Statistical study on childhood malnutrition and anaemia in Angola, Malawi and Senegal

By

Mthobisi Christian Khulu

(210505476)

Submitted in fulfilment of the academic requirements for the degree of

Doctor of Philosophy

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

School of Mathematics, Statistics and Computer Science

University of KwaZulu-Natal

Pietermaritzburg, South Africa

Supervisors: Professor Shaun Ramroop and Dr Faustin Habyarimana

February 2023

PREFACE

The project work stipulated in this thesis was completed under the supervision of Professor Shaun Ramroop and Dr Faustin Habyarimana from the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal (Pietermaritzburg Campus).

[redacted]

Mthobisi Christian Khulu

23/02/2023

Date

As the candidate supervisor

[redacted]

Prof. Shaun Ramroop

23/02/23

Date

and co-supervisor, we agree to the submission of this thesis.

[redacted]

Dr. Faustin Habyarimana

23/02/23

Date

DECLARATION

I, **Mthobisi Christian Khulu** declare that:

1. The research reported under this thesis titled '**Statistical study on childhood malnutrition and anaemia in Angola, Malawi and Senegal** is my original work, except where otherwise indicated.
2. This thesis has not been submitted for the award of other degrees in any other tertiary institutions.
3. This thesis does not contain other persons' data, pictures, graphs or other information unless specifically acknowledged as being sourced from other persons.
4. That my contribution to the project was as follows:

Identification of research topics, design, execution, data analysis and interpretation, manuscript and thesis write-up.

5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source is detailed in the thesis and the reference sections
6. Professor Shaun Ramroop (supervisor) and Co-Supervisor Dr Faustin Habyarimana have reviewed the entire work and the write-up of the manuscripts and thesis.

Mthobisi Christian Khulu

23/02/2023

Date

DEDICATION

To the Almighty God and my family who made this PhD a reality.

ACKNOWLEDGEMENTS

Glory to the Mighty God for giving me strength to complete this study. Secondly, my gratitude goes to my supervisors Dr. F Habyarimana and Prof S Ramroop for believing in me and for academic assistance throughout the duration of this project. Not forgetting the Toyota South Africa Manufacturing (TSAM) management for affording me time and your resources, even in difficult times you showed how much you care for people. The support you showed to me I pray you extend to others.

These four years would not have been so memorable if not for all my friends in the academic and industrial environment. In particular, Statistics PhD candidates at the University of KwaZulu-Natal and the Human Resources team at TSAM, thank you for being the best friends. It goes without saying that I am forever grateful for my family's support. Mom, Nontethelelo and Nompilo- thank you for putting up with me all these years.

Finally, I would like to thank my other family, my small family. My wife and my son, Azokuhle- I'm not sure what I did to deserve such an amazing family, but I'm so grateful for all your support. We have been having late nights together fix manuscripts that need to be submitted to the journals. Please, both feel free to be called Dr.

Table of Contents

List of figures

List of tables

LIST OF ABBREVIATIONS AND ACRONYMS

SGD - Sustainable Development Goal

WHO – World Health Organisation

PCA – Principal Component Analysis

OR – Odd Ratio

DF – Degrees of Freedom

CDF – Cumulative Distribution Function

AIC - Akaike Information Criterion

BIC – Bayesian Information Criterion

GLMM – Generalized Linear Mixed Model

JGLMM – Joint Generalized Linear Mixed Model

WAZ - Weight-for-Age Z-score

EA - Enumeration Area

SEA – Standard Enumeration Area

MLE Maximum likelihood estimate

PQL - Penalized Quasi-likelihood

DHS – Demographic and Health Survey

ADHS – Angola Demographic and Health Survey

MDHS – Malawi Demographic and Health Survey

SDHS – Senegal Demographic and Health Survey

GCV - Generalized Cross Validation

MQL – Marginal Quasi-likelihood

MCMC – Monte Carlo Markov Chain

SI – Standardized Index

KMO- Kaiser-Meyer Olkin

MCAR - Multivariate Conditional Auto Regressive

SEM – Structural Equation Model

GFI – Goodness of Fit Index

RMSEA – Root Mean Square Error of Approximation

CFI – Comparative Fit Index

AGFI – Adjusted Goodness of Fit

PUBLICATIONS (PUBLISHED/UNDER-REVIEW)

The following papers have been published or under-review from this thesis:

1. Chris Khulu and Shaun Ramroop (2020). Determinants of Malnutrition in Under five Children in Angola, Malawi and Senegal. **Published:** The Open Public Health Journal. https://doi.org/10.2174/1874944502013010055, 2020, 13, 55-61
2. Chris Khulu and Shaun Ramroop (2020). Key Determinants of Anaemia among Youngsters under Five Years in Senegal, Malawi, and Angola. **Published:** Int. J. Environ. Res. Public Health, 17, 8538; https://doi.org/10.3390/ijerph17228538
3. Chris Khulu, Shaun Ramroop and Faustin Habyarimana (2022). Copula geo-additive modelling of malnutrition and anaemia in children under five years in Angola, Senegal, and Malawi. **Published:** Int. J. Environ. Res. Public Health, 19, 9080; https://doi.org/10.3390/ijerph19159080
4. Chris Khulu, Shaun Ramroop and Faustin Habyarimana (2022). Modelling factors associated with malnutrition and anaemia in children under five years in Angola, Senegal, and Malawi by using Joint model. **Published:** The Open Public Health Journal. https://DOI: 10.2174/18749445-v15-e221220-2022-82, 2022, 16, 1-8
5. Chris Khulu, Shaun Ramroop and Faustin Habyarimana (2023). Structural equation modelling of malnutrition and anaemia among under five children in Angola, Malawi and Senegal. **Submitted:** African Population.

Abstract

Malnutrition and anaemia continue to be a concern to the future of developing countries. This thesis aimed to examine the risk factors associated with malnutrition and anaemia among under five-year-old children in Angola, Malawi and Senegal. Statistical models and techniques have improved over the years to give more insight into malnutrition and anaemia, in terms of demographic, socio-economic, environmental, and geographic factors. This thesis also assessed the spatial epidemiological overlaps between childhood malnutrition and anaemia diseases which can lead to various advantages in intervention planning, monitoring, controlling and total elimination of such diseases, especially in high-risk regions. This is a secondary data analysis where national representative data from the three countries was used. The Demographic and Health Survey data from Angola, Malawi and Senegal were merged to create a pooled sample which was then used for all the analyses conducted in this study. The relationship between exploratory variables to malnutrition and anaemia was assessed to obtain variables that explain the two outcomes. Consequently, a generalized linear mixed model was used to investigate the significance of the child-level, community-level and household-level factors to malnutrition and anaemia separately. The relationship between the two diseases was further examined using the three joint modelling approaches: (1) a joint generalised linear mixed model; (2) a structural equation model, and (3) a bivariate copula geo-additive model. For each model employed, the significant factors of both malnutrition and anaemia were identified. The GLMM results on malnutrition revealed that children's place of residence, age, gender, mother's level of schooling, wealth status, birth interval and birth order significantly explain malnutrition at the 5% level of significance. Whereas, the GLMM results on anaemia revealed that children's age, gender, mother's level of schooling, wealth status and nutritional status significantly explain anaemia at 5% level of significance. The findings of copula geo-additive modelling of malnutrition and anaemia indicated that there is an association between malnutrition and anaemia. There was a strong association observed between malnutrition and anaemia in the north-west districts of Angola when compared to other districts. The results imply that the policymakers of Angola, Senegal and Malawi can control anaemia through the intervention of malnutrition controlling. The overall findings of this study provide meaningful insight to the policymakers of Angola, Malawi and Senegal which will lead to the implementation of interventions that can assist in achieving the Sustainable Development Goal (SDG) of 25 deaths per 1 000 live births by 2030. To properly eradicate all the causes of malnutrition and anaemia, programs such as parental education, financial education, children's dietary focus programs and mobile health facilities

could add a significant value. The results also highlighted the national priority areas related to child-related factors, household factors and environmental factors for childhood malnutrition and anaemia morbidity control. It also provided policy makers with valuable geographical information for developing and implementing effective intervention. There is a greater need for partnership and collaboration among the studied countries to achieve the SGD target.

KEYWORDS: anaemia; malnutrition; co-morbidity; spatial, probability, joint model; copula functions; geo-additive; latent variables; nonlinear effects; statistical model; chi-square; gamma measure

Chapter 1 : Introduction

Notwithstanding the interventions implemented to address child mortality, anaemia and malnutrition remain a concern to the future of developing countries (World Health Organization and World Bank Group, 2021). Anaemia and malnutrition contribute significantly to childhood morbidity in Sub-Saharan Africa (Ehrhardt, et al., 2020). The report by World Health Organization (WHO) further revealed that most children exposed to malnutrition are residing in Africa and Asia (World Health Organization and World Bank Group, 2021). In 2020, two out of five children in Africa is affected by stunting. Twenty-seven percent of children in Africa are affected by wasting and twenty-seven percent of children is affected by overweight. Statistically, malnutrition is reported to impact children mostly, as approximately three million children die annually due to lack of nutrition (Von Grebmer, et al., 2019). The prevalence of wasting for children under five years of age is estimated to be 8% globally. An estimated 41 million under five children are overweight (6% prevalence). There was a one percent climb in this form of malnutrition from 2000 to 2016. Southern and Northern Africa are the two African regions with greater prevalence of wasting for children under-5 than the average global prevalence (World Health Organization and World Bank Group, 2017).

Anaemia has been shown to be a public health problem that affects low-, middle- and high-income countries, and has significant adverse health consequences, as well as adverse impact on social and economic development. Anaemia may result from several causes, with the most significant contributor being iron deficiency. Approximately 50% of cases of anaemia are due to iron deficiency, but the proportion probably varies among population groups and in different areas (World Health Organization, 2015). It is estimated that half of the children between the age interval of 0 – 5 years are anaemic, globally (Soares, et al., 2020).

Angola is situated on the west coast of Southern Africa with a population of approximately 25 million people, where the 2016 under five mortality rate was 82.5 deaths per 1 000 live births. In Angola, malnutrition and anaemia are the main causes of child death and the number of children affected by malnutrition is reported to be increasing with approximately eighty-five thousand of the total children population being severely malnourished. In the report issued by World Vision (World Vision Angola and Partners, 2014), Angola is ranked as number one for countries that have the weakest commitment to fight malnutrition in children. A survey conducted in Angola between 2015 and 2016 showed that sixty-five percent of children between 6-59 months is anaemic (World Health Organization, 2015). The prevalence of malnutrition was found to higher

for children in the age interval of 6-11 months. A study by (Humbwavali, et al., 2019) showed that collective exposure is likely to play a role in causing malnutrition in Angola. Children experiencing diarrhoea with their mothers as primary caregivers, were found to be among many factors associated with malnutrition (Humbwavali, et al., 2019). A study by (Sartorelli, et al., 2017) in Angola showed that male children and household sources of water such as rivers or lakes increase the chances of children's exposure to malnutrition. In addition, a study to understand the nutritional deficits among children under five in Angola by (Monteiro, et al., 2013) found that children aged between 6-23 months have acute malnutrition, and this may be due to the results of maternal malnutrition and insufficient nutrition from breastfeeding. In contrast, Iron deficiency anaemia was found to be associated with children's age, gender and inflammation in a study conducted on children less than five years in Angola (Barros, et al., 2020). A study conducted by (Sousa-Figueiredo, et al., 2012) defined anaemia as a severe public health threat among children under five in Angola. Anaemia and malaria are the most contributing diseases to under five/ pre-school-aged children in Angola. Approximately 60% and 19 % of pre-school-aged youngsters are suffering from anaemia and malaria, respectively, in Angola (Rosario, et al., 2004). The recent study conducted by (Gasparihno, et al., 2022) to understand wasting, stunting and anaemia among Angolan children using binary longitudinal analysis with different dependence structure indicated a probable advantage of testing and treating children with intestinal parasites for wasting. The trend of these conditions were observed to be decreasing of time, hence the study highlighted the importance of using the longitudinal data to assess the progress.

Senegal, a country populated with 15 million people which has an under five mortality of 47.1 deaths per 1 000 live births is known to be among developing African countries experiencing malnutrition (USAID, 2018). The prevalence of malnutrition in children under five in Senegal was reported to be sitting at sixty-seven percent in 2016 (World Health Organization, 2015). The report further revealed that children under five years that are anaemic average around eighty-two percent from 1990 to 2016. Food insecurity in Senegal differs by place of residence. In the urban areas of Senegal, nine out of a hundred households are food uncertain, whereas, in rural areas of Senegal twenty-one out of a hundred households are food uncertain (World Food Program, 2017). The prevalence of malnutrition in Senegal is lower when compared to other Western countries. Poor complementary feeding and hygiene practices are among many contributors to children's malnutrition in Senegal (National Agency for Statistics and Demography - ANSD / Senegal, and ICF, 2018). The findings of the study identified the age group and sex of the children as key determinants of anaemia. Results from the study conducted by (Diouf, et al., 2013) in Senegal showed that improving the mother's level of literacy and consumption of animal protein (meat, fish and eggs) reduce youngsters' risk of suffering from anaemia. Whereas, the study conducted

by (Tine, et al., 2012) in Senegal showed that malaria parasitaemia, sickle cell disorders, alpha-thalassemia, stunting, an age interval of 2 to 4 years and an age greater than 5 years, are significantly associated with anaemia among youngsters under 10 years. In addition, a study reviewing observational studies of 15 countries including Malawi and Senegal by (McCuskee, et al., 2014) concluded that covariates such as age, type of residence, socio-economic status and maternal education tend to be associated with anaemia. A study further concluded that the association of anaemia with inflammation could suggest the occurrence of nutritional immunity and should be further investigated. In a study conducted by (Samanpreet Brar, et al., 2020) in Senegal revealed that promoting good dietary practices, improving hygiene conditions and monitoring the nutritional status of children could assist in fighting malnutrition. The study further showed that the factors associated with anaemia are the presence of diarrhoea, non-consumption of vegetables, meat and incomplete immunization status. In Senegal, a study that considered the socio-economic determinants of child malnutrition and investigated how programs compensate for the increased risk facing younger mothers and their children, revealed that children of mothers giving birth at a younger age are disadvantaged in nutritional status (Linnemayr & Alderman, 2008). The nutritional status of Senegalese children is very important at hospital admission (Beau, et al., 1987). The mortality of children beneath eighty percent level of weight/height is 2.64 times greater than the mortality of children above eighty percent level of weight/ height. In another study by (Ba, et al., 2021) which was conducted to understand anaemia among children aged 6-59 months in Dakar Suburban Area revealed that anaemia is high on the age group of 13-24 months. This may be due to the food diversity that is afforded to this children age group. The study concluded that there is a requirement for innovative strategies to fight against anaemia which may include feeding practices etc.

In Malawi, poor diets, lack of food and infectious diseases are among contributors to malnutrition (United Nations Children's Fund, World Health Organization, World Bank Group , 2018). Malawi's under five mortality rate was estimated to be 55.1 deaths per 1000 live births with a population of 18 million. Twenty-three percent of all children's deaths in Malawi are related to malnutrition, whereas, four percent of Malawian children are suffering from acute malnutrition. A study conducted by (Ngwira & Kazembe, 2016) suggested that Malawian policymakers and government need to address poverty, maternal anaemia and malnutrition in an attempt to reduce mortality in children under five years. Also, the research by (Brabin, et al., 2004) revealed that improving malaria and anaemia control during pregnancy would reduce the child's risk of suffering from anaemia. Whereas, a study by (Ntenda, et al., 2004) showed that mothers' level of education, household wealth status, children experiencing fever and children stunting are positively associated with anaemia for Malawian children under five. In Malawi, male children

are more likely to be malnourished when compared to their female counterparts (Chirwa, 2009). The study further concluded that children residing in households that have economic empowered females as household heads are less likely to be exposed to malnutrition. In another study conducted by (Doctor, 2017) in Malawi, it was revealed that children residing in wealthy households are less likely to be affected by malnutrition when compared to children residing in poor households. The study further showed that having diarrhoea and fever is associated with malnutrition. It was concluded in the study that community-based interventions that promote access to better water, sanitation and hygiene facilities remain vital. In a study to evaluation the effect of fish-farming on nutritional status in Malawi by (Aiga & Matsuoka, 2009), it was found that a lesser occurrence of malnutrition was found among the children in fish-farming households. It was also revealed in the same study that a more frequent intake of oil and fat other than never/seldom and breast-feeding practices for appropriate duration reduce the children's chances of being exposed to malnutrition. Anaemia in children under five in Malawi is associated with age, fever, mothers' level of education and residing in poor households (Tenda, et al., 2017). The study concluded that individual-level factors have stronger effect than the community-level factors on childhood anaemia; however, community-level education of mothers play an important role in childhood anaemia. Anaemia in Malawian children is multi-factorial (McGann, et al., 2020). In addition to nutritional and other environmental factors, genetic causes of anaemia are common in children of Malawi (Linnemayr & Alderman, 2008). The study investigating the determinants of stunting among under five years children in Malawi by (Afolabi & Palamuleni, 2021) showed that Children who were female from wealthier households whose mothers were overweight/obese aged 25 - 34 years had at least a secondary level education were less likely to be stunted.

Anaemia in Angola, Malawi and Senegal was found to be associated with the mothers' level of education, household wealth status and child's age, whereas, malnutrition was found to be associated with type of residence, household wealth status, mother's level of education and child's age (Ramroop & Khulu, 2020). In a study by (Aynalem, et al., 2022) employed a random-effects model to estimate the polled prevalence of anaemia and malnourished. The results revealed that anaemia was observed in 57.53% of African malnourished pre-school children. Whereas the prevalence in HemoCue and auto-machine diagnosis method of anaemia was found to be 58.52% and 56.81% respectively. The study further concluded that the size of anaemia among African malnourished pre-school children was higher. There is a limited knowledge about the main factors related with malnutrition and anaemia in developing countries. Various statistical techniques and methods have been employed to identify factors associated with malnutrition and anaemia. Fewer studies have been conducted to jointly model malnutrition and anaemia. These studies that have

published on this topic are using different statistical approaches and hence aiming at different objectives (Adeyemi, et al., 2019). A study by (Adeyemi, et al., 2019), jointly model anaemia and malnutrition using the multivariate conditional auto regressive (MCAR) analysis. The study found a positive correlation between malnutrition and anaemia; however, the study did not explore the significance of the correlation or consider other possible outcome combinations. That would have been useful to the policymakers for intervention implementations.

1.1 Statement of research problem and objectives

In Sub-Saharan Africa, malnutrition and anaemia contribute a higher percentage to infant morbidity. Statistics released by (UNICEF, 2019) reveal that the under-five mortality rate has declined by more than half since 1990. However, Sub-Saharan Africa still has the highest rate. The projections released by (UNICEF, 2019) indicate that the under-five mortality rate in Africa in 2030 will be 54 deaths per 1 000 live births. One of the Sustainable Development Goal (SDG) targets is to reduce the under-five mortality rate to 25 deaths per 1 000 live births by 2030 (World Health Organization and World Bank Group, 2017). Based on the projections and the SDG target, Africa will present more than half of the SDG target by 2030. Mayotte, Réunion, Seychelles and Egypt are the only African countries that currently have under five mortality rates below the SDG 2030 target. Angola has an under five mortality rate of 82.5 deaths per 1 000 live births, Senegal has an under five mortality rate of 47.1 deaths per 1 000 live births and Malawi is estimated to have 55.1 deaths per 1000 live births (UNICEF, 2019). Thus, this indicates that Angola, Malawi and Senegal are all expected to reduce their under-five mortality rates by at least half of their current rate. This requires serious interventions and the targeting of significant causes of under-five mortality.

Hence, the specific objectives of this thesis were:

1. To identify the significant sociodemographic, public health and geographic risk factors associated with childhood malnutrition and anaemia in Angola, Malawi and Senegal using GLMM to account for the complexities of the survey's sampling designs. This is the separate modelling of malnutrition and anaemia.
2. To jointly model childhood malnutrition and anaemia in Angola, Malawi and Senegal.
3. To investigate the direct and indirect risk factors of childhood malnutrition and anaemia prevalence in Angola, Malawi and Senegal.
4. To evaluate the spatial patterns and risk factors of childhood malnutrition and anaemia in Angola, Malawi and Senegal.

1.2 Significance of the study

The prime purpose of this study is to identify factors associated with malnutrition and anaemia in Angola, Malawi and Senegal. The results of this study will be beneficial to the health policy of these three countries as well as to their people. The significance of this study underpins investigating the risk factors of malnutrition and anaemia and the spatial variation using robust statistical models and techniques. Furthermore, in this research, malnutrition and anaemia are jointly modelled to provide more understanding about these diseases which includes the overlapping of risk factors. This study will serve as the basis for future plans for health policy concerning the reduction of childhood malnutrition and anaemia in Angola, Malawi and Senegal, and also help with how this can be achieved through the implementation of various strategies. Recommendations to the government and department of health in these three countries will be made based on the results of factors requiring special attention in relation to childhood malnutrition and anaemia.

1.3 Thesis layout

The layout of this thesis is as follows:

Chapter 1 provides an introduction and objectives of the study.

Chapter 2 describes the characteristics of the data and the exploratory analysis.

Chapter 3 presents an overview of the generalized linear mixed model (GLMM) and its application to investigate the risk associated with malnutrition and anaemia.

Chapter 4 provides an overview of the joint generalized linear mixed model and its application to identify factors associated with both malnutrition and anaemia.

Chapter 5 introduce the structural equation method to understand the direct and indirect risk factors of malnutrition and anaemia.

Chapter 6 presents a copula-based model to jointly model childhood malnutrition and anaemia.

Chapter 7 presents the conclusion and recommendations.

Chapter 2 : Data Characteristics and Descriptive Statistics

2.1 Study regions

This study focuses on three countries, namely Angola, Malawi and Senegal. These countries are situated in different parts of the African continent (circled in red in Figure 2-1).

Figure 2-1: Study regions

2.2 Data source

Data employed is from 2016 nationally representative information from Senegal, Malawi and Angola. The survey collected data on the Demographics and Health of the populace. The data was requested from the DHS Micro program. There was no ethical approval required since this study is a secondary data analysis. The Demographic and Health Survey of Angola was collected from the beginning of October 2015 until the end of March 2016 to obtain demographic, socio-economic, health and other information. A household sample of n=16,109 was selected and it was made up of 14,379 women aged between 15 and 49 and 5,684 men aged between 15 and 54. The success rate achieved for interviewing men and women was 94% and 96%, respectively. The 2015-2016 Malawi Demographic and Health Survey had a sample size of 27,516 households of which 26,564 were successfully interviewed. The method used to determine the sample

population was multi-stage sampling. The primary stage of sampling was made up of 173 standard enumeration areas in urban settings and 677 standard enumeration areas in rural settings.

In the secondary stage, 30 families for every urban cluster, and 33 for every rural cluster were chosen with an equal probability of systematic selection. The 2016 Senegal Demographic and Health Survey was executed and conducted by the National Agency of Statistics and Demographics to respond to public health challenges, community challenges, well-being challenges and to monitor the progress of implemented measures. A sample of 3,527 men aged between 15 and 59 and 8,865 women aged between 15 and 49 was interviewed to collect information on demographics and health.

The Demographic and Health Survey data from Senegal, Malawi and Angola were merged to create a pooled sample. The created pool sample was then used for all the analyses conducted in this study. The advantage of using a pooled sample when comparing different data sets includes the ability to generalize the results. Furthermore, its method allows for comparison between countries and this is beneficial to the policymakers. A similar method was used by (Takele, et al., 2019); (Subramanian, et al., 2011); (Neuman, et al., 2013); and (Khulu & Ramroop, 2020). Below are the tables displaying summary statistics of sample size by country and sample according to malnutrition and anaemia (Table 2.1 and Table 2.2 respectively):

Table 2-1: Summary statistics of pooled sample

Country	Year	Age of Children (months)	Community Type	Gender	Sample size
Senegal	2016	0-59	Rural and Urban	Male and Female	2858
Malawi	2015-2016	0-59	Rural and Urban	Male and Female	4229
Angola	2015-2016	0-59	Rural and Urban	Male and Female	4470

Table 2-2: Pooled sample according to malnutrition and anaemia

	Malnourished	Nourished	Total
Anaemic	3916 (33.9%)	793 (6.9%)	4709 (40.7%)
Not anaemic	5677 (49.1%)	1171 (10.1%)	7848 (59.3%)

Total	9593 (83.0%)	1964(17.0%)	11557

2.3 DHS Sampling method

The Demographic and Health Survey is a nationally representative survey with the sample drawn at the national, residential and regional levels. The sampling technique employed to collect the data is two-folded. The first fold is the enumeration area (EA) which is normally drawn from census files. The second fold investigates each selected EA and draws a sample of households to interview. The DHS survey is conducted over 18-20 months, with specific topics to be covered every month to ensure proper collection of the information. The main topics that are covered include the demographic and socio-economic characteristics of the population, marriage and sexual activity, level of fertility and fertility trend, family planning, infant and child mortality, child health, nutrition of women and children, malaria, HIV/AIDS-related knowledge and women's empowerment. These topics collect information on dwelling conditions, sexual debut, beginning of a woman's childbearing, use of contraceptives, perinatal mortality, postnatal care, diarrhoea, fever, nutritional status, anaemia and many more.

2.4 Variables Selection methods

Variable selection is defined as choosing the appropriate variables to include in the final model, this can be achieved through eliminating the irrelevant or redundant variables (Ratner, 2010). Collected data most case have variables additional variables that are not required for data modelling; hence it is important that the variable selection method is managed careful to avoid including noise variables (Lee & Kim, 2016). (Steyerberg & Vergouwe, 2014) suggested that the variable selection should be purely be based on the clinical knowledge and the previous literatures.

Variable selection begins with choosing the candidate variables. This type of variables is selected based on the existing literature on the topic and also consultation with the experts on the subject (Steyerberg, 2008). The candidate variables need to be between 5-20 in order to build an adequate prediction model. Moreover, variables with many missing variables should be excluded from the final model as imputing a large number of missing data will produce results that are not reliable . There is no "one size fit all" approach when it comes to variables selection. There are various literature that recommend the inclusion of all candidate variables into the final model. This approached is called the full model approach. However, due to the practicality and the difficulty of defining the full model, this approach is considered impossible. (Hosmer, et al., 2013)

suggested that the univariate analysis should be conducted to identify the variables to include in the final model. Based on this approach, all the variables that has a p-value less than 0.25 and existing literature have shown that they are significant must be included in the final model.

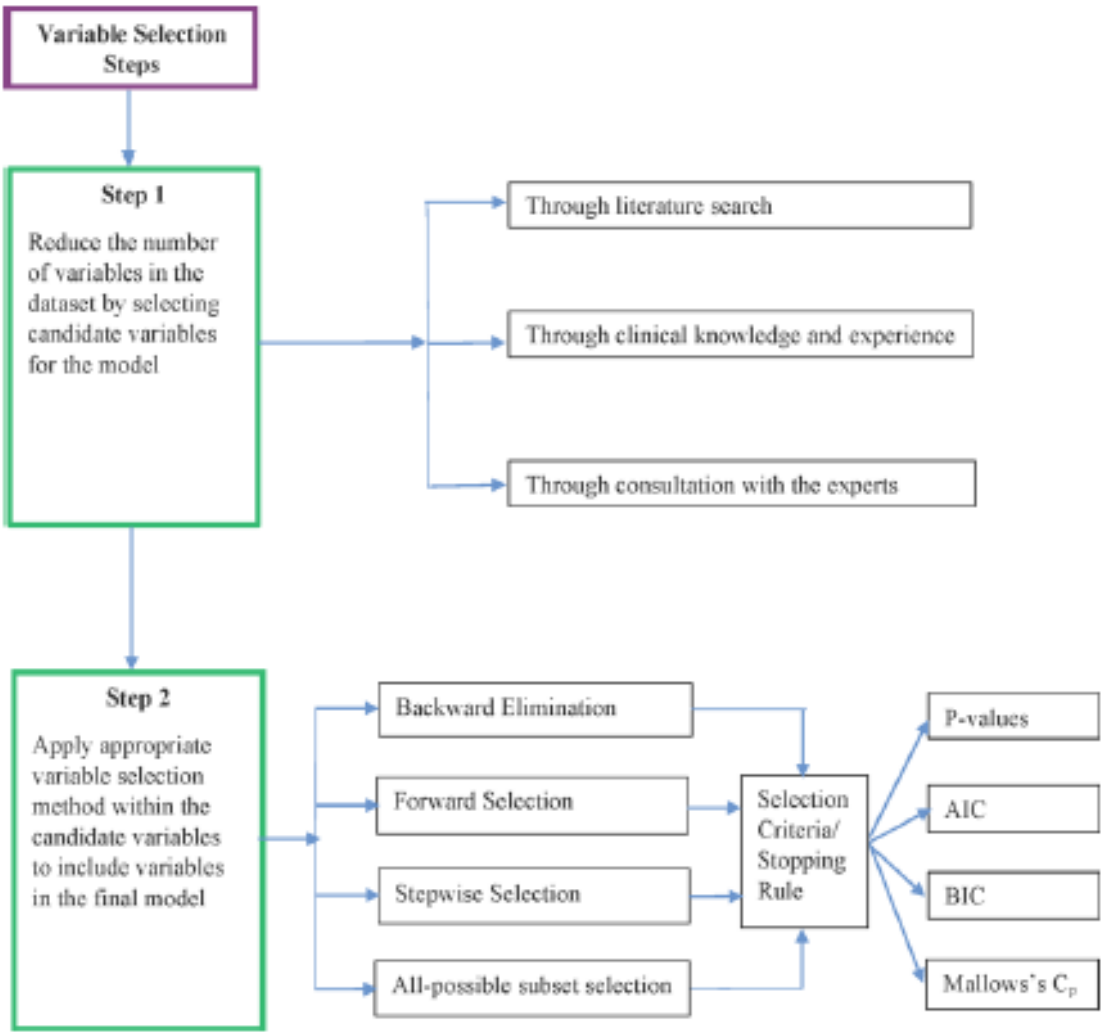

Figure 2-2: Ideal variable selection steps

Figure 2-2 display the steps additional methods that can be used to select the variables for the final model. These methods include backward selection, forward selection, stepwise selection and all-possible subsect selection. These approaches have their own advantages and disadvantages.

2.5 Malnutrition and Aneamia Variables of Interest

In consideration of the different variable selection methods proposed in section 2.4 (including Figure 2-2), this thesis used the clinical, theoretical framework and existing literature to select the candidate variables for malnutrition and anaemia study. In order for the variables to be include in final model, the univariate analysis was conducted on the candidate variables. This method is in

line with ((Steyerberg & Vergouwe, 2014), (Hosmer, et al., 2013)) recommendation of variable selection. This method combines the theoretical framework/ existing literatures and the statistical approach.

2.5.1 Outcome variables

For separate models, as per the recommendation of the WHO, the response variable, weight-for-age Z-score (WAZ) nutritional status of children were divided into three ordinal categories: Severe Malnutrition (<-3.0 WAZ), Moderate Malnutrition (-3.0 to -2.0 WAZ) and Nourished (>-2.0 WAZ). The response variable anaemia status was obtained from the anaemia level variable in the DHS dataset for all three countries. The results of the haemoglobin level from the blood test were used to decide the level of anaemia among youngsters under five. Blood specimens were collected from all eligible children under five, whose parents voluntarily consented to be tested. For the joint modelling, the response variables (children under five years, malnourished and anaemia status) were obtained from the weight-for-age (WAZ) and anaemia level variables in the DHS data. A child is classified as malnourished when the WAZ is less than -3.0 and nourished when the WAZ is greater than -3.0. By contrast, a child is classified as anaemic when the haemoglobin level of a child is less than 9.9 g per dL and a child is classified as not anaemic when the haemoglobin of the child is greater than 9.9 g per dL.

Table 2-3: Response variables level and coding for analysis for separate and joint model

Country	**Levels**	**Coding**	**Coding**
		Separate models	**Joint models**
Malnutrition	Severe	1	1
	Moderate	2	1
	Mild	3	0
Malnutrition	Severe	1	1
	Moderate	2	1
	Mild	3	0

Table 2-3 indicate how the separate model and joint model outcome variable were classified. This method is in line with the WHO classification of these diseases. The outcome variables for separate models have 3 levels, whereas the outcome variables for joint model have tow levels.

2.5.2 Explanatory variables

Socio-economic, demographic, health and environmental elements are known to be the contributing factors to malnutrition and anaemia status. The theoretical framework and literature were used to select the explanatory variables ((Nkurunziza, et al., 2017), (Wirth, et al., 2017)). The community-level variable included in the study is the type of resident (Rural or Urban). Household-level variables included in the study are household size (0-5, 6-10, 11-15 or > 15), sex of household head (Male or Female), mother's level of education (Primary, Secondary or Higher), birth interval (<24, 24-47 or > 47), marital status (unmarried, divorced, married) and wealth index (Not poor, Middle or Poor).

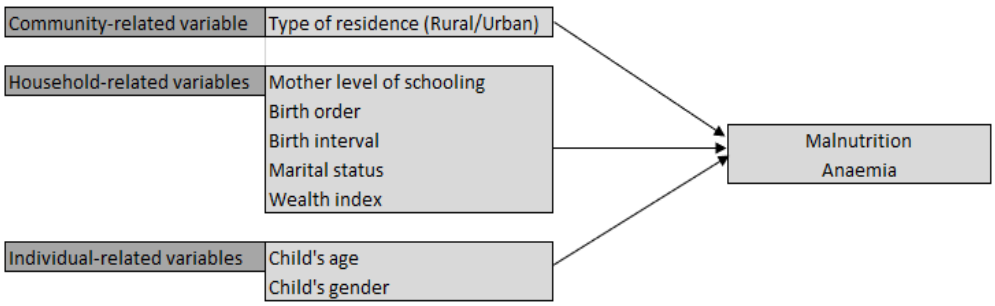

Figure 2-3: Potential risk factors of malnutrition and anaemia among children under five years

The wealth index is a composite measure of household living standards. It is a calculation comprising household's assets, ownership of living stock, household construction material, and water and sanitation facilities. Individual-level variables are the child's age in months (< 12, 12-23, 24-35, 36-47 or 48-59), sex of child (Male or Female), childbirth order (2-3, 4-5 or > 5).

Table 2-4: Potential risk factors of malnutrition and anaemia among children under five years

Explanatory	Levels	Coding
Residence	Rural	1
	Urban	2
Childs's age (months)	0 - 12	1
	13 - 23	2
	24 - 35	3
	36 - 47	4
	48 - 59	5
Child's gender	Male	1
	Female	2
Mother's schooling	Primary	1
	Secondary	2

	Higher	3
Birth interval	< 24	1
	24-47	2
	> 47	3
Wealth status	Poor	1
	Middle	2
	Not poor	3
Birth order	2-3	1
	4-5	2
	> 5	3
Marital status	Married	1
	Living together	2
	Widowed	3

2.5.3 Computation of wealth index using Principal Component Analysis (PCA)

Ideal, household wealth is reflected through its income or expenditure information. Based on the inaccuracy of income data collected and the growing usage of PCA, it is recommended that the household wealth index is measured through asset data (Vyas & Kumaranayake, 2006). Factor and principal component analysis are frequently employed in the measurement of differences in socio-economic status using socio-demographic data (Booysen, 2002). However, factor and principal component analysis are typically used for different reasons; they are not mechanically the same; nor do they have the same underlying linear model. Principal component analysis extracts all the factors underlying a set of variables, and completely explains the variance in each variable. Specifically, with component analyses, unities (values of 1.0) are inserted in the diagonal of the correlation matrix to ensure that full variance is brought into the factor matrix (Journee, et al., 2010). In contrast, factor analysis only analyses the shared variance. To employ only common variance in the estimation of the factors, communities - instead of unities - are inserted in the diagonal. Thus, factors resulting from common factor analysis are based only on the common variance (Jolliffe, 2002). Principal component analysis was used in creating the asset index in this study, since the primary concern is data reduction, that is, focusing on the minimum number of factors needed to account for the maximum portion of the total variance as represented in the original set of variables. According to (Richardson, 2009), Principal Component Analysis (PCA) defined as a multivariate statistical method employed to decrease the number of variables without losing much information in the process. PCA method attains this by creating a fewer number of variables which explains most of the variation in the original data. The new variables which are

created, known as principal components, are linear combinations of the original variables. The first component will explain much of the variation in the original data.

Given n variables $X_1, X_2, X_3, \dots, X_n$ measured in m households, the n principal components $Y_1, Y_2, Y_3, \dots, Y_n$ are induced and are uncorrelated linear combinations of the original variables, $X_1, X_2, X_3, \dots, X_n$ given as

$$\begin{aligned} Y_1 &= p_{11}X_1 + p_{12}X_2 + \cdots + p_{1n}X_n \\ Y_2 &= p_{21}X_1 + p_{22}X_2 + \cdots + p_{2n}X_n \\ Y_3 &= p_{31}X_1 + p_{32}X_2 + \cdots + p_{3n}X_n \\ &\vdots \\ &\vdots \\ &\vdots \\ Y_n &= p_{n1}X_1 + p_{n2}X_2 + \cdots + p_{nn}X_n \end{aligned}$$

This system of equations can be expressed as **y** = **Px**, where **y** = $(Y_1, Y_2, Y_3, \dots, Y_n)$, **x** = $(X_1, X_2, X_3, \dots, X_n)$ and **P** is the matrix of coefficients.

The coefficients of the first principal component, $p_{11},\ p_{12},\ p_{13}, \dots, p_{1n}$, are chosen in such a way that the variance of Y_1 is maximized, subject to the constraint such that $p_{11}^2 + p_{12}^2 + p_{13}^2 + \cdots + p_{1n}^2 = 1$. The variance of this component is equal to λ_1, the largest eigenvalue of matrix **P**. The sum of first principal component and second principal component which are entirely uncorrelated has a variance equal to λ_2 [23]. This component explains additional, but less variation, in the original variable than the first component, subject to the same constraint. Principal components up to at most n are defined in a similar way. Eigenvalues of co-variance or correlation matrix, and scree plot provide a number of components that need to be retained for the analysis of the original data.

2.5.4 Maximum likelihood estimates of the principal components

Let $x_1, x_2, x_3, \dots, x_N$ be $N(> p)$ observations for $N(\mu; \Theta)$, where Σ is a matrix with p different characteristic roots, then a set of maximum likelihood estimates of $\lambda_1,\ \lambda_2,\ \lambda_3, \dots, \lambda_p$ and $\beta^1, \beta^2, \beta^3, \dots, \beta^p$ are the roots $k_1 > k_2 > k_3 > \dots > k_p$ of

$$\left|\hat{\theta} - k\boldsymbol{I}\right| = 0. \qquad \textbf{(2-1)}$$

A set of corresponding vectors $\beta^1, \beta^2, \beta^3, \ldots, \beta^p$ satisfying

$$(\hat{\theta} - k\mathbf{I})b^i = 0,$$

and

$$b^{(i)\prime} b^{(i)} = 0$$

where $\hat{\theta}$ is the maximum likelihood of θ.

The challenging aspect of statistical inference in principal component analysis is the computation of the vectors $\beta^1, \beta^2, \beta^3, \ldots, \beta^p$ and the scalar $\lambda 1, \lambda 2, \lambda 3, \ldots, \lambda p$ (Anderson, 1958).

2.6 Computation of the maximum likelihood estimates of the principal component

There are numerous avenues of calculating the characteristic roots and characteristic vectors (principal components) of a matrix $\hat{\theta}$ (Anderson, 1958). However, only two methods will be discussed in this section.

2.6.1 Determinant expansion

By expanding the determinant equation

$$\left|\hat{\theta} - k\mathbf{I}\right| = 0, \qquad \textbf{(2-2)}$$

Solving the resulting p^{th} degree equation in k for the roots k1 > k2 > k3 >…> kp, then $(\hat{\theta} - k\mathbf{I})b^i$ is of rank $p - 1$, and a solution of $\left(\hat{\theta} - k\mathbf{I}\right)b_j^i = 0$ can be obtained by taking b_j^i as the cofactor of the element in the first column and j^{th} row of $(\hat{\theta} - k\mathbf{I})$.

2.6.2 Iterative method

The corresponding characteristic vector and the characteristic root equation can be written as $\sum \mathbf{x} = \lambda \mathbf{x}$, where we have defined an equation for the population. Let $x_{(0)}$ be any vector not orthogonal to the first characteristic vector and define

$$x_{(i)} = \sum x_{(i-1)}, for\ i = 1, 2, 3, \ldots$$

$$y_{(i)} = \frac{1}{\sqrt{x'_{(i)} x_{(i)}}} x_{(i)}, for\ i = 0, 1, 2, \ldots$$

Then it can be shown that

$$\lim_{i \to \infty} y_{(i)} = \pm \beta'$$

and

$$\lim_{i\to\infty} x'_{(i)}x_{(i)} = \lambda_1^2.$$

2.7 Application of PCA into ADHS, MDHS and SDHS data

Principal Component Analysis was conducted to create an asset index (which will be used to define household wealth status) using 2016 ADHS, 2016 MDHS and 2016 SDHS data. The demographic health survey included information regarding the ownership of durable goods, housing characteristics, access to services along with basic demographics, concerning household size and composition. We firstly re-coded the household variables into dichotomous variables. This assisted in differentiating between households that own the particular asset and others that do not own the asset. Hence, all variables take on a value of zero or one.

Tables 2-5, 2-6 and 2-7 shows the total variance explained by each component when using Angola, Malawi and Senegal DHS data, respectively. The tables only display principal components with eigenvalues greater than 1.00 for interpretation (Kaiser's rule).

Table 2-5: Total variance explained from PCA - Angola DHS data

Component	Eigenvalue	% of variance	Cumulative %
1	5.877	17.284	17.284
2	1.876	5.516	22.800
3	1.421	4.180	26.981
4	1.397	4.110	31.091
5	1.306	3.841	38.390
6	1.176	3.458	43.932
7	1.147	3.374	41.764
8	1.127	3.414	45.078
9	1.112	3.272	48.350
10	1.106	3.252	51.602
11	1.058	3.113	54.715
12	1.047	3.078	57.793
13	1.022	3.005	60.798
14	1.009	2.969	63.767

15	1.005	2.957	66.724

Table 2-6: Total variance explained from PCA - Malawi DHS data

Component	Eigenvalue	% of variance	Cumulative %
1	5.543	14.981	14.981
2	2.336	6.315	21.298
3	1.711	4.624	25.920
4	1.444	3.904	29.824
5	1.287	3.478	33.301
6	1.180	3.191	36.492
7	1.123	3.036	39.528
8	1.108	2.995	42.523
9	1.046	2.827	45.349
10	1.040	2.810	48.160
11	1.024	2.769	50.929
12	1.017	2.749	53.678
13	1.009	2.727	56.405
14	1.004	2.712	59.117
15	1.002	2.709	61.826
16	1.001	2.705	64.531
17	1.000	2.703	67.234

Table 2-7: Total variance explained from PCA - Senegal DHS data

Component	Eigenvalue	% of variance	Cumulative %
1	4.291	13.410	13.410
2	2.148	6.713	20.123
3	1.665	5.202	25.326
4	1.381	4.315	29.641
5	1.269	3.967	33.607
6	1.165	3.642	37.249

7	1.134	3.545	40.794
8	1.104	3.450	42.244
9	1.079	3.371	47.615
10	1.033	3.229	50.844
11	1.031	3.222	54.066
12	1.014	3.169	57.235
13	1.008	3.151	60.386
14	1.003	3.134	63.519
15	1.001	3.129	66.648

The results of PCA presented in Table 5 indicate that the first principal component explains 17,284% of the variation in the original data. The second and third principal components explain 5,516% and 4,180% of the variation in the original data, respectively. Whereas, the results of PCA presented in Table 6 indicate that the first principal component explains 14,981% of the variation in the original data. The second and third principal components explain 6,315% and 4,624% of the variation in the original data, respectively. Table 2.7 indicates that the first principal component explains 13,410% of the variation in the original data. The second and third principal components explain 6,713% and 5,202% of the variation in the original data, respectively. In all the tables above, each subsequent component explains a decreasing proportion of variance.

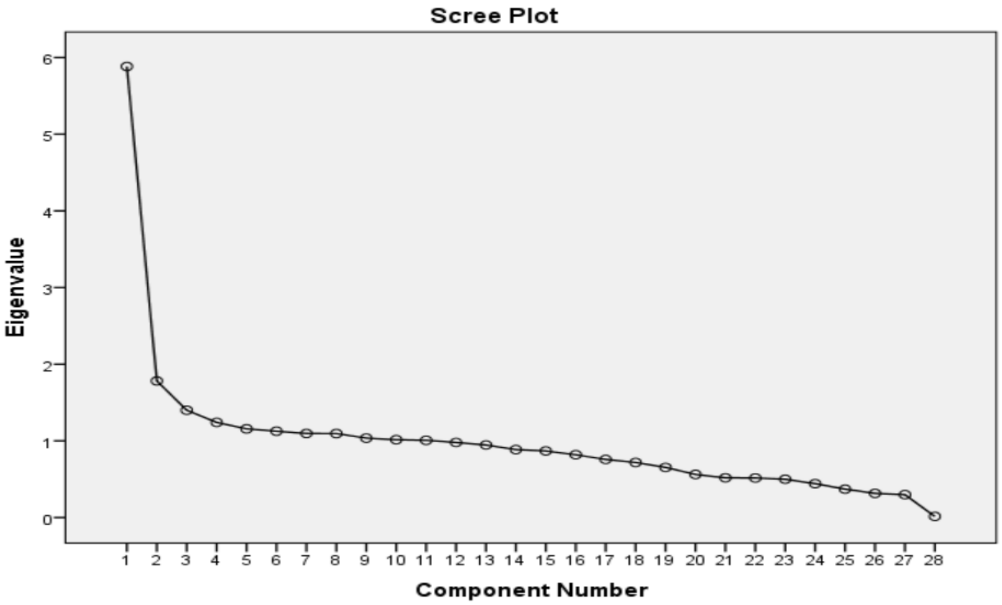

Figure 2-4 Scree plot - Angola DHS data

The scree plots in Figures 2-3, 2-4 and 2-5 display the eigenvalues associated with a component in descending order against component number. The results from three scree plots indicate that the first three components would be enough to explain the original variables. These three

components are declared to be enough in explaining the original variables because the slope of the curve is clearly levelling off (the elbow) after the third component. As a first step in the calculation of a single index, factor score coefficients, also known as component scores, were predicted using the regression method. Factor scores are the scores of each case on each factor.

To compute the factor scores for a given case on a given factor, the case's standardized score on each variable is multiplied by the corresponding factor loading of the variable for the given factor, and these products are summed. This calculation was carried out using SPSS procedure and factor scores were saved as variables in subsequent calculations involving factor scores. The scree plots revealed that only the first three principal components sufficiently explain the original variables.

From the Angola DHS data, the three components explained 26.981 percent of the total variation, with the first, second, and third component, explaining 17.284 percent, 5.516 percent and 4.180 percent, respectively. The results from Malawi DHS data revealed that the first three components explained 25.920 percent of the total variation, with the first, second, and third components, explaining 14.981 percent, 6.315 percent and 4.624 percent respectively.

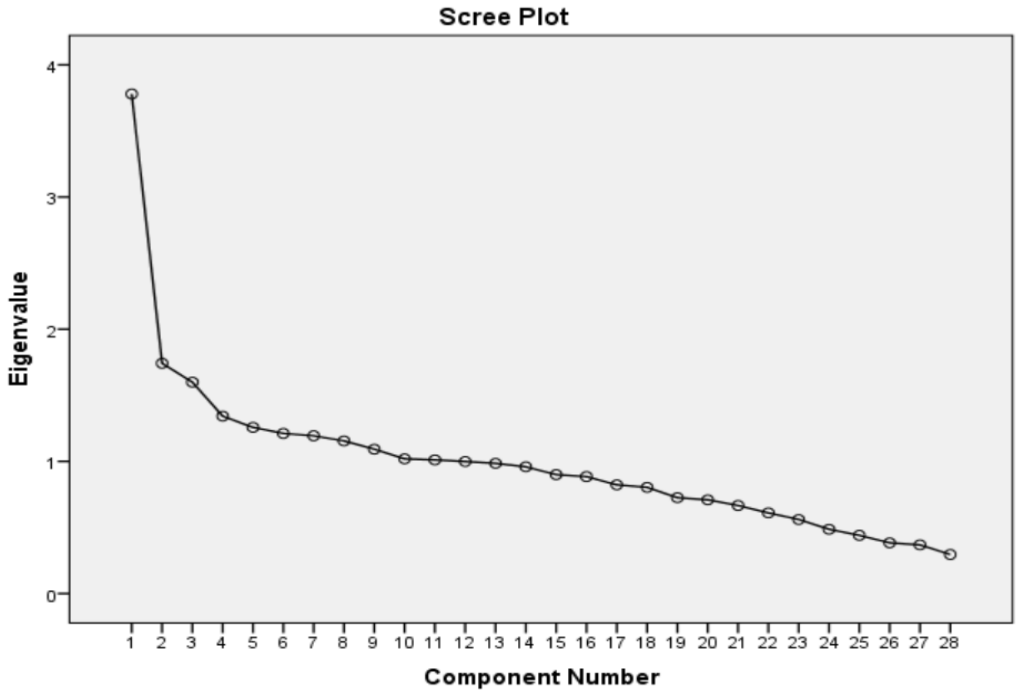

Figure 2-5: Scree plot - Malawi data

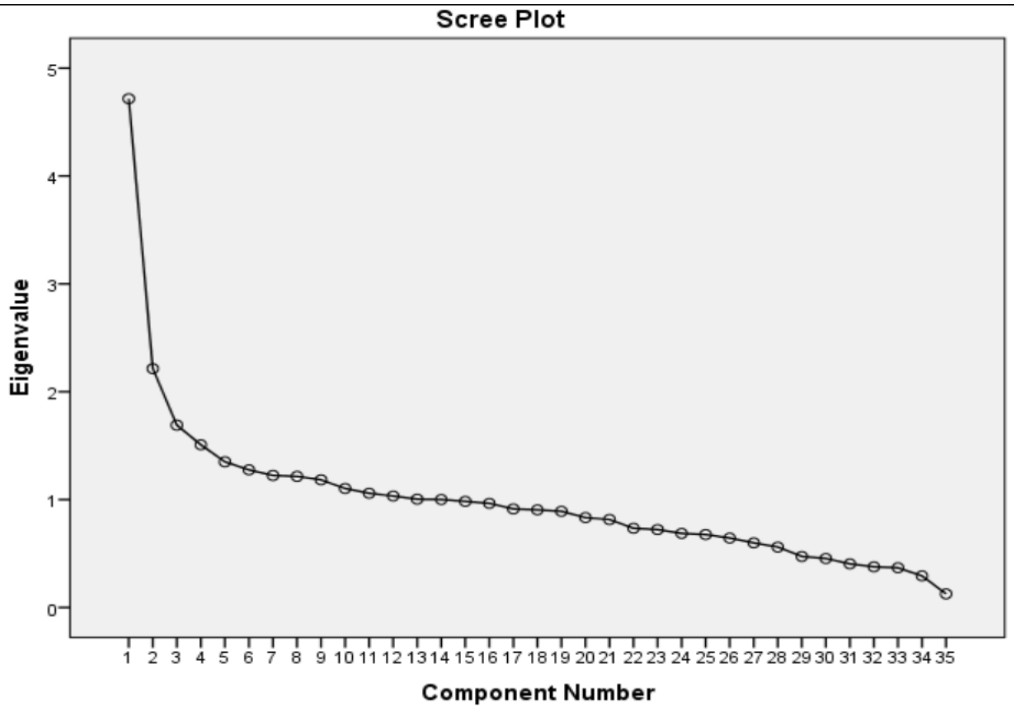

Figure 2-6: Scree plot - Senegal data

The Senegal DHS data results showed that the three components explained 25.326 out of a hundred total variation, with the first, second, and third component, explaining 13.410 out of a hundred, 6.713 out of a hundred and 5.315 out of a hundred, respectively. It can be observed from the results that the significance of the factors in measuring overall socio-economic conditions is different. Using the percentages as weights on the factor score coefficients, a Non-standardized Index (NSI) was developed for each household, using the formula:

$$NSI = \frac{pc_1}{\sum_1^3 pc_i}(factor\ score\ 1) + \frac{pc_2}{\sum_1^3 pc_i}(factor\ score\ 2) + \frac{pc_3}{\sum_1^3 pc_i}(factor\ score\ 3) \qquad \textbf{(2-3)}$$

where pc_1 = first principal component, pc_2 = second principal component and pc_3 = third principal component for each country DHS data. Example using the Angola PCA results:

$$NSI = \frac{17.284}{26.981}(factor\ score\ 1) + \frac{5.516}{26.981}(factor\ score\ 2) + \frac{4.180}{26.981}(factor\ score\ 3)$$

This index measures the socio-economic status of one household relative to the other on a linear scale. The challenge on the interpretation of the results is that the index value can be positive or negative. Therefore, a Standardized Index (SI) was developed, the value of which can range from 0 to 1 (Achia, et al., 2010), using the formula:

$$SI = \frac{NSI\ of\ factor1 - MinNSI}{MaxNSI - MinNSI} \qquad \textbf{(2-4)}$$

From the above formula, factor 1 was used since only the factor scores of the eigenvectors of the first principal component are used in construction of the socio-economic index (Achia, et al., 2010). Households with SI less than 0.30 (Prior rule of thumb, which denoted all loading of 0.30 as having practical significance, this approach is only used when the sample size is greater than 350 (Anderson, et al., 1998)), were classified as poor. Whereas households with SI between 0.30 to 0.60 were classified as middle, and households with SI greater than 0.60 were classified as not poor. Hence, a new categorical variable classifying socio-economic status of each household was created.

The Kaiser-Meyer Olkin (KMO) and Bartlett's sphericity check the interest in the implementation of the principal component analysis on a data set. Bartlett's sphericity test checks if there is a certain redundancy between the variables that can be summarized with a fewer number of factors, whereas the KMO index compares the values of correlation between variables and those of the partial correlation. If the KMO index is high (≈1) the PCA can act efficiently; if KMO is low (≈0) the PCA is irrelevant. In this research, only the value of KMO was used in checking the reliability of applying PCA (Kaiser, 1974).

According to (Tabachnick & Fidell, 2007), a minimum of 0.50 KMO value is recommended for PCA to act efficiently. This acceptance level was also supported by (Hair, et al., 2006), (Wang, et al., 2020) & (Field, 2000). The index of KMO from the three countries' data is presented in Table 2-8 and the results indicate that the PCA can act efficiently on the data set (KMO test statistics for all three countries is above the acceptance value of 0.7 index).

Table 2-8: KMO test

Data	Test Statistics	Index
Angola DHS	KMO	0.717
Malawi DHS	KMO	0.710
Senegal DHS	KMO	0.739

Cronbach's Alpha or coefficient alpha was further conducted to check reliability. Cronbach's alpha is defined as a coefficient that measures a different set of related items as one item. This method has been widely used in the field of social and organizational science (Bonett & Wright, 2014). Coefficient alpha values are defined as excellent and strong (0.91–0.94), reliable and robust (0.81–0.90), fairly high and high (0.73–0.95), good and relatively high (0.70–0.91), slightly low and reasonable (0.67–0.87), adequate and moderate (0.61–0.85), satisfactory and acceptable (0.45–0.98), sufficient (0.45–0.96), not satisfactory and low (0.4–0.55).

Table 2-9: Cronbach's Alpha

Data	Test Statistics	Coefficient
Angola DHS	Cronbach's Alpha	0.811
Malawi DHS	Cronbach's Alpha	0.840
Senegal DHS	Cronbach's Alpha	0.872

The results of the Coefficient Alpha from three data sets are presented in Table 2-9. All the values are above 0.70, which is the acceptance level.

Table 2-10 shows all variables used in the construction of the asset index. In the construction of the socio-economic index, only the factor score, that is eigenvalues of the first principal component, is used because the first principal component explains much of the variation of the original variables. Hence, the first principal component is enough in explaining the original variables. Component scores presented in Table 2-10 are of the first principal component. There are 39 principal components are extracted from Table 2-10. It can be observed that the variables take the value 1 if true, or 0 otherwise.

Table 2-10: Principal component score

Variable	Angola Comp. score	Malawi Comp. score	Senegal Comp. score
Source of water			
Piped into dwelling	0.271	0.538	0.393
Piped into compound	0.370	0.486	0.399
Public tap	0.005	0.053	-0.001
Spring	-0.237	-0.056	-0.027
Rainwater	-0.132	-.006	-0.018
Tanker truck	0.250	-0.002	0.026
Lake	-0.434	-0.107	-0.062
Bottled water	0.146	0.048	0.167
other		-0.100	-0.581
Floor material			
Earth		-0.754	-0.572

Wood	0.098	-0.001	0.008
Ceramic tiles	0.462	0.027	0.217
Cement	0.564	0.187	0.589
Other	0.031	0.739	-0.043
Wall material			
No wall	-0.029	-0.016	-0.136
Stone with mud	-0.012	-0.061	-0.099
Wood	-0.441	-0.086	0.008
Carton	-0.023	0.015	
Stone with cement	0.173	0.343	-0.121
Bricks	0.251	-0.004	0.054
Other			-0.475
Sanitation			
Flush toilet	0.687	0.604	0.592
Latrine	-0.090	-0.290	-0.133
No facility	-0.625	-0.105	-0.443
Bucket	-0.004	0.033	
Other		-0.006	
Cooking fuel			
Electricity	0.125	0.481	0.084
Gas	0.809	0.027	
Paraffin	0.034		
Charcoal	-0.195	0.566	0.339
Wood	-0.683	-0.695	
Other		-0.006	
Other durables			
Has electricity	0.783	0.821	0.806
Has radio	0.520	0.432	0.189
Has television	0.818	0.778	0.801
Has refrigerator	0.807	0.723	0.627
Has bicycle	0.091	0.128	-0.222
Has motorcycle	0.076	0.204	0.106

Has car/ truck	0.475	0.405	0.347

2.8 Descriptive Statistics

This subsection presents the descriptive analysis of each explanatory variable against the two outcomes. This information is presented in graph format reflecting percentages.

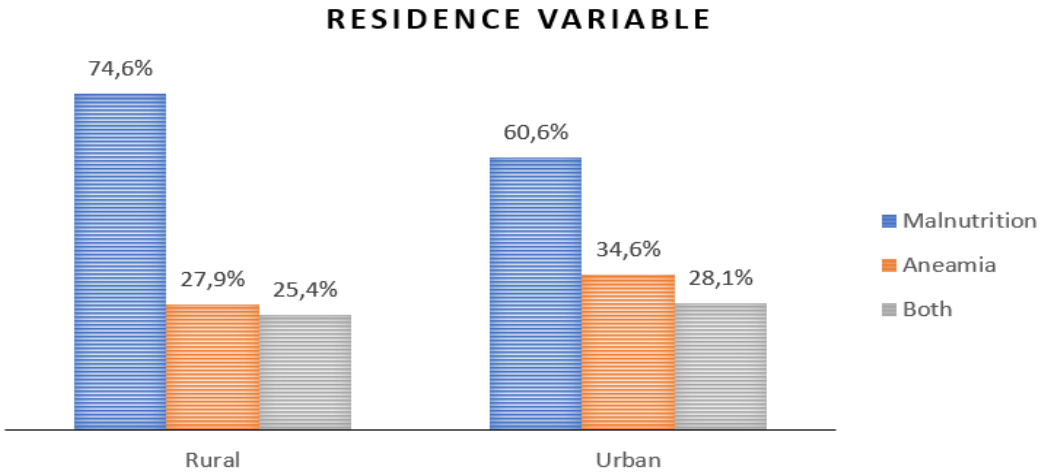

Figure 2-7: Prevalence of malnutrition, anaemia and both related to child's residence

Figure 2-6 shows that there was no substantial difference related to anaemia and both malnutrition and anaemia between children residing in rural or urban areas. A prominent difference was observed linked to the prevalence of malnutrition between children living in rural and urban areas. There was no difference in the prevalence of malnutrition, anaemia and both malnutrition and anaemia between male and female children, displayed in figure 2-7.

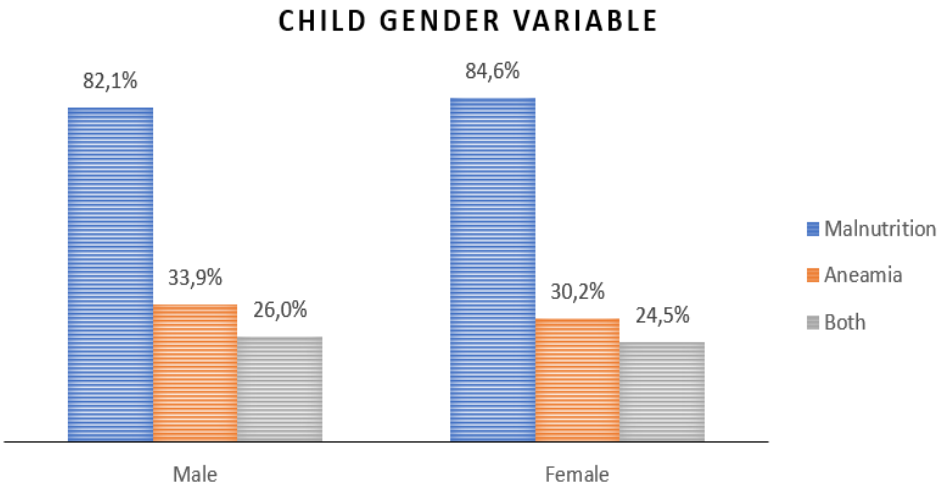

Figure 2-8: Prevalence of malnutrition, anaemia and both according to child's gender

The prevalence of malnutrition, anaemia and both malnutrition and anaemia decrease as the child grows. This is evident from Figure 2-8, where the lowest prevalence of anaemia, malnutrition and both malnutrition and anaemia was observed among children in the age group of 48-59 months.

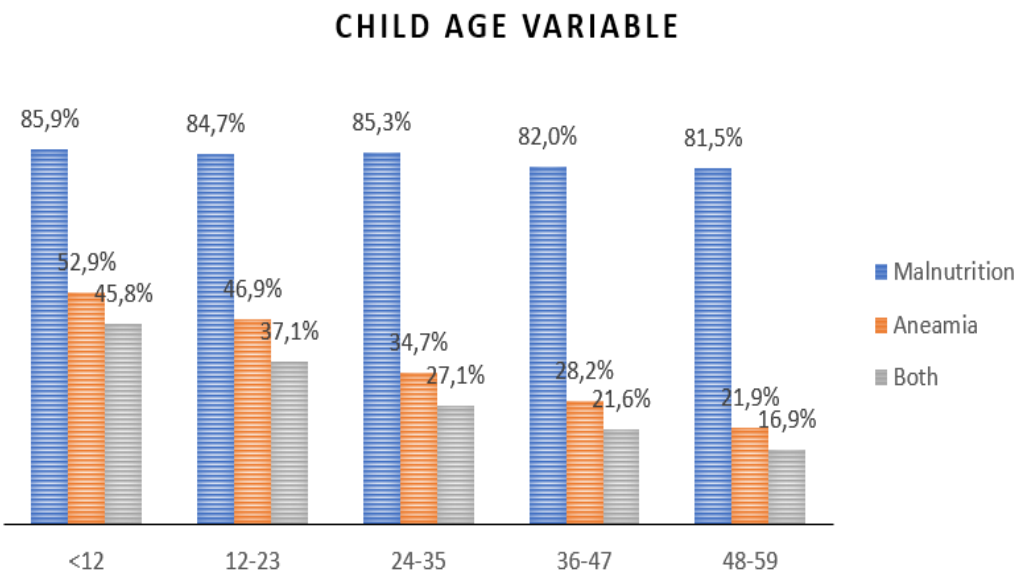

Figure 2-9: Prevalence of malnutrition, anaemia and both according to child's age

Figure 2-9 shows the prevalence of malnutrition, anaemia and both malnutrition and anaemia. It was observed that the prevalence of malnutrition and both malnutrition and anaemia decrease as the mother's level of schooling increases.

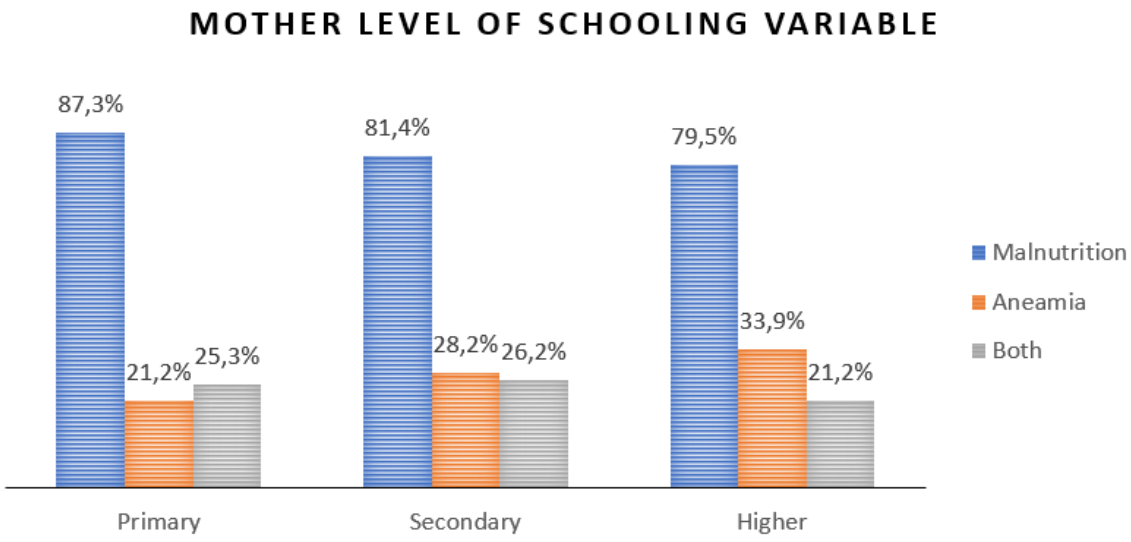

Figure 2-10: Prevalence of malnutrition, anaemia and both according to mother's level of schooling

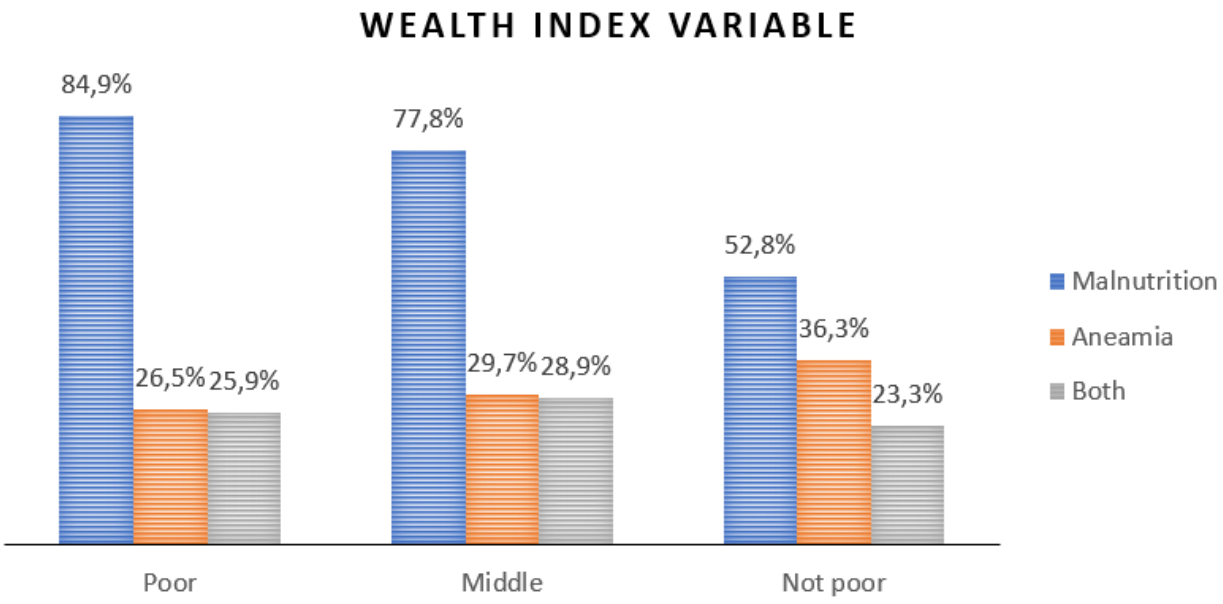

Figure 2-11: Prevalence of malnutrition, anaemia and both according to wealth index

There is a prominent gap in the prevalence of malnutrition between children residing in poor households and not poor households. There is no substantial difference regarding anaemia and both malnutrition and anaemia between children residing in poor or not poor households, as shown in Figure 2-10.

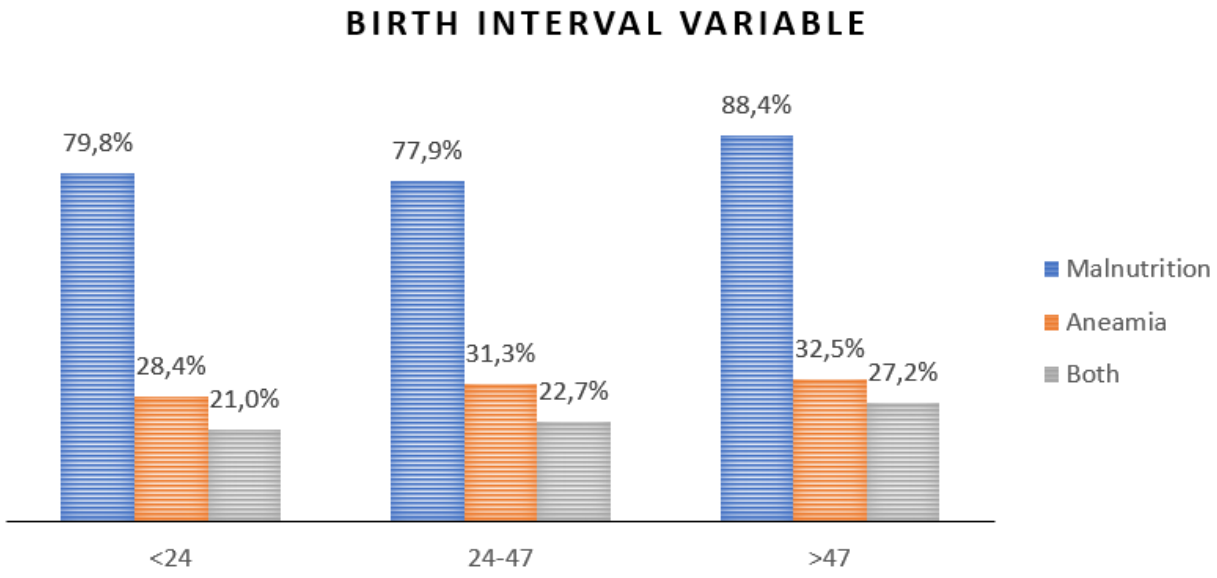

Figure 2-12: Prevalence of malnutrition, anaemia and both according to birth interval

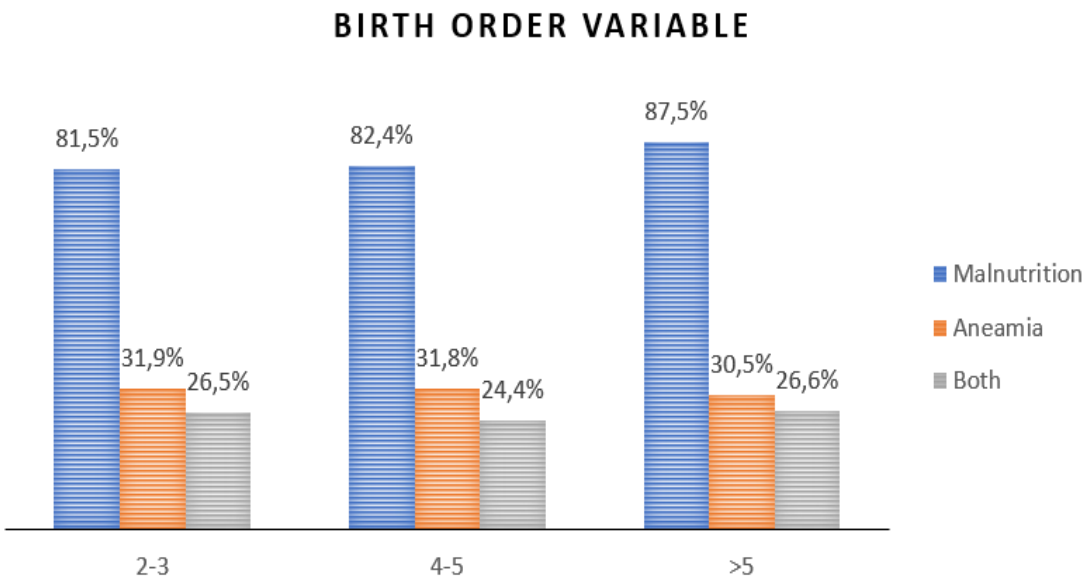

Figure 2.12: Prevalence of malnutrition, anaemia and both related to birth order

Figure 2-11 shows no prominent difference in the prevalence of anaemia and both malnutrition and anaemia between the birth intervals of <24 months and >47 months. In addition, it is evident from Figure 2-12 that there is a substantial difference in the prevalence of malnutrition, anaemia and both malnutrition and anaemia between children of birth order 2-3 and >5.

2.9 Summary

This chapter presented the source data from three countries and was clearly visualized in a map format. The outcome variables and explanatory variables were further identified and defined. The variable - wealth index - was defined based on household assets and ownership, and was carried out using the principal component analysis. Descriptive analysis of the explanatory variables was conducted to explore their relation to child malnutrition, anaemia and both malnutrition and anaemia. There was no substantial difference in terms of the prevalence of malnutrition, anaemia and both malnutrition and anaemia regarding child's gender. In addition, the prevalence of malnutrition and both malnutrition and anaemia was observed to be decreasing as household wealth status improves. The data patterns are very important as they provide an indication of statistical models to employ to achieve this study's objectives. Therefore, advanced statistical models are explored in the next chapters to confirm this chapter's findings. The choice of each model depends on the data exploration results, types of variables, hierarchical nature of the data and the study objectives.

Chapter 3 introduces the first statistical model known as the generalised linear mixed model (GLMM) to identify factors associated with childhood malnutrition and anaemia while

accounting for the complexities of the survey's sampling designs and cluster-specific effects. This is achieved through accounting for sampling design and correlation between observations of the same cluster to ensure correct estimation of standard errors and confident interval. Failure to account for sampling design in any analysis will result in underestimating standard errors and producing ineffective estimators.

Chapter 3 : Modelling risk factors of malnutrition and anaemia using GLMM

3.1 Introduction

In clustered designs, subjects are experimental nested inside larger units, for example, hospitals, hospitals, neighbourhoods and workplaces, whereas in longitudinal designs, recurring observations are nested inside subjects. These are frequently referred to as multilevel or hierarchical data in which the level-1 observations (subjects) are nested inside the higher level-2 observations (clusters). It is possible to have higher levels, for example, a four-level design could have repeated observations (level-1) nested inside subjects (level-2) who are nested inside clusters (level-3) nested in regions (level-4) (Everitt & Howell, 2005).

To analysis this type of multilevel data, random cluster, region, cluster or subject effects can be further included into the regression model to explain the correlation of the data. The resultant model is a mixed model incorporating the traditional fixed effects for the regressors and the random effects. For non-normal data, there have also been many developments. These expansions fall in the group of rubric of generalized linear mixed models (GLMMs), which extend GLMs by the inclusion of random effects in the predictor (Everitt & Howell, 2005).

A generalized linear mixed model (GLMM) is a statistical model that extends the class of generalized linear model by incorporating random effects that follow a distribution. GLMM is useful in accommodating the over-dispersion often observed among outcomes that normally have binomial or Poisson distribution, and for modelling the dependence among outcome variables in longitudinal or repeated measures design. A generalized linear mixed model can be developed for a non-linear distributed response and will allow non-linear link between the mean of the response and predictor. Its broader application is useful in various disciplines, including analysis of clustered data with longitudinal data or repeated measures (Feddag & Mesbah, 2006). We consider the form of the generalized linear mixed model:

$$g[E(Y|y)] = X\beta + Z\epsilon \quad \textbf{(3-1)}$$

where,

$\epsilon \sim N(0, \mathfrak{u})$

Z is an N x q dimension model matrix for the random effect.

β is the p x 1 dimension vector of fixed effect parameters.

X is an N x p dimension design matrix that includes covariates for the fixed effect.

ϵ is the model effect vector from normal distribution.

The distribution of $Y|y$, the conditional-response is specified unlike in the case of the generalized linear model, where the response variable Y distribution was not specified. This formula for specifying the distribution of a conditional response variable is known as conditional model specification.

3.2 Fixed and Random effect

According to (Littell, et al., 2002), the statistically fundamental concerning fixed and random effects is whether the factor effect levels are assumed as being drawn from a probability distribution. Most blocking (sampling design), control and repeated measure factors are usually treated as random, whereas many natural occurring qualitative variables can be rationalized as having fixed effects. In addition, the inference and the analysis of fixed and random factors are different (Littell, et al., 2002). Suppose that the survey was conducted using cluster sampling, and 10 cities were selected at random. If the results of the above survey were to be generalized for only the 10 cities included then the city is treated as a fixed effect, whereas if the results were to be generalized to 10 cities from a large population of cities, then the variable city is treated as a random effect. Based on the fixed effects model statistical assumptions, the responses are observed to be independent. However, this does not hold true for the random and mixed model. With respect to randomized complete sample design including the fixed and random effect, the observations taken within the same population are correlated.

3.2.1 Fixed effect

(Snijders, 2005) defined fixed effect as the average effect in the whole population expressed by the regression coefficient. Fixed effects are deliberately chosen and are the same level that could be used when continuously collecting the data. Furthermore, they can be assumed as factor levels that have been selected for insertion in the study and they are the only levels of the variable in question (Littell, et al., 2002). The aforementioned definition of fixed effects applies to quantitative effects and qualitative effects dataset. In the context of this thesis, the fixed effect describes the population average mean function. The fixed effect is used when the interest is only on analysing the impact of variables that vary over time and are further used in analysing the association between response and explanatory variables within an entity. Each entity has its characteristics that may or may not influence the explanatory variables (Torres-Reyna, 2007).

3.2.2 Random effect

The random effect is the disparity across entities and is expected to be random and uncorrelated with the dependent or independent variables in the model. Furthermore, the random effect is selected to control specific factors which are anticipated to cause random disparity in the coefficients (Kinney and Dunson, 2006). In the case of balanced data, random factors do not cause inferential problems for the test of fixed effects. However, for unbalanced data, improper treatment can lead to mistaken inferences about treatment effect (Littell, et al., 2002). Random effects assumed that its distribution has a mean of zero and unknown variance. (Penny & Holmes, 2003) conducted a study on PET (Positron Emission Tomography) using 12 subjects that were drawn randomly from a large population. Subjects were asked to either repeat a heard letter or respond with a word that began with that letter. The random effect in this case, was the subject variable, hence when drawing inferences about the population, sampling variability must be considered.

3.3 Distribution of exponential family

Following the work of (Clark & Thayer, 2004), suppose a sample of N independent random multivariate response $Yi = (Y_{1n}, \dots, Y_{in})', i = 1, \dots, N$; where y_{ij} is the j^{th} response to the i^{th} cluster or subject. We shall assume that y_{ij} depends on a p x 1 vector of fixed covariates x_{ij} associated with a vector of fixed effect $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ and on a q x 1 vector of fixed covariate z_{ij} associated with the multivariate q x 1 random effect b_i.

Given b_i and suppose we have independent variables, $Y_{1n}, \dots, Y_{in}$, and let $f(y_{ij}|b_i, \beta)$ be the probability mass function or probability density function of y_{ij} given the random and fixed effect, b_i and β respectively, then the probability mass function and probability density function is of the family of exponential distribution if it can be written in the form:

$$f(y_{ij}|b_i, \beta) = exp\left\{\frac{y_{ij}\theta_{ij} - a(\theta_{ij})}{d_{ij}(\varphi)} + c(y_{ij}, \varphi)\right\} \quad \textbf{(3-2)}$$

where θ_{ij} is the canonical parameter and φ is the scale parameters, the functions d_{ij} and c are specific to each distribution. The general expression of the conditional mean and the conditional variance of the exponential family of distribution can be obtained in terms of $a(*), d(*)$, and φ by differentiating both sides of

$$\int f(y_{ij}|b_i, \beta) = \int exp\left\{\frac{y_{ij}\theta_{ij} - a(\theta_{ij})}{d_{ij}(\varphi)} + c(y_{ij}, \varphi)\right\} = 1$$

With respect to θ_{ij}.

Differentiating once the above expression, with respect to θ_{ij} gives the general expression of the mean, similarly with the addition of the second differentiation with respect to θ_{ij} gives the expression of the variance. After a long algebra, it can be proven that the final expression of the conditional mean and the conditional variance of the exponential family of distribution is given by

$$E(y_{ij}|b_i) = \mu_{ij}^{b_i}$$
$$= h^{-1}(x'_{ij}\beta + z'_{ij}b_i)$$

and

$$Var(y_{ij}|b_i) = v(\mu_{ij}^{b_i})\, d_{ij}(\varphi)$$

where h and v are the link and the variance function. The random effect $b_1, \dots, b_N$ are mutually independent with a common underlying distribution G which depends on the unknown parameter α.

3.4 Maximum likelihood for models with both fixed and random effect

The likelihood function has played a significant role in the development of both the theory and the practice of statistics since it was introduced by Fisher in 1921 (Lee, et al., 2007). When dealing with models that include both the fixed and random effect, it is advisable to employ Hierarchical likelihood (h-likelihood) since it permits inference from models that may incorporate both fixed and random parameters (Lee, et al., 2007). The form of hierarchical log-likelihood is given as,

$$h = h(y, \mu|\beta, \Theta, \lambda)$$
$$= log\, f(y|\mu, \beta, \Theta) + log\, f(\mu, \lambda) \quad \textbf{(3-3)}$$

where $f(y|\mu, \beta, \Theta)$ and $f(\mu, \lambda)$denote the conditional density functions of y given random effects μ, and the density function of μ respectively, and β are fixed effects, Θ the dispersion parameters for the conditional distribution of $y|\mu$ and λ the parameters for random effects.

According to (Lee, et al., 2007) the log-likelihood of hierarchical is not the same as the one discovered by Fisher since there is presence of the unobserved, namely, random effect. Fixed and random parameter estimates can be computed by solving,

$$\frac{\partial h}{\partial \beta} = 0$$

and

$$\frac{\partial h}{\partial \mu} = 0$$

respectively.

The variance component can be estimated under adjusted profile likelihood given as,

$$h_p = (h + \frac{1}{2}\, In(2\pi H^{-1}).$$

3.5 Numerical approximation methods for estimates

The expansion of the above function is mathematically difficult and involves assessment of integrals where the integral's dimension is equal to the number of random effects. To work around this problem, there are different methods that have been proposed. The proposed methods include the exact integration-based methods using either numerical integration methods or Monte Carlo Markov chain (MCMC) techniques and approximations to the marginal likelihood or/and marginal moments. The mathematical methods are cantered on the adaptive quadrature. A substitute to the mathematical integration is derived from Monte Carlo Markov chain methods. These techniques become progressively difficult to use when the number of random effects increases, and the computations are intensive.

3.5.1 Penalized Quasi-likelihood estimate (PQL)

Suppose that the mean of the response y_i given the random effect $\vartheta = (\vartheta_1, \dots, \vartheta_n)$ satisfies $E[y_i|\vartheta] = h(x_i'\beta + z_i'\vartheta)$ where $h(*)$ is the inverse function of a known link function $g(*)$ and the error terms have variance given as $var[y_i|\vartheta_i] = \emptyset var(\omega_{ij})$ for $V(*)$ the usual variance function of the exponential family. Under assumption that $y_1, \dots, y_n$ are conditionally independent, given ϑ and that $\vartheta \sim N(0, \vartheta)$, we consider a Taylor series expansion of

$$\begin{aligned} Y_{ij} &= \omega_{ij} + \epsilon_{ij} \\ &= h(x_{ij}'\beta + z_{ij}'\vartheta) + \epsilon_{ij} \end{aligned} \quad \textbf{(3-4)}$$

Around the current estimates of $\hat{\beta}$ and $\hat{\vartheta}$ of the fixed effects and random effect respectively. This result

$$\begin{aligned} y_{ij} &= h\big(x_{ij}'\hat{\beta} + z_{ij}'\,\hat{\vartheta}\big) + h'\big(x_{ij}'\hat{\beta} + z_{ij}'\,\hat{\vartheta}\big)x_{ij}'\big(\beta - \hat{\beta}\big) + h'\big(x_{ij}'\hat{\beta} + z_{ij}'\,\hat{\vartheta}\big)z_{ij}'\big(\vartheta - \hat{\vartheta}\big) + \epsilon_{ij} \\ &= \hat{\omega}_{ij} + v\big(\hat{\omega}_{ij}\big) + x_{ij}'\big(\beta - \hat{\beta}\big) + v\big(\hat{\omega}_{ij}\big)z_{ij}'\big(\vartheta - \hat{\vartheta}\ \ \big) + \epsilon_{ij} \end{aligned}$$

where ϑ equals the current predictor $h\big(x_{ij}'\beta + z_{ij}'\vartheta\big)$ for the conditional mean $E[y_i|\vartheta]$. This becomes,

$$y_i = \hat{\omega} + \hat{v}x_i'\big(\beta - \hat{\beta}\big) + \hat{v}x_i'\big(\vartheta - \hat{\vartheta}\big).$$

For appropriate design matrices X_i and Z_i with $\hat{V}_i$ equals to the diagonal matrices with diagonal entries equal to $V\ (\hat{\vartheta})$. R-arranging the above equation results

$$y_i^* \equiv \hat{V}_i^{-1}\left(Y_i - \hat{\vartheta}\right) + \beta X_i + Z_i\vartheta$$
$$\approx X_i\beta + Z_i\vartheta + \epsilon_i$$

For $\epsilon_i^* = \hat{V}_i^{-1}$ and hence can be observed as linear mixed model for the pseudo data Y_i^* with fixed effect (β), random effect (ϑ) and error term ϵ. This results in an algorithm for fitting the generalized linear mixed model. Obtaining estimates from optimizing a quasi-likelihood function which only includes first and second order conditional moments augment with a penalty term on the random effect is called penalized quasi-likelihood estimates (PQL).

3.5.2 Marginal Quasi-likelihood

Another approximation that is like the PQL but is based on linear Taylor series expansion of the mean μ_i around the current estimates of $\hat{\beta}$ for fixed effects and around $\vartheta = 0$ for the random effects. The linear predictor is now of the form $\rho = h\left(X_i, \hat{\beta}\right)$ instead of $h\left(x_i'\hat{\beta} + z_i'\,\hat{\vartheta}\right)$. The pseudo data is of the form

$$y_i^* \approx X_i\beta + Z_i\vartheta + \epsilon_i^* \qquad \textbf{(3-5)}$$

Model fitting is of the same as in PQL, iterating between the calculation of the pseudo data and the fitting of the approximate linear mixed model for these pseudo data. The resultant approximations are called marginal quasi-likelihood estimates (MQL).

Biased estimates may be obtained using the PQL and MQL approaches (Hedeker, 2005). The resulted estimates may be biased towards zero. Numerous methods for handling these bias estimations have been put forth. The inclusion of bias correction terms was proposed by (Beslow & Lin, 1995), whereas, (Kuk, 1995) proposed the use of iterative bootstrap. According to (Goldstein & Rasbash, 1996), the accuracy of the approximations is increased by incorporating a second order component in the Taylor series expansion.

3.5.3 Gaussian quadrature

The Gaussian quadrature method is an approximate method of simulating the integral $I = \int_a^b y(x)dx$. Suppose that $x = \frac{(b-a)t}{2} + \frac{(a+b)t}{2}$ and $f(t) = \frac{(b-a)y(x)}{2}$, then the integral reduced to the form $\int_{-1}^{1} f(t)dt$.

The Gaussian quadrature formula is thus (Liu & Pierce, 1994):

$$\int_{-1}^{1} f(t)dt \;=\; \sum_{i=1}^{n} A_i f(t_i). \qquad \textbf{(3-6)}$$

The cusps t_i of the Gaussian quadrature formula are the roots of a Legendre polynomial of degree n, $P_n(t)$. The Legendre polynomial has real and various roots in the interval (-1,1). The weights A_i of the Gaussian quadrature formula are defined by

$$A_i = \frac{2}{\left(1 - t_i^2\right)[P_n'(t_i)]^2}$$

Gaussian quadrature formula is exact for an arbitrary polynomial of degree not higher than $2n - 1$. The remainder term of Gauss's formula Rn for the integral $\int_a^b y(x)dx$ is expressed as follows

$$R_n = \frac{(b-a)^{2n-1}n!}{(2n+1)(2n)!^3} y^{(2n)}(\xi), a \le \xi \le b.$$

The Gaussian quadrature method is applied when a sub-integral function is smooth enough and a gain in the number of cusps is essential (for instance, in calculating multiple integrals as iterated integrals).

Newton Raphson and Fisher scoring iterative method can be used to maximize the likelihood of this numerical approximation. It is recommended that these iterative methods are used in the model that has a single random effect or three nested random effects. In a complicated structures, the method will not yield the desired results (McCulloch & Neuhaus, 2005).

3.6 Newton-Raphson and Fisher Scoring Iteration method

According to (Skrondal & Rabe-Hesketh, 2004), the Newton-Raphson procedure can be derived by considering an approximation of the derivatives of the log- likelihood using a first order Taylor series expansion around the current parameter estimates Λ^k:

$$\frac{\partial l(\Lambda)}{\partial \Lambda} \approx \frac{\partial l(\Lambda^k)}{\partial \Lambda} + \frac{\partial l(\Lambda^k)}{\partial \Lambda \partial \Lambda'}(\Lambda - \Lambda^k)$$
$$= g\left(\Lambda^k\right) + H\,(\Lambda^k)(\Lambda - \Lambda^k)$$

where $g\left(\Lambda^k\right)$ is the V -dimension gradient vector and $H\,(\Lambda^k)$ is the Hessian, the vxv matrix of second derivatives of the log-likelihood with respect to the parameters, evaluated at Λ^k. The updated parameters Λ^{k+1} are the parameters for which this first order Taylor expansion is zero, that is

$$g\left(\Lambda^k\right) + H\,\left(\Lambda^k\right)\left(\Lambda^{k+1} - \Lambda^k\right) = 0$$

so that

$$(\Lambda^{k+1}) = \Lambda^k - 1/(H\,(\Lambda^k))g(\Lambda^k)$$

Note that the Taylor expansion is precise if the likelihood is quadratic in the parameters, in which case the maximum is found in a single iteration.

The Fisher scoring algorithm is like the Newton-Raphson algorithm, but the negative of Fisher's information matrix $I(\Lambda^k)$is used instead of the Hessian,

$$(\Lambda^{k+1}) = \Lambda^k + 1/(I\,(\Lambda^k))g(\Lambda^k)$$

where $I(\Lambda^k) = -E(H\,(\Lambda^k))$.

3.7 Advantages of GLMM

Generalized Linear Mixed Models (GLMM) are the best tools for analysing non-normal data that involve random effects. GLMM combine the properties of two statistical frameworks that are widely used: Linear Mixed Models and Generalized Linear Models (Bolker, et al., 2009). When using the GLMM, the researcher needs to specify the distribution, link function and structure of random effect. Linear Mixed Models and Generalized Linear Models can handle multivariate cases; however the difference to the advantage of GLMM to these models is that it releases the Gaussian assumption of error term and closely represents the real network distribution by modelling it generally with E(e) and Var(e), and it also adopts the random effect that particularly addresses the subgroup variance in the data. Random effect must not be ignored; if the researcher misses to account for it, the often-significant correlation between data coming from experimental items is not modelled, leading to invalid standard error estimates. Any conclusion may only be true of your random sample of items, not of another random sample. Modelling random effects as fixed effects makes it impossible to derive conclusions about fixed effect because unlimited variation can be attributed to a subject.

3.8 Type 3 test of fixed effects

Type 3 test of fixed effects tests the significant of each fixed effect specified when modelling. The null hypothesis is that the fixed effect variable does not significantly explain the response variable, against the alternative that it does. PROC GLIMMIX computes type 3 test of fixed effects by first constructing a type 3 L matrix for each effect. The L matrix is then used to construct the test statistic shown below;

$$\frac{\hat{\beta}L'(LQL')^{-1}L\hat{\beta}}{rank(LQL')}$$

In this chapter the null hypothesis is that the fixed effect variables do not significantly explain malnutrition or anaemia, against the alternative that it does.

3.9 Test of association definition

The Chi-square and Goodman gamma measures were employed to test for association between the outcome variable and the explanatory variable. The strength between the outcome variable and the ordinal explanatory variable was tested by employing Goodman gamma measure (Talukder, 2017). The estimator of Goodman gamma is calculated using the formula below;

$$\gamma = \frac{Ns-Nd}{Ns+Nd} \quad \textbf{(3-7)}$$

where Ns is the total number of concordant pairs and Nd is the total number of discordant pairs, whereas, the strength between the outcome variable and the nominal explanatory variable was tested by employing chi-square (Talukder, 2017). The formula for chi-square is given by:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad \textbf{(3-8)}$$

where O represents the observed frequency and E is the expected frequency under the null hypothesis. E is calculated as follows:

$$E = \frac{row\ total\ x\ column\ total}{sample\ size}$$

The association measure uses the fact that the statistic, gamma follows a normal distribution with mean = μ and standard error (SE) calculated from the delta method. The Chi-square use the fact that it follows a chi-square distribution with $(r-1)(c-1)$ degrees of freedom (df), where r is the number of categories of the covariates and c is the number of categories of response variable.

3.10 Random Intercept Model (RIM)

In a random intercepts model, the intercepts are permitted to change, and as a result, the intercept that varies across groups predicts the scores on the dependent variable for each given observation. This model considers slopes to be fixed (the same across different contexts) (Gomes, 2022). Additionally, this model offers data on intraclass correlations that are useful in deciding if multilevel models are actually necessary in the first place.

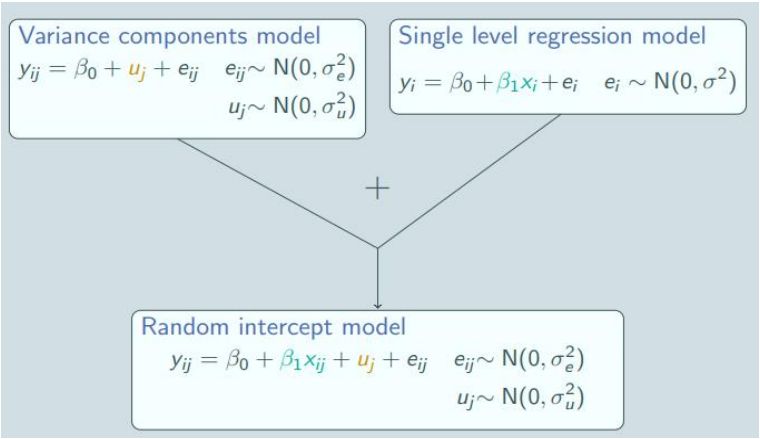

Figure 3-1: Random Intercept Model (RIM)

Figure 3-1 shows how the variance components and regression models are combined to produce a random intercept model. The random intercept model has two parts, there is a fixed part and a random part. Under fixed part, the parameters that are estimated are the coefficients $\beta_0, \beta_1, \beta_2, \ldots$ whereas the random part estimate the variances σ_u^2 and σ_e^2.

3.11 Model diagnostics

There are various statistical methods to test the goodness of fit. Testing for the goodness of fit is very important as it indicates whether the model is the best model for the data. In this thesis, we consider three types of the goodness of fit and all the models that will be discussed in the next chapters will be selected based on them.

1. **Akaike Information Criterion (AIC):** This is a traditionally used measure of goodness of model fit. The AIC attempts to avoid overfitting of the data by penalising high model complexities. AIC was primary established by (Akaike, 1973) as a method to differentiate models on a given outcome. Researchers interested in variables impacting the ranking of a certain drink and how these variables impact the ranking of the drink, can assess this through several different regression models. The type of flavour used, the price of the drink or the region the drink was produced play a role in influencing the ranking of the drink. Regression equations may incorporate price and region information, only price information, only region information, or any other combination of variables. The model reveals the association between the dependent and a specific variable. Model selection is significant, underfitting a model might not detect the true variability in the independent variable, while an overfitted model loses generality. AIC assist in selecting the model of the data. Post selecting the best model, a traditional null-hypothesis testing is used to determine the relationship between specific variables and the independent variable. An estimated parameter vector denoted as the AIC $\emptyset$ is defined as:

$$AIC = -2[l(\hat{\emptyset}) + g],$$

where $l(\hat{\emptyset})$ = log $(f(y|\emptyset))$ is the maximal log-likelihood value and g is the number of estimated parameters in the model.

2. **Bayesian Information Criterion (BIC):** The Bayesian Information Criterion originates from the Bayesian context. This approach initiates a penalty term for the number of parameters in the model when overfitting occurs. It only happens when the likelihood is increased by adding parameters to the model. The BIC is given as:

$$AIC = -2l(\hat{\emptyset}) + glog(n),$$

where $l(\hat{\emptyset})$ = log $(f(y|\emptyset))$ is the maximal log-likelihood value, g is the number of estimated parameters in the model and n is the sample size (Demyonovs, et al., 2012). This goodness testing is suitable for model comparison when the sample size is not the same. The initiated penalty term is dependent on the sample size, where the larger the sample size, the larger the penalty. However, the smaller the sample size the smaller the penalty.

Characteristics of the Bayesian information criterion

- It can measure the effectiveness of the parameterized model.
- It manages the complexity of the model where complexity is the number of parameters in the model.
- It is almost identical to the minimum description length criterion.
- It can be employed to select the number of clusters present in a dataset.
- It is almost the same as the other penalized likelihood criteria such as BIC and the Akaike information criterion.

3. **Generalized Cross Validation (GCV):** Generalised cross-validation is not an information or likelihood-based criterion like the AIC and BIC. The GCV focuses on model optimization instead of complexity and can adapt the sum of squares for residuals, based on the scoring algorithm suitable for non-Gaussian responses. Furthermore, the GCV uses square Pearson or deviance residuals for measurements. In the GLMM framework, the goodness of fit measures can be defined in terms of the deviance residuals of the form

$$D_i = D(y_i, \mu_i) = 2[l_i(y_i) - l_i(\mu_i)]$$

Where $l_i(y_i)$ is the log-likelihood of observations i, and $l_i(\mu_i)$ is the log-likelihood of the observations i assessed for the predicted mean μ_i.

3.12 Results of the GLMM on malnutrition

In this section, we firstly apply statistical methods to identify the relationship between explanatory variables and malnutrition. We further use the GLMM to identify the risk factors for malnutrition. We used the WHO Stunting Framework in choosing a set of socio-economic and demographic variables related to child malnutrition, and these variables were considered as covariates in the development of the GLMM model. The WHO Stunting Framework build upon the traditional UNICEF framework that focuses on the causes of malnutrition (Takele K., et al., 2019). The variables that were selected for this study are mother's level of education (Primary, Secondary or Higher), type of residence (Rural or Urban), household size (0-5, 6-10, 11-15 or 15), child's age in months (<12, 12-23, 24-35, 36-47 or 48-59), sex of child (Male or Female), wealth index (Poor, Middle or Not poor), birth interval (<24, 24-47 or > 47) and birth order (2-3, 4-5 or > 5).

3.12.1 Test of association

The results of the test of association between explanatory variables and the outcome variable for children under five is summarised in Table 3-1. The p-values of the covariates were obtained from chi-square test and gamma measure.

Table 3-1: Assessing the association between selected covariate and malnutrition status

Factors	Severe (%)	Moderate (%)	Mild (%)	$\gamma(p\ value)$	$\chi 2(p\ value)$
Residence setting					
Rural	412 (7.9%)	805 (15.4%)	4014 (76.7%)	-	**<0.001**
Urban	241 (3.8)	1191 (18.8%)	4894 (77.4%)		
Child's age (months)					
0 - 12	199 (14.1%)	144(10.2%)	1014 (75.8%)	**<0.001**	-
13 - 23	278 (15.3%)	199 (10.9%)	1343 (73.8%)		
24 - 35	380 (14.7%)	356 (13.8%)	1843 (71.5%)		
36 - 47	619 (21.1%)	496 (16.9%)	1819 (62.0%)		
48 - 59	643 (21.7%)	478 (16.1%)	1846 (62.2%)		
Child's gender					
Male	1061 (18.5%)	809 (14.1%)	3876 (67.4%)	-	**<0.001**
Female	997 (17.2%)	764 (13.1%)	4050 (69.7%)		

Mother's level of schooling					
Primary	1869 (20.9%)	1276 (14.3%)	5786 (64.8%)	**<0.001**	-
Secondary	124(5.1%)	258 (10.6%)	2054 (84.3%)		
Higher	1 (0.5%)	5 (2.6%)	184 (96.8%)		
Birth interval					
< 24	1182 (31.7%)	644 (17.3%)	1898 (51.0%)	**<0.001**	-
24-47	922 (22.1%)	556 (13.3%)	2698 (64.6%)		
> 47	424 (11.6%)	435 (11.9%)	2798 (76.5%)		
Wealth status					
Poor	540 (11.1%)	614 (12.7%)	3676 (76.1%)	**<0.001**	-
Middle	66 (2.6%)	657 (25.9%)	1816 (71.5%)		
Not poor	47 (1.1%)	221 (5.3%)	3920 (93.6%)		
Birth order					
2-3	466 (11.0%)	979 (23.2%)	2773 (65.7%)	**<0.001**	-
4-5	451 (11.6%)	834 (21.4%)	2604 (67.0%)		
> 5	626 (18.1%)	529 (15.3%)	2295 (66.5%)		

The results in Table 3-1 indicate that type of residence, sex of the child, child's age, mother's level of education, birth interval, wealth index and birth order are significantly associated with malnutrition at 5% level of significant. There is a high prevalence of children under five residing in rural areas who are suffering from severe malnutrition when compared to children under five residing in urban areas (7.9%, p-value <0.001). Male children are more likely to have severe malnutrition when compared to female counterparts (18.5%, p-value <0.001). Moreover, it is observed that a high proportion of children under five that have severe malnutrition condition are among the age group 36-47 and 48-59 (21.1% and 21.7% respectively, p-value <0.001). Children under five with illiterate mothers are more likely to suffer from severe malnutrition than those with literate mothers (20.9%, p-value <0.001). Furthermore, children with short birth spacing are more exposed to severe malnutrition when compared to those with long birth spacing (31.7%, p-value <0.001). In addition, children under five from poor households are more likely to suffer from severe malnutrition when compared to children under five from middle and not poor household (11.1%, p-value <0.001). Similarly, children of birth order greater than 5 are more likely to suffer from severe malnutrition when compared to children of birth order 2-3 and 4-5 (18.1% respectively, p-value <0.001).

3.12.2 GLMM application on malnutrition risk factors

In this subsection we used the GLMM to identify key risk factors associated with malnutrition. The data was analysed in SAS using the Dual Quasi- Newtown optimization technique and Gauss-Hermite Quadrature likelihood approximation. Method = Gauss-Hermite Quadrature was used because of convergence; other methods: MMPL, MSPL, QUAD, RMPL and RSPL did not converge resulting in not obtaining the parameter estimates. In this thesis we used the default convergence criterion of TOLERANCE (0.0001). The convergence is achieved when the maximum changes in the ratio of variance component and residual and the change in the residual variance are less that the specified TOLERANCE.

The GLMM that was fitted in SAS is give as:

$$logit(p_{ij}) = ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_7 X_{7ij} + \vartheta_i + \vartheta_{ij}$$
$$= \beta_0 + \beta_1 Place\ of\ residence + \beta_2 Child'sgender + \cdots + \beta_7 Birth\ order + \vartheta_i + \vartheta_{ij}$$

Where p_{ij} is the conditional probability of the response, $\beta_0, \beta_1, \ldots, \beta_7$ are the unknown parameter coefficients of fixed effect, ϑ_i and ϑ_{ij} are country and cluster level random effects, respectively.

Table 3-2: Information criteria for the comparison of two random intercept models

Model	AIC	BIC	log link	Deviance
One RIM	2663.7	2776.1	-1315.8	2631.7
Two RIM	2665.7	2785.2	-1315.8	2631.7

Table 3-2 shows the information criteria for the comparison of the two random intercept models. The results reveal that the AIC of one random intercept model is lower than that of two random intercept models (2663.7). These findings indicate that the one random intercept model is a parsimonious one.

Table 3-3: Parameter estimates and odds ratios for malnutrition

Covariates	Estimates	Standard Error	Odds Ratio	P-Value
Intercept	0.666	1.509		0.659
Residence (Ref: Urban)				
Rural	0.033	0.145	1.034	0.820

Childs's age (months) (Ref: <12)				
13 - 23	-0.560	0.249	0.571	0.823
24 - 35	0.031	0.234	1.031	0.895
36 - 47	0.099	0.228	1.104	0.662
48 - 59	0.206	0.228	1.229	0.336
Child's gender (Ref: Male)				
Female	0.041	0.109	1.042	0.705
Mother's schooling (Ref: Primary)				
Secondary	-0.446	0.196	0.640	0.023
Higher	-0.451	0.808	0.637	0.667
Birth interval (Ref: < 24)				
24-47	0.085	0.167	1.089	0.023
> 47	0.350	0.183	1.419	0.056
Wealth status (Ref: Poor)				
Middle	-0.119	0.204	0.888	<0.001
Not poor	-0.438	0.254	0.645	<0.001
Birth order (Ref: 2-3)				
4-5	-0.081	0.132	0.922	0.553
> 5	-0.385	0.132	0.680	0.004
Random effect	**Variance**	**Standard Deviation**		
Country	6.6	2.569		

The final model incorporated one random intercept (country), excluding cluster random effects. Table 3-3 presents odds ratio estimates associated with the type of residence, sex of the child, child's age, mother's level of education, birth interval, wealth index and birth order.

Table 3-4: Parameter estimates and odds ratios

Covariates	Num DF	Den DF	F Value	P-Value

Residence	1	7737	14.87	<0.001
Childs's age	4	7737	13.45	<0.001
Child's gender	1	7737	20.43	<0.001
Mother's level of schooling	2	7737	7.04	<0.001
Birth interval	2	7737	16.33	<0.001
Wealth status	2	7737	9.76	<0.001
Birth order	2	7737	10.94	<0.001

*Num DF = Numerator degrees of freedom. Den DF = Denominator degrees of freedom

A type 3 test of fixed effects tests the significance of each fixed effect specified when modelling. The null hypothesis is that the fixed effect variable does not significantly explain the response variable, against the alternative that it does. The results in Table 3-4 show that children's place of residence, age, gender, mother's level of schooling, wealth status, birth interval and birth order significantly explain malnutrition at the 5% level of significance.

3.13 Results of the GLMM on anaemia

The response variable (anaemia status children under five) was obtained from the anaemia level variable in the DHS dataset for all three countries. The results of the hemoglobin level from the blood test were used to decide the level of anaemia among children under five years. Blood specimens were collected from all eligible children under five whom their parents voluntarily offered to be tested. After obtaining blood samples, hemoglobin was carried out using a portable Hemocue analyzer. In this study, the hemoglobin level of children under five was employed to classify mild, moderate and severe anaemia levels. Children under five with a hemoglobin level from 10.0 g/dl to 10.9 g/dl was classified as mild. In addition, children under five with a hemoglobin level from 7.0 g/dl to 9.9 g/dl and children with hemoglobin less than 7 g/dl were classified as moderate and severe respectively. Similar anemia categories were used in studies by ((Muchie, 2016), (Kawo, et al., 2018), (Habyarimana, et al., 2014)).

Socio-economic, demographic, health and environmental elements of living are known to be the contributing factors regarding anemia status. The framework used to select the explanatory variables is similar to that used by ((Gaston, et al., 2018), (Kawo, et al., 2018), (Ray, et al., 2016)). The community-level variables included in the study are the setting of rural or urban residences. Household-level variables included in the study are household size (0-5, 6-10, 11-15 or > 15), sex of household head (Male or Female), mother's level of schooling (Primary, Secondary or Higher),

wealth status (Poor, Middle or Not poor), marital status (unmarried, divorced, married) and birth interval (<24, 24-47 or > 47). Wealth status is a combined measure of living standards. It is a calculated index using households' assets, ownership of livestock, household construction material and water and sanitation facilities.

3.13.1 Test of association

The Bonferroni correction was applied to make corrections for the number of tests performed. The results of the descriptive analysis were obtained from SPSS. The explanatory variables that were considered for test of association were selected based on the conceptual framework for the analysis of anaemia risk and literature. The p-value of all the factors were obtained through Goodman gamma measure and chi-square. The Goodman gamma measure was used for the ordinal covariate, whereas the chi-square test was used for the nominal covariate.

Table 3-5: Assessing the association between selected covariate and anaemia status

Factors	**Severe (%)**	**Moderate (%)**	**Mild (%)**	$\gamma(p\ value)$	$\chi 2(p\ value)$
Residence setting					
Rural	180 (2.4%)	2396 (32.0%)	4923 (65.6%)	-	**<0.001**
Urban	61 (1.5%)	1072 (26.4%)	2925 (72.1%)		
Child's age (months)					
0 - 12	22 (3.1%)	348 (49.8%)	329 (47.1%)	**<0.001**	-
13 - 23	50 (2.9%)	769 (44.0%)	927 (53.1%)		
24 - 35	60 (2.4%)	800 (32.3%)	1620 (65.3%)		
36 - 47	71 (2.2%)	860 (26.1%)	2366 (71.8%)		
48 - 59	38 (1.1%)	691 (20.7%)	2606 (78.1%)		
Child's gender					
Male	147 (2.5%)	1831 (31.4%)	3851 (66.1%)	-	**<0.001**
Female	94 (1.6%)	1637 (28.6%)	3997 (69.8%)		
Mother level of schooling					
Primary	199 (2.4%)	2615 (31.5%)	5498 (66.1%)	**<0.001**	-
Secondary	349 (11.3%)	859 (27.8%)	1881 (60.9%)		
Higher	4 (2.6%)	29 (18.6%)	123 (78.8%)		

Birth interval					
< 24	70 (3.4%)	349 (17.1%)	1616 (79.4%)	**0.046**	-
24-47	125 (2.7%)	1306 (25.7%)	3649 (71.8%)		
> 47	111 (2.5%)	1127 (25.3%)	3204 (72.1%)		
Wealth status					
Poor	178 (3.1%)	1921 (33.2%)	3680 (61.6%)	**<0.001**	-
Middle	35 (1.4%)	701 (28.3%)	1739 (70.3%)		
Not poor	28 (0.8%)	846 (25.6%)	2429 (73.5%)		
Birth order					
2-3	72 (1.7%)	1578 (36.7%)	2644 (68.1%)	0.221	-
4-5	54 (1.4%)	1258 (32.8%)	2525 (65.8%)		
> 5	48 (1.4%)	976 (28.5%)	2402 (70.1%)		
Nutritional status					
Severe	76 (3.9%)	717 (36.5%)	1171 (59.6%)	**<0.001**	-
Moderate	39 (2.5%)	477 (30.0%)	1073 (67.5%)		
Nourished	126 (1.6%)	2274 (28.4%)	5602 (70.0%)		
Marital status					
Married	16 (2.0%)	219 (28.0%)	548 (70.0%)	-	0.762
Living together	200 (2.1%)	2855 (30.0%)	6468 (67.9%)		
Widowed	23 (1.8%)	352 (28.1%)	876 (70.1%)		

The results from Table 3-5 show that the child's age, type of residence, child's gender, birth interval, mother's level of schooling, wealth status and nutritional status are associated with anaemia at a 5% level of significance. There is a high prevalence of children under five residing in rural settings in Senegal, Malawi and Angola that are affected by severe and moderate anaemia when compared to children under five residing in urban settings affected by severe and moderate anaemia (2.4% and 32.0% respectively, p-value <0.001). It can be observed that a high proportion of children in the age interval of 0 - 12 months is moderate (49.8%, p-value <0.001), whereas a high proportion of children in the age interval of 13 - 59 months is mildly anaemic. There is a higher proportion of male children that are affected by severe and moderate anaemia in Senegal, Malawi and Angola when compared to their female counterparts (2.5% and 31.4% respectively, p-value <0.001).

There is a high prevalence of under five-year children residing with primary educated mothers that are moderate anaemic when compared to under five-year children residing with secondary or higher educated mothers (31.5%, p-value <0.001). In addition, children with a shorter birth interval (< 24 months) are more exposed to severe anaemia conditions when compared to children with a longer birth interval (3.4%, p-value = 0.046). Furthermore, children under five from less fortunate households tend to be affected by severe and moderate anaemia when compared to children under five from a middle and not poor household (3.1% and 33.2% respectively, p-value <0.001). There is a high prevalence of children under five with a severe nutritional condition that are suffering from severe and moderate anaemia when compared to children with moderate and mild nutritional and anaemia status (3.9% and 36.5% respectively, p-value = <.001).

3.13.2 GLMM application on anaemia risk factors

The GLMM that was fitted in SAS is give as

$$logit(p_{ij}) = ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_7 X_{7ij} + \vartheta_i + \vartheta_{ij}$$
$$= \beta_0 + \beta_1 Place\ of\ residence + \beta_2 Child'sgender + \cdots$$
$$+ \beta_7 Nutrition\ status + \vartheta_i + \vartheta_{ij}$$

Where p_{ij} is the conditional probability of the response, $\beta_0, \beta_1, \ldots, \beta_7$ are the unknown parameter coefficients of fixed effect, ϑ_i and ϑ_{ij} are country and cluster level random effects, respectively.

The AIC and BIC information criteria were used to identify whether one random intercept model or two random intercept models would be appropriate. The information criteria (Table 3-6) results showed that one random intercept model is appropriate. Hence, in the final model country is stated as the random factor.

Table 3-6: Information criteria for the comparison of two random intercept models

Model	AIC	BIC	log link	Deviance
One RIM	2653.3	2966.1	-1305.8	2621.7
Two RIM	2765.7	3769.2	-1311.8	2631.7

Table 3-7: Parameter estimates and odds ratios

Factors	Estimate	Standard error	Odds ratio	P-Values
Intercept	-0.685	0.132		0.035

Mild				
Intercept	2.466	0.148		0.003
Moderate				
Residence				
Ref: Urban				
Rural	-0.067	0.062	0.935	0.280
Child's age				
Ref :0 - 12				
13 - 23	0.350	0.120	1.419	0.004
24 - 35	0.825	0.117	2.282	<0.001
36 - 47	1.155	0.116	3.174	<0.001
48 - 59	1.584	0.118	4.874	<0.001
Child's gender				
Ref: Male				
Female	0.180	0.050	1.197	<0.001
Mother level of schooling				
Ref: Higher				
Primary	0.105	0.273	1.111	0.702
Secondary	0.213	0.076	1.237	0.005
Birth interval				
Ref: 0- 24 months				
24-47	- 0.022	0.091	0.978	0.809
> 47	- 0.083	0.057	0.920	0.144
Wealth status				
Ref: Not poor				
Poor	0.270	0.076	1.310	<0.001
Middle	0.194	0.073	1.214	0.008
Nutritional status				
Ref: Nourished				
Severe	0.400	0.069	1.492	<0.001

Moderate	0.194	0.093	1.214	0.037
McFadden R Squared	0.254			

The results of the multilevel analysis were obtained from SAS. The multilevel analysis table (Table 3-7) presents the findings correlated with a child's age, residence setting, a child's gender, birth interval, mother's schooling, wealth status, and nutritional status. Regarding children age less than 12 months, it is observed that children in the age interval 13–23, 24–35, 36–47 and 48–59 are more likely to be affected by anaemia (OR = 1.419, 2.282, 3.174 and 4.874 respectively). Furthermore, the odds ratio of being affected by anaemia among under five-year children is observed to be increasing as the age interval increases. Female children under five in Senegal, Malawi, and Angola are 1.197 times at risk of being affected by anaemia when compared to their male counterparts (p-value ≤ 0.001). Children under five residing in a household with a mother that attained a secondary level of schooling are 1.237 times at risk of being affected by anaemia when compared to children under five residing in the household with a mother that obtained a higher level of schooling. Whereas, children under five living in the household with a mother that obtained primary level of schooling are 1.111 times at risk of being affected by anaemia when compared to children under five living in the household with a mother that obtained a higher level of schooling. Children under five residing in a poor household are 1.310 times more likely to be affected by anaemia when compared to children residing in not poor households (p-value ≤ 0.001), whereas, children under five residing in a middle-classified household are 1.214 times more likely to be affected by anaemia when compared to children residing in not poor households (p-value = 0.008). Severely and moderately malnourished children from Senegal, Malawi, and Angola are at a higher risk of being affected by anaemia when compared to children that are nourished (OR = 1.492 and 1.914 respectively).

Table 3-8: Type 3 analysis results

Factors	Num DF	Den DF	F Value	P Value
Residence	1	7737	1.17	0.280
Child's age	4	7737	80.02	<0.001
Child's gender	1	7737	12.88	<0.001
Mother's Level of Schooling	2	7737	3.89	0.021
Birthing Interval	2	7737	1.13	0.322
Wealth Status	2	7737	7.15	<0.001

Nutritional Status	2	7737	18.08	<0.001

*Num DF = Numerator degrees of freedom. Den DF = Denominator degrees of freedom.

Type 3 test of fixed effects tests the significance of each fixed effect specified when modeling. The null hypothesis is that the fixed effect variable does not significantly explain the response variable, against the alternative that it does. The results from Table 3-8 show that children's age, gender, mother's level of schooling, wealth status and nutritional status significantly explain anaemia at 5% level of significance.

3.14 Summary

In this chapter, the GLMM was adopted to assess the malnutrition and anaemia using the 2016 DHS data from Angola, Malawi and Senegal. In the analysis of identifying the risk factors of malnutrition, the sampling design was incorporated in the model. Based on malnutrition results, the study showed that children's place of residence, age, gender, mother's level of schooling, wealth status, birth interval and birth order significantly explain malnutrition at the 5% level of significance. The analysis showed that female children in Angola, Malawi and Senegal are at a higher risk of being affected by malnutrition when compared to their male counterparts. Results also reveal that an increase in children's age is positively associated with being exposed to malnutrition. Children aged 24-35 months are at a higher risk of being affected. These findings have indicated the need for the health institutes of Angola, Malawi and Senegal to monitor the wellbeing of children of all ages, not only younger children, as malnutrition affects older children as well. A mother's level of education is associated with malnutrition. An improved mother's level of schooling reduces child's risk of being affected by malnutrition. Hence, Angola, Malawi and Senegal need to strengthen the educational system and develop strategies for channelling educational services to the designated populations. The results further indicated that children with mothers having attained at least secondary level education are less likely to be affected by malnutrition. These results are congruent with (Takele K., et al., 2019). The household wealth index is another determinant found to be associated with malnutrition. Improving household wealth reduces the risk of children's exposure to malnutrition. This relates to the affordability of accessing the health services and food consumption. This finding is similar to the results of the studies that were conducted in Burkina Faso, Bangladesh and Ethiopia: ((Talukder, 2017) (Poda, et al., 2017) (Abuka & Tsegaw, 2017)). Hence, the household wealth index factor can be thought to be affecting all developing countries.

The results have shown the necessity for collaboration among the three countries in order to achieve the SGD target. African continent is a developing continent with limited resources for equity/ uplifting communities and for meeting their daily requirements. Therefore, it is recommended that Angola, Malawi and Senegal work jointly to eradicate child malnutrition.

Based on the anaemia results, the study showed that children's age, gender, mother's level of schooling, wealth status and nutritional status significantly explain anaemia in Senegal, Malawi and Angola. Multilevel analysis findings show that female children are more likely to be affected by anaemia when compared to their male counterparts. Results further reveal that a unit increase in children's age, positively increases the chances of being exposed to anaemia. Children of age greater than 13 months are more likely to be affected by anaemia. This finding indicates that children transitioning from breastfeeding to normal diets are more exposed to anaemia. Mother's level of schooling is found to be significantly linked to anaemia. Based on these findings, it can be observed that improving mother's level of schooling reduces the risk of children being affected by anaemia. The results revealed that children living with mothers who obtained primary and secondary levels of schooling are at a higher risk of being exposed to anaemia. Similar findings were obtained by (Takele K., et al., 2019). Household wealth status results show that children under five living in a household that is poor or middle are more exposed to anaemia. This variable is found to be significant in explaining anaemia. The governments of Senegal, Malawi and Angola need to focus on how to uplift the population of poor and middle households in the economic perspective through the introduction of financial programs. The findings of malnutrition status indicate that in order to control children under five anaemic condition, the focus needs be directed to severely and moderately affected children. This agrees with the results of the study conducted by ((Khan, et al., 2016), (Zhoa, et al., 2012)).

To achieve the 2030 SGD target, there is a need for Senegal, Malawi and Angola to initiate programs that will focus on improving the economy of all communities to close the gap between poor and not poor households and improve women's education. Furthermore, with limited resources on the African continent, this suggests the necessity of collaboration between Senegal, Malawi and Angola to achieve the 2030 SGD target. Collaboration among the three countries is pivotal in ensuring that all the identified factors are controlled.

Chapter 3 explored the utilization GLMM to model malnutrition and anaemia, respectively. The findings provided risk factors associated with malnutrition and anaemia; however there is a need to jointly model malnutrition and anaemia. This is necessary as it will provide the shared or identifying overlapping risk factors of malnutrition and anaemia. The joint modelling has the

greater advantage over univariate analysis which includes enhancing control over Type I error rates during multiple testing, easy interpretation and efficiency in estimating parameters. Also, the correlation between the outcomes can be quantified and controlled. Therefore, the subsequent chapter concerns joint modelling of malnutrition and anaemia.

Chapter 4 : Joint modelling of malnutrition and anaemia risk factors

4.1 Introduction

Joint modelling, also known as multivariate analysis, is modelling of two or more related response variables. The joint modelling of diseases is growing in the area of science. Joint modelling is an upgrading over traditional existing modelling since it considers the longitudinal observations of variables that are predictive of an event. Predictions from joint models have greater accuracy because they are custom-made to account for individual variability. These predictions provide relatively accurate characterizations of individual disease progression, which might be significant in the timing of interventions (Broatch & Karl, 2018).

The subsequent chapters of this thesis will be exploring the benefits of modelling malnutrition and anaemia jointly, assuming conditional independence given correlation between distinct diseases effects for each outcome. For some outcomes, the joint models profit from significantly enhanced median log-loss and complete residuals of cross validation predictions. The joint model offers the ability to assess for significant associations between high-level hierarchical effects (e.g. random team effects), since significant predictors for outcomes at the community level may not be important at the household level.

In the previous chapter (Chapter 3) we adopted the generalised linear mixed model (GLMM) to identify the risk factors of malnutrition and anaemia through the univariate analysis. The following chapters will be discussing the commonly used joint modelling of malnutrition and anaemia. The relationship between malnutrition and anaemia can be defined by several factors which stem from no access to health services and education on dietary programs for example. Childhood malnutrition as an indicator of anaemia, would only be suitable and valuable in areas where anaemia is the key driver of malnutrition. It is very significant to assess the association between malnutrition and anaemia and how it changes over time.
The joint model approach has the following advantages over univariate analyses:

1. Enhanced control over Type I error and efficiency in estimating parameters.
2. The correlation between the responses can be quantified and controlled.

The most used approach to jointly model two or more response variables include: (1) the joint generalized linear mixed model, where the generalized linear mixed model for each response is

joint through the description of a joint multivariate distribution for the random effects (Zewotir, et al., 2016). (2) The structural equation method; and (3) the copula regression, where a copula function is used to separate the marginal distributions from the dependence structure of a given multivariate distribution. In this chapter, we will employ the joint generalized linear mixed model (JGLMM) to model both malnutrition and anaemia.

4.2 Model overview

Let Y_{ij} be a dichotomous random variable and take the values of 1 or 0 which denotes the i^{th} outcome ($i = 1, 2$) of the j^{th} household ($j = 1, 2, \dots, 16708$) subject with $i = 1$ for nutrition status and $i = 2$ for anaemia status. The two dichotomous outcomes formulate to a bivariate response vector $Y = \binom{Y_{1j}}{Y_{2j}}$. If the two outcomes (nutrition status and anaemia status) are independent, then the generalized linear mixed model is given as

$$g_1(\mu_{1j}) = \beta_1 X_{1J} + Z_{1J}\vartheta_{1j} + \epsilon_{1j}$$

$$g_2(\mu_{2j}) = \beta_2 X_{2J} + Z_{2J}\vartheta_{2j} + \epsilon_{2j} \qquad \textbf{(4-1)}$$

where $\epsilon_{1j} \sim N(0, \sigma_1^2)$, $\epsilon_{2j} \sim N(0, \sigma_2^2)$, $\vartheta_{1j} \sim N(0, G_1)$, $\vartheta_{2j} \sim N(0, G_2)$, β_1, β_2 are the vectors of fixed effects, ϑ_1, ϑ_2 are the vectors of the random effects, X_{1J}, X_{2J}, Z_{1J} and Z_{2J} are the design matrices for the fixed effects and random effects, respectively (Gebregziabher, et al., 2013). To consider the relationship between nutrition status and anaemia status, one could use the following bivariate linear mixed model

$$Y_j = \beta X_j + Z_j\vartheta_j + \epsilon_j \qquad \textbf{(4-2)}$$

with $\epsilon_j \sim N(0, \text{'H})$, $\vartheta_j \sim N(0, G_i)$ where Y_j is the response variable, $X_j = \begin{pmatrix} X_{1J} & 0 \\ 0 & X_{2J} \end{pmatrix}$, $Z_j = \begin{pmatrix} Z_{1J} & 0 \\ 0 & Z_{2J} \end{pmatrix}$, $\beta = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ and $\vartheta = \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \end{pmatrix}$.

The covariance matrix of measurement of errors is defined by $\text{'H} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$.

The covariance matrix of the random effects is the matrix $G = \begin{pmatrix} G_1 & G_{12} \\ G_{21} & G_2 \end{pmatrix}$.

For the purpose of this chapter, matrix G was used as the covariance matrices of the joint model where G_1 and G_2 are the variance components of nutrition status and anaemia status, respectively.

G_{12} and G_{21} are the correlation components between nutrition status and anaemia status. Based on the assumption of conditional independence, likelihood function of the joint responses can be determined (Zewotir, et al., 2016). In this chapter, a maximum likelihood approach was employed to obtain the estimates of the parameters. As the integrals do not have closed form, we used an approximation based on Gaussian quadrature as it is consistent and asymptotically normally distributed.

4.3 Gaussian quadrature method

The Gaussian quadrature method is an approximate method of solves the integral $I = \int_a^b y(x)dx$. Suppose that $x = \frac{(b-a)t}{2} + \frac{(a+b)t}{2}$ and $f(t) = \frac{(b-a)y(x)}{2}$, then the integral reduced to the form $\int_{-1}^{1} f(t)dt$.

The Gaussian quadrature formula is thus:

$$\int_{-1}^{1} f(t)dt \ = \ \sum_{i=1}^{n} A_i f(t_i). \qquad \textbf{(4-3)}$$

The cusps t_i of the Gaussian quadrature formula are the roots of a Legendre polynomial of degree n, $P_n(t)$. The Legendre polynomial has real and various roots in the interval (-1,1). The weights A_i of the Gaussian quadrature formula are defined by

$$A_i = \frac{2}{\left(1 - t_i^2\right)[P_n'(t_i)]^2}$$

Gaussian quadrature formula is exact for an arbitrary polynomial of degree not higher than $2n - 1$. The remainder term of Gauss's formula Rn for the integral $\int_a^b y(x)dx$ is expressed as follows

$$R_n = \frac{(b-a)^{2n-1}n!}{(2n+1)(2n)!^3} y^{(2n)}(\xi), a \le \xi \le b.$$

The Gaussian quadrature method is applied when a sub-integral function is smooth enough and a gain in the number of cusps is essential (for instance, in calculating multiple integrals as iterated integrals).

4.4 Covariance test

Let A_K denote the list of active variables and S_{A_K} denote the sign list. The covariance test provides a significant test on the underlying linear model (Tibshirani, et al., 2017),

$$\partial = X\beta^*$$

where β^* is the set of parameter estimates and X is the set of covariates. The null hypothesis test is given by

$$H_0 : A_{K-1} \supseteq supp\ (\beta^*)$$

where $supp\ (\beta^*)$ denotes the support of the set of β^*.

The covariance statistic at step *k* can be written as;

$$C_k = [\rho_k^2 \text{ x } \lambda_k(\lambda_k - \lambda_{(k+1)})]/\sigma^2$$

Where λ_K and λ_{K+1} are the knots at step *k* and *k+1* of the path respectively. ρ_K is the weight.

Based on the correlation restrictions on the predictors *X* and other conditions, the covariance statistics has a conservative Exp(1) limiting distribution under the null hypothesis,

$$\lim_{n,p\to\infty} P_{A_{(k-1)} \supseteq supp(\beta^*)}\ (C_k > t|\hat{A}_k(y) = A_k, \hat{S}_{A_k} = S_{A_k})$$

For all $t \geq 0$.

The p-value of the covariance statistics under Exp(1) satisfies;

$$exp\left(\frac{\rho_K^2}{\sigma^2}\ \lambda_k\left(\lambda_k - \lambda_{(k+1)}\right)\right)$$

$$= \left(\frac{\emptyset\left(\lambda_{(k-1)}\frac{\rho_k}{\sigma}\right) - \emptyset\left(\lambda_k\frac{\rho_k}{\sigma}\right)}{\emptyset\left(\lambda_{(k-1)}\frac{\rho_k}{\sigma}\right) - \emptyset\left(\lambda_{(k+1)}\frac{\rho_k}{\sigma}\right)}\right) x\ \left(1 + \gamma_p(1)\right)$$

Where $\gamma_p(1)$ denotes the converging terms towards zero probability and $\emptyset(*)$ denotes the standard normal cumulative distribution function (CDF).

4.5 Covariance structures

From the traditional linear model, the expected value of $\boldsymbol{y}$ is modelled through the fixed effects $\boldsymbol{\beta}$. The extension provided by the mixed model is that the variance of $\boldsymbol{y}$ can be modelled through $\boldsymbol{Z}$, $\boldsymbol{G}$, and $\boldsymbol{R}$. The classical mixed model has $\boldsymbol{R} = \sigma^2\boldsymbol{I}$, where **I** is the nxn identity matrix and $\boldsymbol{G}$ is a diagonal matrix containing variance components. This model is very important when dealing with randomized block and split-plot designs (Wolfinger, 1993). Below in Table 4-1 are the types of covariance structures,

Table 4-1: Covariance structure notation and examples

Structure	Notation	Example
Diagonal	1	$\begin{bmatrix} \sigma^2 & & \\ & \sigma^2 & \\ & & \sigma^2 \end{bmatrix}$
Variance Components	VC (A B)	$\begin{bmatrix} \sigma_A^2 & & \\ & \sigma_A^2 & \\ & & \sigma_B^2 \end{bmatrix}$
Compound Symmetry	CS	$\begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 \\ & \sigma^2+\sigma_1 & \sigma_1 \\ & & \sigma^2+\sigma_1 \end{bmatrix}$
Unstructured	UN	$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{13} \\ & & \sigma_{33} \end{bmatrix}$
First-order Autoregressive	AR (1)	$\sigma^2\begin{bmatrix} 1 & \rho & \rho^2 \\ & 1 & \rho^2 \\ & & 1 \end{bmatrix}$
Banded	UN (2)	$\begin{bmatrix} \sigma_1 & \sigma_4 & \\ & \sigma_2 & \sigma_5 \\ & & \sigma_3 \end{bmatrix}$
Toeplitz	TOEP	$\begin{bmatrix} \sigma_1 & \sigma_2 & \sigma_3 \\ & \sigma_1 & \sigma_2 \\ & & \sigma_1 \end{bmatrix}$
Banded Toeplitz	TOEP (2)	$\begin{bmatrix} \sigma_1 & \sigma_2 & \\ & \sigma_1 & \sigma_2 \\ & & \sigma_1 \end{bmatrix}$

4.6 Results of the JGLMM on malnutrition and anaemia

The analysis of the data was performed using SAS 9.4 PROC GLIMMIX procedure. Concerning the diagnostics of the fitted model, we examined the normality of the random effect using the Shapiro-Wilks test at a 5% level of significance. The Shapiro-Wilks test yields a p-value of 0.591; this indicated a non-significant deviation from normality. The joint generalized linear mixed model of the nutritional status and anaemia status of under five-year children:

$$\begin{pmatrix} Nutrition \\ Anemia \end{pmatrix} = \begin{pmatrix} b_{10} + b_{11}residence + b_{12}age + \cdots + b_{16}wealth \\ b_{20} + b_{21}residence + b_{22}age + \cdots + b_{26}wealth \end{pmatrix} + \begin{pmatrix} \gamma_1 country \\ \gamma_2 country \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

The authors also considered different covariance structures (autoregressive AR (1), variance component (VC), compound symmetry (CS), heterogeneous compound symmetry (CSH) and unstructured (UN)), but only unstructured covariance variance was found to be suitable for the analysis based on convergence criteria. The analysis to check for possible interaction effect was also conducted but none was found to be significant. To confirm the need for the random intercept in the model, the COVTEST statement in SAS was used. The COVTEST provides a mechanism to acquire statistical interpretations for the covariance parameters. The ration of residual likelihood produces the significant test; the confidence limits are computed as likelihood ratio. The results of this test are shown in Table 4-2. The results indicated that the covariance parameter was highly significant and thereby confirmed the necessity of including the random count effect in the model. Table 4-3 findings revealed that there is a significant positive correlation between nutritional status and anaemia status (est. = 0.471, p-value = 0.024). Positive correlation indicates that nutrition and anaemia change in the same direction.

Table 4-2: Test covariance parameters

Label	DF	-2 log likelihood	χ^2	P-Value
No G-side effects	1	7768	33.86	<0.001

Table 4-3: Variance components

Label	Estimate	Standard error	P-Value
Malnutrition status	0.546	0.135	<0.001
Anaemia status	1.832	0.249	0.045
Correlation	0.471	0.115	0.024

Table 4-4: Parameter estimates and odds ratios

Covariates	Malnutrition			Anaemic		
	Estimates	OR	P- Value	Estimate	OR	P-Value
Intercept	3.500	-	0.000	2.378	-	<0.001

Resident						
Ref: Urban						
Rural	0.387	1.473	0.000	0.561	1.752	<0.001
Child's age						
Ref: < 12						
12-23	0.012	1.012	0.945	0.994	2.702	<0.001
24-35	0.017	1.017	0.914	0.884	2.421	<0.001
36-47	-0.045	0.956	0.766	0.948	2.581	<0.001
48-59	-0.068	0.934	0.654	0.781	2.184	<0.001
Sex of child						
Ref: Male						
Female	0.093	1.097	0.193	0.266	1.304	<0.001
Mother's level of schooling						
Ref: Higher						
Primary	1.097	2.995	0.047	1.888	6.606	<0.001
Secondary	0.217	1.242	0.032	0.442	1.556	0.002
Birth interval						
Ref: >47 months						
<24	-0.403	0.668	0.001	0.208	1.231	0.278
24-47	-0.330	0.719	<0.001	0.398	1.489	0.001
Wealth index						
Ref: Rich						
Poor	0.880	2.411	<0.001	0.258	1.294	0.084
Middle	0.204	1.226	<0.001	-0.089	0.915	0.533

Table 4-4 displays the parameter estimates and the odds ratios of malnutrition and anaemia. Factors such as birth interval, mother's level of education, wealth index and type of residence were found to be statistically significant in influencing nutritional status. Children under five years residing in the rural setting of Angola, Senegal and Malawi are 1.473 times more likely to be malnourished when compared to children under five years residing in the urban setting of Angola, Senegal and Malawi. It can be observed from the results that as the age of the child increases, the odds of being malnourished reduces. This finding indicates that children under five

years from Angola, Senegal and Malawi need to be protected from malnourished exposure at an early stage of their lives. Child gender was found to be significantly associated with the nutrition status of the child. A female under five years in Angola, Senegal and Malawi is more likely to be malnourished (1.097 times) when compared to a male under five. In addition, the results showed that improving the mother's level of education reduces the chance of children under five being malnourished. Children under five residing with a mother who has attained a primary level of education are 2.995 times more likely to be malnourished when compared to children under five residing with a mother who has attained a tertiary level of education. Household wealth status was found to significantly impact the nutrition status of children under five. The odds of being malnourished reduce as the household wealth improves. Children under five in Angola, Senegal and Malawi residing in a poor household are 2.411 times more likely to be malnourished when compared to children under five residing in a rich household. In the same table, factors such as child's age, child's gender, type of residence and mother's level of education are statistically significant in influencing anaemia status. Children under five years residing in the rural setting of Angola, Senegal and Malawi are 1.572 times more likely to be anaemic when compared to children under five residing in the urban setting of Angola, Senegal and Malawi. In addition, the child's gender was found to be significantly linked to anaemia status. A female child from Angola, Senegal and Malawi was found to be 1.304 times more likely to be anaemic when compared to a male child counterpart. Whereas, children under five residing with a mother who has attained a primary level of education are 6.606 times more likely to be affected by anaemia when compared to children under five residing with a mother who has attained a tertiary level of education. In contrast, children under five in Angola, Senegal and Malawi residing in a poor household are 1.294 times more likely to be anaemic when compared to children under five years residing in a rich household.

4.7 Summary

In this chapter, a joint model under a generalized linear mixed model was employed to obtain the determinants of malnutrition and anaemia among children under five in Angola. The results of the joint generalized linear mixed model were obtained from SAS 9.4 software. Covariance components results revealed that there is a significant positive correlation between malnutrition status and anaemia status. The results further revealed that type of residence, mother’s level of schooling, birth interval and wealth index are statistically significant in influencing malnutrition status. Whereas, type of residence, the child's age, the child's gender and mother's level of schooling is statistically significant in influencing anaemia status. A child from a rural setting is found to be more likely to be malnourished and anaemic when compared to a child from an urban setting. This finding indicates that there is a gap in the health resources available to rural and

urban areas. The department of health and the policymakers of healthcare in Angola, Senegal and Malawi need to supply mobile health services to the rural areas.

The fore-mentioned intervention is one of many interventions to be implemented in an attempt to close the gap. In contrast, a child residing with a mother who has attained a primary level of schooling is more likely found to be malnourished and anaemic when compared to a child residing with a mother who has a higher level of education. Similar results were found in the study by ((Takele, et al., 2019), (Sop, et al., 2015), (World Health Organization and World Bank Group, 2015)). The education system and policymakers in Angola, Senegal and Malawi need to propose strategic interventions for the education system. Among many interventions considered through other literature, there needs to be the introduction of an educational program that will cover all age groups, since most of the households in Angola, Senegal and Malawi are occupied by the older generation that did not have an opportunity to attain schooling (Poda, et al., 2017). These strategic interventions need to reach the rural settings of Angola, Senegal and Malawi.

Another factor that is found to be significantly associated with malnourishment and anaemia is household wealth status. Many studies conducted in developing countries have found this factor to be a common factor among many factors that contribute significantly to under five children's mortality rate ((Fang, et al., 2018); (World Health Organization and World Bank Group, 2015), (Abuka & Tsegaw, 2017)). A child from a poor household of Angola, Senegal and Malawi is more likely to be malnourished and anaemic when compared with a child from a rich household. Studies conducted in developing countries further revealed that improving household wealth status and mother's level of education reduces the child's risk of being exposed to malnourishment and anaemia (Takele, et al., 2019), (Kuziga, et al., 2017), (Sop, et al., 2015), (Gebreegziabiher, et al., 2014), (Kejo, et al., 2016). This, in general, indicates that the government and policymakers of developing countries need to put more effort and resources into improving all the programs that are targeting poor and not educated communes.

A joint modelling framework is more appealing than a univariate response analysis in terms of ease of interpretation, improved underlying dependence structure and the ability to assess overlapping risk among two or more disease outcomes. In this chapter we discussed the joint generalised linear mixed model where malnutrition and anaemia were jointly modelled. The subsequent chapter explores another statistical method that allows for joint modelling of two or more diseases known as structural equation model (SEM). This statistical method will be explored to assess the risk factors that might directly or indirectly influence the co-morbidity of childhood malnutrition and anaemia in Angola, Malawi and Senegal. The SEM can assess the correlation

between multiple variables and relate unobservable and observable variables. These variables will be socio-economic and child inherent factors.

Chapter 5 : Structural equation model for malnutrition and anaemia assessment

5.1 Introduction

Structural Equation Modelling is very generic, very influential multivariate analysis method that incorporate specialized versions of several other analysis methods as special cases. The Structural Equation Model (SEM) is a multivariate model that relates traits and unmeasured concepts known as latent variables. To simply define a latent variable, (Bollen, 1989) used an example and said they are representations of the concepts in a measurement model of socio-economic status. The SEM can assess the complexity of the interrelationships between multiple variables and relate unobservable and observable variables with estimation in terms of the sample covariance matrix of the observed variables and the population covariance matrix generated by the SEM framework. The applications of structural equation modelling include:

1. Causal modelling- assumes causal associations among variables and tests the causal models with a linear equation system.
2. Confirmatory factor analysis- an extension of factor analysis that test the intercorrelations and structure of the factor loadings.
3. Second-order factor analysis- factor analysis with the correlation matrix of the common factors analysed to provide second-order factors.
4. Regression models- an extension of linear regression analysis where the regression weights may be constrained to be equal.
5. Covariance structure models- assumes that a covariance matrix has a particular form.
6. Correlation structure models- assumes that a correlation matrix has a particular form.

In this chapter we are discussing causal modelling which assumes causal associations among variables and tests the causal models with a linear equation system. The method is used to examine the complex association between socio-economic and child inherent factors, and their direct or indirect association with childhood malnutrition and anaemia co-morbidity. To achieve this, four SEM analysis steps were satisfied:

1. The development of a conceptual and theoretical model
2. Model estimation
3. Determining the model's identifiability
4. Determining the model fit.

5.2 Conceptual and theoretical model

A general structural equation model (SEM) entails of measurements and structural sub-models. SEM incorporates multivariate data analysis methods that combine features from multiple regression and factor analysis to concurrently estimate a series of relatives of dependency that allow defining procedures aimed at directly incorporating measurement errors into the model (Amorim, et al., 2010). The structural sub-model is given by:

$$\theta = \beta\eta + \gamma\varphi + \epsilon \quad \textbf{(5-1)}$$

where η represents a vector mx1 of latent endogenous variables; φ represents a vector kx1 of latent exogenous variables; β is a matrix zxz of coefficients relating the latent endogenous variables to each other; γ is a matrix zxk of coefficient linking the endogenous variables to the exogenous variables; and ϵ is a vector zx1 of structural disturbances. The fundamental step in SEM is the confirmation of the model latent variables.

The main diagonal elements of β are zeros and φ with ϵ are mutually independent and normally distributed. Through the measurement model, the observed variables and latent variables are related with the mathematically separation defined as:

$$Y = \Pi_y\eta + \alpha$$

and

$$X = \Pi_x\varphi + \psi$$

where Π_y (p x m) and Π_x(q x z) are the coefficient matrices showing the relation of observed variables to the latent endogenous and exogenous variables, α and φ are the (p x 1) and (p x 1) vectors of measurement errors in Y and X, respectively. It is anticipated that the measurement errors α and ψ have zero expectations, each with a multivariate normal distribution. The errors assume independence of each other, independent of latent endogenous variables (η), latent exogenous variables (φ) as well as independent of the disturbances (ϵ).

The structural errors (ϵ) have zero expectation with a multivariate normal distribution, and the disturbances are independent of the latent exogenous variables (φ). Likewise, the observed indicators Y and X have a multivariate normal distribution,

$$\binom{X}{Y} \sim N_{p+q}(0,\cup) \quad \textbf{(5-2)}$$

Where $\cup$ is the indicators' covariance matrix which is a function of the model's parameter $\Lambda =$

$(\beta, \gamma, \Pi_y, \Pi_x, \rho_\alpha, \rho_\psi, \sigma)$ and can be expressed as

$$\cup = \begin{pmatrix} \cup_{xx} & \cup_{xy} \\ \cup_{yx} & \cup_{yy} \end{pmatrix} = \begin{pmatrix} \Pi_y(1-\beta)^{-1}(\gamma\sigma\,\gamma')[(1-\beta)^{-1}]'\Pi_y' + \rho_\alpha & \Pi_y(1-\beta)^{-1}\,\gamma\sigma\Pi_x' \\ \Pi_y\sigma\,\gamma[(1-\beta)^{-1}]'\Pi_y' & \Pi_x\sigma\Pi_x' + \rho_\psi \end{pmatrix}$$

where σ is the (z x z) covariance matrix of the latent exogenous variables (φ), ρ_α and ρ_ψ denote the covariance matrices of the measurement errors α and ψ.

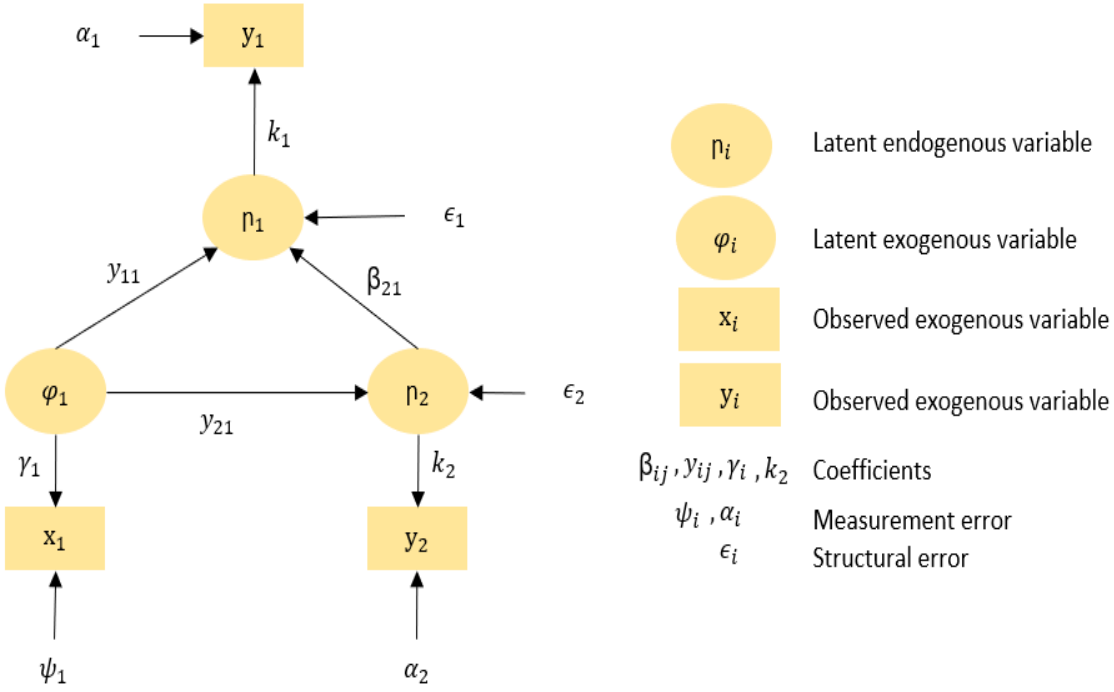

Figure 5-1: Graphical expression of a structural model

SEM includes complex models and it is very suitable to display such models in diagram system. The path diagram allows the visualization of the relation of interdependency in the theoretical model. This diagram has a set of geometric figures and arrows showing the types of variables and the relationship between them. Refer to Figure 5-1 for a graphical example of a structural equation.

The theoretical framework of SEM indicates the importance of latent variables which are inferred through observable variables since they are not directly measured, to ascertain their contribution to childhood co-morbidity of malnutrition and anaemia. Based on Figure 5-2, the proposed hypotheses for testing are as follows;

H1: Socio-economic status is associated with Child's Inherent

H2: Socio-economic status is associated with Co-morbidity

H3: Child Inherent is associated with Co-morbidity

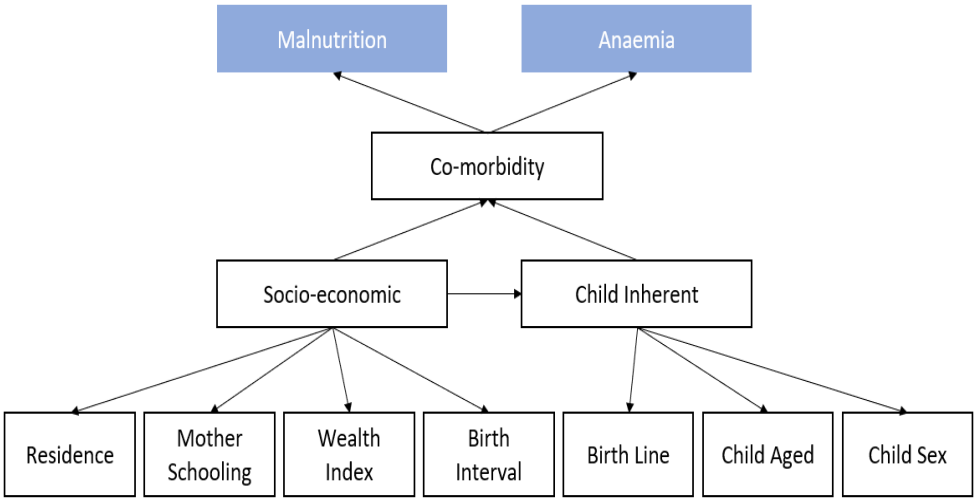

Figure 5-2: Structural model theory

The following assumptions are made on the theoretical model, (1) the socio-economic status is directly influential on the co-morbidity of malnutrition and anaemia, (2) the child inherent is directly influential on the co-morbidity of malnutrition and anaemia, (3) the socio-economic variable including type of residence, mother level of schooling, wealth index and birth interval are indirectly influential on the co-morbidity of malnutrition and anaemia and (4) the child inherent variables including birth line, child age and child sex are indirectly influential on the co-morbidity of malnutrition and anaemia. The three latent variables, namely socio-economic, child inherent and co-morbidity were examined with all the observed factors to identify the causal relationship. The goodness of fit test was used to determine whether the models should be accepted or not.

Figure 5-2 can be translated to the equation(s). Consider the equation for the latent socio-economic variable for wealth index,

$$Wealth\ Index\ =\ \beta \text{ x } socio\ economic + error\ term$$

where β is the parameter estimate for the relationship of wealth index and socio-economic latent variable. Similarly, the equation for the child inherent latent variable for child sex is given as

$$Child\ sex\ =\ \beta \text{ x } Child\ Inherent + error\ term$$

where β is the parameter estimate for the relationship of wealth index and socio-economic latent variable. All other equations can be extracted using the path diagram.

5.3 Model estimation

To obtain the matrix ∪ associated with the confirmatory factor analysis, it must be assumed that

$(\beta = 0, \gamma = 0, \Pi_y = 0, \sigma = 0, \rho_\alpha = 0)$. For enough restriction in some of the components of the ∪ matrix, the maximum likelihood (MLE) may be obtained for the parameters of the model and the log-likelihood analogous to the model can be expressed as a function of the model parameters. In structural equation modelling, this procedure aims to estimate the parameter $\sum$ so that the discrepancy function $G(S;\cup)$ can be minimalized. The $G(S;\cup)$ is the discrepancy function which is a scalar measure of the distance between the sampling covariance matrix and the covariance matrix adjusted.

The MLE is the widely used method for parameter estimation in the SEM. The method of MLE define parameters in a way that the two matrices (covariance matrix and sampling covariance matrix) are as close as possible, where the likelihood logarithm measures the proximity between the two matrices.

5.4 Model modification

An essential step in model estimation is the confirmation model's identifiability of latent variables, which is a complex challenge in SEM without a simple solution (Steiger, 1990). A model is described as not being identifiable when no solution is found for the equation system. A necessary condition for identifiability is usually the number of permitted parameters in the model, which must not exceed the number of variances and covariances between the observable variables. This is mathematically defined as the counting rule:

$$\frac{(P+Q)\mathrm{x}(P+Q+1)}{2} \qquad \textbf{(5-3)}$$

where P, is the number of the endogenous variables and Q stands for the number of exogeneous variables in the model (Kline, 2015).

Below are the conditions requiring satisfaction for a model to be identifiable:

1. Unrelated measurement errors.
2. A minimum of two exclusive indicators occur for each of the latent variables.
3. There is a chance of a single indicator for latent variables equally assumed without any error.
4. Structural model includes only observed variables.

Statistical method detects any attempt to adjust the under-identifiable model and a unique information matrix is obtained. Another statistical approach to identify under- or non-identifiable

models is through the observation of variance estimates; when the variance estimates are very large, the model is under-identifiable and requires readjustment. A lower number of observations is required to adjust a structural equation model.

5.5 Model diagnostics

The model diagnostic of the SEM is complex to determine since many goodness-of-fit criteria have been developed to evaluate the model under different assumptions (Schumacker & Lomax, 2004). In this chapter, the model fit indicators such as chi-square χ^2, Relative Fit Index (RFI) and Root Mean Square Errors of Approximation (RMSEA) were considered (Baldwin, 1989).

The χ^2 statistic offers a test of the null hypothesis to ensure the theoretical model is the best fit for the data. A good fit is identified based on non-significant p-values. A root means square error of approximation (RMSEA) values close to 0.06, comparative fit index (CFI) values close to 0.9, and adjusted goodness-of-fit index (AGFI) indicate acceptable fit of the model (Steiger, 1990).

The commonly used model-fit criteria in SEM modelling are as follows:

1. Goodness of fit index $(GFI) = 1 - \frac{\chi^2_{model}}{\chi^2_{null}}$, where 1 indicates perfect fit.
2. Root mean square error of $(RMSEA) = \sqrt{\frac{\chi^2_{model} - df_{model}}{(N-1)df_{model}}}$, where a value of 0.05 or less indicate the model's good fit.
3. Comparative fit index $(CFI) = 1 - \frac{\chi^2_{model} - df_{model}}{\chi^2_{null} - df_{null}}$, where a value of 0.9 or greater are expected.
4. Parsimony comparative fit index $(PCFI) = CFI\ x\ \left(\frac{2df_{model}}{k(k-1)}\right)$
5. Parsimony goodness of fit index $(PGFI) = GFI\ x\ \left(\frac{2df_{model}}{k(k+1)}\right)$

Table 5-1 display all the SEM goodness of fit techniques and their general rule for acceptance.

Table 5-1: Fit indices of structural equation model

Goodness of fit measure	Categorical data Value	General rule for acceptable fit
Chi-square value ($\chi 2$)	918.42	
Degree of freedom (df)	282	
Chi-square ($\chi 2$)/df	3.26	Ratio of $\chi 2$ to df $\leq$ 2 or 3
Root mean square error of approximation (RMSEA)	0.05	< .08
Root Mean Square Residual (RMR)	0.03	$\leq$ .08
Tucker–Lewis index (TLI)	0.94	$\geq$ .90
Normed Fit Index (NFI)	0.92	$\geq$ .90
Incremental fit index (IFI)	0.95	$\geq$ .90
Comparative fit index (CFI)	0.95	$\geq$ .90
Goodness of fit index (GFI)	0.91	$\geq$ .90
Adjusted goodness of fit index (AGFI)	0.89	$\geq$ .90

5.6 Results of the SEM on malnutrition and anaemia for pooled data

The response variables are as per the recommendation of the WHO, weight-for-age Z-score (WAZ). Nutritional status of children was divided into two ordinal categories: Malnutrition (≤-2.0 WAZ) and Nourished (>-2.0 WAZ) (Khulu & Ramroop, 2020); (Amorim, et al., 2010). In contrast, a child is categorized as anaemic when the haemoglobin level of a child is less than 9.9 g per dL and a child is classified as not anaemic when the haemoglobin of the child is greater than 9.9 g per dL (Ramroop & Khulu, 2020).

For the explanatory variables' selection, the socio-economic and demographic variables related to children's malnutrition and anaemia were considered. These variables were identified as covariates in the development of the structural model. The selection also follows the WHO's Stunting Framework, which develops from the traditional UNICEF framework that focuses on the identifiers of malnutrition and anaemia ((Khulu & Ramroop, 2020), (Neuman, et al., 2013)). The variables selected for this study were the mother's level of education (primary, secondary or higher), type of resident (rural or urban), child's age in months (<12, 12-23, 24-35, 36-47 or 48-59), sex of child (male or female), wealth index (poor, middle or not poor), the birth interval in months (<24, 24-47 or >47) and birth order (2-3, 4-5 or >5).

The results of the distribution of childhood covariates' prevalence were obtained from SPSS and the SEM results were obtained from R 3.6.3 software. The descriptive analysis results for the pooled sample of children under five are summarized in Table 5-2. The majority of the under five children reside in rural areas and age between 36 to 59 months. The level of education for mothers

was observed to be mostly skewed towards primary education, whereas, 34.7% of the total households were found to be poor.

Table 5-2: Children under five years in Angola, Malawi and Senegal covariates prevalence for pooled data

Covariates	Frequency	Proportion	SE
Type of residence			
Rural	31316	60.9	0.002
Urban	15592	30.3	
Sex of child			
Male	6389	50.1	0.004
Female	6351	49.9	
Child's age (months)			
< 12	1418	11.1	0.012
12-23	1824	14.3	
24-35	2584	20.3	
36-47	3437	27.0	
48-59	3477	27.3	
Mother's level of schooling			
Primary	9119	77.6	0.004
Secondary	2445	20.8	
Higher	191	1.6	
Birth interval (months)			
< 24	908	10.4	0.007
24-47	4182	47.8	
> 47	3663	41.8	
Wealth index			
Poor	16254	34.7	0.004
Middle	9675	20.6	
Not poor	20978	44.7	
Birth order			
2-3	3724	39.5	0.009
4-5	2398	25.4	
>5	3314	35.1	

The model fit indicators result in Table 5-3 suggest that the data are well fitted to the SEM (GFI= 0.897, CFI =0.904, RMSEA = 0.008). The PCFI and PGFI values for the full model were observed to be 0.64 and 0.59 respectively.

Table 5-3: The goodness-of-fit indices in the two models (CFA and SEM)

Model	GFI	CFI	RMSEA	PCFI	PGFI
Conceptual CFA model	0.768	0.504	0.036	0.34	0.44

Full structural model	0.897	0.904	0.008	0.64	0.59

Figure 5-3 displays the results of the full SEM, whereas Table 5-4 and Table 5-5 summarizes the direct and indirect interrelationships between contextual factors and their influence on malnutrition and anemia. This study used two phases to evaluate partial mediation effects. Phase one was the direct effect model test assessing the effect of the predictors (Socio-economic status (SES) and Child Inherent factors) on childhood malnutrition and anemia. The results showed that SES is associated with child inherent ($\beta = 0.19, p < 0.001$). Furthermore, the direct path coefficient from the SES factors on malnutrition and anaemia co-morbidity was statistically significant ($\beta = -1.16, p < 0.001$); this indicated a negative direct effect of SES factors on childhood co-morbidity.

Phase two involved testing the indirect relationship between SES factors and co-morbidity. The estimated indirect path for the effect of SES factors on childhood malnutrition and anaemia as mediated by child-inherent factors was statistically significant and positive ($\beta = 0.02, p < 0.001$). In this model, child inherent factors such as age ($p < 0.001$) and sex ($p < 0.001$) were associated with child's malnutrition and anaemia prevalence. Whereas, SES factors such as type of residence ($p < 0.001$), mother's level of education ($p < 0.001$) and household wealth index ($p < 0.001$) were associated with child's malnutrition and anaemia prevalence.

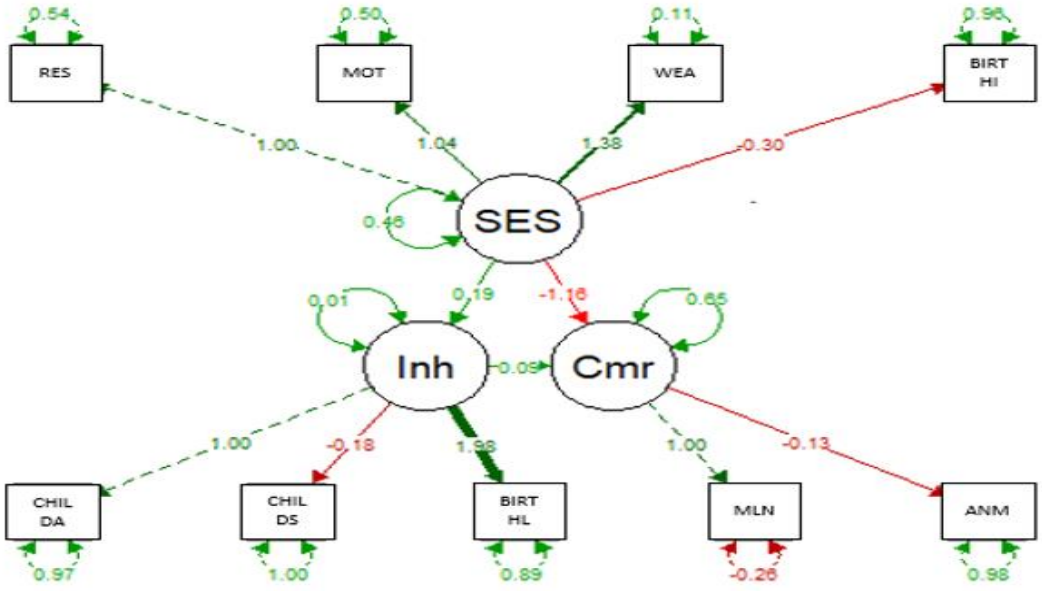

Figure 5-3: Full structural model results for pooled data

Table 5-4: Standardized direct and indirect effects of factors on childhood malnourishment and anaemia for pooled data

Factors	Total effect	Direct effect	Indirect effect
SES → Child Inherent	0.19*	0.19*	-
Inherent → Co-morbidity	0.09*	0.09*	-
SES → Co-morbidity	-1.14*	-1.16*	0.02*
P*<0.05			

*SES=Socio-economic status

Table 5-5: Standardized Regression Coefficients for the Full Structured Model for pooled data

Parameters	Estimate	Standard Error	P- value
Cmr ← SES	-1.164	0.131	<0.001
Cmr ← Inh	0.091	0.678	<0.001
Inh ← SES	0.185	0.021	<0.001
RES ← SES	0.981	0.312	<0.001
MOT ← SES	1.040	0.025	<0.001
WEA ← SES	1.384	0.031	<0.001
BIRTHI ← SES	-0.303	0.021	<0.001
CHILDA ← Inh	0.978	0.022	<0.001
CHILDS ← Inh	-0.181	0.129	<0.001
BIRTHL ← Inh	1.976	0.242	<0.001

*SES=Socio-economic status, Cmr=Co-morbidity, Inh=Child Inherent, RES=Type of resident, MOT=Mother's level of education, WEA=Wealth Index, BIRTHI=Birth Interval, CHILDA=Child Age, CHILDS=Child Sex, BIRTHL=Birth Line

5.7 Results of the SEM on malnutrition and anaemia for each country data

Section 5.6 provided the SEM results of the pool data. However, it is pivotal check whether there is any different in the parameter estimates for each country. The below table display the results of the parameter estimate for each country.

Table 5-6: Standardized Regression Coefficients for the Full Structured Model for each country data

Parameters	Angola		Malawi		Senegal	
	Estimate	P- value	Estimate	P- value	Estimate	P- value
Cmr ← SES	-1.054	<0.001	-0.034	<0.001	-0.025	<0.001
Cmr ← Inh	0.401	<0.001	0.851	<0.001	0.202	<0.001
Inh ← SES	0.155	<0.001	0.012	<0.001	0.197	<0.001
RES ← SES	0.081	<0.001	0.157	<0.001	0.077	<0.001

MOT ← SES	0.040	<0.001	0.011	<0.001	0.785	<0.001
WEA ← SES	0.384	<0.001	0.527	<0.001	0.257	<0.001
BIRTHI ← SES	-0.003	<0.001	-0.101	<0.001	-0.241	<0.001
CHILDA ← Inh	0.678	<0.001	0.508	<0.001	0.388	<0.001
CHILDS ← Inh	-0.221	<0.001	-0.041	<0.001	-0.351	<0.001
BIRTHL ← Inh	1.564	<0.001	1.002	<0.001	1.452	<0.001

*SES=Socio-economic status, Cmr=Co-morbidity, Inh=Child Inherent, RES=Type of resident, MOT=Mother's level of education, WEA=Wealth Index, BIRTHI=Birth Interval, CHILDA=Child Age, CHILDS=Child Sex, BIRTHL=Birth Line

Table 5-6 display the parameter estimates for standardized regression coefficients for the full structured Model for each country data (i.e. Angola, Malawi and Senegal). Looking at the results, there is no difference in the parameter estimates between countries. Factors such as Mother level of education, wealth index, child's age, birth interval, child's sex and child's birth line were observed to be associated with child's malnutrition and anaemia prevalence.

5.8 Summary

In this chapter, the structural equation model (SEM) was explored to examine the interrelationships between contextual factors (demographic and socio-economic) and the co-morbidity of childhood malnutrition and anaemia among children under five years in Angola, Senegal and Malawi. The 2016 Demographic Health Survey (DHS) data from Angola, Senegal and Malawi was combined to create a pooled sample. The study assumed that the is no difference on the effect of the countries. This method was used in other literature ((Khulu & Ramroop, 2020), (Ramroop & Khulu, 2020), (Takele, et al., 2019), (Subramanian, et al., 2011), (Neuman, et al., 2013)).

The results of the SEM were obtained from R 3.6.3 software. The results of the covariance components revealed that there is a significant positive correlation between nutritional status and anaemia status. The model fitted our data well and has demonstrated its potential for examining the interrelationships between variables in a dataset.

The study findings revealed that socio-economic factors (child place of residence, mother's level of education and household wealth index) have a direct effect on childhood co-morbidity prevalence. These findings are similar to the results of the studies conducted in Burkina Faso, Bangladesh and Ethiopia ((Abuka & Tsegaw, 2017), (Poda, et al., 2017), (Feddag & Mesbah, 2006), (Talukder, 2017)). The estimated indirect path for the effect of socio-economic factors on childhood malnutrition and anaemia as mediated by child inherent factors, was statistically significant and positive. This result indicates that child-inherent factors are important risk factors

for childhood co-morbidity prevalence. Child inherent factors such as age of the child and sex of the child were statistically significant predictors of childhood co-morbidity of malnutrition and anaemia. Similar results were observed in the study conducted in Ethiopia (Wirth, et al., 2016).

The limitation of the SEM includes modeling two diseases without accounting for dependence structure spatial variability. The SEM could also not reveal the trends of association between the predictors and response variables, whether linearly or nonlinearly related. Hence the subsequent chapter looks at the statistical method that include the accountability of dependence structure and spatial variation of co-morbidity in childhood malnutrition and anaemia. This type of statistical method is called copula-based generalized joint regression model.

Chapter 6 : Copula-based modelling of malnutrition and anaemia risk factors

6.1 Introduction

The literature on copula theory was firstly discussed back in 1959. In the past three decades, the concept of the copula model attracted attention, mostly in the field of science. In simple English, 'copula' is any term that link subjects with the predicate. In statistical terms, the copula model represents multivariate joint distribution function that links to one-dimension, uniform margins over unit range. The significance of the copula model in research and analysis, includes learning scale-free measures of statistical dependency and building classes of joint distributions (Burney & Ajaz, 2020). The geo-additive models geographically convert the reference outcomes to maps by directing the confounding effects of other covariates, such as demographic and socio-economic factors which are known to be associated with malnutrition and anaemia (Wand, et al., 2011). Extending the copula model by introducing the geo-additive model will support in implementing more effective interventions for childhood malnutrition and anaemia. Moreover, based on the results and mapping, this will support in ensuring that the limited resources are distributed to the needy areas.

In this chapter we will discuss the copula geo-additive model to jointly model and map the risk factors of malnutrition and anaemia. The advantage of the copula geo-additive model over the joint generalized linear mixed model and the structural equation model, is that it spreads to disease mapping through spatial modelling. This then can assist in our understanding of how the association between the outcomes varies, according to the geographical location.

6.2 Copula-based model overview

Copula model is commonly used when modelling the dependence especial in the field of financial risk management and price determination. This model was originally developed with the intention of illustrating the dependence between random variables independently of their marginal distribution. This type of modelling binds the marginals distribution to form an appropriate joint distribution function. Spatial statistics and epidemiology are now exploring the use of this method.

6.3 Bivariate copula functions

Let C represent a Z- dimensional copula of joint distributed random variables $K_1, K_2, K_3, \dots, K_z$ with standard uniform marginal distribution $(0,1)$. The copula function C: $[0,1]^k \rightarrow [0,1]$ satisfies the property

$$C_\delta(k_1, k_2, k_3, \dots, k_z) = Pr(K_1 < k_1, K_2 < k_2, K_3 < k_3, \dots, K_z < k_z) \quad \textbf{(6-1)}$$

where δ denotes a parameter vector of the copula function, often referred to as the dependence parameter vector (Haugh, 2016). The copula C_δ permits the generalization of the joint multivariate distribution function. Given Z random variables $Y_1, Y_2, Y_3, \dots, Y_z)$, where each of the random variable has a univariate continuous marginal distribution function $F_z(y_z) = \Pr(Y_z < y_z)$, $z = 1, 2, 3, \dots, Z$ and F_z^{-1} denote the inverse univariate cumulative distribution function, by the use of integral transformation; we write that for each $z(z = 1, 2, 3, \dots, Z)$

$$F_z(y_z) = \Pr(Y_z < y_z) = \Pr(F_z^{-1}(K_z) < y_z) = \Pr\big(K_z < F_z(y_z)\big).$$

The joint Z-dimensional distribution function of the random variable with continuous marginal distribution function $F_z(y_z)$ can be computed as

$$F(y_1, y_2, y_3, \dots, y_z) = \Pr(Y_1 < y_1, Y_2 < y_2, Y_3 < y_3, \dots, Y_z < y_z)$$

$$= \Pr(K_1 < F_1(y_1), K_2 < F_2(y_2), K_3 < F_3(y_3), \dots, K_z < F_z(y_z))$$

$$= C_\delta\big(k_1 = F_1(y_1), k_2 = F_2(y_2), k_3 = F_3(y_3), \dots, k_z = F_z(y_z)\big).$$

Copula can be generated using different methods which include geometric comparison, algebra methods and inverse methods. The commonly used method is the inverse method. Suppose a multivariate distribution $F(y_1, y_2, y_3, \dots, y_z)$ has a continuous margin $F_z(y_z)$, then the inverse method tries to invert the relation of $F(y_1, y_2, y_3, \dots, y_z) = C_\delta\big(k_1 = F_1(y_1), k_2 = F_2(y_2), k_3 = F_3(y_3), \dots, k_z = F_z(y_z)\big)$ to generate a copula

$$C_\delta(k_1, k_2, k_3, \dots, k_z) = \Pr(K_1 < k_1, K_2 < k_2, K_3 < k_3, \dots, K_z < k_z)$$

$$= \Pr\left(Y_1 < F_1^{-1}(k_1), Y_2 < F_2^{-1}(k_2), Y_3 < F_3^{-1}(k_3), \dots, Y_z < F_z^{-1}(k_z)\right)$$

$$= F(y_1 < F_1^{-1}(k_1), y_2 < F_2^{-1}(k_2), y_3 < F_3^{-1}(k_3), \dots, y_z < F_z^{-1}(k_z)).$$

Let $F(y_1, y_2, y_3, \dots, y_z)$ be a Z variate joint cumulative distribution function with univariate marginal cdfs $F_1,\ F_2,\ F_3, \dots,\ F_k$, each marginal distribution takes any value in the rage [0,1]. The joint cumulative distribution is bound below and above by the Frechet-Hoeffding lower and upper bounds, $\forall_L$ and $\forall_U$. The definitions of $\forall_L$ and $\forall_U$ are given below

$$\forall_L(y_1, y_2, y_3, \dots, y_z) = max\left[\sum_{j=1}^{z} F_j - d + 1{,}0\right] = W$$

$$\forall_U(y_1, y_2, y_3, \dots, y_z) = min[\ F_1,\ F_2,\ F_3, \dots,\ F_z] = M$$

such that

$$max\left[\sum_{j=1}^{z} F_j - d + 1{,}0\right] \leq F(y_1, y_2, y_3, \dots, y_z) \leq min[\ F_1,\ F_2,\ F_3, \dots,\ F_z] \qquad \textbf{(6-2)}$$

Where the upper bound $\forall_U$ is a consistent cumulative function and the lower bound $\forall_L$ is a cumulative distribution function for d=2.

Consider any bivariate copula with the cumulative distribution $\mathrm{C}_\delta(\mathrm{k}_1, \mathrm{k}_2)$, the copula C_δ must satisfy Equation 6-2. This equation can be expressed in terms of Frechet-Hoeffding bounds for bivariate distribution $F(y_1, y_2)$ such that

$$M(k_1, k_2) = \min\ (k_1, k_2)$$

$$W(k_1, k_2) \leq\ \mathrm{C}_\delta(\mathrm{k}_1, \mathrm{k}_2) \leq M(k_1, k_2)$$

$$\max(F_1(y_1), + F_2(y_2) - 1, 0) \leq\ F(y_1, y_2) \leq \min\big(F_1(y_1), F_2(y_2)\big).$$

If $\mathrm{C}_\delta(\mathrm{k}_1, \mathrm{k}_2) = W(k_1, k_2)$, then the random variable Y_1 and Y_2 yield a decreasing functions. Whereas, $\mathrm{C}_\delta(\mathrm{k}_1, \mathrm{k}_2) = M(k_1, k_2)$ are the random variables that has an increasing function.

For a given bivariate copula, the distribution function $F(y_1, y_2)$ can computed from the two random variables $Y_1 = F_1(Y_1)$ and $Y_2 = F_2(Y_2)$ with F_1 and F_2 being the margins. For specific functional forms of margins, the random variables Y_1 and Y_2 exact bivariate dependence structure is $\mathrm{C}_\delta(\mathrm{k}_1, \mathrm{k}_2)$ with the dependence parameter.

6.4 Copula families

The current existing literature on copulas belong to a certain family of copula. Their strength and properties of dependence is controlled by one or more parameters. In this section will discuss the different types of copulas, specifically, five families of copulas. The first part of this section will

be discussing the Elliptical copulas and the second part will be discussing the Archimedean copulas.

6.4.1 Elliptical copulas

The copula for a specific Elliptical distribution is called Elliptical copula. All Elliptical distributions are radially symmetric, that is $\emptyset_U(X,Y) = \emptyset_L(X,Y)$. The approach used on Elliptical copulas is different from the approached used into Archimedean copulas since the Elliptical copulas do not have a closed form expression for calculating dependencies (Manner, 2007). This thesis will discuss the Gaussian and Student T copulas.

I. Gaussian copula

The bivariate Gaussian copula is defined as

$$C_{Gaussian} = (k_1, k_2) = \vartheta S_{12}(\vartheta^{-1}(k_1), \vartheta^{-1}(k_2))$$

Where ϑS_{12} is the bivariate cumulative function with correlation matrix R_{12} of the standard normal distribution function and ϑ^{-1} is the inverse cumulative distribution function of the standard normal function. The copula for Gaussian can be written as

$$C_{Gaussian}(k,l) = \int_{-\infty}^{\vartheta^{-1}(k)} \int_{-\infty}^{\vartheta^{-1}l} \frac{1}{2\pi(1 - R_{12}^2)^{1/2}} exp\left(-\frac{s^2 - 2R_{12}st + t^2}{2(1 - R_{12}^2)}\right) dsdt.$$

The probability distribution function generator of the bivariate gaussian distribution is given as;

$$g(k) = (2\pi)^{-1} exp\left(\frac{-1}{2}k\right).$$

II. Student- T copula

The bivariate student-T copula is defined as

$$C_T(k_1, k_2) = t_{\delta,m}\left(t_m^{-1}(k_1, k_2)\right) = \int_{-\infty}^{t_v^{-1}(k_2)} \frac{1}{2\pi\sqrt{1-\omega^2}} \left(1 + \frac{s^2 + s^2 - 2\omega st}{m(1-p^2)}\right)^{-\frac{m+2}{2}} dsdt$$

Where t_v^{-1} denotes the quantile function of a standard univariate t_v distribution, m is the degrees of freedom and ω is the correlation coefficient defined as;

$$\omega = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

6.4.2 Archimedean copulas

Archimedean copulas form a large family of copulas with a number of convenient properties and they allow for many dependence structures. Let φ denote the generator function of a copula with the following properties:

1. $\varphi(1) = 0$
2. $\varphi'(t) \leq 0, for\ all\ t\ \epsilon\ (0,1)$
3. $\varphi''(t) \geq 0, for\ all\ t\ \epsilon\ (0,1)$.

Now let $\varphi^{(-1)}$ denote the pseudo-inverse, which is equal to the normal inverse for $t\ \epsilon\ [0, \varphi(0)]$ and is equal to zero for $t \geq \varphi(0)$. Then the Archimedean copula is given by:

$$C(u,v) = \varphi^{(-1)}\big(\varphi(u) + \varphi(v)\big).$$

The Archimedean copulas are symmetric, that is $C(u,v) = C(v,u)$ and they are associative, that is, $C(C(u,v),w) = C\big(u, C(v,w)\big)$.

The most used Archimedean copulas are:

I. Clayton copula

$$\varphi(t) = \frac{t^{-\theta} - 1}{\theta}, \theta\ \epsilon\ [-1, \infty].$$

For $\theta \geq 0$, the Clayton copula is strict and has lower tail dependence.

II. Gumbel copula

$$\varphi(t) = (-\ln(t))^{\theta}, where\ \theta \geq 1.$$

The Gumbel copula has upper tail dependence.

III. Frank copula

$$\varphi(t) = -\ln\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right), where\ \theta \neq 1$$

Frank copulas display the properties of radial symmetry and do not have a tail dependence.

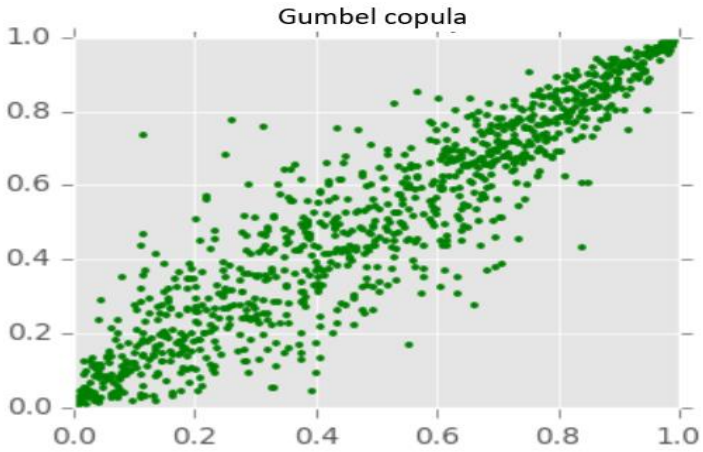

Figure 6-1: Scatter plot of Gumbel Archimedean copula

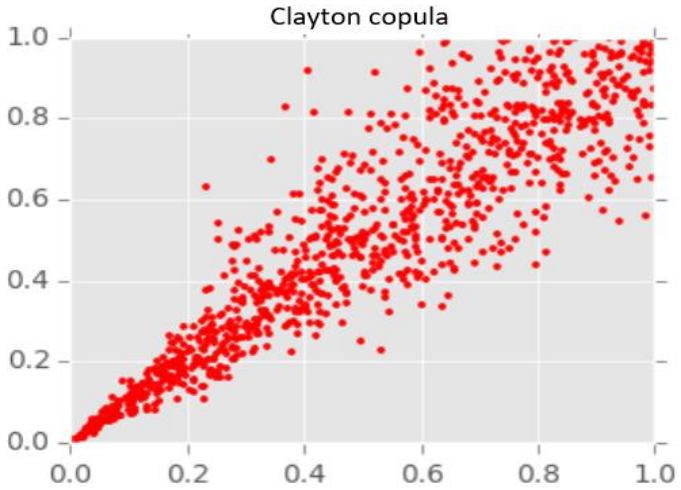

Figure 6-2: Scatter plot of Clayton Archimedean copula

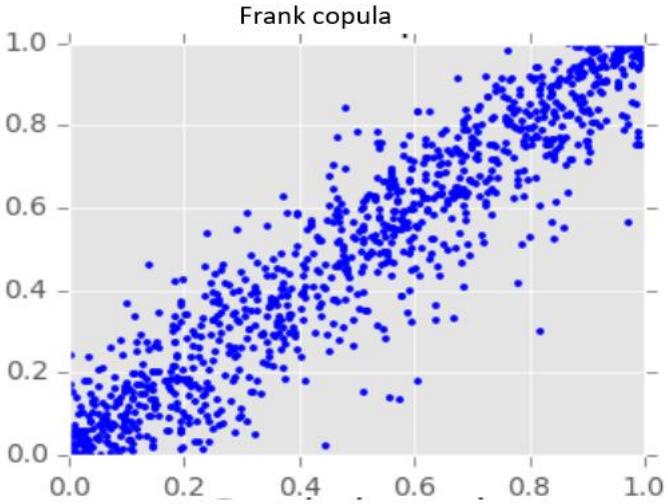

Figure 6-3: Scatter plot of Frank Archimedean copula

These simple visualisations of the most-used Archimedean copulas are displayed in Figure 6-1, 6-2 and 6-3. These figures provide an easy understanding of the difference among Archimedean copulas (Krylov & Zerubia, 2009).

6.5 Kendall's τ measure of dependence

Kendall's τ measure of dependence for two random variables (Y_1, Y_2) is defined as the difference in the probability of the concordance and probability of discordance. Below is the mathematical format:

$$\tau(Y_1, Y_2) = P\big((Y_1, Y_1')(Y_2, Y_2') > 0\big) - P\big((Y_1, Y_1')(Y_2, Y_2') < 0\big),$$

where the two variables (Y_1, Y_2) are known to be concordant when the large values of one variable is related with large values of the other, small values of one variable are related with small values of the other. In addition, (Y_1, Y_2) are known to be discordant when the large values of one variable are related with small values of the other; small values of one variable are related with large values of the other.

6.6 Parameter estimates

There are 4 existing methods of estimating copula models (Manner, 2007). These models have their own advantages and disadvantages. The six methods are;

1. Exact Maximum likelihood – This estimator allows to attain the minimum asymptotical variance bound when the size of the data for the two series is equal. However, the obtaining of the estimates may be difficult.
2. Inference for Margins – The computational of the estimates for the estimates is less demanding compared to Exact Maximum Likelihood. The inference of margins estimates the parameters for margins and the copula parameter in two stages. The inference for margins is easy to compute, however the efficiency is lost through the process.
3. The semi parametric two-step estimator – If the copula is mis-specified, this estimator converges to the pseudo true parameter. This ensures that the model is closest to the data generating process.
4. Nonparametric estimator by Genest and Rivest – Marginal distribution of this approach do not need to be specified. Furthermore, it needs to be noted that this estimator can only be applied to a limited number of one parameter per models.

From the above, there are two frequently used maximum likelihood-based methods to calibrate the copula model. These are exact maximum likelihood method and inference for margins

method. From the current literatures, the exact maximum likelihood is the commonly employed method for parameter estimation (Kaziannka & Pilz, 2011).

Based on the advantages and disadvantages stated in the literature, this thesis will consider te exact maximum likelihood for parameter estimation.

The dependence structure $\emptyset$, the copula parameter γ and the parameter of the marginal distribution are the parameters desired to be estimated in a copula – based model. The likelihood function is given by:

$$L((\theta, M(x)) = c_{\emptyset,\gamma}\big(F_\alpha(M_1), F_\alpha(M_2), F_\alpha(M_3), \dots, F_\alpha(M_N)\big)X \prod_{i=1}^{n} f_\alpha(M_i) \qquad \textbf{(6-3)}$$

where $\theta = (\emptyset, \gamma, \alpha)$, $c_{\emptyset,\gamma}$ is the copula density, f_α is the marginal density and F_α is the distribution function. There are additional MLE estimations in the literature including the pseudo maximum likelihood estimation (PMLE). The PMLE can be employed for the estimation of copula parameters. The pseudo-observation estimation confirms the removal of the marginal feature of the variables, and only the information about the dependence structure remains. The estimator of the maximum likelihood is efficient as it obtains minimum asymptotic variance bound when the amount of data available for two series is equal. The standard errors can be obtained through the use of the inverse Fisher information matrix.

6.7 Link functions selection

For the selection of a suitable copula-based link function, the traditional AIC and the BIC statistical technique are widely employed, this technique is recommended by various literature including (Siciliano, et al., 2014), (Sauerbrei, et al., 2007), (Roberts & Zewotir, 2020)and (Johnson & Omland, 2004).

The visualisation comparing the copula-based link functions is presented in Figure 6-4. Image (a) in Figure 6-4 presents the visual of the standardised link function and image (b) presents the visual of the standardised inverse link function (Mesfioui, et al., 2022). For a simple comparison of the copula-based link functions, different colours were adopted. The black color is for the clog log link function; the green color is for the probit link function; and the blue color is for the logit link function.

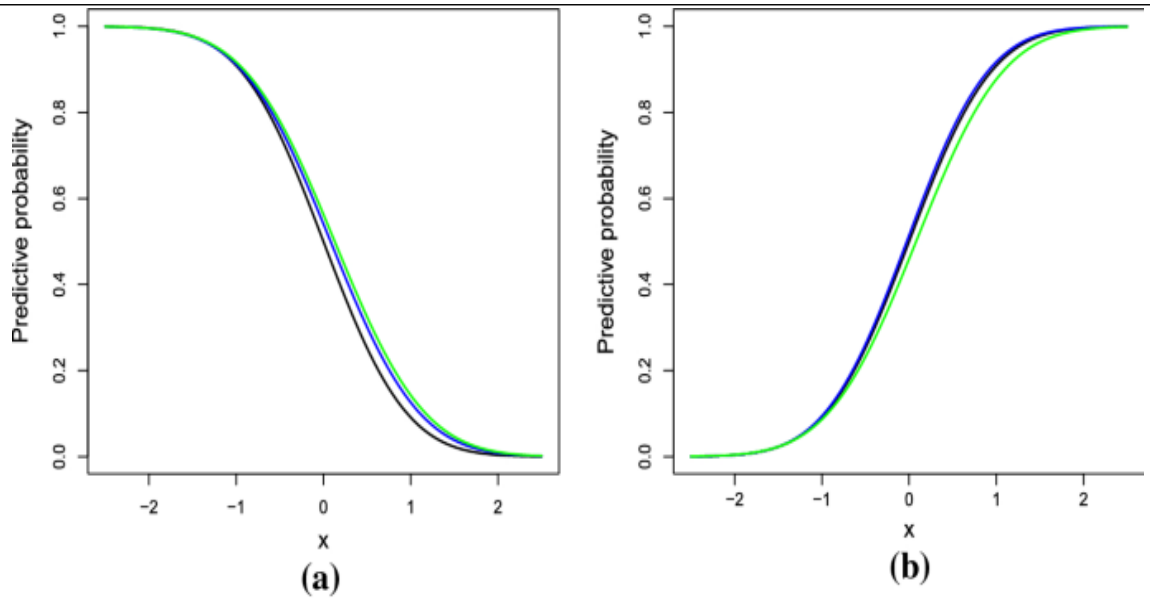

Figure 6-4: Visualization of the Copula-based link function in bivariate regression

6.8 Results of the Copula-based model on malnutrition and anaemia

The response variables (children under five years malnourished and anaemia status) were obtained from the weight-for-age (WAZ) and anaemia level variable in the DHS data. A child is classified as malnourished when the WAZ is less than -3.0 and nourished when the WAZ is greater than -3.0. In contrast, a child is classified as anaemic when the haemoglobin level of a child is less than 9.9 g per dL and a child is classified as not anaemic when the haemoglobin of the child is greater than 9.9 g per dL.

Socio-economic, demographic, health, and environmental elements of living are known to be the contributing factors to malnutrition and anaemia status. The framework used to select the explanatory variables is similar to that discussed in Chapter 3, 4 and 5. The community-level variable included in the study is type of resident (Rural or Urban). Household-level variables included in the study are household size (0-5, 6-10, 11-15 or > 15), sex of household head (Male or Female), mother's level of education (Primary, Secondary or Higher), birth interval (<24, 24-47 or > 47), marital status (unmarried, divorced, married) and wealth index (Not poor, Middle or Poor). The wealth index is a composite measure of household living standards. It is a calculated household's assets, ownership of livestock, household construction material, and water and sanitation facilities. The full method used is explained in Chapter 2. Individual level variables are child's age in months (< 12, 12-23, 24-35, 36-47 or 48-59), sex of child (Male or Female) and childbirth order (2-3, 4-5 or > 5).

The multivariate analysis was performed in R 3.63 package *GJRM* (Generalized Joint Regression Modelling). For the mapping, each country's boundaries were obtained from DHS program and were thereafter saved as shapes in QGIS 3.4 software. These shapes were imported into R 3.63 software for results mapping. The Akaike information criterion (AIC) and BIC were used to select the best model; it is a good criterion for finding the best fitting copula model ((Haugh, 2016), (Burnham, et al., 2011), (Siciliano, et al., 2014)).

Table 6-1: The AIC results of the copula selection

Type	**df**	**AIC Value**	**BIC Value**
Clayton	31	14993	15471
Gumbel	31	14962	15453
Frank	31	14963	15455
Gaussian	31	14965	15459
Student- T	31	14969	15460

Table 6-2: The AIC results of the link function

Type	**df**	**AIC Value**	**BIC Value**
c ("logit", "logit")	31	14947.86	17057.11
c ("logit", "cloglog")	31	14951.24	17156.78
c ("logit", "probit")	31	14947.31	17167.91
c ("cloglog", "probit")	31	14987.34	17197.39

Table 6-1 show the AIC and BIC results of the Copulas that were considered to jointly model the response variables. Based on the above table, the Gumbel Copula is selected to jointly model our responses (Anaemia and Malnourishment). Post selecting the copula, we further selected the link function using the AIC and BIC. The results from Table 6-2 indicated that the c ("logit", "probit") function is the best suited for the final model.

The results of the fixed effects are based on the Gumbel Copula and c ("logit", "probit") link function. The R 3.63 package *GJRM* (Generalized Joint Regression Modelling) was used to obtain the results. All the factors that were not significant in the initial step of model selection were excluded from the final model. Seven factors were included in the last copula model:

$$Y_{i1} = \beta_0 + \beta_1 residence + \beta_2 child's\ age + \cdots + \beta_7 birth\ order$$

$$Y_{i2} = \beta_0 + \beta_1 residence + \beta_2 child's\ age + \cdots + \beta_7 birth\ order$$

where Y_{i1} and Y_{i2} are the child's anaemic status and child's malnutrition status. The $\beta_0, \beta_1, \ldots, \beta_7$ are the parameter estimates. Hence, the joint copula model of the two response variables is:

$$P(Y_{i1} = 1, Y_{i2} = 1 \mid x_{i1}, x_{i2}) = C[P(Y_{i1} = 1|x_{i1}); P(Y_{i2} = 1|x_{i2});\ \theta]$$

where x_{i1} & x_{i2} are the child's independent factors.

Table 6-3: Parameter estimates of the fixed effects for the bivariate copula regression model of malnourishment and anaemia

	Malnourished			**Anaemic**		
Covariates	**Estimates**	**St. error**	**P- Value**	**Estimate**	**St. error**	**P-Value**
Intercept	3.500	-	0.982	1.065	-	
Resident						
Ref: Urban						
Rural	-0.155	0.856	0.104	0.053	1,054	0.152
Child's age						
Ref :< 12						
12-23	-0.058	0.944	0.765	-0.226	0,798	0.003
24-35	0.078	1.081	0.676	-0.529	0,589	<0.001
36-47	-0.037	0.964	0.840	-0.726	0,484	<0.001
48-59	-0.073	0.930	0.685	-0.105	0,900	<0.001
Child's gender						
Ref: Male						
Female	0.144	1.155	0.027	-0.105	0,900	<0.001
Mother's level of schooling						
Ref: Higher						
Primary	1.217	2.995	0.027	1.888	6,606	0.047
Secondary	1.097	3.377	0.014	0.442	1,556	0.024
Birth interval						
Ref: <24 months						
24-47	-0.266	0.766	<0.001	0.044	1,045	0.393
>47	0.265	1.303	0.020	-0.019	0,981	0.727
Wealth index						
Ref: Poor						
Middle	-0.553	0.575	0.032	0.074	1,077	0.001
Not poor	-3.444	0.032	<0.001	0.226	1,254	<0.001

Birth order						
Ref: 2-3						
4-5	-0.419	0.658	<0.001	-0.017	0,983	0.641
>5	-0.695	0.499	0.000	-0.032	0,969	0.339

Table 6-3 display the results of the fixed effects of each marginal model. In this section, the level of significance used is 5%. Based on the results, children's place of residence had no significant effect on malnourishment and anaemia (rural estimate = -0.155, p-value = 0.104 for malnourishment; rural estimate = 0.053, p-value = 0.152 for anaemia). Child's age was significantly associated with child's anaemia status at all age categories but had no significant effect on the child's malnourishment status. A female child was found to be significantly associated with both anaemia status and malnourished status (female estimate = 0.144, p-value = 0.027 for anaemia; female estimate = -0.105, p-value <0.001). Furthermore, the likelihood of each outcome significantly decreased with an improvement on the mother's level of education. A unit increase in the household's wealth index Z-score was significantly associated with the decrease in the likelihood of malnourishment and anaemia. Table 6-4 shows the significance of the non-linear and spatial effects for Anaemia and Malnutrition response variable.

Table 6-4: Approximate significance for the non-linear and spatial effects

Variable	**Anaemia**		**Malnourished**	
	Chi-Square Value	**P-value**	**Chi-Square Value**	**P-value**
Childs' age	251.01	0.003	875.09	0.011
Unstructured effect	379.32	<0.001	685.44	<0.001
Structured effect	457.77	0.045	475.33	<0.001

The structured spatial effect and unstructured spatial effect had a significant effect on the likelihood of each outcome. The child's age had a significant non-linear effect on the likelihood of each response. Non-linear effect results of child's age on anaemia and malnourishment are displayed in Figure 6-5 and Figure 6-6, respectively. The likelihood of anaemia among children under five increases from the age 0-30 months, and thereafter decreases, whereas, the likelihood of malnourishment decreases as the child grows. The region/district level structured special effect for both malnourishment and anaemia is presented in Figure 6-7 and Figure 6-8.

Based on the data and maps available, the Senegal and Malawi spatial effect was conducted at regional level. The districts or regions in light maroon resemble to a negative estimated effect and

therefore are associated with a lower likelihood of the occurrence. Whereas, districts or regions in dark maroon resemble to a positive effect and were associated with a higher likelihood of the occurrence.

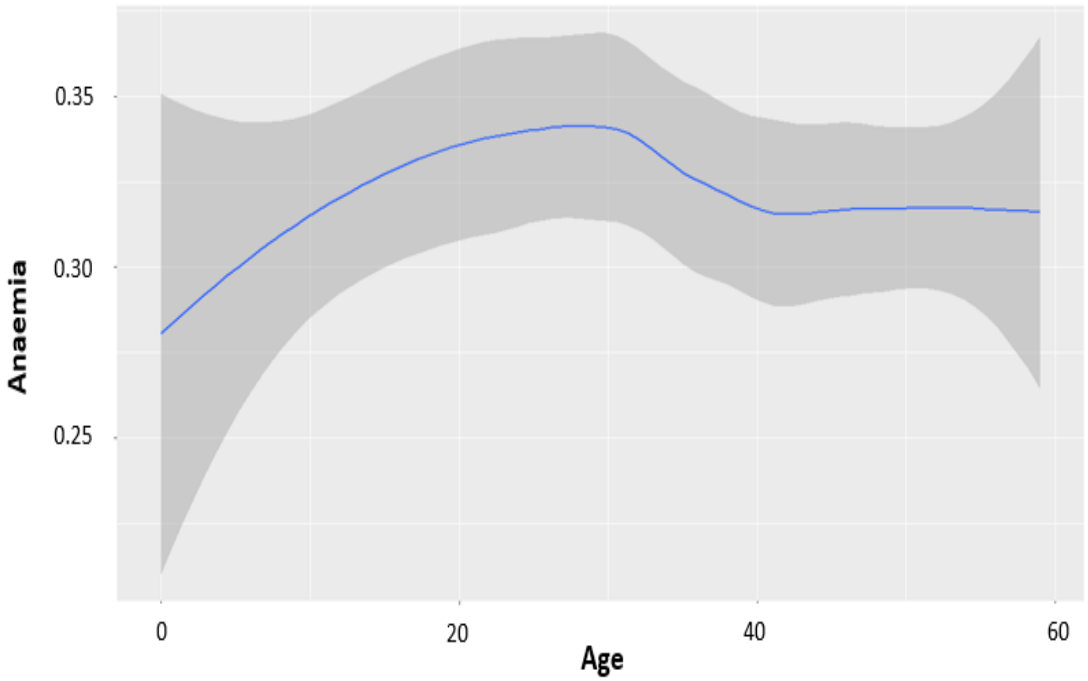

Figure 6-5: Estimated non-linear effect of the child's age on anaemia

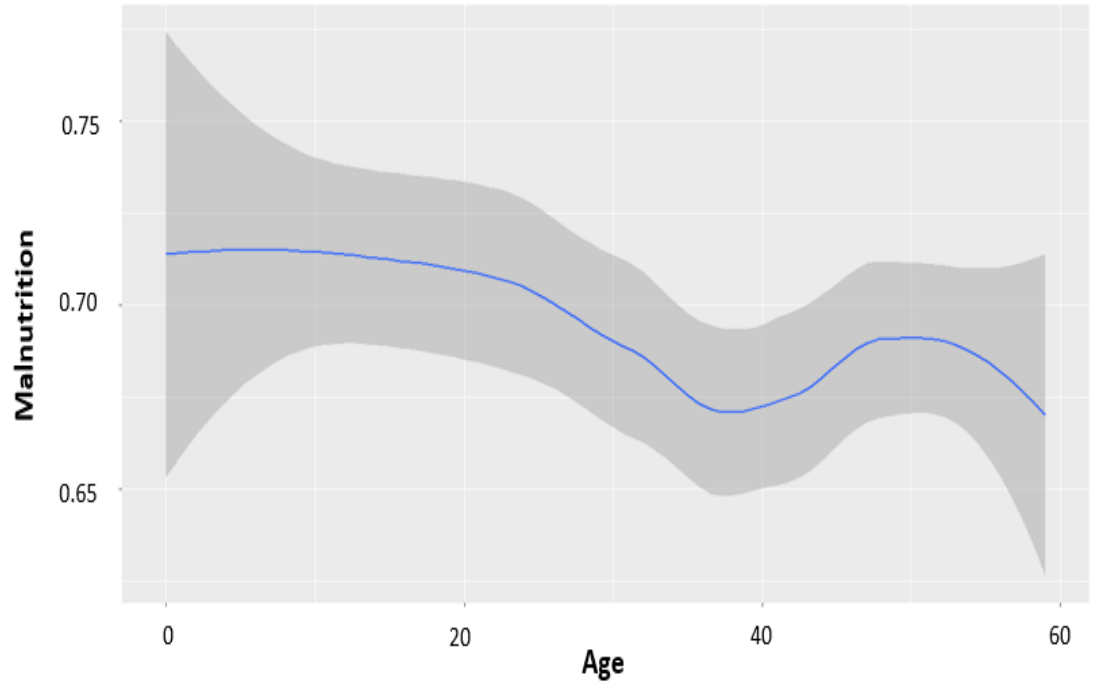

Figure 6-6: Estimated non-linear effect of the child's age on malnourishment

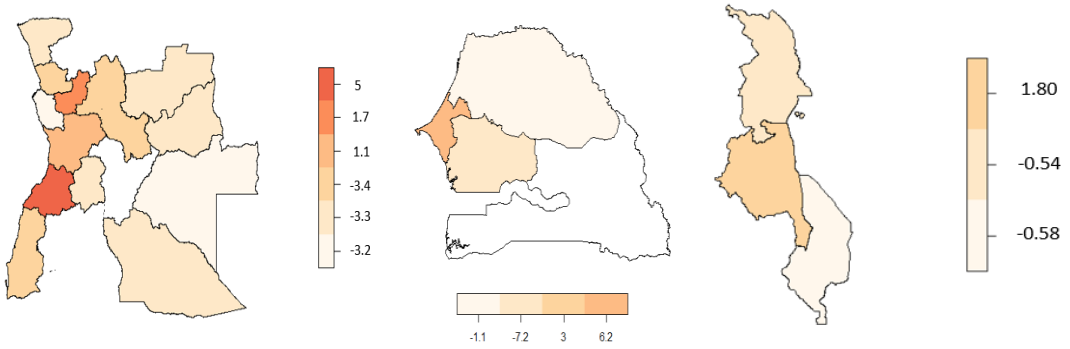

Figure 6-7: Estimated effect of the structured spatial effect on malnourishment. Left: Angola; middle: Senegal and right: Malawi

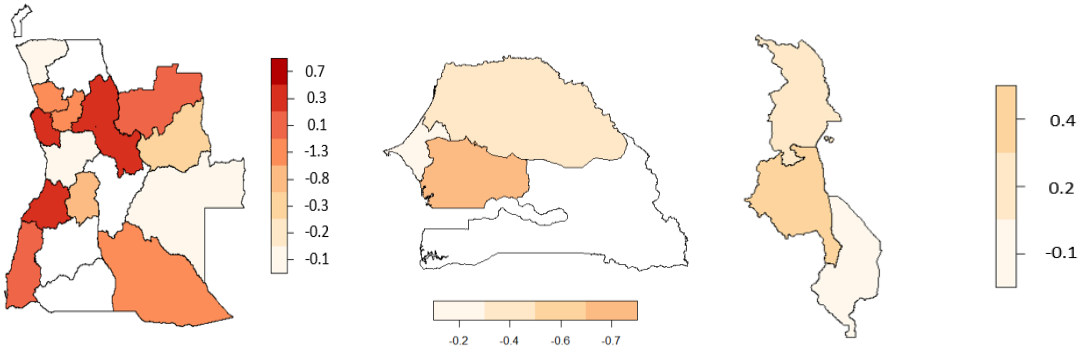

Figure 6-8: Estimated effect of the structured spatial effect on anaemia. Left: Angola; middle: Senegal and right: Malawi

The structured spatial effect for malnourishment revealed that Angola consisted of districts associated with a lower likelihood of malnourishment as well as districts associated with a higher likelihood of anaemia. This spatial variation suggests that it was imperative to control in order to avoid the reduction in the statistical power of inference in the model, and that could have led to incorrect results.

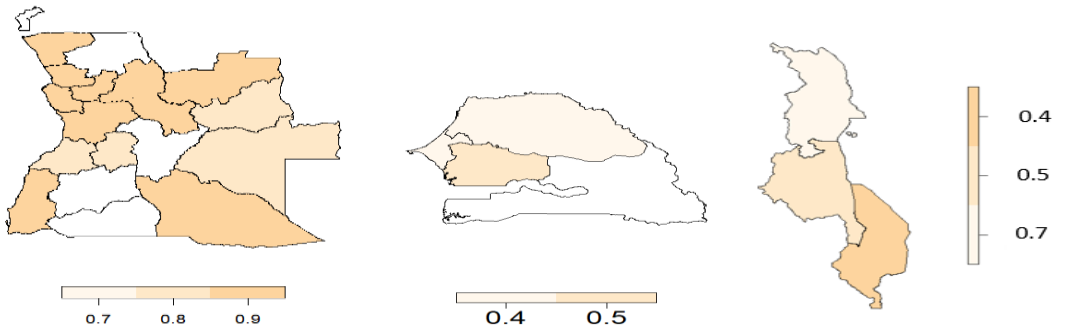

Figure 6-9: Joint probability of a child being malnourished and anaemic. Left: Angola; middle: Senegal and right: Malawi

From the fitted bivariate copula regression model, the estimated joint probabilities were extracted and averaged over the districts or regions. Figure 6-9, 6-10, 6-11 and 6-12 shows these joint probabilities for each combination of the malnourishment and anaemia outcome. Looking at Figure 6-9, most of the districts in Angola showed a considerably high joint probability of a child being malnourished and anaemic. Senegal and Malawi had few regions that showed high joint probability of both malnourishment and anaemia in children. From Figure 6-10, fewer districts or regions in Angola and Senegal have a high probability of having a nourished or not anaemic child. Whereas, in Senegal there is a low probability of observing a child that is not malnourished, but anaemic (Figure 6-11). Thus, this indicates that in Senegal there is a high likelihood of children being malnourished when they have anaemia. Considering Figure 6-12, most of the districts or regions in Angola, Senegal and Malawi have a small probability of children being malnourished, but not anaemic.

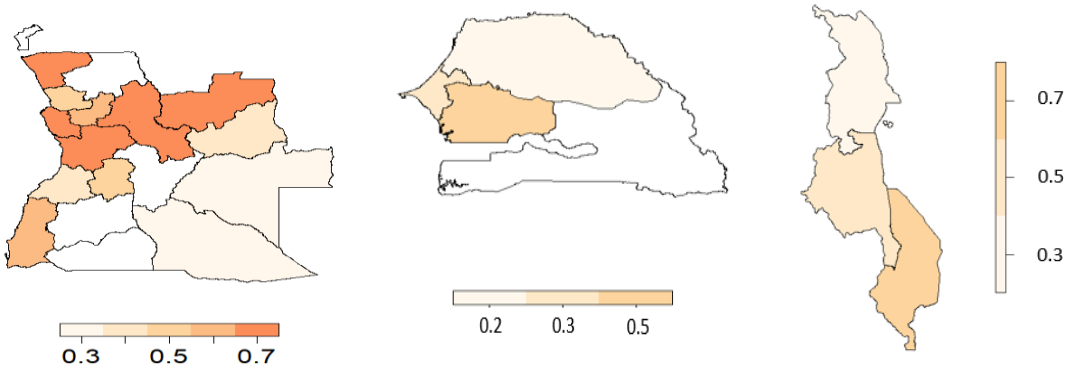

Figure 6-10: Joint probability of a child being not malnourished and not anaemic. Left: Angola; middle: Senegal and right: Malawi

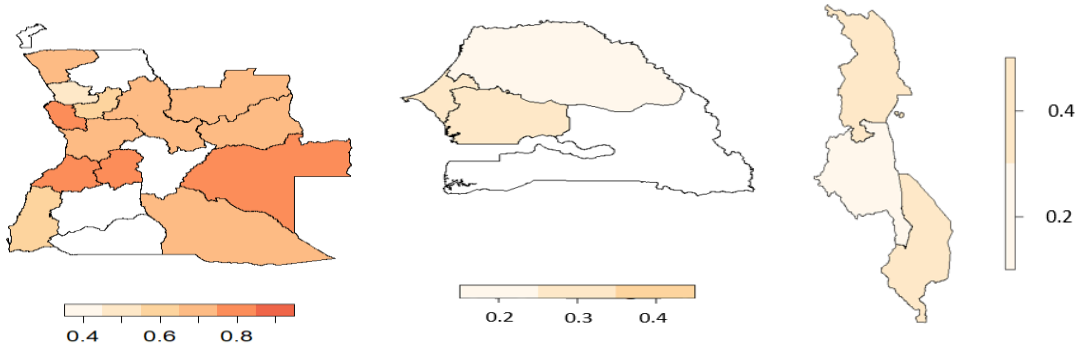

Figure 6-11: Joint probability of a child being not malnourished and anaemic. Left: Angola; middle: Senegal and right: Malawi

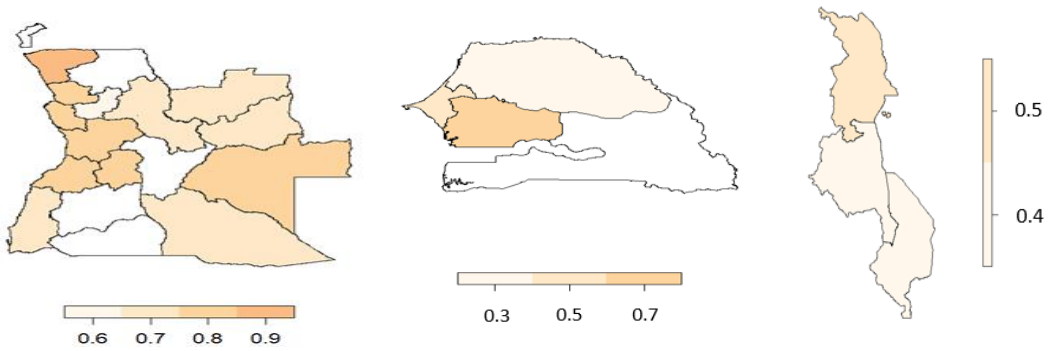

Figure 6-12: Joint probability of a child being malnourished and not anaemic. Left: Angola; middle: Senegal and right: Malawi

6.9 Summary

This chapter aimed to evaluate the association between malnutrition and anaemia among children under five years in Angola, Senegal and Malawi. The model that was used to analyse the data is the joint bivariate copula regression model. This type of model allows the correlation between the two responses to be estimated while controlling the linear and non-linear effects, as well as the effect of spatial variation. The advantage of copula regression over multivariable analysis is that the normality and the linearity of the dependence between the response variables are not assumed and replicate the dataset through simulating with any type of marginal distribution. The copula-based model can model the behaviour of skewed data. The data analysis of this study was completed in R 3.63 software. For the mapping, each country's boundaries were obtained from the DHS program and were thereafter saved as shapes in QGIS 3.4 software. These shapes were imported into R 3.63 software for results mapping. The joint probability mapping results showed that the association between malnutrition and anaemia differs by district or region. Some regions had a stronger association between the two responses when compared to other districts or regions. The focus of this paper was on the likelihood of a child being malnourished and anaemic; however exploring other possible outcome combinations were evaluated to provide more insight in an attempt to understand the relationship between malnutrition and anaemia. Furthermore, differentiating the association at the district or regional level will ensue to more direct intervention implementation to control the response variables. In the districts or regions where there is a stronger correlation between malnutrition and anaemia, the success of the anaemia control program can be an indicator of the success of malnutrition control. In the districts or regions with a higher probability of malnourished but not anaemic children, it would be rational to assume that there are other drivers of malnutrition. Thus, implementing the interventions for malnutrition in such a district or region to control anaemia would not be effective. The northwest districts of Angola showed a higher probability of a child being malnourished and anaemic. The government

need to allocate health and educational resources or implement intervention programmes in these districts to control malnutrition and anaemia in Angola. The regions in Senegal and Malawi showed the least amount of variation in the joint probabilities of a child being malnourished and anaemic. This might be due to the programmes or interventions that have been implemented. More attention or resource distribution between these countries must be given to Angola as a priority. The fixed effects results are consistent with the results from other malnutrition and anaemia separate modelling. The results revealed that the sex of the child, mother's level of schooling and household wealth index are significantly associated with malnutrition and anaemia ((Roberts & Zewotir, 2020); (Semedo, et al., 2014)). Another factor that was found to be significantly associated with malnutrition is a child's birth interval. This result is consistent with the results of the study by (Govender, et al., 2021). A child's age was found to be significantly associated with anaemia; similar results were observed in a study by (Semedo, et al., 2014).

Chapter 7 : Conclusion and Recommendation

This thesis's main purpose was to contribute to malnutrition and anaemia epidemiology by identifying the factors associated with childhood malnutrition and anaemia in Angola, Malawi and Senegal based on the national representative survey data. The objectives of the study were motivated by the projections released by UNICEF (2018) indicating that under five mortality in Africa is 54 deaths per 1000 live births. However, the Sustainable Development Goal (SDG) target is to reduce under five mortality to 25 deaths per 1000 live births by 2030. Based on the projections, Africa will be more than half of the SDG target by 2030. Mayotte, Reunion, Seychelles and Egypt are the only African countries that currently have under five mortality below the SDG 2030 target.

For any analysis, we considered a set of covariates that were suggested based on theoretical frameworks or literature; these covariates include demographic, socio-economic and environmental factors. In Chapter 2 of this thesis, we presented a comprehensive exploratory analysis which gave insight onto the covariates and models to be considered for the determinants of risk factors related to malnutrition and anaemia. In each chapter that has multivariate analysis, the covariates used were mentioned and tested for their association with the outcome variable(s). Further in chapter 2, the wealth of the household was defined based on asset ownership. This was done in SPSS using the principal component analysis.

To deeper understand the risk factors of malnutrition and anaemia, the generalized linear mixed model (GLMM) approach was adopted, which allowed for the complexities of the survey designs and correlation between observations of the same cluster. Controlling these data characteristics improves the accuracy of the estimates for the fixed effects. The GLMM was used on the data for malnutrition and anaemia risk factors; the findings are presented in Chapter 3. On the malnutrition risk factors, the results of the GLMM showed that type of residence, sex of the child, age of the child, mother's level of schooling, birth interval, wealth index and birth order are the correlates of malnutrition in Angola, Malawi and Senegal. Mother's level of schooling results suggested that children from Angola, Malawi and Senegal who resides with mothers who attained primary education, are at higher risk of being affected by malnutrition. It was further concluded that children from the rural communities, from a poor household with a mother who attained primary education, who are female and who are between the age of 24-59 months are associated with malnutrition. The results show the necessity of collaboration among the three countries in order to achieve the SGD target. It is evident that the African continent is a developing continent with

limited resources for equity amongst communities and meeting the needs of the communities; thus, it is recommended that Angola, Malawi and Senegal work jointly to eradicate child malnutrition.

Further in Chapter 3, seven factors were included in the final model for identifying risk factors of anaemia. However, only five were found to be significant in explaining anaemia at a 5% level of significance. A generalized linear mixed model identified child's age, child's gender, mother's level of schooling, wealth status and child malnutrition status as determinants of anaemia among children under five years in Angola, Malawi and Senegal. It was asserted that female children under five between the age of 0 - 59 months with a severe or moderate malnutrition status from a poor or middle household and with primary or secondary educated mothers, are more exposed to anemia in Senegal, Malawi and Angola. To achieve the SGD target, there is a need for Angola, Malawi and Senegal to initiate programs that will focus on improving the economy of all communities to close the gap between poor and not poor households and to improve women's education. This finding agrees with the findings of the research conducted by (Habyarimana, et al., 2014), (Zhoa, et al., 2012), (Woldie, et al., 2014).

The findings from Chapter 3 provided risk factors associated with malnutrition and anaemia separately; however there was a need to jointly model malnutrition and anaemia. This is necessary as it will provide the shared risk factors of malnutrition and anaemia. Other advantages of joint modelling include (1) enhanced control over Type I error rates during multiple testing and efficiency in estimating parameters; (2) the correlation between the outcomes can be quantified and controlled. In this thesis we discussed three joint model approaches: the joint generalized linear mixed model, structural equation model and the copula-based model.

The results of the joint generalized linear mixed model are presented in Chapter 4 and showed that the variance components test revealed a positive correlation between nutrition status and anaemia status. Based on the multivariate analysis, it was revealed that type of residence, sex of the child, age of the child, mother's level of schooling, birth interval and wealth index are the correlates of malnourishment and anaemia in Angola, Malawi and Senegal. These variables are also identified in other studies conducted in Africa or developing countries (Ray, et al., 2016), (Muchie, 2016). It was concluded from the study that female children between the age of 24-59 months from the rural communities with the mother having attained primary or secondary education and residing in the poor or middle households, are associated with malnutrition and anaemia problems. Therefore, for Angola, Senegal, and Malawi to achieve the SGD target, the government and policymakers of these countries need to improve strategic programs that will address the issues inducing childhood mortality. Such programs include parental education,

financial education, children's dietary focus programs and mobile health facilities. There is a greater need for partnership and collaboration among the studied countries to achieve the SGD target.

To assess the complexity of the interrelationships between multiple variables and relate unobservable and observable variables with estimation in terms of the sample covariance matrix of the observed variables and the population covariance matrix, we adopted the SEM. In addition, the SEM examined the interrelationships between contextual factors (demographic and socio-economic) and the co-morbidity of childhood malnutrition and anaemia among children under five years in Angola, Senegal and Malawi. The results in Chapter 5 showed that the socio-economic factors and child inherent factors have a direct effect on childhood co-morbidity. Whereas, the indirect path of socio-economic factors on childhood malnutrition and anaemia as mediated by child inherent factors was found to be statistically significant. The study found that place of residence, mother's level of education, household wealth index, age of the child and sex of the child are the important risk factors for childhood co-morbidity prevalence. The results indicate that there is a need for urgent intervention from policy makers of Angola, Malawi and Senegal to reduce the childhood co-morbidity prevalence. The interventions to be implemented include improving resources that are distributed to the rural areas, introducing education program for old people and setting aside the budget that will support people who want to start small businesses.

However, the SEM did not account for dependence structure spatial variability of the two diseases. Also, the SEM could also not reveal the trends of the association between the predictors and response variables, whether linearly or nonlinearly related. Hence the statistical method that includes the accountability of dependence structure and spatial variation of co-morbidity in childhood malnutrition and anaemia was adopted. This type of statistical method is called copula geo-additive model. Copula geo-additive model extension aimed at responding to the questions about how the association between the responses varies according to the geographical location.

The findings of copula geo-additive modelling of malnutrition and anaemia are presented in Chapter 6. The results of the mapping showed that there is an association between malnutrition and anaemia. This implies that the policymakers of Angola, Senegal and Malawi can control anaemia through the intervention of malnutrition controlling. This will save the countries' resource for implementing interventions for both malnutrition and anaemia. The stronger association between malnutrition and anaemia was observed in the north-west districts of Angola when compared to other districts. This suggests that for better control of malnutrition and

anaemia, the government need to target that part of the country. The Senegal and Malawi regions have lower probabilities and are distributed fairly across the regions. Based on the fixed effects analysis, sex of the child, mother's level of education and household wealth index are significant factors of malnutrition and anaemia. These results are consistent with other studies on child mortality (Semedo, et al., 2014), (Mbabazi & Kanyamuhunga, 2021). Angola, Malawi and Senegal can form a focus group that will target poor households, parental education and the importance of breastfeeding, for example. All these will assist in reducing the prevalence of malnutrition and anaemia among young children in Angola, Senegal and Malawi. The statistical literature on copula modelling is still growing; however the application is notably in survival analysis, actuarial science and finance. This study was hoped to foster the use of copula methodology in this field of science with the use of cross-sectional data.

This thesis revealed important insights about malnutrition and anaemia, epidemiologically. It highlighted the national priority areas related to child-related factors, household factors and environmental factors for childhood malnutrition and anaemia morbidity control. It also provided policy makers with valuable geographical information for developing and implementing effective intervention.

There are many variables that can result in co-morbidity of childhood malnutrition and anaemia in any country including, Angola, Malawi and Senegal. This is the debate which is not easy to solve; as a result this limits our conclusions of this thesis. This study used few explanatory variables which were based on theoretical frameworks and literature on childhood malnutrition and anaemia. In addition, we only adopted three types of joint modelling in this study; however more studies can be done on the analysis of malnutrition and anaemia with more variables explaining malnutrition and anaemia through several statistical models and methods, including shared component modelling of childhood malnutrition and anaemia especially for the joint modelling of these two diseases. Malnutrition and anaemia have been global public health concerns for decades and different literature have proposed several methods/ interventions to control them. Its complexity issue and multidimensionality require rigorous analysis and further studies.

Therefore, the future research consists of:

- Analysing malnutrition and anaemia using more explanatory variables including dietary variables, episodes of diarrhoea etc.

- Consider DHS data from other African countries. This would add to the knowledge of factors contributing to childhood mortality and broaden the landscape of countries that

can collaborate. It is therefore suggested that a similar study should be carried out including more than three countries.

- Changing the response variables from being binary to multiple on the joint models.

- Since the DHS data become available after 5 years, future research includes monitoring the progress made since this thesis proposed interventions.

- To introduce the use of Copula methodology in the field of science we used the cross-sectional data. Hence, the future research from this study is to consider the use of longitudinal data.

Bibliography

Abuka, T. & Tsegaw, D., 2017. Determinants for acute malnutrition among under-five children at public health facilities in Gedeo Zone, Ethiopia: Case-Control Study. *Pediatr Ther,* pp. 317-328.

Achia, N. O., Khadioli, N. & Wangombe, A., 2010. A Logistic Regression Model to Identify Key Determinants of Poverty Using Demographic and Health Survey Data. *European Journal of Social Science,* pp. 38-45.

Adeyemi, R. A., Zewotir, T. & Ramroop, S., 2019. Joint spatial mapping of childhood anaemia and malnutrition in sub-Saharan Africa: a cross-sectional study of small-scale geographical disparities. *Afri Health Sci.,* pp. 2692-2712.

Afolabi, R. F. & Palamuleni, M. E., 2021. Exploring determinants of under-5 stunting in Malawi using a generalised linear mixed model.. *SAJCH,* pp. 18-24.

Aiga, H. & Matsuoka, S. K. C. &. Y. S., 2009. Malnutrition among children in rural Malawian fish-farming households.. *Elsevier,* pp. 827-833.

Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In:Petrov,B.N.,Csaki,F.(Eds.),Second International Symposium on Information Theory.. *Akademiai Kiado,Budapest,* pp. 267-281.

Amorim, L. et al., 2010. Structural equation modeling in epidemiology. *Cad Saude Publica.,* pp. 2251-2262.

Anderson, R., Black, W., Hair, J. & Tatham, R., 1998. Multivariate data analysis. 5th edition. In: New York: Prentice Hall.

Anderson, T., 1958. An introduction to multivariate Statistical Analysis. In: New York: John Wiley and Sons Inc..

Aynalem, M., Shiferaw, E. A. T., Gelaw, Y. & Enawgaw, B., 2022. Anemia in African malnourished pre-school children: A systematic review and meta-analysis. *SAGE,* pp. 1-11.

Ba, I. D. et al., 2021. Anemia in among Children Aged 6-59 Months at the Institute of Social Pediatric (ISP) in Dakar Suburban Area. *J Pediatr Perinatol Child Health,* pp. 158-163.

Baldwin, B., 1989. A primer in the use and interpretation of structural equation models.. *Measurement and Evaluation in Counseling and Development,* pp. 100-112.

Barros, H. et al., 2020. Iron deficiency anaemia among 6-to-36 month children from northern Angola. *BMC Pediatrics,* pp. 1-13.

Beau, J., Garenne, M., Diop, B. & Briend, A. &. D. M. I., 1987. Diarrhoea and nutritional status as risk factors of child mortality in Dakar hospital (Senegal). *Journal of tropical paediatrics,* pp. 4-9.

Beslow, N. & Lin, X., 1995. Bias correction in generalized linear mixed models with single component of dispersion. *Biometrka,* Volume 82, pp. 81-91.

Bolker, B. et al., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol,* pp. 127-135.

Bollen, K. A., 1989. *Structural equations with latent variables.* USA: John Wiley and Sons.

Bonett, D. & Wright, T., 2014. Cronbach's alpha reliability: Interval estimation,hypothesis testing, and sample size planning. *Journal of Organizational Behavior, J. Organiz. Behav.,* pp. 1-12.

Booysen, F. R., 2002. Using demographic and health survey to measure poverty: An application to South Africa. *J. Stud. Econ. Econometrics,* pp. 53-70.

Brabin, B. J. et al., 2004. Risk factors for fetal anemia in a malarias area of Malawi. *Annals of tropical paediatrics,* pp. 311-321.

Broatch, J. & Karl, A. T., 2018. Multivariate Generalized Linear Mixed Models for Joint Estimation of Sporting Outcomes. *Italian Journal of Applied Statistics,* pp. 189-211.

Burney, A. & Ajaz, O., 2020. *Copulas: A Historical Literature Review and Major developments Abstract.* s.l., s.n.

Burnham, K. P., Anderson, D. R. & Huyvaert, K. P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol,* Volume 65, pp. 23-35.

Chirwa, E. &. N. H. P., 2009. Determinants of child nutrition in Malawi. *SAJE,* pp. 628-640.

Clark, D. & Thayer, C., 2004. *A primer on the exponential family of distribution,* s.l.: Paper program on Generalized Linear Models..

Demyonovs, S., Bailey, J., Ramanohararao, K. & Leckie, C., 2012. AIC and BIC based pproaches for SVM parameter value estimation with RBF kernels. *JMLR,* pp. 97-112.

Diouf, S. et al., 2013. Anemia among apparently health Senegalese children aged 9-15 months.. *International Journal of Child Health and Nutrition,* pp. 9-14.

Doctor, H. &. N.-S. S., 2017. Trends and Determinants of Child Growth Indicators in Malawi and Implications for the Sustainable Development Goals. *Public health,* pp. 590-614.

Ehrhardt, S. et al., 2020. Malaria, Anaemia, and Malnutrition in African ChildrenDe ning Intervention Priorities. *Paediatr Int Child Health,* pp. 108-114.

Everitt, B. & Howell, D., 2005. *Encyclopedia of statistics in behavoiral science.* s.l.:Hohn Wiley & Sons, Ltd.

Fang, D., Sun, R. & Wilson, J. R., 2018. Joint modeling of correlated binary outcomes: The case of contraceptive use and HIV knowledge in Bangladesh. *Plos one,* pp. 1-9.

Feddag, M. & Mesbah, M., 2006. Approximate estimation in Generalized Linear Mixed Models with applications to the Rasch Model. *Comput Math Appl,* pp. 269-278.

Field, A., 2000. *Discovering statistics using SPSS windows,* s.l.: Sage publications Ltd.

Gasparihno, C. et al., 2022. Wasting, Stunting, and Anemia in Angolan Children after Deworming with Albendazole or a Test-and-Treat Approach for Intestinal Parasites: Binary Longitudinal Models with Temporal Structure in a Four-Arm Randomized Trial. *MDPI,* pp. 2-21.

Gaston, T. R., Ramroop, S. & Habyarimana, F., 2018. Determinants of factors associated with anemia among children under ve years in Lesotho. *African Population Studies,* pp. 3893-3902.

Gebreegziabiher, G., Etana, B. & Niggusie, D., 2014. Determinants of anemia among children aged 6-59 months living in Kilte Awulaelo Woreda, Northern Ethiopia. *Hindawi,* pp. 1-9.

Gebregziabher, M. et al., 2013. Joint modelling of multiple longitudinal cost outcomes using multivariate generalized linear mixed models. *Health Serv Outcomes Res Methodol,* pp. 39-57.

Goldstein, H. & Rasbash, J., 1996. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society,* pp. 505-513.

Gomes, D. G. E., 2022. Should I use fixed effects or random effects when I have fewer than five levels of a grouping factor in a mixed-effects model?. *PeerJ.,* Volume 10, pp. 1-16.

Govender, I., Rangiah, S., Kaswa, R. & Nzaumvila, D., 2021. Malnutrition in children under the age of 5 years in a primary health care setting. *S Afr Fam Pract.,* pp. 1-6.

Habyarimana, F., Zewotir, T. & Ramroop, S., 2014. A proportional odds model with complex sampling design to identify key determinants of malnutrition of children under ve years in Rwanda. *MCSER,* pp. 1642-1648.

Habyarimana, F. Z. T. &. R. S., 2016. Key determinants of malnutrition of children under five years of age in Rwanda: Simultaneous measurement of three anthropometric indices. *African Population Studies,* pp. 2328-2340.

Hair, J. et al., 2006. *Multivariate Data Analysis..* New Jersey: Pearson University Press.

Haugh, M., 2016. An introduction to Copulas. Quantitative risk management. In: s.l.:Spring.

Hedeker, D., 2005. *Generalized linear mixed models.* s.l.:John Wiley & Sons, Inc..

Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X., 2013. *Applied logistic regression.* New York: John Wiley & Sons, Incorporated.

Humbwavali, J. B. et al., 2019. Malnutrition and its associated factors: a cross-sectional study with children under 2 years in a suburban area in Angola. *BMC Public Health,* pp. 1-11.

Johnson, J. & Omland, K. S., 2004. Model selection in ecology and evolution. *Trends Ecol Evol,* Volume 19, pp. 101-108.

Jolliffe, I., 2002. *Principal Component Analysis, 2nd edition.,* New York, USA: Springer-Verlag and New York Inc..

Journee, M., Nesterov, Y., Richtarik, P. & Sepulchre, R., 2010. Generalized Power Method for Sparse Principal Component Analysis. *Journal of Machine Learning Research,* pp. 517-553.

Kaiser, H. F., 1974. An index of factorial simplicity.. *Phychometrika,* pp. 31-36.

Kawo, K. N., Asfaw, Z. G. & Yohannes, N., 2018. Multilevel Analysis of Determinants of Anemia Preva-lence among Children Aged 659 Months in Ethiopia: Classical and Bayesian Approaches. *handaiwi,* pp. 1-13.

Kaziannka, H. & Pilz, J., 2011. Bayesian spatial modeling and interpolation using copulas. *Computers & geosciences,* pp. 310--319.

Kejo, D. et al., 2016. Prevalence and predictors of anemia among children under 5 years of age in Arusha District, Tanzania. *Prediatric Health, Medicine and Therapeutic ,* pp. 9-18.

Khan, J. R., Awan, N. & Misu, F., 2016. Determinant of anemia among 6-59 months aged children in Bangladesh: evidence from natially representative data. *BMC pediatrics,* pp. 1-10.

Khulu, C. & Ramroop, S., 2020. Determinants of malnutrition in under five children in Angola, Malawi and Senegal.. *The open public health journal,* pp. 3-9.

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

Krylov, V. & Zerubia, J., 2009. *High resolution SAR image classification,* s.l.: Research Report 7108.

Kuk, A., 1995. Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society,* Volume 57, pp. 395-407.

Kuziga, F., Adoke, Y. & Ayenyenze, R. K., 2017. Prevalence and factors associated with anemia among children aged 6 to 59 months in Namutumba district, Uganda: across- sectional study. *BMC Pediatrics,* pp. 1-9.

Lee, A. et al., 2007. Stress and psychological distress among SARS survivors 1 year after the outbreak. *Can J Psychiatry,* pp. 233-240.

Lee, Y.-h. & Kim, D. J., 2016. How to establish clinical prediction models. *Endocrinol Metab,* Volume 31, pp. 38-44.

Linnemayr, S. & Alderman, H. &. A. K., 2008. Determinants of malnutrition in Senegal: individual, household and community variables, and their interaction. *Elsevier,* pp. 252-263.

Littell, R., Stroup, W. & Freud, R., 2002. SAS for Linear Models, 4th edn.. In: Cary, NC.: Institute, Inc.,.

Liu, Q. & Pierce, D., 1994. A note on gauss-hermite quadrature. *Biometrika,* Volume 81, pp. 624-629.

Manner, H., 2007. *Estimating and model selection of Copulas with an application to Exchange rates,* s.l.: s.n.

Mbabazi, E. & Kanyamuhunga, A., 2021. Prevalence of Anaemia and Associated Socio-Economic Determinants amongst Malnourished Children Aged 6-59 Months, Centre Hospitalier Universitaire de Kigali (CHUK) – A Retrospective Observational Study. *Rwanda Medical Journal,* pp. 29-36.

McCulloch, P. & Neuhaus, J., 2005. *Generalized linear mixed models.* s.l.:John Wiley & Sons,Inc..

McCuskee, S., Brickley, E. B., Wood, A. & Mossialos, E., 2014. Malaria and Macronutrient De ciency as Correlates of Anemia in Young Children: A Systematic Review of Observational Studies. *Annal s o f Globa l Health,* pp. 458-465.

McGann, P. T. et al., 2020. Genetic Causes of Anaemia in Malawian Children Less Than 5 Years of Age: Results from the Malawi Demographic and Health Survey. *American Society of Hematology,* pp. 1-3.

Mesfioui, M., Bouezmarni, T. & Belalia, M., 2022. Copula-based link functions in binary regression models. *Stat Papers ,* pp. 1-10.

Monteiro, R., Fernandes, E., Monteiro, R. & Ribeiro, L. &. S. D., 2013. Prevalence of nutritional deficits among children under five years of age in Angola. *Rev Saude Publica,* pp. 1-4.

Muchie, K., 2016. Determinants of severity levels of anemia among children aged 6-59 months in Ethiopia: further analysis of the 2011 Ethiopia Demographic and Health Survey. *BMC Nutrition,* pp. 1-8.

Muchie, K. F., 2016. Determinants of severity levels of anemia among children aged 6-59 months in Ethiopia: further analysis of the 2011 Ethiopia Demographic and Health Survey. *BMC Nutrition,* pp. 1-8.

National Agency for Statistics and Demography - ANSD / Senegal, and ICF, 2018. *Senegal: Continuous Demographic and Health Survey - EDS-Continue 2017,* USA: ANSD and ICF: Rockville, Maryland.

Ncogo, P. et al., 2017. Prevalence of anemia and associated factors in children living in urban and rural settings from Ba. *PlosOne,* pp. 173-181.

Neuman, M., Kawachi, I., Gortmaker, S. & Subramanian, S. V., 2013. Urban-rural di erences in BMI in low- and middle-income countries: the role of socioeconomic status. *Am J Clin Nutr,* pp. 428-436.

Ngwira, A. & Kazembe, L. N., 2016. Analysis of severity of childhood anemia in Malawi: a Bayesian ordered categories model.. *Open Access Medical Statistics,* pp. 9-20.

Nkurunziza, S., Meessen, B., Van geertruyde, J. & Korachias, C., 2017. Determinants of stunting and severe stunting among Burundian children aged 6–23 months: Evidence from a national cross-sectional household survey. *BMC Pediatr,* Volume 17, pp. 1-14.

Ntenda, P. A. M., Chaung, K. Y., Tiruneh, F. N. & Chaung, Y. C., 2004. Multilevel analysis of the e ects of individual and community level facotrs on childhood anemia, severe anemia and homoglobin conentration in Malawi. *Journal of Tropical Pediatrics,* pp. 267-278.

Penny, W. & Holmes, A., 2003. Random effects analysis. *Academic Press,* p. 843–850.

Poda, G. G., Hsu, C. Y. & Chao, J. C., 2017. Factors associated with malnutrition among children <5 years old in Burkina Faso: evidence from the Demographic and Health Surveys IV 2010. *Int J Qual Health Care,* pp. 901-908.

Poda, G., Hsu, C. & Chao, J., 2017. Factors associated with malnutrition among children <5 years old in Burkina Faso: evidence from the Demographic and Health Surveys IV 2010.. *Int J Qual Health Care,* pp. 901-908.

Ramroop, S. & Khulu, C., 2020. Key Determinants of Anaemia among Youngsters under Five Years in Senegal, Malawi, and Angola. *Int. J. Environ. Res. Public Health,* pp. 1-12.

Ratner, B., 2010. Variable selection methods in regression: Ignorable problem, outing notable solution.. *Journal of Targeting, Measurement and Analysis for Marketing,* Volume 18, pp. 65-75.

Ray, S. et al., 2016. Determinants of nutritional anemia in children less that ve years age. *International Journal of Contemporary Pediatrics,* pp. 403-408.

Richardson, M., 2009. *Principal Component Analysis.* [Online]
Available at: http://people.maths.ox.ac.uk/richardsonm/SignalProcPCA.pdf.

Roberts, D. J. & Zewotir, T., 2020. Copula geaoadditive modelling of anaemia and malaria in younf children in Kenya, Malawi, Tanzania and Uganda. *Journal of health, population and nutrition,* Volume 39, pp. 1-14.

Roberts, D. J. & Zewotir, T., 2020. Copula geoadditive modelling of anaemia and malaria in young children in Kenya, Malawi, Tanzania and Uganda. *Journal of Health, Population and Nutrition ,* pp. 1-14.

Rosario, E. V. N., Costa, D., Francisco, D. & Brito, M., 2004. DHSS pro le: The Dande Health and Demographic Surveillance System (Dande HDSS, Angola). *International Journal of Epidemiology,* pp. 1094a-21094g.

Samanpreet Brar, S. et al., 2020. Drivers of stunting reduction in Senegal: a country case study. *Am J Clin Nutr,* pp. s860-s874.

Sartorelli, D., Fernandes, E. & Castro, T., 2017. Associated factors of malnutrition among African children under five years old,Bom Jesus, Angola.. *Rev. Nutr,* pp. 33-44.

Sauerbrei, W., Royston, P. & Binder, H., 2007. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med,* Volume 26, pp. 5512-5528.

Schumacker, R. E. & Lomax, R. G., 2004. *A beginner's guide to structural equation modeling, psychology press.,* s.l.: s.n.

Semedo, R. et al., 2014. Prevalence of Anaemia and Associated Factors among Children below Five Years of Age in Cape Verde, West Africa. *J health Popul Nutr.,* pp. 646-657.

Siciliano, S. et al., 2014. Soil fertility is associated with fungal and bacterial richness, whereas pH is associated with community composition in polar soil microbial communities. *Soil Biol Biochem,* Volume 78, p. 10–20.

Skrondal, A. & Rabe-Hesketh, S., 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* s.l.:Chapman and Hall/CRC.

Snijders, T., 2005. Fixed and random effects. In: Everitt BS, Howell DC (eds) Encyclopedia of statistics in behavioral science. *John Wiley & Sons, Inc.,* p. 664–665.

Soares, A. et al., 2020. Iron deficiency anaemia among 6-to-36 month children from northern Angola.. *BMC Pediatrics,* pp. 1-13.

Sop, S. M. K., Mananga, M. J., Tetanye, E. & Gouado, I., 2015. Risk factors of anemia among young children in rural Cameron.. *International Journal of Current Microbiology and Applied Sciences,* pp. 925-935.

Sousa-Figueiredo, J. C. et al., 2012. Epidemiology of malaria, schistosomiasis, geohelminths, anemia and malnutrition in the context of a Demographic s. *Plos one,* pp. 1-9.

Steiger, J. H., 1990. Structural model evaluation and modification: An interval estimation approach. *Multivariate behavioral research,* pp. 173-180.

Steyerberg, E. W., 2008. *Clinical prediction models: a practical approach to development, validation, and updating.* s.l.:Springer Science & Business Media.

Steyerberg, E. W. & Vergouwe, Y., 2014. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J,* Volume 35, pp. 1925-1931.

Subramanian, S. V., Perkins, J. M., Ozaltin, E. & Smith, G. D., 2011. Weight of nations: a socioeconomic analysis of women in low-to middle-income countries. *Am J Clin Nutr,* pp. 413-421..

Tabachnick, B. G. & Fidell, L. S., 2007. *Using Multivariate Statistics.* Boston, MA: Pearson Education Inc..

Takele K., Zewotir T. & D., N., 2019. Understanding correlates of child stunting in Ethiopia using generalized linear mixed models.. *BMC Public Health,* pp. 626-636.

Takele, K., Zewotir, T. & Ndanguza, D., 2019. Understanding correlates of child stunting in Ethiopia using generalized linear mixed models. *BMC Public Health,* pp. 1-8.

Talukder, A., 2017. Factors associated with malnutrition among under five children: Illustration using Bangladesh Demographic and Health Survey, 2014 data. *MDPI,,* pp. 1-8.

Tenda, P. A. M., Chaung, K. & Tiruneh, F. N. &. C. Y., 2017. Multilevel Analysis of the Effects of Individual and Community-Level Factors on Childhood Anaemia, Severe Anaemia, and Hemoglobin Concentration in Malawi. *Journal of Tropical Pediatrics,* pp. 267-278.

Tibshirani, R. J., Taylor, J., Lockhart, R. & Tibshirani, R., 2017. Exact post-selection inference for sequential regression procedures. *Journal of the American statistical association,* Volume 111, pp. 600-620.

Tine, R. C. K. N. M. H. H. H. et al., 2012. The association between malaria parasitaemia, erythrocyte polymorphisms, malnutrition and anaemia in children less than 10 years in Senegal: a case control study. *BMC Research,* pp. 1-10.

Torres-Reyna, O., 2007. *Panel Data Analysis Fixed and Random Effects Using Stata (v. 4.2),* Priceton University: Data & Statistical Services.

UNICEF, 2019. *Global annual results reports 2018,* s.l.: s.n.

United Nations Children's Fund, World Health Organization, World Bank Group , 2018. *Levels and trends in child malnutrition: Key findings of the 2018 Edition of the Joint Child Malnutrition Estimates,* s.l.: s.n.

USAID, 2018. *Senegal: Nutrition Profile.* [Online]
Available at: https://www.usaid.gov/sites/default/files/documents/1864/Senegal-Nutrition-Profile-Mar2018-508.pdf

Von Grebmer, K. et al., 2019. *2019 Global Hunger Index: The Challenge of Hunger and Climate Change,* s.l.: Bonn: Welthungerhilfe; and Dublin: Concern Worldwide.

Vyas, S. & Kumaranayake, L., 2006. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy and Planning,* pp. 459-468.

Wand, H., Whitaker, C. & Ramjee, G., 2011. Geoadditive models to assess spatial variation of HIV infections among women in Local communities of Durban, South Africa. *International Journal of Health Geographics,* pp. 1-9.

Wang, C. et al., 2020. DNA methylation of insulin-like growthfactor 2 and H19 cluster in cord blood andprenatal air pollution exposure to fineparticulate matter. *Environmental Health,* pp. 1-12.

Wirth, J. P. et al., 2017. Assessment of the WHO Stunting Framework using Ethiopia as a case study. *Matern Child Nutr.,* Volume 13, pp. 1-10.

Wirth, J. et al., 2016. Assessment of the WHO Stunting Framework using Ethiopia as a case study. *Maternal & Child Nutrition,* pp. 1-16.

Woldie, H., Kebede, Y. & Tariku, A., 2014. Factors associated with Anemia among children aged 6-23 months attending growth monitoring at Tsitsika health center, Wag-Himra Zone, Northern Ethiopia. *Journal of Nutrition and Metabolism,* pp. 1-9.

Wolfinger, R., 1993. Covariance structure selection in general mixed models. *Commun. statist.-simula,* pp. 1079-1106.

World Food Program, 2017. *Senega.* [Online]
Available at: http://www1.wfp.org/countries/senegal

World Health Organization and World Bank Group, 2015. *World health statistics,* Washington, DC: s.n.

World Health Organization and World Bank Group, 2017. *Levels and trends in child malnutrition,* Washington, DC: s.n.

World Health Organization and World Bank Group, 2021. *World health statitsics,* Washington, DC: s.n.

World Health Organization, 2015. *Global Health Observatory Data Repository/World Health Statistics. ,* Geneva, Switzerland 2015: s.n.

World Health Organization, 2015. *Prevalence of Anaemia in 2011,* Geneva, Switzerland : s.n.

World Vision Angola and Partners, 2014. *Community-based Management of Acute Malnutrition Using Community Health Activists in Angola,* s.l.: s.n.

Zewotir, T., Habyarimana, F. & Ramroop, S., 2016. Joint modelling 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,* pp. 108-114.

Zhoa, A. et al., 2012. Prevalence of anemia and its risk factors among children 6-36 months old in Burma. *The American journal of tropical medicine and hygiene,* pp. 306-311.