado, S.R., Nielsen, C.F., Troell, P., Jenda, G., Ye, Y., 2017. Malaria control interventions contributed to declines in malaria parasitemia, severe anaemia, and all-cause mortality in children less than 5 years of age in Malawi, 2000–2010. Am. J. Trop. Med. Hyg. 97 (3Suppl), 76–88.

Jenkins, R., Omollo, R., Ongecha, M., Sifuna, P., Othieno, C., Ongeri, L., Kingora, J., Ogutu, B., 2015. Prevalence of malaria parasites in adults and its determinants in malaria endemic area of Kisumu County, Kenya. Malar. J. 14 (1), 263.

Kabaghe, A.N., Phiri, M.D., Phiri, K.S., van Vugt, M., 2017. Challenges in implementing uncomplicated malaria treatment in children: a health facility survey in rural Malawi. Malar. J. 16 (1), 419.

Kazembe, L.N., Kleinschmidt, I., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int. J. Health Geogr. 5 (1), 41.

Kazembe, L.N., 2007. Spatial modelling and risk factors of malaria incidence in northern Malawi. Acta Trop. 102 (2), 126–137.

Kazembe, L.N., Mathanga, D.P., 2016. Estimating risk factors of urban malaria in Blantyre, Malawi: a spatial regression analysis. Asian Pac. J. Trop. Biomed. 6 (5), 376–381.

Lazzerini, M., Seward, N., Lufesi, N., Banda, R., Sinyeka, S., Masache, G., Nambiar, B., Makwenda, C., Costello, A., McCollum, E.D., Colbourn, T., 2016. Mortality and its risk factors in Malawian children admitted to hospital with clinical pneumonia, 2001–12: a retrospective observational study. Lancet Global Health 4 (1), e57–e68.

Lin, X., Zhang, D., 1999. Inference in generalized additive mixed models by using smoothing splines. J. Roy. Stat. Soc. B 61 (2), 381–400.

Lindsay, S.W., Martens, W.J., 1998. Malaria in the African highlands: past, present and future. Bull. World Health Organ. 76 (1), 33.

Lwetoijera, D.W., Kiware, S.S., Mageni, Z.D., Dongus, S., Harris, C., Devine, G.J., Majambere, S., 2013. A need for better housing to further reduce indoor malaria transmission in areas with high bed net coverage. Parasites Vectors 6 (1), 57.

Mathanga, D.P., Walker, E.D., Wilson, M.L., Ali, D., Taylor, T.E., Laufer, M.K., 2012. Malaria control in Malawi: current status and directions for the future. Acta Trop. 121 (3), 212–217.

Mhalu, F.S., 2005. Burden of diseases in poor resource countries: meeting the challenges of combating HIV/AIDS, tuberculosis and malaria. Tanzan. J. Health Res. 7 (3), 179–184.

Minakawa, N., Dida, G.O., Sonye, G.O., Futami, K., Njenga, S.M., 2012. Malaria vectors in Lake Victoria and adjacent habitats in western Kenya. PloS One 7 (3), e32725.

National Malaria Control Programme (NMCP) [Malawi] and ICF International, 2010. Malawi Malaria Indicator Survey (MIS). Lilongwe, Malawi & Calverton, Maryland, USA.

National Malaria Control Programme (NMCP) [Malawi] and ICF International, 2012. Malawi malaria indicator survey (MIS) 2012. Lilongwe: NMCP ICF Int.

National Malaria Control Programme (NMCP) [Malawi] and ICF International, 2014. Malawi malaria indicator survey (MIS) 2014. Lilongwe: NMCP ICF Int.

National Malaria Control Programme (NMCP) [Malawi] and, I.C.F., 2018. Malawi Malaria Indicator Survey 2017. Lilongwe: NMCP and ICF.

Ohrt, C., Sutamihardja, M.A., Tang, D., Kain, K.C., 2002. Impact of microscopy error on estimates of protective efficacy in malaria-prevention trials. J. Infect. Dis. 186 (4), 540–546.

Parvin, H., Beygi, S., Helm, J.E., Larson, P.S., Van Oyen, M.P., 2018. Distribution of medication considering information, transshipment, and clustering: malaria in Malawi. Prod. Oper. Manag. 27 (4), 774–797.

Perkins, D.J., Were, T., Davenport, G.C., Kempaiah, P., Hittner, J.B., Ong'echa, J.M., 2011. Severe malarial anaemia: innate immunity and pathogenesis. Int. J. Biol. Sci. 7 (9), 1427.

Roberts, D, Matthews, G, 2016. Risk factors of malaria in children under the age of five years old in Uganda. Malaria Journal 15 (1), P.246.

Ruppert, D., Wand, M.P., Carroll, R.J., 2003. *Semiparametric Regression* (No. 12). Cambridge University Press, New York.

Semakula, H.M., Song, G., Achuu, S.P., Zhang, S., 2016. A Bayesian belief network modelling of household factors influencing the risk of malaria: a study of parasitaemia in children under five years of age in sub-Saharan Africa. Environ. Model. Software 75, 59–67.

Snyman, K., Mwangwa, F., Bigira, V., Kapisi, J., Clark, T.D., Osterbauer, B., Greenhouse, B., Sturrock, H., Gosling, R., Liu, J., Dorsey, G., 2015. Poor housing construction associated with increased malaria incidence in a cohort of young Ugandan children. Am. J. Trop. Med. Hyg. 92 (6), 1207–1213.

Sultana, M., Sheikh, N., Mahumud, R.A., Jahir, T., Islam, Z., Sarker, A.R., 2017. Prevalence and associated determinants of malaria parasites among Kenyan children. Trop. Med. Health 45 (1), 25.

Team, M.C., 2018. Systematic country diagnostic: breaking the cycle of low growth and slow poverty reduction. Available at. http://documents.worldbank.org/curated/en/723781545072859945/Breaking-the-Cycle-of-Low-Growth-and-Slow-Poverty-Reduction. (Accessed 10 March 2020).

Wang, Y., 1998. Mixed effects smoothing spline analysis of variance. J. Roy. Stat. Soc. B 60 (1), 159–174.

Wongsrichanalai, C., Barcus, M.J., Muth, S., Sutamihardja, A., Wernsdorfer, W.H., 2007. A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT). Am. J. Trop. Med. Hyg. 77 (6 Suppl), 119–127.

Wood, S.N., 2006. Generalized additive models: an introduction with R, 67. Chapman and Hall/CRC. *Texts in Statistical Sciences*, p. 391.

Wood, S.N., 2017. Generalized Additive Models: an Introduction with R, second ed. Chapman and Hall/CRC, London, UK.

World Health Organization, 2013. WHO Global Malaria Programme. World Malaria Report 2013. World Health Organization, Geneva, p. 255.

World Health Organization, 2015. World Malaria Report 2015. World Health Organization, Geneva.

World Health Organization, 2016. World Malaria Report 2016. World Health Organization, Geneva.

World Health Organization, 2017. World Malaria Report 2017. World Health Organization, Geneva.

Worrall, E., Basu, S., Hanson, K., 2002. The relationship between socio-economic status and malaria: a review of the literature. In: Background Paper for Ensuring that Malaria Control Interventions Reach the Poor, 56. School of Hygiene and Tropical Medicine, London.

Wu, H., Zhang, J.T., 2006. Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches, 515. John Wiley & Sons, New York.

Zgambo, M., Mbakaya, B.C., 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), e0175537.

Heliyon 7 (2021) e06899

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Joint modelling of malaria and anaemia in children less than five years of age in Malawi

Rugiranka Tony Gaston [a,b,*], Shaun Ramroop [a], Faustin Habyarimana [a]

[a] School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Pietermaritzburg Campus, Private Bag X01, Scottsville, 3209, South Africa
[b] Health Economics and HIV/AIDS Research Division (HEARD), University of KwaZulu-Natal, Westville Campus, Private Bag X01, Westville, 3629, South Africa

ARTICLE INFO

Keywords:
Malaria
Anaemia
Correlation
Children under five years of age
Multivariate joint model
MMIS 2017
Public health
Infection disease
Nutrition

ABSTRACT

Background: Malaria and anaemia jointly remain a public health problem in developing countries of which Malawi is one. Although there is an improvement along with intervention strategies in fighting against malaria and anaemia in Malawi, the two diseases remain significant problems, especially in children 6–59 months of age. The main objective of this study was to examine the association between malaria and anaemia. Moreover, the study investigated whether socio-economic, geographic, and demographic factors had a significant impact on malaria and anaemia.

Data and methodology: The present study used a secondary cross-sectional data set from the 2017 Malawi Malaria Indicator Survey (MMIS) with a total number of 2 724 children 6–9 months of age. The study utilized a multivariate joint model within the ambit of the generalized linear mixed model (GLMM) to analyse the data. The two response variables for this study were: the child has either malaria or anaemia.

Results: The prevalence of malaria was 37.2% of the total number of children who were tested using an RDT, while 56.9% were anaemic. The results from the multivariate joint model under GLMM indicated a positive association between anaemia and malaria. Furthermore, the same results showed that mother's education level, child's age, the altitude of the place of residence, place of residence, toilet facility, access to electricity and children who slept under a mosquito bed net the night before the survey had a significant effect on malaria and anaemia.

Conclusion: The study indicated that there is a strong association between anaemia and malaria. This is interpreted to indicate that controlling for malaria can result in a reduction of anaemia. The socio-economic, geographical and demographic variables have a significant effect on improving malaria and anaemia. Thus, improving health care, toilet facilities, access to electricity, especially in rural areas, educating the mothers of children and increasing mosquito bed nets would contribute in the reduction of malaria and anaemia in Malawi.

1. Introduction

Malaria and anaemia jointly remain a public health problem world wide in both developed and developing countries (Kanchana et al., 2018). Malaria and anaemia are life threatening diseases, mainly in developing countries, where children and pregnant women are more vulnerable (McLean et al., 2009; WHO, 2017; Kejo et al., 2018). The two conditions are known to add to the tremendous weight of morbidity and mortality, especially in children under five years of age (Wanzira et al., 2017; White, 2018). The significant progress in fighting both malaria and anaemia has been improved worldwide. However, both diseases remain a health problem especially in children from developing countries, and Malawi is among the developing countries (WHO, 2015; Kejo et al., 2018; Roberts and Zewotir, 2020).

Malaria is caused by different *Plasmodium* parasites, such as *P. falciparum, P. vivax, P. ovale, P. malariae* and *P. knowlesi, with P. falciparum* the most prevalent parasite in the African region (Perkins et al., 2011; WHO, 2015; Seyoum, 2018). Malaria is transmitted in the human body through the bites of infected *Anopheles* mosquitoes, where the malaria parasites infect the red blood cells and reduce the amount of the blood cells, which can lead to severe anaemia (Noland et al., 2012; Seyoum, 2018). The combination of rainfall, humidity and high temperature in low altitudes creates favourable conditions for breeding and development of malaria vectors. Thus, for this reason the transmission of

* Corresponding author.
E-mail address: rugiratony@gmail.com (R.T. Gaston).

https://doi.org/10.1016/j.heliyon.2021.e06899
Received 30 October 2020; Received in revised form 6 March 2021; Accepted 21 April 2021

malaria becomes higher when the climate is wet, hot and more humid (Chirombo et al., 2014; Gaston and Ramroop, 2020).

The World Health Organization (WHO) estimated approximately 219 million cases of malaria in 2017, with 200 million from Africa (World Malaria Report, 2018). In addition, the WHO reported the total number of 43 000 deaths from malaria globally in 2017, and 93% of cases were from the African Region. Children under five years accounted for 61% of deaths globally and 93% were from Africa (World Malaria Report, 2018). In 2013, malaria in Malawi accounted for 20% of children under five years who died in hospital, and the prevalence of malaria is still high (WHO, 2016; NMCP and ICF, 2018; Gaston and Ramroop, 2020).

The World Health Organization defines anaemia as a reduction of hemoglobin in blood cells, which causes the body tissues to not have enough oxygen. In women, men or children, anaemia can be categorized as mild, moderate, and severe. In children 6 59 months of age, when the hemoglobin concentration level is less than 11 g/dl, the child is consid ered anaemic (Korenromp et al., 2004; WHO, 2015; NMCP and ICF, 2018). However, in the case of malaria related anaemia, the cutoff for the hemoglobin concentration level is below 8 g/dl (NMCP and ICF, 2018).

Globally, approximately 1.62 billion people are affected by anaemia and this accounts for more than 24.8% of the world population, around 43% from developing countries, and 47.4% of these are children (McLean et al., 2009; Kanchana et al., 2018). In Malawi, the prevalence of anaemia in children was 63% and this shows that anaemia remains a health problem and more care is needed in the country (NSO and ICF, 2017). Anaemia is detrimental to a child's health in that it can affect children's physical and mental development which can also affect socio economics (Abegunde and Stanciole, 2006; Magalhaes and Clements, 2011; WHO, 2011; Gaston et al., 2018).

The main causes of anaemia are nutritional deficiencies and infection diseases such as HIV, intestinal worms, intake of iron, folate, vitamin B12, malaria and other parasitic infections (McCuskee et al., 2014; Gaston et al., 2018). However, in the region of malaria endemic, the main contributor to anaemia is malaria (Brabin et al., 2004; Crawley, 2004; World Malaria Report, 2018). It is known that malaria is the major contributor to anaemia, and a huge amount of mortality and morbidity is caused with both malaria and anaemia (Bjorkman, 2002; Carneiro et al., 2006; Wanzira et al., 2017). In the area of high prevalence of malaria, anaemia is held accountable for about half of malaria related deaths (Korenromp et al., 2004; Adebayo et al., 2016; Seyoum, 2018). These two diseases are associated and this means that controlling malaria can reduce anaemia, and controlling anaemia can results in reduction of deaths related to malaria (Korenromp et al., 2004; Noland et al., 2012; Reithinger et al., 2013; Hershey et al., 2017). The study by McLean et al. (2009) revealed that more than half of the reduction in malaria resulted in a reduction of 60% in the risk of having anaemia. This confirms that anaemia and malaria are correlated diseases, which can increase mor tality and do more damage in children if no actions are taken timeously (Gaston et al., 2018).

There are very few studies on modelling both childhood anaemia and malaria simultaneously as many studies showed that the young children are more vulnerable to both (Hershey et al., 2017; Kuziga et al., 2017; Yimgang et al., 2021). In addition, the health of a child should be prioritized as they are the posterity of the country (Gaston et al., 2018). Hence, in light of the aforementioned reasons, the current study focused on the modelling of anaemia and malaria in children 6 59 months of age in order to understand the link between the two conditions diseases so that they can be controlled and eliminated. Furthermore, it assists in policy making and planning of interventions strategies from different donors. In Malawi, many researchers were interested in modelling the prevalence of anaemia and malaria in children separately (Chitunhu and Musenge, 2015; Mathanga et al., 2015; Calis et al., 2016; Kabaghe et al., 2017; Ntenda et al., 2017; Zgambo et al., 2017; Hajison et al., 2018; Nkoka et al., 2019). The separated model has its benefits but cannot address the possible association between the two diseases jointly. The joint model is needed to simultaneously model anaemia and malaria to address the association between the two diseases along with identifying factors associated with the diseases. The multivariate joint model under a GLMM has focal points when compared to separate models, for instance, better control of type I error rates in the various tests. Besides this, the multivariate joint model is better for expansion in the capability of the parameter estimate and the ability to address distinctively multivariate questions. Furthermore, the GLMM includes the random effect in the model in order to model the correlation between two or more observa tions (Gueorguieva, 2001; Hedeker, 2005; Agresti, 2015; Habyarimana et al., 2016; Gaston and Ramroop, 2020).

Relevant literature reveals that numerous researchers proposed different statistical models to analyze the association between malaria and anaemia in children (Safeukui et al., 2015; Adebayo et al., 2016; Seyoum, 2018).

In Malawi, the study by Kabaghe et al. (2017) used a year repeated cross sectional survey from a rural area in Malawi to analyse the short change in anaemia and malaria under five years' children. The study by McGann et al. (2018) also used the 2015 DHS data set to describe the prevalence and distribution of inherited blood disorders among young children in Malawi and explore their associations with malaria and anaemia. The recent study by Roberts and Zewotir (2020) used Geo spatial maps to visualize the relationship between malaria and anaemia in Malawi, Uganda, Tanzania and Kenya. In addition, the study by Yimgang et al. (2021) evaluated the population attributable fraction of anaemia due to malaria in children between 5 15 years in Southern Malawi. However, according to our knowledge, no study in literature utilized the generalized linear mixed model (GLMM) to simultaneously join malaria and anaemia in children 6 59 months of age in Malawi and this highlights the novelty of the current research. In addition, the data set used is different in comparison to the 2017 Malawi Malaria Indicator Survey (MMIS) data set.

Therefore, the current study aimed to simultaneously model the as sociation between malaria and anaemia and identify factors associated with the two diseases by utilizing the joint model for a multivariate generalized linear mixed model (GLMM) using the 2017 MMIS.

2. Methodology and material

2.1. Study area

Malawi is one of the African countries and is among the Sub Saharan African nations located south of the equator and surrounded by Tanzania in the North and Northeast; by Mozambique toward the East and Southwest; and Zambia toward the West and Northwest (NMCP and ICF, 2018). Malawi is divided into three regions and twenty eight districts; with the Northern region split into six districts, whilst the Central region consists of nine districts, and the Southern region has thirteen districts.

The country has a tropical climate that is conducive for the breeding of *Anopheles* mosquitos and the mosquitos increase in the rainy season from November to April. The climate is cool and dry from May to August during which the transmission of malaria reduces compared to the rainy season (Kazembe, 2007; Zgambo et al., 2017). The economy of the country is based on agriculture and is among the poorest countries in the world with poor healthcare in comparison with other African countries (WHO, 2016).

2.2. Data sources

The study utilized secondary cross sectional data set from the 2017 Malawi Malaria Indicator Survey (MMIS). The data was gathered be tween 15 April and June 2017 and executed by the Malawi National Malaria Control Program (NMCP) through help from the President's Malaria Initiative (PMI). The United States Agency for International Development (USAID) offered money related help through the Presi dent's Malaria Initiative (PMI). They likewise supported the undertaking by offering specialized help with the administration of community and

wellbeing studies as they do in nations around the world (NMCP and ICF, 2018). The overseeing group of Malawi gave staff office space, and key assistance. From there on, the ICF offered specialized help through the Demographic and Health Survey (DHS) program. The ethical approval was evaluated and granted by the Malawian Ministry of Health Research and Ethics Committee with the support of the Institutional Review Board of ICF International.

2.3. Data sampling and design

Women between 15 49 years of age and children from 6 59 months who stayed in or visited the selected households the night before the survey were included in the interview. The 2017 MMIS was a population based on a household cluster survey and the sampling survey followed the two stage sample design. Furthermore, the 2017 survey enables as sessments of key malaria indicators for the nation as a whole, for urban and rural areas independently, and for all of the three administrative territories in Malawi: Northern, Central, and Southern.

The first stage of sampling comprised a selection of 150 clusters from the enumeration areas (EAs) outlined in the 2008 Population and Housing Census. Among the 150 clusters, 60 were from urban and 90 from rural areas. The second stage of sampling involved the systematic selection of a sample of 3 750 households. Of these households, 25 households were selected from each enumeration area (EA) (NMCP and ICF, 2018).

The study used a total weighted number of 2 724 children 6 59 months of age to establish a national level portrait (NMCP and ICF, 2018). The study used the weighted sample to gain insights that were illustrative of the nation and to account for the complex sample design from the data set. In the sampling procedure, the individuals surveyed in each region should contribute proportionally to the size of the total sample in the region. In any case, some regions may have small pop ulations, and this unweighted appropriation does not represent the exact population. Therefore, the region with a small population is oversampled to overcome these issues and is, for this reason, the weighted sample used in this study (NMCP and ICF, 2018; Gaston and Ramroop, 2020).

2.4. Blood collection and laboratory method

Children aged 6 59 months were tested for both anaemia and malaria with the parents' or guardians' consent. Trained nurses were responsible for the testing and the children who tested positive were given medica tion on the spot, according to the national rules.

2.4.1. Anaemia testing

Blood samples were collected from each child aged 6 59 months using a spring loaded, sterile lancet to make a finger or heel prick. The drop of blood was collected in a microcuvette, and the Haemoglobin level analysed using a portable HemoCue analyzer. The results were available within 10 min and were given to the child's parent or guardian verbally and in writing (NMCP and ICF, 2018). Parents were encouraged to take the children with haemoglobin level less than 8 g/dl to the nearest health care facility for the follow up. The parents were given a referral letter with the haemoglobin examination to show the staff at the health care office. The results from the anaemia test were captured on the Biomarker Questionnaire and on the handout left in the household that contained information on the causes and prevention of anaemia (NMCP and ICF, 2018).

2.4.2. Malaria testing

The blood sample was collected from children's finger or heel prick using the SD Biolne Malaria Ag P.f/P, a rapid diagnostic test (RDT). Malaria testing can be done using microscopy or rapid diagnostic test; however, in this study the RDT was considered.

Microscopic diagnosis is helpful in testing malaria, although it does have some limitations; these include inadequately trained microscopists, lack of quality control, the chance of misdiagnosis because of low para sitemia or blended diseases, and in some cases it is hard to determine the types of plasmodium. Moreover, microscopy services are not accessible, for example, in a remote area or after standard laboratory hours (Ohrt et al., 2002; Wongsrichanalai et al., 2007; Gaston and Ramroop, 2020). Thus, for that reason, the RDT is used and is relevant in the detection of the histidine rich protein II (HRP II). Furthermore, the RDT detects an antigen of Plasmodium falciparum and common Plasmodium lactate dehydrogenase (PLDH) of Plasmodium species in human blood (NMCP and ICF, 2018). The diagnostic test incorporates an expendable example tool that arrives in a standard bundle. A tiny volume of blood is captured on the applicator and placed in the well of the testing device. All field laboratory experts were upskilled to use the RDT in the field as per the producer's directions. The RDT results were accessible quickly and recorded as either positive or negative, with blackout test lines viewed positive. Likewise, with the anaemia testing, malaria RDT results were given to the child's parent or guardian in oral and composed structure and were recorded on the Biomarker Questionnaire. Moreover, children who tested positive for malaria were provided with a full course of medication as indicated by the standard system for uncomplicated ma laria treatment in Malawi (NMCP and ICF, 2018).

3. Data analysis

3.1. Dependent variable

The present study considered two response variables or dependent variables. The first one was anaemia status for children 6 59 months of age. The anaemia status in children is determined based on the hemo globin concentration level in the blood measured in grams per deciliter (g/dl). When the hemoglobin concentration level adjusted for altitude is less than 11 g/dl, the child is reviewed as anaemic, otherwise not anaemic (WHO, 2015; NMCP and ICF, 2018). The second one is malaria status using RDT to check if the child has malaria (positive) or not (negative).

3.2. Independent variables

The exploratory covariate or independent variables used in this study were also used in previous literature and involved a number of socio economic, demographic, and geographic factors (Bennett et al., 2013; Alegana et al., 2014; Buchwald et al., 2016; Zgambo et al., 2017). The present study used independent variables assumed to be linked with malaria and (or) anaemia such as the child's age in months; the sex of the child; the type of residence; region of the dwelling; wealth quantile; mother's highest education level; source of drinking water; type of toilet facility; household sharing the toilet; household using electricity or not; children under 5 slept under a mosquito bed net the night before the survey; the main material of wall, floors; roofs of the rooms and residence altitude. In Malawi the studies which modelled anaemia and malaria also used the same independent variables found in the present study (Kabaghe et al., 2017; McGann et al., 2018; Yimgang et al., 2021). However, these studies did not include some variables assumed to be associated with anaemia and (or) malaria (Gaston and Ramroop, 2020). The independent variables which were not included in their studies are source of drinking water; type of toilet facility; household sharing the toilet; household using electricity or not; children under 5 slept under a mosquito bed net the night before the survey; the main material of wall, floors; roofs of the rooms and residence altitude.

4. Statistical analysis

The present study used Statistical Package for the Social Sciences (SPSS) to clean the data. In addition, the analysis of bivariate method was done in SPSS, with the application of cross tabulation techniques. Pear son's chi square test and p values were used to investigate whether the

independent variables were associated with each one of the responses variables or not. To summarize the data, the frequencies and percentages were used, and p values to check the relationship between exploratory variables and response variables. Any exploratory variables with a p value of less than 0.2 were included in the multivariate generalized linear mixed model (GLMM) (El Kishawi et al., 2015; Gaston et al., 2018). Then any exploratory variable found to have a statistically significant associ ation with response variables at p value less than 0.05 in multivariate was reported. The analysis of multivariate used SAS 9.4 PROC GLIMMIX procedure. The procedure enabled us to jointly model two outcomes (response) variables with similar distributions, link functions, or different link functions. However, this study assessed a similar distribu tion and link functions for the two outcome variables. Furthermore, numerous covariance structures were considered based on lowest value of Akaike information criteria (AIC) and Unstructured (UN) were found to be suitable to our analysis. In addition, we checked the possible interaction and none was statistically significant (Molenberghs and Verbeke, 2005; Habyarimana et al., 2016).

4.1. Model formulation

The study considered two response variables, which were malaria and the anaemic status of a child. Suppose the response variable y_{i1} to be malaria RDT status, where one (1) is assigned as positive status and zero (0) negative status. The second response variable as y_{i2} to be anaemia status, where one (1) is assigned as an anaemic child and zero (0) non anaemic. When the distinguished results emerge from a bivariate Ber noulli distribution, with p_{i1} as the likelihood of malaria occurring in a child i and p_{i2} as the probability of anaemia occurring in a child i. In this way, the binary generalized linear model can be written as:

Table 1. Childhood anaemia by categorical variable.

Variables	Category	Anaemia status		P-value
		Anaemic	Not anaemic	
Child's age in months	6–23 months	609 (78.6%)	166 (21.4%)	0.000
	24–31 months	466 (58.0%)	337 (42.0%)	
	32–59 months	474 (51.0%)	455 (49.0%)	
Sex of the child	Male	771 (61.0%)	492 (39.0%)	0.432
	Female	779 (62.6%)	466 (37.4%)	
Sleep under mosquito bed net	All children	956 (58.0%)	691 (42.0%)	0.002
	Some	136 (52.7%)	122 (47.3%)	
	None	458 (55.9%)	361 (41.1%)	
Region	North	145 (53.5%)	126 (46.5%)	0.004
	Central	686 (64.4%)	379 (35.6%)	
	South	718 (61.3%)	453 (38.7%)	
Place of residence	Rural	1366(63.0%)	801 (37.0%)	0.001
	Urban	184 (54.0%)	157 (46.0%)	
Wealth index	Poorer	793 (67.7%)	379 (32.3%)	0.000
	Middle	284 (59.2%)	196 (40.8%)	
	Richer	473 (55.2%)	384 (44.8%)	
Mother's education level	No education	226 (68.9%)	102 (31.1%)	0.000
	Primary	1125(55.8%)	891 (44.2%)	
	Post-primary	198(52.2%)	181 (47.8%)	
Altitude	0–500 m	240 (72.9%)	89 (27.1%)	0.000
	501–1000 m	604 (60.0%)	402 (40.0%)	
	>1000 m	706 (60.2%)	467 (39.8%)	
Toilet facility	Toilet with flush	198 (51.3%)	188 (48.7%)	0.000
	Pit latrine	1235(63.3%)	717 (36.7%)	
	No facility	117 (68.8%)	53 (31.2%)	
Type of drinking water	Tap water	339 (53.6%)	294 (46.4%)	0.024
	Protected water	290 (54.3%)	244 (45.7%)	
	Unprotected	920 (59.1%)	636 (40.9%)	
Main roof material	Thatch/Palm leaf	920 (61.4%)	579 (38.6%)	0.000
	Corrugated metal	557 (50.7%)	542 (49.3%)	
	Stick & mud	73 (57.5%)	54 (42.5%)	
Main wall material	Wood/Mud	475 (61.9%)	292 (38.1%)	0.001
	bricks	624 (53.6%)	540 (46.4%)	
	Cement/Block	451 (56.9%)	342 (43.1%)	
Main floor material	Earth/Sand	1143(59.2%)	789 (40.8%)	0.000
	Mud block/Wood	110 (58.5%)	78 (41.5%)	
	Cement/Block	297 (49.2%)	307 (50.8%)	
Electricity	Yes	122 (42.5%)	456 (44.6%)	0.000
	No	1427(58.6%)	654 (42.9%)	
Household share toilet	Yes	566 (55.4%)	456 (44.6%)	0.380
	No	872 (57.1%)	654 (42.9%)	

$$g_1(\mu_{t1}) \quad X_{i1}\beta_1 + Z_{i1}v_1 \tag{1}$$

$$g_2(\mu_{t1}) \quad X_{t2}\beta_2 + Z_{t2}\,v_2 \tag{2}$$

Where β_1 and β_2 are assumed to be the vectors of fixed effects, v_1 and v_2 are the vectors of the random effects, while X_{t1}, X_{t2}, Z_{t1} and Z_{t2} are the designed matrices for fixed effects and random effects respectively. Thus, equation of the variance covariance matrices model is shown as follows:

$$v \quad \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \sim i.i.d.MVN\left(0, \sum\right) \quad MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{bmatrix}\right) \tag{3}$$

where the $\sum_{11}$ and $\sum_{22}$, are the variance components of malaria of children under five years and anaemia respectively, while $\sum_{12}$ and $\sum_{21}$ are the correlation components between malaria and anaemia are the same. When the correlation components from Eq. (3), $\sum_{12} \quad \sum_{21} \quad 0$, the multivariate joint model under generalized linear mixed model be comes a separate model (Molenberghs and Verbeke, 2005; Habyarimana et al., 2016).

4.2. Results and interpretation

4.2.1. Univariate results

The results from Table 1 and Table 2 indicated the frequency distri bution and percentages of childhood anaemia and malaria respectively with the associated factors. Cross tabulation techniques were used to analyse the data and summarise the results in both Tables. Pearson's chi square test and p values were used to investigate whether the indepen dent variables were statistically significant associated with each one of the responses variables or not. The results from Table 1 indicated that all independent variables were significantly associated with childhood anaemia with the p value less than 0.05, except for the sex of the child,

Table 2. Childhood malaria by categorical variable.

Variables	Category	Malaria status		P-value
		Positive	Negative	
Child's age in months	6–23 months	200(25.6%)	581(74.4%)	0.000
	24–31 months	411(40.6%)	602(59.4%)	
	32–59 months	403(43.4%)	526(56.6%)	
Sex of the child	Male	463(36.5%)	804(63.5%)	0.716
	Female	447(35.8%)	800(64.2%)	
Sleep under mosquito bed net	All children	553(33.6%)	1093(66.4%)	0.000
	Some	88 (34.1%)	170 (65.9%)	
	None	372(45.4%)	447 (54.6%)	
Region	North	62 (20.8%)	236 (79.2%)	0.000
	Central	452(39.8%)	683 (60.2%)	
	South	500(38.7%)	791 (61.3%)	
Place of residence	Rural	987 (41.9%)	1369(58.1%)	0.000
	Urban	27 (7.3%)	341 (92.7%)	
Wealth index	Poorer	611(48.0%)	662(52.0%)	0.000
	Middle	204(39.6%)	311(60.4%)	
	Richer	199(21.3%)	737(78.7%)	
Mother's education level	No education	152(46.3%)	176(53.7%)	0.000
	Primary	799(39.6%)	1218(60.4%)	
	Post-primary	63 (16.6%)	216(83.4%)	
Altitude	0–500 m	119(34.3%)	228 (65.7%)	0.000
	501–1000 m	492(43.8%)	631 (56.2%)	
	>1000 m	403(32.2%)	850 (67.8%)	
Toilet facility	Toilet with flush	88 (21.1%)	330(78.9%)	0.000
	Pit latrine	832(39.3%)	1287(60.7%)	
	No facility	95 (50.8%)	92 (49.2%)	
Type of drinking water	Tap water	164(25.9%)	470 (74.1%)	0.000
	Protected water	122(31.9%)	260 (68.1%)	
	Unprotected	728(42.6%)	980 (57.4%)	
Main roof material	Thatch/Palm leaf	686(45.8%)	812 (54.2%)	0.000
	Corrugated metal	283(25.8%)	816 (74.2%)	
	Stick & Mud	45 (35.7%)	81 (64.3%)	
Main wall material	Wood/Mud	331(43.2%)	436 (56.8%)	0.000
	bricks	467(40.1%)	697 (59.9%)	
	Cement/Block	216(27.2%)	577 (72.8%)	
Main floor material	Earth/Sand	840(43.5%)	1092(56.5%)	0.000
	Mud block/Wood	74 (39.4%)	114 (60.6%)	
	Cement/Block	100(16.6%)	504 (83.4%)	
Electricity	Yes	21 (7.3%)	265 (92.7%)	0.000
	No	992(40.7%)	1445(59.3%)	
Household share toilet	Yes	329(32.2%)	692 (67.8%)	0.000
	No	600(39.3%)	927(60.7%)	

and households who share a toilet. Table 2 also showed that all inde pendent variables were significantly associated with childhood malaria with the p value less than 0.05, except for the sex of the child. The prevalence of anaemia and malaria from this study was 56.9 % and 37.2% respectively.

The results from Table 1 revealed that the prevalence of anaemia was higher in children from mothers with no education (68.9%) and lower with mothers with primary (55.8%) and post primary (52.2%) respec tively. The same results indicated a decrease in the prevalence of anaemia in children from wealthier classes (55.2%), and increase in middle (59.2%), and poorer classes (67.7%). The prevalence of anaemia in children from a rural area (63.0%) was higher compared to those from an urban area (54.0%). The results also indicated a decrease in prevalence of anaemia as the age for a child increased. The prevalence of anaemia was 78.6%, 58.0%, and 51.0% among children aged between 6 23, 24 41, and 42 59 months respectively.

The results from Table 2 revealed that the prevalence of malaria was higher in children from mothers with no education (46.3%), and reduced by mothers with primary (39.6%) and post primary education (16.6%) respectively. The same results indicated that the prevalence of malaria was lower in children from the wealthier (21.3%), and increased to middle (39.6%) and poorer (48.0%) classes respectively. The prevalence of malaria in children from rural area was higher (41.9%); while for those from the urban area it was lower (7.3%). Table 2 also showed that the prevalence of malaria increased as the child's age increased, which was 25.6%, 40.6%, and 43.4% among children aged between 6 23, 24 41, and 42 59 months respectively.

4.2.2. Multivariate results

The analysis of multivariate used SAS 9.4 PROC GLIMMIX procedure to assess the correlation between anaemia and malaria and any other covariates factors which might be associated with the two diseases. In addition, we checked all possible interaction effects between the exploratory variables and none was statistically significant, hence these were not included in the results.

The results from Table 3 indicated the parameter estimates, odds ratio (OR), 95% confidence intervals (CI) and p values. The study reported only the exploratory variables with statistically significant impact on malaria and anaemia ($p < 0.05$). The variables with a significant effect on both malaria and anaemia were the child's age, mother's education level, availability of electricity, toilet facilities, and children under five who slept under a mosquito bed net the night before the survey. The residence of household and altitude of residence had only a statistically significant effect on malaria.

The results from Table 3 indicated that children aged 6 23 months were 0.367 (OR: 0.367, 95% CI: 0.274; 0.490) times less likely to test positive for malaria using an RDT when compared with those in the reference group (42 59 months). In contrast, the children aged 6 23 months were 4.289 (OR: 4.289, 95% CI: 3.418; 5.382) times more likely to have anaemia compared with those in the age group 42 59 months.

The same results showed that children from mothers with post primary levels were 0.505 (OR: 0.505, 95% CI: 0.305; 0.835) times less likely to have malaria compared to those from the mother with no edu cation. The children from mothers with primary school were 0.710 (OR: 0.710, 95% CI: 0.506; 0.997) times less likely to have anaemia compared with those in the reference category group. The results revealed that the odds of testing positive malaria in an RDT for children from households with no access to electricity were 2.296 (OR: 2.296, 95% CI: 1.415; 3.745) times more likely than those from households with access to electricity. The same results indicated that children from households with no access to electricity were 1.279 (OR: 1.279, 95% CI: 1.005; 1.732) times more likely to have anaemia compared to those from households with access to electricity.

The study also revealed that children from a household with pit la trines were 0.625 (OR: 0.625, 95% CI: 0.401; 0.975) times less likely to test positive for malaria than those with no toilet facilities, while those from households with flush toilet were 0.470 (OR: 0.470, 95% CI: 0.271; 0.815) times less likely to have malaria compare to those with no toilet facilities. Furthermore, children from households with flush toilets were 0.580 (OR: 0.580, 95% CI: 0.369; 0.913) times less likely to have anaemia, compared to those with no toilet facilities.

The results indicated that the odds of testing positive malaria in an RDT was 1.586 (OR: 1.586, 95% CI: 1.045; 2.406) times more likely in households with no children slept under a mosquito bed net the night

Table 3. Parameter estimates for a joint marginal model for malaria and anaemia.

Indicator variables	Malaria				Anemia			
	Estimate	OR[a]	95% CI[b] Upper; Lower	P-value	Estimate	OR	95% CI Upper; Lower	P-value
Child's age (Ref: 42–59)	-	-	-	-	-	-	-	-
6–23 months	-1.003	0.367	0.274; 0.490	0.000	1.456	4.289	3.418; 5.382	0.000
24–41 months	-0.069	0.933	0.728; 1.197	0.589	-0.008	0.992	0.815; 1.207	0.939
Residence (Ref: Urban)	-	-	-	-	-	-	-	-
Rural	1.284	-3.611	2.111; 6.178	0.000	-0.149	0.856	0.649; 1.145	0.303
Altitude (Ref: 501–1000 m)	-	-	-	-	-	-	-	-
>1000 m	-0.866	0.421	0.244; 0.725	0.002	-0.090	0.914	0.698; 1.197	0.512
0–500 m	-0.497	0.608	0.295; 1.256	0.179	0.280	1.323	0.919; 1.906	0.131
Education (Ref: No education)	-	-	-	-	-	-	-	-
Primary	-0.062	0.940	0.649; 1.361	0.743	-0.342	0.710	0.506; 0.997	0.048
Post-primary	-0.684	0.505	0.305; 0.835	0.008	-0.325	0.723	0.486; 1.075	0.109
Electricity (Ref: Yes)	-	-	-	-	-	-	-	-
No	0.831	2.296	1.415; 3.745	0.001	0.277	1.279	1.005; 1.732	0.046
Toilet facilities (Ref: No facility)	-	-	-	-	-	-	-	-
Pit latrine	-0.470	0.625	0.401; 0.975	0.038	-0.222	0.801	0.537; 5.930	0.277
Toilet with flush	-0.755	0.470	0.271; 0.815	0.007	-0.544	0.580	0.369; 0.913	0.019
Child sleeping under net (Ref: All children)	-	-	-	-	-	-	-	-
Some children	0.024	1.024	0.691; 1.516	0.906	0.364	1.439	1.064; 1.946	0.018
No children	0.461	1.586	1.045; 2.406	0.030	0.055	1.057	0.763; 1.462	0.742

[a] OR: odds ratio.
[b] 95% CI: confidence interval.

before the survey, compared to households with all children slept under a mosquito bed net.

However, the odds of having anaemia in some children who slept under a mosquito bed net the night before the survey, were 1.439 (OR: 1.439, 95% CI: 1.064; 1.946) times more likely than all children who slept under a mosquito bed net.

Table 3 showed that children living in rural areas were 3.611 (OR: 3.611, 95% CI: 2.111; 6.178) times more likely to test positive for malaria compared to those living in urban areas. The same results indicated that children who live in a residence with an altitude of more than 1000 m were 0.421 (OR: 0.421, 95% CI: 0.244; 0.725) times less likely to have malaria than those living in residence with an altitude between 501 and 1000 m.

However, the children living in rural and altitude residence areas were not statistically significant to anaemia, hence we did not interpret the results.

Table 4 shows the variance components and covariance between anaemia and malaria. The coefficient estimate of 0.700 indicates the relationship between anaemia and malaria. This means that there is a statistically significant positive correlation between malaria and anaemia.

The overall fitted model was highly significant as the coefficient of covariance parameter indicated the p value 0.000. Hence, including the random cluster effect in the model was necessary (Molenberghs and Verbeke, 2005; Zhang and Lin, 2008).

The pseudo likelihood test rejected the null hypothesis of zero correlation with Pearson chi squared test 4591.22 and the ratio of Pearson and degree of freedom of 0.91 with p value 0.000. This revealed that there is a good variability in the dataset and there was no residual over dispersion (Molenberghs and Verbeke, 2005). Furthermore, this indicated that the association between the prevalence of malaria and anaemia was highly significant. In addition, the odds ratio (2.014) also confirmed that anaemia and malaria are highly associated. The results from the present study are in line with the study by Seyoum (2018).

5. Discussion

The present study utilized the joint model for a multivariate generalized linear mixed model (GLMM) to simultaneously model the association between malaria and anaemia and identify factors associated with the two diseases. The study indicated that anaemia and malaria are highly associated. This means that malaria and anaemia move in the same direction, where any increase of malaria in children, will also result in an increase of anaemia. This finding is consistent with existing literature (Noland et al., 2012; Zgambo et al., 2017). The same change can be in a negative direction; when malaria reduces in children, so does anaemia. Therefore, the change between both malaria and anaemia can be interpreted to mean that controlling malaria can result in effectively reducing anaemia (Reithinger et al., 2013; Hershey et al., 2017; Yimgang et al., 2021). In addition, controlling anaemia in the area of high prevalence of malaria can result in reduction of deaths related to malaria (Korenromp et al., 2004; Seyoum, 2018).

The findings from this study revealed that children from mothers with primary and post primary education level were less likely to have both malaria and anaemia compared to those from mothers with no education. This shows that the risk of having malaria or anaemia reduces as the education levels of their mothers increases. The findings from this study are aligned with the studies by Adebayo et al. (2016), Seyoum (2018), and Gaston and Ramroop (2020).

The results indicated that children from households with no access to electricity have more risk of having malaria and anaemia. Furthermore, households with no toilet facilities are more likely to increase the rates of positive malaria RDT and anaemia. Access to electricity and toilet facilities are indicators of socio economics. Therefore, households with access to socio economics can easily access healthcare, are able to eat healthy food and can easily afford the treatment (Ayele et al., 2014b; Gaston and Ramroop, 2020).

The findings from the present study also indicated that children who did not sleep under a mosquito bed net the night before the survey had more risk of having malaria and anaemia. This might be due to the children who sleep under mosquito bed nets being more protected from being bitten by *Anopheles* mosquitoes which is the cause of malaria (Gaston and Ramroop, 2020). The same results were found in previous studies, such as those by Ayele et al. (2014) and Zgambo et al. (2017).

The research showed that the children living in high altitude residence are less likely to have malaria. This is due to the markedly higher temperature at lower altitude residence area, where the mosquitoes develop in hotter areas (Chirombo et al., 2014; Teh et al., 2018; Gaston and Ramroop, 2020). The results also revealed that children from rural area are more likely to test positive for malaria in an RDT. The findings from this study are in line with the studies by Adebayo et al. (2016), and Gaston and Ramroop (2020).

The findings from this study revealed that the probability of a positive malaria increased as the child's age increased. The children aged 6 23 months were less likely to test positive for malaria. These results are in contrast with the results found in the study by Seyoum (2018). However, the same results are in line with the studies by Zgambo et al. (2017), Gaston and Ramroop (2020), and Yimgang et al. (2021).

In contrast, the probability of having anaemia reduced as the child's age increased. Children aged 6 23 months were more likely to have anaemia. This might be due to the fact that anaemia is also associated with other factors such as nutritional deficiencies, diseases infections such as HIV, intestinal worms, intake of iron, folate, vitamin B12, and other parasitic infections (Ayoya et al., 2013; Gaston et al., 2018). Furthermore, the immune system of young children is not strong enough to fight against different diseases but becomes stronger as they grow older. Thus, for this self same reason young children are highly at risk of getting anaemia, and this reduces as they grow older. The results from this study are consistent with the studies by Hershey et al. (2017), Kuziga et al. (2017), Gaston et al. (2018), and Yimgang et al. (2021).

6. Conclusion

The main objective of this cross sectional study was to examine the association between malaria and anaemia using a multivariate joint model under the GLMM in children 6 59 months of age in Malawi. The current scientific setting also checked other factors which might be associated with both malaria and anaemia. Finally, we examined all possible interaction effects between the exploratory variables and these were not included in the results since none was statistically significant.

The findings from this study indicate that there is an association between anaemia and malaria and any change in one disease has a similar effect on the other disease. This means that as malaria increases so does anaemia and vice versa. Therefore, the study suggests that any policy change to malaria will impact on anaemia. Furthermore, the two diseases

Table 4. Variance components.

Variables	Estimates	Odds Ratio (OR)	95% CI (Upper, Lower)	P-value
Variance (Malaria)	1.014	2.757	0.610; 1.418	0.000
Variance (Anaemia)	0.118	1.125	0.020; 0.216	0.009
Correlation between Anaemia and Malaria	0.700	2.014	0.333; 1.067	0.000

are associated with socio economic, demographic, and geographical factors, which makes these diseases a persistent and a current problem.

Based on the findings from this study, there is a need for educating the population, particularly those from rural areas, on how to prevent ma laria and anaemia in children under five years of age. The policy makers and Malawian government should focus on improving the toilet facilities, access to electricity and providing more mosquito bed nets, mostly for the individuals who live in rural area and at low altitude. In addition, edu cation of the mothers should be prioritized so that they are able to treat and take care of their children, especially those in the age group 6 23 months, as they are more vulnerable. Understanding the relationship between anaemia and malaria together with other factors associated with the two diseases can provide useful insights to the government and policy makers in planning, controlling and the elimination of both diseases. In addition, the statistical model used in this study will help other re searchers to compare findings and referencing. Future research could make use of spatial joint models for malaria and anaemia simultaneously in order to investigate the correlation between the two diseases by geographical location.

7. Limitation

The MMIS 2017, under household member dataset, did not provide the nutritional status of the child, cough and diarrhea effects, which might be associated with anaemia. It would be helpful to use a dataset that included those factors. The second limitation is that the dataset was cross sectional. It would have been ideal to have a longitudinal data set to study the change in factors and prevalence over time.

8. Recommendations

In order to develop effective intervention strategies to help reduce or alleviate anaemia and malaria in children aged 6 59 months, it is rec ommended that the policy makers and government make a concerted effort in educating the mothers on malaria and anaemia, and improving health care, toilet facilities and mosquito bed nets, especially for in dividuals from rural areas. This can be done through various platforms such as social media, television, radio, roadshows and even rural based workshops.

Declarations

Author contribution statement

Rugiranka Tony Gaston: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Faustin Habyarimana and Shaun Ramroop: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by the Health Economics and HIV/AIDS Research Division (HEARD).

Data availability statement

The data that has been used is confidential.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

The authors acknowledge the National Malaria Control Programme and the ICF through the DHS programme for providing the data.

References

Abegunde, D., Stanciole, A., 2006. An Estimation of the Economic Impact of Chronic Non Communicable Diseases in Selected Countries. World Health Organization, Department of Chronic Diseases and Health Promotion, 2006.

Adebayo, S.B., Gayawan, E., Heumann, C., Seiler, C., 2016. Joint modeling of Anaemia and Malaria in children under five in Nigeria. Spat. Spatio-Temporal Epidemiol. 17, 105–115.

Agresti, A., 2015. Foundations of Linear and Generalized Linear Models. John Wiley & Sons, Inc., Hoboken, New Jersey.

Alegana, V.A., Wright, J.A., Nahzat, S.M., Butt, W., Sediqi, A.W., Habib, N., Snow, R.W., Atkinson, P.M., Noor, A.M., 2014. Modelling the incidence of Plasmodium vivax and Plasmodium falciparum malaria in Afghanistan 2006–2009. PloS One 9 (7), e102304.

Ayele, D.G., Zewotir, T.T., Mwambi, H.G., 2014. Modelling the joint determinants of a positive malaria Rapid Diagnosis Test result, use of mosquito nets and indoor residual spraying with insecticide. Occupat. Healt. Souther. Afric. 20 (4), 20–27.

Ayoya, M.A., Ngnie-Teta, I., Séraphin, M.N., Mamadoultaibou, A., Boldon, E., Saint-Fleur, J.E., Koo, L., Bernard, S., 2013. Prevalence and Risk Factors of Anemia Among Children 6–59 Months Old in Haiti. Anemia, p. 502968, 2013.

Bennett, A., Kazembe, L., Mathanga, D.P., Kinyoki, D., Ali, D., Snow, R.W., Noor, A.M., 2013. Mapping malaria transmission intensity in Malawi, 2000–2010. Am. J. Trop. Med. Hyg. 89 (5), 840–849.

Bjorkman, A., 2002. Malaria associated anaemia, drug resistance and antimalarial combination therapy. Int. J. Parasitol. 32 (13), 1637–1643.

Brabin, B.J., Kalanda, B.F., Verhoeff, F.H., Chimsuku, L.H., Broadhead, R.L., 2004. Risk factors for fetal anaemia in a malarious area of Malawi. Ann. Trop. Paediatr. 24 (4), 311–321.

Buchwald, A.G., Walldorf, J.A., Cohee, L.M., Coalson, J.E., Chimbiya, N., Bauleni, A., Nkanaunena, K., Ngwira, A., Kapito-Tembo, A., Mathanga, D.P., Taylor, T.E., 2016. Bed net use among school-aged children after a universal bed net campaign in Malawi. Malar. J. 15 (1), 127.

Calis, J.C., Phiri, K.S., Faragher, E.B., Brabin, B.J., Bates, I., Cuevas, L.E., de Haan, R.J., Phiri, A.I., Malange, P., Khoka, M., Hulshof, P.J., 2016. Research article (new england journal of medicine) severe anaemia in Malawian children. Malawi Med. J. 28 (3), 99–107.

Carneiro, I.A., Smith, T., Lusingu, J.P., Malima, R., Utzinger, J., Drakeley, C.J., 2006. Modeling the relationship between the population prevalence of Plasmodium falciparum malaria and anaemia. Am. J. Trop. Med. Hyg. 75 (2 suppl), 82–89.

Chirombo, J., Lowe, R., Kazembe, L., 2014. Using structured additive regression models to estimate risk factors of malaria: analysis of 2010 Malawi malaria indicator survey data. PloS One 9 (7), e101116.

Chitunhu, S., Musenge, E., 2015. Direct and indirect determinants of childhood malaria morbidity in Malawi: a survey cross-sectional analysis based on malaria indicator survey data for 2012. Malar. J. 14 (1), 265.

Crawley, J., 2004. Reducing the burden of anemia in infants and young children in malaria-endemic countries of Africa: from evidence to action. Am. J. Trop. Med. Hyg. 71 (2 suppl), 25–34.

El Kishawi, R.R., Soo, K.L., Abed, Y.A., Muda, W.A.M.W., 2015. Anaemia among children aged 2–5 years in the Gaza Strip-Palestinian: a cross sectional study. BMC Publ. Health 15 (1), 319.

Gaston, R.T., Ramroop, S., Habyarimana, F., 2018. Determinants of factors associated with anaemia among children under five years in Lesotho. Afr. Popul. Stud. 32 (1).

Gaston, R.T., Ramroop, S., 2020. Prevalence of and factors associated with malaria in children under five years of age in Malawi, using malaria indicator survey data. Heliyon 6 (5), e03946.

Gueorguieva, R., 2001. A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. Stat. Model. Int. J. 1 (3), 177–193.

Habyarimana, F., Zewotir, T., Ramroop, S., 2016. Joint modeling of poverty of households and malnutrition of children under five years from demographic and health survey data: case of Rwanda. J. Econom. Behav. Studi. 8 (2), 108–114.

Hajison, P.L., Feresu, S.A., Mwakikunga, B.W., 2018. Malaria in children under-five: a comparison of risk factors in lakeshore and highland areas, Zomba district, Malawi. PloS One 13 (11), e0207207.

Hedeker, D., 2005. Generalized Linear Mixed Models. Encyclopedia of Statistics in Behavioral Science. John Wiley and Sons, Ltd.

Hershey, C.L., Florey, L.S., Ali, D., Bennett, A., Luhanga, M., Mathanga, D.P., Salgado, S.R., Nielsen, C.F., Troell, P., Jenda, G., Ye, Y., 2017. Malaria control interventions contributed to declines in malaria parasitemia, severe anaemia, and all-cause mortality in children less than 5 years of age in Malawi, 2000–2010. Am. J. Trop. Med. Hyg. 97 (3 Suppl), 76–88.

Kabaghe, A.N., Chipeta, M.G., Terlouw, D.J., McCann, R.S., Van Vugt, M., Grobusch, M.P., Takken, W., Phiri, K.S., 2017. Short-term changes in anaemia and malaria parasite prevalence in children under 5 years during one year of repeated cross-sectional surveys in rural Malawi. Am. J. Trop. Med. Hyg. 97 (5), 1568–1575.

Korenromp, E.L., Armstrong-Schellenberg, J.R., Williams, B.G., Nahlen, B.L., Snow, R.W., 2004. Impact of malaria control on childhood anaemia in Africa–a quantitative review. Trop. Med. Int. Health 9 (10), 1050–1065.

Kanchana Sr, M., Ahuja, S., Nagaraj, N., 2018. Prevalence and risk factors of anemia in under five-year-old children in children's hospital. Int J Contemp Pediatr 5, 499–502.

Kazembe, L.N., 2007. Spatial modelling and risk factors of malaria incidence in northern Malawi. Acta Trop. 102 (2), 126–137.
Kejo, D., Petrucka, P.M., Martin, H., Kimanya, M.E., Mosha, T.C., 2018. Prevalence and predictors of anemia among children under 5 years of age in Arusha District, Tanzania. Pediatr. Health Med. Therapeut. 9, 9.
Kuziga, F., Adoke, Y., Wanyenze, R.K., 2017. Prevalence and factors associated with anaemia among children aged 6 to 59 months in Namutumba district, Uganda: a cross-sectional study. BMC Pediatr. 17 (1), 25.
Magalhaes, R.J.S., Clements, A.C., 2011. Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition, malaria, and helminth infections in West Africa. PLoS Med. 8 (6), e1000438.
Mathanga, D.P., Mwandama, D.A., Bauleni, A., Chisaka, J., Shah, M.P., Landman, K.Z., Lindblade, K.A., Steinhardt, L.C., 2015. The effectiveness of long-lasting, insecticide-treated nets in a setting of pyrethroid resistance: a case–control study among febrile children 6 to 59 months of age in Machinga District, Malawi. Malar. J. 14 (1), 457.
McCuskee, S., Brickley, E.B., Wood, A., Mossialos, E., 2014. Malaria and macronutrient deficiency as correlates of anemia in young children: a systematic review of observational studies. Annal. Global Heal. 80 (6), 458–465.
McGann, P.T., Williams, A.M., Ellis, G., McElhinney, K.E., Romano, L., Woodall, J., Howard, T.A., Tegha, G., Krysiak, R., Lark, R.M., Ander, E.L., 2018. Prevalence of inherited blood disorders and associations with malaria and anemia in Malawian children. Blood advances 2 (21), 3035–3044.
McLean, E., Cogswell, M., Egli, I., Wojdyla, D., De Benoist, B., 2009. Worldwide prevalence of anaemia, WHO vitamin and mineral nutrition information system, 1993–2005. Publ. Health Nutr. 12 (4), 444–454.
Molenberghs, G., Verbeke, G., 2006. Models for Discrete Longitudinal Data. Springer-Verlag, New York.
Nkoka, O., Chipeta, M.S., Chuang, Y.C., Fergus, D., Chuang, K.Y., 2019. A comparative study of the prevalence of and factors associated with insecticide-treated nets usage among children under 5 years of age in households that already own nets in Malawi. Malar. J. 18 (1), 43.
National Statistical Office (NSO) [Malawi] and ICF, 2017. Malawi Demographic and Health Survey 2015-16. Zomba, Malawi, and Rockville. NSO and ICF, Maryland, USA.
National Malaria Control Programme (NMCP) [Malawi] and ICF, 2018. Malawi Malaria Indicator Survey 2017. Lilongwe: NMCP and ICF.
Noland, G.S., Ayodo, G., Abuya, J., Hodges, J.S., Rolfes, M.A., John, C.C., 2012. Decreased prevalence of anaemia in highland areas of low malaria transmission after a 1-year interruption of transmission. Clin. Infect. Dis. 54 (2), 178–184.
Ntenda, P.A.M., Chuang, K.Y., Tiruneh, F.N., Chuang, Y.C., 2017. Multilevel analysis of the effects of individual-and community-level factors on childhood anaemia, severe anaemia, and hemoglobin concentration in Malawi. J. Trop. Pediatr. 64 (4), 267–278.
Ohrt, C., Sutamihardja, M.A., Tang, D., Kain, K.C., 2002. Impact of microscopy error on estimates of protective efficacy in malaria-prevention trials. J. Infect. Dis. 186 (4), 540–546.
Perkins, D.J., Were, T., Davenport, G.C., Kempaiah, P., Hittner, J.B., Ong'echa, J.M., 2011. Severe malarial anemia: innate immunity and pathogenesis. Int. J. Biol. Sci. 7 (9), 1427.
Reithinger, R., Ngondi, J.M., Graves, P.M., Hwang, J., Getachew, A., Jima, D., , Ethiopia Malaria Indicator Survey Working Group, Amena, M., Bergeron, L., Bilak, H., Chirwa, B., 2013. Risk factors for anaemia in children under 6 years of age in Ethiopia: analysis of the data from the cross-sectional Malaria Indicator Survey, 2007. Trans. R. Soc. Trop. Med. Hyg. 107 (12), 769–776.
Roberts, D.J., Zewotir, T., 2020. Copula geoadditive modelling of anaemia and malaria in young children in Kenya, Malawi, Tanzania and Uganda. J. Health Popul. Nutr. 39 (1), 1–14.
Safeukui, I., Gomez, N.D., Adelani, A.A., Burte, F., Afolabi, N.K., Akondy, R., Velazquez, P., Holder, A., Tewari, R., Buffet, P., Brown, B.J., 2015. Malaria induces anaemia through CD8+ T cell-dependent parasite clearance and erythrocyte removal in the spleen. mBio 6 (1), e02493–e02514.
Seyoum, S., 2018. Analysis of prevalence of malaria and anaemia using bivariate probit model. Annals Data Sci. 5 (2), 301–312.
Teh, R.N., Sumbele, I.U.N., Meduke, D.N., Ojong, S.T., Kimbi, H.K., 2018. Malaria parasitaemia, anaemia and malnutrition in children less than 15 years residing in different altitudes along the slope of Mount Cameroon: prevalence, intensity and risk factors. Malar. J. 17 (1), 336.
Wanzira, H., Katamba, H., Okullo, A.E., Agaba, B., Kasule, M., Rubahika, D., 2017. Factors associated with malaria parasitaemia among children under 5 years in Uganda: a secondary data analysis of the 2014 Malaria Indicator Survey dataset. Malar. J. 16 (1), 191.
White, N.J., 2018. Anaemia and malaria. Malar. J. 17 (1), 1–17.
Wongsrichanalai, C., Barcus, M.J., Muth, S., Sutamihardja, A., Wernsdorfer, W.H., 2007. A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT). Am. J. Trop. Med. Hyg. 77 (6 Suppl), 119–127.
World Health Organization, 2011. Haemoglobin Concentrations for the Diagnosis of Anaemia and Assessment of Severity (No. WHO/NMH/NHD/MNM/11.1). World Health Organization.
World Health Organization, 2015. Investing to Overcome the Global Impact of Neglected Tropical Diseases: Third WHO Report on Neglected Tropical Diseases 2015. World Health Organization. Vol. 3.
World Health Organization, 2016. World Malaria Report 2016. World Health Organization, Geneva. Geneva.
WHO, 2017. Nutritional Anaemias: Tools for Effective Prevention and Control. World Health Organisation, Geneva, Switzerland.
World Malaria Report, 2018. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO. https://apps.who.int/iris/bitstream/handle/10665/275867/9789241565653-eng.pdf. (Accessed 15 December 2019).
Yimgang, D.P., Buchwald, A.G., Coalson, J.E., Walldorf, J.A., Bauleni, A., Kapito-Tembo, A., Mathanga, D.P., Taylor, T.E., Laufer, M.K., Cohee, L.M., 2021. Population Attributable Fraction of Anemia Associated with P. Falciparum Infection in Children in Southern Malawi. The American Journal of Tropical Medicine and Hygiene, tpmd201120.
Zgambo, M., Mbakaya, B.C., 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), e0175537.
Zhang, D., Lin, X., 2008. Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. In: Random Effect and Latent Variable Model Selection. Springer, New York, NY, pp. 19–36.

BMC Public Health

RESEARCH

Open Access

Joint modelling of anaemia and stunting in children less than five years of age in Lesotho: a cross-sectional case study

Rugiranka Tony Gaston[1,2*], Faustin Habyarimana[1] and Shaun Ramroop[1]

Abstract

Background: Anaemia and stunting remain jointly a serious health issue worldwide especially in developing countries. In Lesotho, their prevalence is high, particularly among children less than 5 years of age.

Objectives: The primary objective was to determine the association between anaemia and stunting, and identify factors relating to both conditions among children younger than 5 years in Lesotho.

Methods: This cross-sectional study used secondary data from 3112 children collected during the 2014 Lesotho Demographic Health Survey (LDHS). Haemoglobin (Hb) levels were adjusted for altitude and a level less than 11 g per deciliters (11 g/dl) was determined as the cutoff for being anaemic. A child with the height-for-age z score (HAZ) below minus two standard deviations (SD) was considered to have stunting. We linked factors relating to anaemia and stunting using a multivariate joint model under the scope of the generalized linear mixed model (GLMM).

Results: The prevalence of anaemia and stunting in children younger than 5 years were 51% and 43% respectively. The multivariate results revealed a strong association between anaemia and stunting. In addition, maternal education, urban vs. rural residence, wealth index and childbirth weight significantly impacted childhood stunting or malnutrition, while having fever and/or diarrhoea was linked to anaemia. Lastly, age was shown to have a significant effect on both stunting and anaemia.

Conclusion: Anaemia and stunting or malnutrition showed linked longitudinal trajectories, suggesting both conditions could lead to synergetic improvements in overall child health. Demographic, socio-economic, and geographical characteristics were also important drivers of stunting and anaemia in children younger than 5 years. Thus, children living in similar resources settings as Lesotho could benefit from coordinated programs designed to address both malnutrition and anaemia.

Keywords: Malnutrition, Anaemia, Correlation, Multivariate joint model, Children less than 5 years of age and LDHS

Introduction

Malnutrition and anaemia remain major health problems worldwide especially in developing countries [5, 26, 48]. These conditions intersect and are linked to morbidity and mortality worldwide, particularly in pregnant women and children [6, 10, 22]. Despite progress in the fight against childhood malnutrition and anaemia, several challenges remain. Globally, in 2017, 151 million children younger than 5 years. Among those children, three quarters were from the South-East Asian and African regions [42]. In 2011, the global prevalence of anaemia in children younger than 5 years was approximately 43%, with

*Correspondence: rugiratony@gmail.com
[1] School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Private Bag X01, Pietermaritzburg CampusScottsville 3209, South Africa
Full list of author information is available at the end of the article

highest percentages also found in the African and South Asian regions [24, 30, 47].

In Lesotho, childhood anaemia and malnutrition remain a concerning health problem with 51% of children being anaemic and 53% malnourished [31]. Improving both the nutritional and anaemic status in children younger than 5 years is critical to ensure high quality of life to future contributors and leaders of the country [14].

Anaemia is defined by low blood level of haemoglobin (Hb). According to the World Health Organization (WHO), children younger than 5 years are considered anaemic when their Hb level, adjusted for altitude, is less than 11 g/dl [17, 24], WHO 2015. The cause of anaemia is multifactorial and iron deficiency is considered the most fundamental cause in about 50% of cases. Insufficient folate, vitamin B12, protein deficiencies, nutrients and some diseases, such as malaria and diarrhoea among others, can also increase the risk of anaemia [12, 14]. Childhood anaemia can negatively impact cognitive development, performance in school, physical growth and immunity [14, 25, 40].

Malnutrition develops with either over- or under consumption of food but herein, it is defined as an insufficient intake of nutrients and/ or other minerals. In developing countries, a low nutritional status of a child is an indicator of health problems [29]. Childhood malnutrition leads to long-term negative effects, such as poor performance at school, delayed psychomotor development, lower capacity for work and reduced quality of life in adulthood [10, 21, 34]. The nutritional status of a child is mainly evaluated based on different anthropometric indicators in reference to WHO growth standards [31]. Stunting (height-for-age) indicates chronic or long-term malnutrition, wasting (low-weight-for-height) is linked to low food intake and/or illness and is described as acute malnutrition, while an underweight child (weight-for-age) can be either stunted, wasted or both [31]. The study examines the association between anaemia and stunting as an indicator of long-term malnutrition [3].

According to the WHO, an individual whose z-score falls below -2 SD is considered malnourished [46]. Stunting is categorised as moderate acute malnutrition (MAM) when the height-for-age z score (HAZ) is less than minus two SD, and severe acute malnutrition (SAM) when HAZ is below minus two SD [7, 8, 23, 31]. Literature indicates that most studies have evaluated anaemia and stunting independently [14, 28, 38, 45]. The independent models are sufficient in modelling anaemia and stunting but are inadequate for addressing the association between the two conditions. The generalized linear mixed model (GLMM) were extended to evaluate joint trajectories of repeated measures [11, 33],where the random effect can be used to evaluate the correlation structure between several response variables and can better control for type 1 error [4, 15, 20]. In addition, the multivariate joint model has the ability to address multivariate questions [18, 33].

The studies aimed to understand the association between anaemia and stunting in children younger than 5 years are limited [2, 32, 36, 37, 39, 43, 48]. We found no study in literature that has utilized the joint model for anaemia and stunting in children younger than 5 years of age in Lesotho. Therefore, we expanded existing models by examining the longitudinal interdependent relationship of anaemia and stunting among children younger than 5 years in Lesotho using joint multivariate GLMM. Understanding the link between anaemia and stunting and other factors will help prioritize efforts for policy-makers and donors that aim to improve global child health.

Materials and methods

Study area

Lesotho is small landlocked country with a surface area of 30355km^2 and a population of about of 2.2 million and is surrounded by only South Africa. Lesotho has ten politico-administrative districts with Maseru as the capital city, and is ruled by a King as the head of state and the Prime Minister as head of the government. The kingdom of Lesotho is known for its abundant water resources and high altitude. However, the country has high unemployment rates, high prevalence of HIV and AIDS, poverty, food insecurity, and the burden from other diseases [31]. The country is also vulnerable to natural disasters and climate change such as droughts and heavy rain and flooding (Renzaho 2015).

Data source and sampling techniques

This study was cross-sectional and used secondary data from the LDHS, conducted from September to December 2014. The ethical approval for the 2014 LDHS was assessed and confirmed by the Lesotho Ministry of Health Research and Ethics Committee together with support of the Institutional Review Board of ICF International.

The 2014 LDHS was representative at national, urban and rural areas as well as four ecological zones, and each of Lesotho's 10 districts [31]. The sample was stratified and selected in two stages. The stratification was executed by separating each district into urban and rural areas. The overall of 20 sampling strata were designed, and thereafter samples were selected independently in each sampling stratum by following a two-stage sampling method. The first stage included a

random selection of 400 clusters from the enumeration area (EAs). Out of the 400 clusters, 118 clusters were from urban areas and 282 clusters from rural areas. The second stage involved systematic sampling of 9942 randomly selected households covering all EAs. Out of these, 25 households were then selected from each enumeration area [31].

The 2014 survey included all residents or visitors who were in the selected household the night before the interview. All women were included in the survey based on the condition that they had never been married, not currently pregnant, and had not given birth in the previous 2 months. Children aged between 0–59 months from mothers living at or visiting the households the night before were included in survey. All surveyed children were measured for height and tested for anaemia, under the supervision of their parents or guardians. Exclusion criteria, comprised the anomaly in BMI, height, and weight-for-age measurements. With respect to anaemia testing, all children less than 6 months were not included in the survey. The present study used a total weighted number of 1138 children for anaemia, and 1297 for stunting. The exclusion and system missing values were considered as missing values and were consequently ignored [31].

The calculation of sample size was determined using the statistical formula:

$$n = \frac{z^2 p(1-p)}{d^2}$$

where n = sample size, p = prevalence of anaemia, z = z-value at 90% confidence (= 1.96), and d = level of significance (= 5%). Out of the total weighted of 3112 children drawn as a sample, only 1292 children were used in this study, and 1816 were considered as the missing value.

In the sampling procedure, the number of women surveyed in each region should present the size of the total sample in the proportion to the size of the region [31]. However, some regions such as Qacha's Nek, Quthing are less populated, while Maseru and Leribe are heavily populated. Since the population in each region was not of equally weighting, individuals surveyed in each region should contribute equivalently to the total per region. Hence, this unweighted distribution does not accurately represent the exact population. Therefore, the weighted samples were used to infer the national status and account for the complex sample design from the LDHS data set as well as to account for the lack of adequate representation in the sample. More details on calculation of sampling size and sampling weight can be found in Lesotho's report of demographic and health survey [31].

Anaemia and nutrition assessment

Anaemia testing

All children younger than 5 years surveyed were tested for anaemia under the supervision of their parents or guardians. Blood collection by finger-or heel-prick was performed by professional nurses using a spring-loaded sterile lancet. A blood drop was gathered in microcuvette, and the Hb was measured using a HemoCue analyser (company, city of equipment). The lancet, microcuvette, gloves and alcohol swabs were used once for hygiene safety. Results were obtained within 10 min and shared verbally with the children's parents or guardians and recorded as a hard copy, captured on the Biomarker questionnaire and indicated on handout that explains causes and counteractions of anaemia, which were left with the family [31].

All parents or guardians of children with a Hb level less than 7 g/dl were told to take the children to the nearest healthcare facility for follow-up [31].

Nutrition status assessment

Height was measured with a tape board and weight with an electronic balance (model of equipment, city, country) provided by UNICEF. The weight of the children was estimated utilizing a Seca gauging scale (model of equipment, city, country), which aligned to zero. For weighting, parents or guardians unclothed their children or keep them in light clothing. For children unable to stand, child weight was calculated based on difference in parent weight compared to weight of parent holding child. The height measurement was carried out using a short board, which was lying down or standing on a level ground surface. Children were measured without shoes and headgear, standing against a board. Children less than 87 cm were measured in supine position. From the children's weight, height and age was calculated their nutritional status (i.e. weight-for-age, height-for-age, and weight-for-height) based on the WHO guidelines [9, 12, 31].

Data analysis

Dependent variable

The two response variables for the study were measures of anaemia and stunting. Anaemia in children can be grouped as severe anaemia when Hb level is less than 7 g/dl; moderate anaemia with Hb level between 7 and 8.9 g/dl; and mild anaemia with Hb level between 9 and 10.9 g/dl. However, in the scope of this study, a child was determined as either as anaemic or not (i.e., Hb level above or below altitude adjusted threshold of 11 g/dl) (WHO 2015). Nutrition status of a child cab be described as non-malnourished; moderate acute malnutrition; or severe acute malnutrition. Yet, the second response variable

was recorded as binary exposure since the interest of the study was to check if the child is stunted or normal [23, 31].

Independent variables

The independent variables include socio-economic and demographic factors, and were used in previous studies on childhood anaemia or stunting [14, 15, 28, 39]. Demographic variables included sex of child (male or female),whether the child had fever, cough or diarrhoea in the 2 weeks prior to the survey; the birth order of the child; the childbirth weight; and the age in months. Socio-economic variables included reception of Vitamin A supplementation in the 6 month prior to the survey; whether the child visited a health facility in past 2 weeks prior to the survey; maternal education (no education, primary, post primary); place of residence (urban, rural); whether a household used electricity; main material of floor, wall, and roof; wealth index of the household; mother's body mass index (BMI); toilet facility; and source of drinking water.

Statistical analysis

Multivariate joint GLMM models were built to identify the correlation between two response variables (anaemia and stunting) and assess their association with demographic and socio-economic factors. Using Statistical Package for the Social Sciences (SPSS version 25.0), we cross tabulated and summarised the data with frequencies and percentages. We then conducted univariate analysis to selected variables showing relationship to either responses variables based having a p-value less than 0.2; this selection helps to account for multicollinearity and confounders between covariates [12–14].

Selected variables were then included in the multivariate joint GLMM conducted with SAS 9.4 using PROC GLIMMIX. This procedure is able to join models with two response variables that have similar distribution and link function. Based on the convergence criteria, the unstructured (UN) convergence was chosen seemed to be the best for the analysis. The final multivariate joint analysis retained independent variables showing p-value < 0.05. Interactions were evaluated based on Akaike information criteria (AIC); but were not significant [15, 33].

Model formulation

Two responses variables, stunting and anaemia in children less than 5 years of age were examined: the first variable x_{i1} was stunting, where one (1) indicate the presence stunting and zero its' absence; and the second x_{i2} allocate a one to the presence of anaemia, versus zero (0) for its' absence. We assume the outcome to be from a bivariate Bernoulli distribution, with p_{i1} as the likelihood of stunting occurring in a child i and p_{i2} as the probability of anaemia occurring in a child i. Therefore, the binary generalized linear model can be expressed as follow:

$$f_1(\psi_{i1}) = Y_{i1}\theta_1 + Z_{i1}u_1 \tag{1}$$

$$f_2(\psi_{i1}) = Y_{i2}\theta_2 + Z_{i2}u_2 \tag{2}$$

where, θ_1 and θ_2 are assumed to be the vectors of fixed effects, u_1 and u_2 are the vectors of the random effects. In addition,Y_{i1},Y_{i2}, Z_{i1} and Z_{i2} are the designed matrices for fixed and random effects, respectively. Hence, the model's equation of the variance–covariance matrices for multivariate normal distribution (MVN) is shown as follows:

$$U = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim i.i.d.MVN(0,\sigma) = MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}\right) \tag{3}$$

where the σ_{11} and σ_{22}, are the variance components of stunting and anaemia respectively, while σ_{12} and σ_{21} are the covariance components between these conditions. When the covariance components from Eq. (3), σ_{12}= σ_{21} =0, the multivariate joint model becomes standard separated generalized linear mixed [19, 33].

Model testing and and goodness-of-fit

Covariance structure of the random effects

The test of random effect checks whether specific covariates should be included within the random effect structure of a model by testing the model against the hypothesis that the variance of population distribution is zero. To illustrate, assume that we want to test the variance parameter σ^2 for a GLMM with a single random effect: Since the variance cannot be negative, zero is the boundary of the parameter space, and model variance can be tested against zero using a one-sided hypothesis [41, 44].

The null hypothesis to be tested is $H_0 : \sigma^2 = 0$, against $H_1 : \sigma^2 \neq 0$, and this can tested using the likelihood ratio test, the Wald test or the score test. These are asymptotically comparable under the null hypothesis and follow a distribution with a given degree of freedom. In addition, these tests can determine if the random effects in a GLMM contribute [41, 44].

Goodness-of-fit tests

The goodness-of-fit can be evaluated by a likelihood ratio test against an alternative saturated model, which is the most complex model explained by the data. The saturated model can be expressed as $\boldsymbol{E(Y_{ni}/\beta,\theta_n) = Y_{ni}}$ for all n and i. When the dependent variables are categorical, the goodness-of-fit test for the data is Pearson chi-squared

test [41]. The Pearson chi-squared test can be written as follows:

$$\chi^2 = \sum_{i=1}^{I} n_i \sum_{j=0}^{J} \frac{\left(y_{ij} - \widehat{\pi}_{ij}\right)^2}{\widehat{\pi}_{ij}} \quad (4)$$

where n_i is the number of participants with the same values on the covariates, y_{ij} and $\widehat{\pi}_{ij}$ are respectively the observed and expected quantity of the participants in the group i responding in category j [41].

Results and interpretation

Univariate results

This study included a total weighted number of 3112 children aged between 0–59 months. The prevalence of anaemia and stunting was 51% and 43% respectively, with 35.2% of children having both conditions. Tables 1 and 2 indicate the frequency and percentage of anaemia with corresponding independent variables. Childhood stunting and anaemia were both related to child age but in opposing direction (i.e., stunting increased with age, while anaemia decreased with age), while stunting itself was further associated with sex, having visited healthcare facilities, maternal education, wealth index, access to electricity, drinking water and waste water management, and dwelling characteristics (wall and floor material); while anaemia was more specifically related to fever in last 2 weeks, recent diarrhoea, and roof characteristics of the dwelling. The prevalence of anaemia was higher in children aged between 0–19 months (62.7%), and then decreased in children aged 20–39 months (55.0%) and then again in those aged 40–59 months (44.4%). The anaemia prevalence was higher in children who experienced fever in the last 2 weeks (62.8%) compared to children who did not (37.2%). The prevalence of stunting was lower in younger children (age group 0–19 months, 17.8%), The prevalence of stunting was higher from mothers with low education (72.7%), but increased as children aged (20–39 months, 33.3%; and 40–59 months, 31.2%).

The prevalence of stunting was higher in children from mothers with low education (72.7%), and reduced as the level of mother's education increased by primary (31.3%) and post-primary (21.7%) respectively.

Multivariate results

Table 3 presents the estimation results for the fixed effects of the joint GLMM several socio-economic and demographics factors showed significant relationships with both stunting and anaemia. Child's age had a significant effect on both anaemia and stunting. Children aged less than 20 months were less likely to be stunted (OR: 0.44, 95% CI: 0.298; 0.638) but did not differ in their risk of anaemia compared to the reference group (40–59 months). However, children aged 20–39 months did not differ in their risk of stunting but were more likely to be anaemic (OR: 1.7, 95% CI: 1.207; 2.396) compared to the reference. Fever and diarrhoea were not linked to stunting but were both associated with a higher risk of anaemia (OR: 0.491, 95% CI: 0.341; 0.707 and OR: 0.609, 95% CI: 0.410; 0.905, respectively). High birth weight (≥ 2500 g) and living in a rural area were protective against stunting (OR: 0.24, 95% CI: 0.182; 0.452 and OR: 0.52, 95% CI: 0.333; 0.814, respectively) compared to lower birthweight children; while the odds of being stunted increased with higher levels of poverty (OR: 3.5, 95% CI: 2.149; 5.703) compared to those from wealthier homes. Children from the middle wealth tertile were 2.9 times more likely to be stunted compared to those from the top wealth tertile. Lastly, lower maternal education was also associated with stunting.

The variance components and covariance between anaemia and stunting are presented in Table 4. The covariance coefficient estimate of 1.000 indicated a positive relationship between the two conditions, meaning that changes in either nutrition or anaemia in a child impacts the likelihood of both diseases. In addition, the odds ratio of 2.718 confirmed that the two conditions are highly associated. The overall fitted model was highly significant as the coefficient of covariance parameter indicated the p-value < 0.001. Hence, including the random effect in the model was shown to be very important [16, 33], Zhang and Lin 2008.

The test covariance parameters based on pseudo-likelihood rejected the null hypothesis of zero correlation with Pearson chi-squared test = 2644.470 and p-value < 0.001. This revealed that the association between anaemia and stunting was significant and not zero [41]. Furthermore, the results from the fitted statistics for conditional distribution indicated the Pearson chi-squared = 2123.070 with 0.94 ° of freedom. This is an indication of a good variability in the dataset and residual over dispersion was not present [16, 33].

Discussion

This cross-sectional study used secondary data from 2014 LDHS. To our knowledge, this was the first study to simultaneously model the association between anaemia and stunting in children less than 5 years of age in Lesotho. The study utilized a multivariate joint model under the scope of GLMM to association both diseases and explore their associated socio-economic and demographic factors. Anaemia and stunting show a significant positive association confirming that the two diseases should be considered interrelated health problems in children where these diseases are more likely to coexist

Table 1 Childhood anaemia by categorical variable

Variables	Categories	Anaemic	Not anaemic	*p*-value
Sex of the child	Male	56.2% (301)	43.8% (235)	0.185
	Female	52.2% (315)	47.8% (288)	
Child's age in months	6–19	62.7% (227)	33.3% (135)	0.000
	20–39	55.0% (224)	45.0% (183)	
	40–59	44.4% (164)	55.6% (205)	
Child's birth weight	< 2500 g	58.5% (48)	41.5% (34)	0.359
	≥ 2500 g	53.2% (467)	46.8% (410)	
Had fever in last 2 weeks	Yes	62.8% (120)	37.2% (71)	0.008
	No	52.3% (493)	47.7% (449)	
Had diarrhoea recently	Yes	62.7% (94)	37.3% (54)	0.024
	No	52.8% (520)	47.2% (465)	
Had cough in last 2 weeks	Yes	54.4% (184)	45.6% (154)	0.883
	No	54.0% (429)	46.0% (366)	
Received vitamin A in last 6 months	Yes	54.3% (357)	45.7% (301)	0.858
	No	53.7% (253)	46.3% (218)	
Birth order	1st	53.2% (231)	46.8% (203)	0.728
	2–3	56.0% (261)	44.0% (205)	
	4–5	52.1% (85)	47.9% (78)	
	≥ 6	51.3% (39)	48.7% (37)	
Mother's BMI	< 18.5	56.0% (14)	44.0% (11)	0.830
	≥ 18.5	53.8% (596)	46.2% (511)	
Mother's education level	No education	53.8% (7)	46.2% (6)	0.193
	Primary	51.1% (267)	48.9% (255)	
	Post primary	56.6% (341)	43.4% (262)	
Visited health facility	Yes	52.7% (479)	47.3% (430)	0.069
	No	59.4% (136)	40.6% (93)	
Wealth Index	Poor	58.0% (280)	42.0% (203)	0.059
	Middle	53.1% (127)	46.9% (112)	
	Rich	50.1% (209)	49.9% (208)	
Place of residence	Rural	55.5% (457)	44.5% (366)	0.104
	Urban	50.2% (158)	49.8% (157)	
Household with electricity	Yes	50.0% (150)	50.0% (150)	0.102
	No	55.5% (465)	44.5% (373)	
Toilet facility	Toilet with flush	53.7% (306)	46.3% (264)	0.148
	Pit latrine	48.3% (87)	52.7% (93)	
	No facility	57.1% (222)	42.9% (167)	
Type of drinking water	Tap water	54.0% (274)	46.0% (233)	0.166
	Protected water	49.1% (109)	50.9% (113)	
	Unprotected	57.0% (233)	43.0% (176)	
Main roof material	Thatch/Palm leaf	58.8% (248)	41.2% (174)	0.014
	Corrugated metal	52.3% (335)	47.7% (306)	
	Stick &mud	42.7% (32)	57.3% (43)	
Main wall material	Wood/Mud	56.6% (233)	43.4% (179)	0.270
	Bricks	51.5% (202)	48.5% (190)	
	Cement /Block	52.6% (120)	47.4% (108)	
Main floor material	Earth/Sand	52.8% (152)	47.2% (136)	0.441
	Mud block/Wood	56.6% (233)	43.4% (179)	
	Cement/ Block	52.5% (229)	47.5% (207)	

and interinfluence their manifestations. Thus, coordinated interventions aiming to improve both stunting and anaemia are likely to produce synergetic effects on child health. The association between the two conditions can be interpreted as an indication of chronic malnutrition which might cause iron deficiency. Similar results were described in studies by Yang et al. [48], Gari et al. [13], Mohammed et al. [32], Rahman et al. [37] and Rivadeneira et al. [39]. Our findings also indicate that child age has a significant effect on both anaemia and malnutrition but impact different age groups. The chance of having anaemia or stunting reduced as the children aged.

A possible explanation to this issue is due to the fact that the immune systems of children are still developing

Table 2 Childhood stunting by categorical variable

Variables	Categories	Stunted (Malnourished)	No stunted (Nourished)	*p*-value
Sex of the child	Male	30.7% (184)	69.3% (416)	0.002
	Female	23.0% (160)	77.0% (537)	
Child's age in months	0–19	17.8% (92)	82.2% (426)	0.000
	20–39	33.3% (136)	66.7% (273)	
	40–59	31.2% (115)	68.8% (254)	
Child's birth weight	< 2500 g	50.6% (45)	49.4% (44)	0.000
	≥ 2500 g	22.4% (227)	78.6% (787)	
Had fever in last 2 weeks	Yes	21.2% (44)	78.8% (164)	0.054
	No	27.6% (299)	72.4% (785)	
Had diarrhoea recently	Yes	27.2% (44)	72.8% (118)	0.855
	No	26.5% (299)	73.5% (830)	
Had cough in last 2 weeks	Yes	26.1% (97)	73.9% (275)	0.807
	No	26.7% (246)	73.3% (674)	
Received vitamin A in last 6 months	Yes	26.0% (176)	74.0% (500)	0.633
	No	27.2% (166)	72.8% (444)	
Birth order	1st	24.5% (126)	75.5% (388)	0.529
	2–3	27.0% (145)	73.0%(392)	
	4–5	29.8% (51)	70.2%(120)	
	≥ 6	28.4% (21)	71.6% (53)	
Mother's BMI	< 18.5	29.6% (8)	70.4% (19)	0.702
	≥ 18.5	26.3% (333)	73.7% (931)	
Mother's education level	No education	72.7% (8)	27.3% (3)	0.000
	Primary	31.3% (186)	68.7% (408)	
	Post primary	21.7% (150)	78.3% (542)	
Visited health facility	Yes	24.4% (257)	75.6% (796)	0.000
	No	35.7% (87)	64.3% (157)	
Wealth Index	Poor	33.8% (184)	66.2%(361)	0.000
	Middle	30.1% (80)	69.9%(186)	
	Rich	16.3% (79)	83.7% (406)	
Place of residence	Rural	27.3% (257)	72.7% (683)	0.246
	Urban	24.2% (86)	75.8% (270)	
Household with electricity	Yes	18.2% (64)	81.8% (288)	0.000
	No	29.6% (279)	70.4% (665)	
Toilet facility	Toilet with flush	26.4% (170)	73.6% (473)	0.027
	Pit latrine	19.8% (41)	80.2% (166)	
	No facility	29.8%(133)	70.2% (133)	
Type of drinking water	Tap water	29.9% (173)	70.1% (406)	0.006
	Protected water	19.2% (48)	80.2% (202)	
	Unprotected	26.1% (122)	73.9% (345)	
Main roof material	Thatch/Palm leaf	34.4% (165)	65.6% (314)	0.000
	Corrugated metal	23.2% (169)	76.8% (558)	
	Stick & mud	11.1% (10)	88.9% (80)	
Main wall material	Wood/Mud	33.0% (194)	67.0% (394)	0.000
	Bricks	20.1% (89)	79.9% (354)	
	Cement /Block	22.6% (60)	77.4% (205)	
Main floor material	Earth/Sand	20.9% (72)	79.1% (272)	0.000
	Mud block/Wood	33.8% (158)	66.2% (309)	
	Cement/ Block	23.5% (114)	76.5% (372)	

and consequently weak, hence need more nutrients to support the rapid body growth. In addition, many children at an early age are not breastfeed which makes them susceptible to various illness. Some of these conditions reduce the haemoglobin level within the blood which may lead to anaemia and stunting. Furthermore, as children age and are introduced to foods, and are able to consume a variety of nutrition, this would aid in putting them at less risk of being anaemic or stunting. Similar results we found in previous studies [3, 7, 13, 14, 24, 37]. However, the studies by Anticona and Sebastian [7] and Oliveira et al. [35] showed that stunting increased as the children grew older.

Table 3 Parameter estimates for a joint marginal model for stunting and anaemia

Indicator variables	Stunting				Anaemia			
	Estimate; SE	OR	95% CI	*P*-value	Estimate; SE	OR	95% CI	*P*-value
Child's age								
Ref: > 39 months	-	-	-	-	-	-	-	-
20–39 months	0.226;0.186	1.254	0.870;1.806	0.225	0.531;0.175	1.701	1.207;2.396	0.003
< 20 months	0.829;0.194	0.436	0.298;0.638	< 0.001	0.096;0.167	0.908	0.655;1.260	0.565
Child had fever								
Ref: Yes	-	-	-	-	-	-	-	-
No	0.153;0.217	1.165	0.762;1.782	0.482	-0.712;0.186	0.491	0.341;0.707	< 0.001
Child had Diarrhoea								
Ref:Yes	-	-	-	-	-	-	-	-
No	-0.227; 0.231	0.797	0.507;1.254	0.326	-0.496; 0.202	0.609	0.410;0.905	0.014
Child's birth weight								
< 2500 g	-	-	-	-	-	-	-	-
≥ 2500 g	-1.248; 0.232	0.240	0.182;0.452	< 0.001	0.038; 0.230	1.039	0.662;1.631	0.868
Residence								
Ref: Urban	-	-	-	-	-	-	-	-
Rural	-0.653; 0.228	0.520	0.333;0.814	0.004	0.102; 0.197	1.107	0.753;1.629	0.603
Wealth Index								
Ref: Richer	-	-	-	-	-	-	-	-
Middle	1.065; 0.247	2.901	1.788;4.707	< 0.001	0.224; 0.199	1.251	0.847;1.848	0.262
Poorer	1.253; 0.249	3.501	2.149;5.703	< 0.001	0.282; 0.200	1.326	0.896;1.962	0.158
Education level								
Ref:No education	-	-	-	-	-	-	-	-
Primary	-1.473; 0.783	0.229	0.049;1.064	0.060	-0.772; 0.812	0.462	0.094;2.270	0.342
Post Primary	-1.841; 0.787	0.159	0.121;2.933	0.020	-0.519; 0.814	0.595	0.121;2.933	0.524

Table 4 Variance Components and covariance between anaemia and stunting

Variables	Estimate; SE	OR	95% CI	*P*-value
Variance (stunting)	0.104; 0.010	1.110	1.088; 1.132	0.149
Variance (anaemia)	0.314; 0.170	1.369	0.981; 1.910	0.033
Covariance between anaemia and stunting	1.000; 0.141	2.718	2.063; 3.582	0.001

The findings from this study revealed that risk of anaemia was related to having experienced recent fever and diarrhoea. This may be due to the fact that fever and diarrhoea are commonly accompanied by the number of diseases and morbidity which are associated with anaemia. This has also been previously described [14, 19, 39]. The probability of being stunted reduced with increasing level of maternal education. The aforementioned could be linked to socio-economic status, where educated individuals are more likely to have a better standard of living, and knowledge of balanced diet. In addition, educated individuals can easily access and improve the nutritional status as most of them have a monthly income. This is also consistent with previous studies such as Kavosi et al. [22], Aheto et al. [5], Adebayo et al. [1], Aheto et al. [6], and Adhikari et al. [3].

Childbirth weight significantly impacts stunting in children, with lower risk in children born with a higher weight (≥ 2500 g). This relationship can be explained by the fact that children with low birth weight are more likely to have other co-morbidity illnesses that might be associated with stunting. Similar findings were found in studies by Yang et al. [48], Habyarimana et al. [19], and Kejo et al. [24].

We found that children living in rural areas have a lower risk of stunting, an effect that is debated in the field. A possible explanation to this is due to the fact that some individuals in rural areas are educated and they eat fresh food and fruits with more nutrients. Also, individuals from rural areas are breastfeed for a long period of time, which can contribute to fighting stunting at an early age. Some studies have described similar results, such as the study by Kavosi et al. [22], while others have described contrasting results Yang et al. [48] and El Kishawi et al. [12].

Lastly, children living within families from the middle and top tertile wealth index have a lower risk of having malnutrition. This further undergirds the fact that malnutrition is linked to socio-economic conditions, where the children from the lower wealth index cannot afford proper food, maintain hygiene and access to health care services. Similar results were found in previous studies such as the study by Gari et al. [13], Mohammed et al. [32], and Rivadeneira et al. [39].

Strengths and limitation of the study

We used joint modelling to assess the association between anaemia and stunting in children less than 5 years of age. However, the present study has some limitation, the first being that the dataset was cross-sectional and it would be good to assess changes in disease trajectories and associated over time. The study used stunting variable as a longer term indicator of malnutrition [37] but other variables that reflect nutritional status could have also been assessed. Also, it would be interesting to use spatial joint models to assess the association between stunting and anaemia by geographical location. In addition, the food records would be mentioned in the study. Lastly, the study did not include the maternal anaemia levels, which are usually the biggest predictor of child anaemia. Hence, all the areas of focus not covered in this study, especially those mentioned, are considered limitations that will be addressed in future studies.

Conclusion

This study aimed to determine the association between anaemia and stunting in children less than 5 years of age in Lesotho using multivariate joint model under GLMM. The study also assessed the association of socio-economic and demographic factors with anaemia and stunting. Lastly, we evaluated possible interaction effect between independent variables but none passed significance threshold. We found a significant positive association between anaemia and stunting which indicates that, when malnutrition increases in children less than 5 years, anaemia also increases and vice-versa. Thus, a change in childhood stunting can have a significant impact on anaemia status. In addition, several socio-economic and demographic factors impact both malnutrition and anaemia such as family wealth, maternal education, urban vs. rural living environments. In addition, children that were low birthweight or who have recently experienced fever, or diarrhoea should be prioritized for intervention. Knowledge on the relationship between anaemia and malnutrition together with other determinants can provide useful insights to policy makers, donors and government in planning and fighting to improve child through tailored public health messages and interventions.

Recommendations

To improve anaemia and malnutrition in children less than 5 years of age in Lesotho, policy makers, donors and government should focus on improving nutrition status especially, in children from rural area, with diarrhoea, fever, low birthweight, from poorer quantile index household and uneducated mother.

Abbreviations

OR: Odds ratio; SE: Standard error; 95% CI: Confidence Interval; LDHS: Lesotho demographic health survey; GLMM: Generalized linear mixed model; MVN: Multivariate normal distribution; AIC: Akaike information criterion; Hb: Haemoglobin concentration; UN: Unstructured; WHO: World Health Organization; SD: Standard deviations; HAZ: Height-for-age z score, SAM: severe acute malnutrition; MAM: Moderate acute malnutrition; UNICEF: United nations international children's emergency fund; MOHSW: Ministry of health and family welfare; HIV: Human immunodeficiency virus; AIDS: Acquired immunodeficiency syndrome; EAs: Enumeration areas.

Acknowledgements

The authors acknowledge the Ministry of Health (Lesotho) and ICF International through the DHS programme for providing the data.

Authors' contributions

R.T. Gaston: designed the study; acquired the data; performed the analysis; interpreted the results; and wrote the manuscript. F. Habyarimana: Revised the manuscript and approved the final manuscript. S. Ramroop: Revised the manuscript; and approved the final manuscript.

Funding

This work was supported by the Health Economics and HIV/AIDS Research Division (HEARD).

Availability of data and materials

The present study utilized existing dataset and is available from https://www.dhsprogram.com/data/dataset_admin/login_main.cfm with the permission from the DHS program.

Declarations

Ethics approval and consent to participate

The ethical approval for the 2014 LDHS was approved by the Lesotho Ministry of Health Research and Ethics Committee together with support of the Institutional Review Board of ICF International. Verbal informed consent was obtained from children parents or guardians before conducting the surveys. The interviewers explained the procedure, the willingness to participate in the survey and the confidentiality of the data.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

Author details

[1]School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Private Bag X01, Pietermaritzburg CampusScottsville 3209, South Africa. [2]Health Economics and HIV/AIDS Research Division (HEARD), University of KwaZulu-Natal, Westville Campus, Private Bag X01, Westville 3629, South Africa.

Received: 9 April 2021 Accepted: 27 January 2022
Published online: 11 February 2022

References

1. Adebayo SB, Gayawan E, Heumann C, Seiler C. Joint modeling of Anaemia and Malaria in children under five in Nigeria. Spatial and spatio-temporal epidemiology. 2016;17:105–15.
2. Adeyemi RA, Zewotir T, Ramroop S. Joint spatial mapping of childhood anaemia and malnutrition in sub-Saharan Africa: a cross-sectional study of small-scale geographical disparities. Afr Health Sci. 2019;19(3):2692–712.
3. Adhikari RP, Shrestha ML, Acharya A, Upadhaya N. Determinants of stunting among children aged 0–59 months in Nepal: findings from Nepal Demographic and Health Survey, 2006, 2011, and 2016. BMC nutrition. 2019;5(1):1–10.
4. Agresti A. Foundations of linear and generalized linear models. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2015.
5. Aheto JMK, Keegan TJ, Taylor BM, Diggle PJ. Childhood Malnutrition and Its Determinants among Under-Five Children in Ghana. Paediatr Perinat Epidemiol. 2015;29(6):552–61.
6. Aheto JMK, Taylor BM, Keegan TJ, Diggle PJ. Modelling and forecasting spatio-temporal variation in the risk of chronic malnutrition among under-five children in Ghana. Spatial and spatio-temporal epidemiology. 2017;21:37–46.
7. Anticona C, San Sebastian M. Anemia and malnutrition in indigenous children and adolescents of the Peruvian Amazon in a context of lead exposure: a cross-sectional study. Glob Health Action. 2014;7(1):22888.
8. Bardosono S, Sastroamidjojo S, Lukito W. Determinants of child malnutrition during the 1999 economic crisis in selected poor areas of Indonesia. Asia Pacific journal of clinical nutrition. 2007;16(3):512-26.
9. Bechir M, Schelling E, Hamit MA, Tanner M, Zinsstag J. Parasitic infections, anaemia and malnutrition among rural settled and mobile pastoralist mothers and their children in Chad. EcoHealth. 2012;9(2):122–31.
10. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, De Onis M, Ezzati M, Grantham-McGregor S, Katz J, Martorell R, Uauy R. Maternal and child undernutrition and overweight in low-income and middle-income countries. The lancet. 2013;382(9890):427–51.
11. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc. 1993;88(421):9–25.
12. El Kishawi RR, Soo KL, Abed YA, Muda WAMW. Anaemia among children aged 2–5 years in the Gaza Strip-Palestinian: a cross sectional study. BMC Public Health. 2015;15(1):319.
13. Gari T, Loha E, Deressa W, Solomon T, Atsbeha H, Assegid M, Hailu A, Lindtjørn B. Anaemia among children in a drought affected community in south-central Ethiopia. PloS one. 2017;12(3):e0170898.
14. Gaston RT, Ramroop S, Habyarimana F. Determinants of factors associated with anaemia among children under five years in Lesotho. African Population Studies. 2018;32(1):3893-902.
15. Gaston RT, Ramroop S. Prevalence of and factors associated with malaria in children under five years of age in Malawi, using malaria indicator survey data. Heliyon. 2020;6(5):e03946.
16. Gaston RT, Ramroop S, Habyarimana F. Joint modelling of malaria and anaemia in children less than five years of age in Malawi. Heliyon. 2021;7(5):e06899.
17. Gayawan E, Arogundade ED, Adebayo SB. Possible determinants and spatial patterns of anaemia among young children in Nigeria: a Bayesian semi-parametric modelling. Int Health. 2014;6(1):35–45.
18. Gueorguieva R. A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. Stat Model. 2001;1(3):177–93.
19. Habyarimana F, Zewotir T, Ramroop S. Joint modeling of poverty of households and malnutrition of children under five years from demographic and health survey data: case of Rwanda. Journal of Economics and Behavioral Studies. 2016;8(2):108–14.
20. Hedeker D. Generalized linear mixed models. In: Everitt, B., Howell, D. (Eds.), Encyclopedia of statistics in behavioral science. New York: Wiley; 2005. p. 729-738.
21. Kassebaum N, Kyu HH, Zoeckler L, Olsen HE, Thomas K, Pinho C, Bhutta ZA, Dandona L, Ferrari A, Ghiwot TT, Hay SI. Child and adolescent health from 1990 to 2015: findings from the global burden of diseases, injuries, and risk factors 2015 study. JAMA Paediatrics. 2017;171(6):573–92.
22. Kavosi E, Rostami ZH, Kavosi Z, Nasihatkon A, Moghadami M, Heidari M. Prevalence and determinants of under-nutrition among children under six: a cross-sectional survey in Fars province. Iran International Journal of Health Policy and Management. 2014;3(2):71.
23. Kazembe LN. An additive regression model for investigating the relationship between childhood health and socio-economic status. Spatial and Spatio-temporal Epidemiology. 2013;6:71–84.
24. Kejo D, Petrucka PM, Martin H, Kimanya ME, Mosha TC. Prevalence and predictors of anemia among children under 5 years of age in Arusha District, Tanzania. Pediatric health, medicine and therapeutics. 2018;9:9.
25. Kounnavong S, Sunahara T, Mascie-Taylor CN, Hashizume M, Okumura J, Moji K, Boupha B, Yamamoto T. Effect of daily versus weekly home fortification with multiple micronutrient powder on haemoglobin concentration of young children in a rural area, Lao People's Democratic Republic: a randomised trial. Nutr J. 2011;10(1):129.
26. Leal LP, Batista Filho M, Lira PICD, Figueiroa JN, Osório MM. Prevalence of anaemia and associated factors in children aged 6–59 months in Pernambuco, Northeastern Brazil. Rev Saude Publica. 2011;45:457–66.
27. Letsie MM, Grab SW. Assessment of social vulnerability to natural hazards in the Mountain Kingdom of Lesotho. Mountain Research and Development. 2015;5(2):115-25.
28. Letuka T, Frade S. Household and individual risk factors of anaemia among under-5 children in Lesotho. Afr Health Sci. 2020;20(3):1478–86.
29. Masibo PK, Makoka D. Trends and determinants of undernutrition among young Kenyan children: Kenya Demographic and Health Survey; 1993, 1998, 2003 and 2008–2009. Public Health Nutr. 2012;15(9):1715–27.
30. Milman N. Anaemia-still a major health problem in many parts of the world! Ann Hematol. 2011;90(4):369–77.
31. Ministry of Health [Lesotho] and ICF International. Lesotho Demographic and Health Survey 2014. Maseru: Ministry of Health and ICF International; 2016.
32. Mohammed SH, Larijani B, Esmaillzadeh A. Concurrent anemia and stunting in young children: prevalence, dietary and non-dietary associated factors. Nutr J. 2019;18(1):1–10.
33. Molenberghs G, Verbeke G. Models for discrete longitudinal data. New York: Springer-Verlag; 2005.
34. Okoromah CA, Ekure EN, Lesi FE, Okunowo WO, Tijani BO, Okeiyi JC. Prevalence, profile and predictors of malnutrition in children with congenital heart defects: a case–control observational study. Arch Dis Child. 2011;96(4):354–60.
35. Oliveira D, Ferreira FS, Atouguia J, Fortes F, Guerra A, Centeno-Lima S. Infection by intestinal parasites, stunting and anemia in school-aged children from southern Angola. PloS one. 2015;10(9):e0137327.
36. Olney DK, Kariger PK, Stoltzfus RJ, Khalfan SS, Ali NS, Tielsch JM, Sazawal S, Black R, Allen LH, Pollitt E. Development of nutritionally at-risk young children is predicted by malaria, anaemia, and stunting in Pemba. Zanzibar The Journal of Nutrition. 2009;139(4):763–72.
37. Rahman MS, Mushfiquee M, Masud MS, Howlader T. Association between malnutrition and anaemia in under-five children and women of reproductive age: Evidence from Bangladesh Demographic and Health Survey 2011. PloS one. 2019;14(7):e0219170.
38. Renzaho AMN. Mortality, malnutrition, and the humanitarian response to the food crises in Lesotho. Journal of Emergency Primary Health Care. 2006;4(4):1-15.
39. Rivadeneira MF, Moncayo AL, Tello B, Torres AL, Buitrón GJ, Astudillo F, Fredricks TR, Grijalva MJ. A Multi-Causal Model for Chronic Malnutrition and Anaemia in a Population of Rural Coastal Children in Ecuador. Matern Child Health J. 2020;24(4):472–82.
40. Sachdev HPS, Gera T, Nestel P. Effect of iron supplementation on mental and motor development in children: systematic review of randomised controlled trials. Public Health Nutr. 2005;8(2):117–32.
41. Tuerlinckx F, Rijmen F, Verbeke G, De Boeck P. Statistical inference in generalized linear mixed models: A review. Br J Math Stat Psychol. 2006;59(2):225–55.
42. UNICEF. State of the World's children: Children in a digital World. New York: UNICEF; 2017.
43. Varghese JS, Stein AD. Malnutrition among women and children in India: limited evidence of clustering of underweight, anemia, overweight, and stunting within individuals and households at both state and district levels. Am J Clin Nutr. 2019;109(4):1207–15.
44. Verbeke G, Molenberghs G. The use of score tests for inference on variance components. Biometrics. 2003;59(2):254–62.

45. Worku Z. Malnutrition among rural and urban children in Lesotho: related hazard and survival probabilities. Health SA Gesondheid. 2003;8(3):98–109.
46. World Health Organization. WHO child growth standards: head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: methods and development. Geneva: Switzerland World Health Organization; 2007. p. 2007.
47. World Health Organization. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Geneva: World Health Organization; 2011. p. 2011.
48. Yang W, Li X, Li Y, Zhang S, Liu L, Wang X, Li W. Anemia, malnutrition and their correlations with socio-demographic characteristics and feeding practices among infants aged 0–18 months in rural areas of Shaanxi province in north western China: a cross-sectional study. BMC Public Health. 2012;12(1):1–7.

Publisher's Note