MAKERERE UNIVERSITY

COLLEGE OF HEALTH SCIENCES

SCHOOL OF PUBLIC HEALTH

COMPARING PERFORMANCE OF GENERALIZED LINEAR MIXED MODEL AND GENERALIZED ESTIMATING EQUATIONS IN MODELLING UNDER-FIVE MORTALITY: A CASE STUDY OF IGANGA-MAYUGE DISTRICTS UGANDA.

BY

TRYPHENA NAREEBA

2020/HD07/19450U

SUPERVISORS:

Prof. Nazarius Mbona Tumwesigye

Dr. Vincent Micheal Kiberu

A dissertation submitted to Makerere University School of Public Health in partial fulfilment of the requirements for the award of the Master of Biostatistics of Makerere University, Kampala

DECLARATION

I hereby declare that this research proposal is my original work and has never been submitted to any academic or non-academic institution for the award of Master of Biostatistics or any other Degree. Where other sources of information have been used, they have been correctly acknowledged by referencing.

TRYPHENA NAREEBA

Signature: … …….…………….

Date: …1st September 2023………………………….

APPROVAL

This dissertation has been written under the supervision of Prof. Nazarius Mbona Tumwesigye and Dr. Vincent Micheal Kiberu.

Prof. Nazarius Mbona Tumwesigye

Signature: ....

Date: 01-09-2023

Dr. Vincent Micheal Kiberu

ru

Signature: ....

Date: .... 06/Sept/2023 ....

DEDICATION

I dedicate this work to my husband, parents and my fellow biostatisticians and mentors in this field. You were truly inspirational to me.

ACKNOWLEDGEMENT

I thank the Almighty God who has given me the strength and wisdom to get this work done, I also thank my supervisors who have mentored me and guided me through this research. Dr. Dan Kajungu who availed me with the data from Makerere University Center for Health and Population Research (MUCHAP) and my course mates, from whom we have always encouraged each other in working hard.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

GEE: Generalized Estimating Equations

OR: Odds Ratio

CI: Confidence Interval

SDG: Sustainable Development Goal

MDGs: Millennium Development Goals

GLMs: Generalized Linear Models

MAR: Missing at Random

MCAR: Missing completely at Random

MNHR: Multi-national Global Network for Women's and Children's Health Research Maternal and Neonatal Health Registry

GLMMs: Generalized Linear Mixed Models

AIC: Akaike's information criteria

QIC: Quasi-information criterion

IMHDSS: Iganga-Mayuge Health and Demographic Surveillance Site

PCA: Principal component Analysis

TBA: Traditional birth attendant

RAs: Research Assistants

HDSS: Health and Demographic Surveillance Site

HRS: Household Registration System

MakSPH: Makerere School of Public Health

IRB: Institutional Regulatory Board

UNCST: Uganda National Council of Science and Technology

ABSTRACT

Introduction

Reduction of childhood mortality is a global health priority and an indicator of child survival. Over 5 million in 2021 children under five years died every day from preventable and curable diseases which is undesirable. Between 1990 and 2020, the global mortality rate for children under five years declined by 61 per cent, from 93 deaths per 1,000 live births to 37 deaths per 1,000 live births. Over the last three decades, under-five mortality has steadily been declining in Uganda. Between 1990 and 2020, the rate declined by more than 70%, from 183 to 43.3 deaths per 1,000 live births. This study aimed at determining the association between the predictor variables and under-five mortality rates in the Iganga and Mayuge districts.

Methods

Data from Iganga-Mayuge Health and Demographic Surveillance Site (IMHDSS) of 2010 to 2015 collected on women of reproductive age (13 to 49 years) were used. The outcome was death before the fifth birthday and the independent variables were gender, place of residence, birth order, mother's age, education, wealth index, marital status, place of delivery, attend antenatal care, occupation and year. Village was used as a random effect. For descriptive analysis, proportions were used. Means and standard deviation were used for continuous variables. The explanatory variables were checked for multicollinearity to ensure validity for further analysis. AIC and QIC were used for model selection. Data management and analysis was done using excel and STATA 18.

Results

Between 2010 and 2015, 2011 had the highest number of under-five deaths. Mothers from both the rural and Peri_urban areas had an average of one child. The odds of dying before the age of five reduced among the children whose mothers were cohabiting and those whose highest level of education was below primary. Being male, residing in the rural areas and children whose mothers were teenagers increased the odds of dying before the age of five.

Conclusion

The choice between these methods should be guided by the underlying data structure, assumptions, research objectives, and the practical significance of the differences observed. Interventions like healthcare access could improve maternal and child healthcare services, enhance education, thus improving future research, guide policy development aimed at improving overall child well-being and reduce under-five mortality in Iganga and Mayuge districts.

CHAPTER ONE: INTRODUCTION AND BACKGROUND

1.1 Introduction

Reduction of childhood mortality is a global health priority and an indicator of child survival. Over 5 million in 2021 children under five years died every day from preventable and curable diseases which is undesirable. Having a good life after celebrating five years of age is a right for every child and therefore a wealth of a country. Between 1990 and 2020, the global mortality rate for children under five years declined by 61 per cent, from 93 deaths per 1,000 live births to 37 deaths per 1,000 live births (UNICEF, 2021a, 2021c).

However, significant disparities still exist in childhood mortality across country income levels. In 2020, the mortality rate among children under five years old in low-income countries was 61 deaths per 1,000 live births, more than 14 times the average in high-income countries (Europe and Northern America). In the highest mortality country, the risk of a child dying before 60 months was 65 times higher than in the lowest mortality country (Cha & Jin, 2019; Chao et al., 2018; UNICEF, 2021c). The highest under-five mortality rates are in the sub-Saharan Africa with 2.7 million children and Southern Asia with 1.4 million children dying before their fifth birthday. Nigeria, India, Pakistan, the Democratic Republic of the Congo and Ethiopia contributed to half of all the death among children aged below 60 months in 2020 (Kitui et al., 2013; UNICEF, 2021c).

Ending preventable deaths of newborn and children under 5 years of age by 2030 Sustainable Development Goal (SDG) 3 is a target for all countries (UNICEF, 2020). Under-five mortality rates in 69 countries (mostly low-income) are currently higher than the Sustainable Development Goal target of 25 deaths per 1,000 live births (Das et al., 2021; UNICEF, 2020). At current rates of progress, only 16 of the 195 countries are set to achieve the SDG target by 2030 and the 54 countries are to meet the target on time if the pace of mortality decline quickens (UNICEF, 2021c). Among the East African countries, Uganda is among the few countries in the sub-Saharan Africa whose under-five mortality rates have consistently reduced and have that met the Millennium Development Goal (MDGs) 4 target (Agho et al., 2020; MoFPED, 2015).

Over the last three decades, under-five mortality has steadily been declining in Uganda. Between 1990 and 2020, the rate declined by more than 70%, from 183 to 43.3 deaths per 1,000 live births (Hirose et al., 2020; UNICEF, 2021c).

1.2 Background

Under-five mortality rate that measures child survival is the probability per 1,000 that a new born baby will die before reaching age five, if subject to age-specific mortality rates of the specified year (UNICEF, 2021b). Although there has been a decline in the under-five mortality rate globally, Uganda's under-five mortality rates are still high at 43 deaths per 1000 live births (Nambuusi et al., 2019; UNICEF, 2021a). Therefore, understanding the determinants of the under-five mortality rates is important in a way that it will end preventable diseases and hence attaining the SDG 3.2 (UNICEF, 2020).

Different factors affect the wellbeing of children aged 0 to 59 months leading to their death before celebrating their fifth birthday. In Uganda, more than half of the children aged below 60 months that died in 2015 died of communicable diseases like Pneumonia, Malaria and diarrhoea. Among these, malaria accounted for the largest proportion of these mortalities (Awor et al., 2018; UNICEF, 2021b). Despite the implementation of various government strategies like immunisation, integrated community case management of childhood illnesses, micronutrient supplementation, prevention of mother-to-child transmission of HIV among others aimed to reduce under-five mortality in Uganda, the eastern region's under-five mortality rate is still higher than the national average (Awor et al., 2018).

Generalized Estimating Equations (GEE) is a statistical approach based on the quasi-likelihood function used to fit a marginal model for clustered data that cannot be correctly analysed using Generalized Linear Models (GLMs) in clinical trials and biomedical studies (Awor et al., 2018). GLMs do not take into account correlation leading to the regression estimates (Betas) being less efficient. This implies that the Betas would be more widely scattered around the true population value (Ballinger, 2004; Hubbard et al., 2010). Unlike GLMs, GEE method is based on few assumptions and directly generates population level parameters. It also incorporates within-subject and between-subject variations into model fitting hence improves efficiency of the estimation and the power of the study (Zorn, 2001). Majority of the clinical trials and biomedical studies have commonly used GLMMs on data that might be clustered, other than GEE while analysing longitudinal data that this study therefore seeks to perform a risk factor analysis to identify determinants of under-five survival in the Iganga and Mayuge districts using GEE.

CHAPTER TWO: LITERATURE REVIEW

2.1 Under-five mortality

Understanding under-five mortality rate is important in a way that it is an indicator of a child's wellbeing which is an indicator for SDG 3: "Ensure healthy lives and promote well-being for all at all ages" and is a proxy of social and economic development. Globally, majority of the death of children below 60 months is due to preventable diseases with communicable diseases taking the lead (Exemplars, 2020; UNICEF, 2015).

Pneumonia, Malaria and Diarrhoea are the common causes of death among children below the age of five years. These deaths mainly occurred in malnourished children more particularly among those with acute malnutrition (World Health Organization, 2022). Malaria is the leading cause of death and diarrhoea attributing to increase in number of deaths (Kantorova et al., 2021; Nambuusi et al., 2019).

Maternal factors, nutrition related factors and socio-economic factors are important to study because they impact the child's survival (World Health Organization, 2022). With this knowledge, better intervention strategies are applied and better policies are developed which in return reduce the morbidity and mortality among children below five years.

2.1.1 Determinants of Under-five mortality

Place of residence is associated with the mortality of the children. In line with other studies, according to Hirose et.al (Hirose et al., 2020) more of the deaths of children under the age of five were from the rural areas as compared to the Peri-urban areas (Ahinkorah et al., 2022).

Households' income is one of the commonly identified social determinants of health. Previous studies have established an inverse relationship between household income and under-five mortality. Children from poorer households face a higher mortality rate compared to the children from rich households (Ahinkorah et al., 2022; UNICEF, 2021c).

Nasejje JB et.al (Nasejje et al., 2015) states that mothers who had their highest level of education as primary level and below had a higher under-five mortality rate compared to those whose education level was above primary level of education. Better maternal education increases the under-five survival. Educated mothers are more likely to have better income, and higher health literacy levels (Abir et al., 2017).

Among other studies, Gebremichael et.al (Gebremichael & Fenta, 2020) found out that women that didn't have jobs or carrying out any kind of occupation experienced more of the under-five mortality rates unlike women that were carrying out a kind of occupation (Woldeamanuel, 2019).

Previous studies that assessed the impact of father's involvement on child development, functioning and quality of life, showed that children born to a mother whose partner was absent had an increased risk of dying as compared to that woman that had a partner (Sartorius et al., 2011).

Worku et.al (Worku et al., 2021) among other scholars found out that children that are born to mothers whose previous birth interval of less than two years were more prone to death because short birth intervals increase the risk of preterm birth, under-nutrition and low birth weight (Islam et al., 2022).

In other studies, it is indicated that under-five male children have a better survival in the neonatal period compared to girls. This may be explained by biological reasons other than social and cultural factors (Khadka et al., 2015; Nisar & Dibley, 2014).

Abir et.al (Abir et al., 2017) found out that young mothers that is those that were below the age of 24 years had majority of their children die before celebrating five years. This was the same with what other scholars found out (Mani et al., 2012).

It is justified that child mortality is associated with the mother attending antenatal care. Children born to mothers who don't attend antenatal care are likely to die before celebrating their fifth birthday compared to those born to mothers who attend antenatal care. These miss out on the education and supplements that are availed during the different antenatal visits (Abir et al., 2017; Oduse et al., 2021).

Filippi V RC et.al (Filippi et al., 2006) mentioned that outside health facility deliveries have been associated with high under-five mortality. This could be associated with limited access to complete immunization and postnatal care that would improve the health of the child and hence reduce morbidities and mortality among the children born outside the health facility.

2.2 Description of Generalized Estimating equations and Generalized Linear Mixed Model

Some children in the dataset used shared a mother and since the mother had the same information across the different years 2010 to 2015, this led to repeated measurements and correlation with in the mothers and therefore statistical approaches GEE and GLMM were applied in analysing the data in order to make correct inferences.

GEE is a statistical approach used to fit a marginal model basing on quasi-likelihood function and produces parameter estimates with correct standard errors for marginal models of the exponential class like Gaussian, Gamma, Poisson, binomial (Ghisletta & Spini, 2004; Ditlhong et al., 2018).

A marginal model is one with fixed effects only and population parameters are those that are estimated across all subjects (Wang, 2014; Zorn, 2001).

GEE model accounts for the correlation between observations in generalized linear regression models by use of empirical (sandwich/robust) variance estimator.

The quasi-likelihood under independence criterion' (QIC) introduced by Pan (2001) for variable selection and the Wald-Wolfowitz run test (Chang (2000)) are used to assess the goodness-of-fit (Pan, 2002).

Generalized Linear mixed model (GLMM) is a conditional mean regression model whose estimates are obtained from a likelihood function. GLMM model combines both Generalized Linear Models (GLMs) and Linear Mixed Effects Models to handle data with non-normal distributions and account for correlated or clustered data. It is conditioned on both the fixed design matrix and the random effects. Correlation arises among repeated observations within a given cluster because of the shared random effects yet the measurements are considered to be conditionally independent given the random effects (Muff et al., 2016).

2.2.1 Specifications/Notations

Model parameters of GEE model are estimated from the extension of GLM approach by incorporating correlations.

Suppose the data consists of N subjects. For subject $i,\{i=1,2,\cdots,N\}$, with T observations (repeated measurements) $t=\{1,2,...,T\}$ and let Y_{ij} denote the j^{th} response $\{j=1,2,\cdots,n_i\}$, and let x_{ij} denote the $p\times 1$ vector of explanatory variables.

Suppose $y_i=[y_{i1},y_{i2},\cdots,y_{in_i}]'$ denote the corresponding column vector of response variable for the i^{th} subject with the mean vector $\mu_i=[\mu_{i1},\mu_{i2},\cdots,\mu_{iT}]$ where μ_{ij} is the corresponding j^{th} mean.

The marginal model specifies that a relationship between E (y_{ij}) =μ_{ij} and the covariates x_{ij} is as follows:

$y_{ij}|x_{ij}$ ~ Bernoulli (μ_{ij}) and $\log\left[\frac{\mu ij}{1-\mu ij}\right]$ =$x^{T}_{ij}\beta$…………………………………. (1)

where $\log\left[\frac{\mu ij}{1-\mu ij}\right]$ is a link function that identifies the logit function and β is the vector of regression coefficients.

The variance of y_{ij} is specified as;

$$\text{var}\ (y_{ij}) = \phi\ V\ (\mu_{ij}) \quad \text{...(2)}$$

where V is a known variance function of μ_{ij} and ϕ is a scale parameter which may need to be estimated. V and ϕ mostly depend on the distributions of outcomes.

The regression coefficient estimates, β are defined by the solution of GEE as;

$$\sum_{i=1}^{N} \frac{d\mu_i}{d\beta^T} V_i^{-1}(Y_i - \mu_i)=0 \quad \text{... (3)}$$

The covariance matrix for y_i is noted by;

$$V_i = \phi A^{1/2} R_i(\alpha)\ A_i^{1/2} \quad \text{... (4)}$$

where A_i is a diagonal matrix of variance functions, and the so-called "working" correlation structure $R_i(\alpha)$ describes the pattern of measures within subject, which is of size $n_i \times n_i$ and depends on a vector of association parameters denoted by α (Liang & Zeger, 1986; Wang, 2014).

For GLMM model, let y be the binary outcome (child living status) from n_i children from N mothers and x be the matrix of fixed effect predictors, β is the vector of fixed effect coefficients. The fixed-effects component of the model is given as;

$$g\ (\mu) = x\beta \quad \text{...(1)}$$

where g (.) is the link function and μ is the mean of the response variable.

Including random effects in the model. Let Z be the matrix of random-effect predictors, b be the vector of random effects, D be the variance-covariance matrix of the random effects. The random-effects component of the model is expressed as;

$$b \sim N\ (0, D) \quad \text{... (2)}$$

The combined model is given as;

$$g\ (\mu) = x\beta + Zb \quad \text{... (3)}$$

Since it's a binary outcome the likelihood function follows a Bernoulli distribution and the link function being the logistic function (logit link);

P(Y=1) $= \frac{1}{1 + e^{x\beta + zb}}$

2.2.2 Assumptions of GEE ad GLMM

In the presence of missing data, GEE model is only valid under the strong assumption of missing completely at random (MCAR) (Ditlhong et al., 2018).

The working correlation structure that represents the assumed correlation between observations within the same cluster under GEE model has to be specified.

The variance of the response variable is assumed to be constant within each cluster across different levels of the predictors.

GLMM model estimates are subject-specific. The inferences are valid if the assumptions are met and it also assumes missing data is missing at random (MAR).

The GLMM model assumes a linear relationship of the mean with the covariates and the random effects $b_{i.}$

GLMM model depends on parametric assumptions to model the random effects and sampling variability. If either of the assumptions like normality of random effects is violated, they provide biased estimates (Kain et al., 2015).

GLMM model assumes the residuals to have a constant variance or in some cases, exhibit a pattern of heteroscedasticity that can be accounted for and the residuals should follow the distribution assumed by the chosen GLM.

2.2.3 Advantages of GEE and GLMM

The main advantage of the GEE model is that one is only required to specify correctly the mean structure of the response for the parameter estimator to be consistent and asymptotically normal (Ditlhong et al., 2018).

The variance function is potential aspect affecting the goodness-of-fit of the GEE model and therefore correctly specified variance function can assist in the selection of covariates and an appropriate correlation structure.

GEE model is capable of providing asymptotically unbiased parameter estimates of primary interest even where the exact nature of intracluster independence is not known (Zorn, 2001).

GEE model estimator has a robustness property, this provides a consistent estimator of β even if there is misspecification of the within-subject associations among the repeated measures (Wang, 2014).

GEE model relies on the mean and variance relationships specified by the chosen GLM and the link function and therefore does not require assumptions about the distribution of the response variable.

GEE model provides inferences at the population level and computationally efficient at handling relatively large data sets without excessive computational burden.

Unlike the GEE model, GLMMs can handle complex designs like nested groups.

GLMMs can model a wide range of response variables beyond normally distributed data like count, binary among others.

GLMMs use likelihood estimation methods and therefore good at simulations hence giving insights into both general trends and cluster-specific variations (Kain et al., 2015).

GLMMs incorporate in random effects that capture individual-level variability that is not explained by fixed predictors.

2.2.4 Uses of Generalized Estimating Equations

To evaluate their primary outcome of death or neurodevelopmental impairment, Juul et al (Juul et al., 2020) used GEE with robust standard errors to provide valid statistical inference and to fully account for the potential correlation of outcomes in siblings from the same pregnancy and for comparisons between groups to account for potential correlation within siblings from multiple gestations. They found no significant difference between groups in the primary outcome of death or severe neurodevelopmental impairment at 2 years of age.

Piecewise Generalized estimating equations have been used in determining to obtain adjusted relative risks with 95% confidence intervals in a study by Vaucher et al (Vaucher et al., 2012) that was assessing whether early, non-invasive CPAP with a limited ventilation strategy, as compared with early surfactant administration, and a lower, as compared with higher, target range of oxygen saturation would each decrease the incidence of death before assessment at 18 to 22 months or neurodevelopmental impairment at 18 to 22 months of corrected age. This study found out that mortality did not differ significantly between the CPAP and surfactant groups, and mortality remained significantly higher in the lower-oxygen-saturation group than in the higher-oxygen-saturation group (Vaucher et al., 2012).

Hierarchically, GEE models with an identity link and an unstructured correlation were to account for the correlations between 3 follow-up times within subjects. The goal of these analyses was to examine whether the PN-LCNS program improved HRQOL for each cancer type at 3 follow-up times (cancer groups female colorectal cancer, male colorectal cancer, breast cancer, and prostate cancer) PN-LCNS demonstrated a significant improvement in HRQOL in comparison with PN only for colorectal cancer survivors but not for breast and prostate cancer survivors (Ramirez et al., 2020) .

A GEE approach for Poisson regression was used to compare age-specific rates of severe malaria and high-density infection. Bronner et.al found out that the incidence of severe malaria decreased considerably after infancy, whereas the incidence of high-density infection was similar among all age groups. Infections before and after episodes of severe malaria were associated with similar parasite densities (Gonçalves et al., 2014).

Panahi et.al (Panahi et al., 2022) used GEE to evaluate the effects of time and groups on the outcomes. In their study they assessed the effectiveness of a fathers' educational program on their support for breastfeeding, mothers' breastfeeding practice, and exclusive breastfeeding status and found out that the results showed the interaction effects of time and group were significant in the GEE test for the fathers' support for breastfeeding and the mothers' breast-feeding practice.

While fitting a Poisson model with generalized estimating equations (GEEs), Grantz et.al (Grantz et al., 2016) estimated the association between social factors and the risk of influenza and pneumonia mortality. The Poisson model showed that influenza and pneumonia mortality increased, on average, by 32.2% for every 10% increase in illiteracy rate adjusted for population density, homeownership, unemployment, and age.

Bauserman et.al (Bauserman et al., 2015) whose objective was to describe maternal mortality rates in a large, multicountry dataset from women enrolled in the multinational Global Network for Women's and Children's Health Research Maternal and Neonatal Health Registry (MNHR) while using GEE found evaluated social factors and measures of care and pregnancy complications would influence maternal death.

Using data from the Stockton-On-Tees prospective cohort study, Akhter et.al (Akhter et al., 2021) used GEE to assess the associations between behaviours and the explanatory factors. All health behaviours, except for frequent physical activity, varied significantly by deprivation. Material factors (like employment, education and housing tenure) were the most-important and environmental factors the least-important explanatory factors.

CHAPTER THREE: PROBLEM STATEMENT, JUSTIFICATION AND CONCEPTUAL FRAMEWORK

3.1 Statement of the problem

Under-five mortality rate in eastern Uganda is 84 deaths per 1,000 live births (UBOS, 2016) that is higher than the national average. Vast scholars over the years have used traditional methods like GLMs to determine factors associated with under-five mortality for longitudinal data i.e data that is measured repeatedly over time of the same subject but these methods violate the assumption of independence, assume errors to be normally distributed, and homoscedasticity. These methodologies lead to production of biased estimates hence underestimation or overestimation the mortality rates.

It is therefore of importance to carry out a risk factor analysis using more sound statistical methods like GEE that produce population-averaged estimates which produce asymptotically correct estimates and inferences if model is correctly specified instead of subject-specific parameter estimates (Ghisletta & Spini, 2004; Pekár & Brabec, 2018). This will eventually better inform the best strategies of intervention which will improve on the child wellbeing and thus attain the SDG3 (UNICEF, 2020).

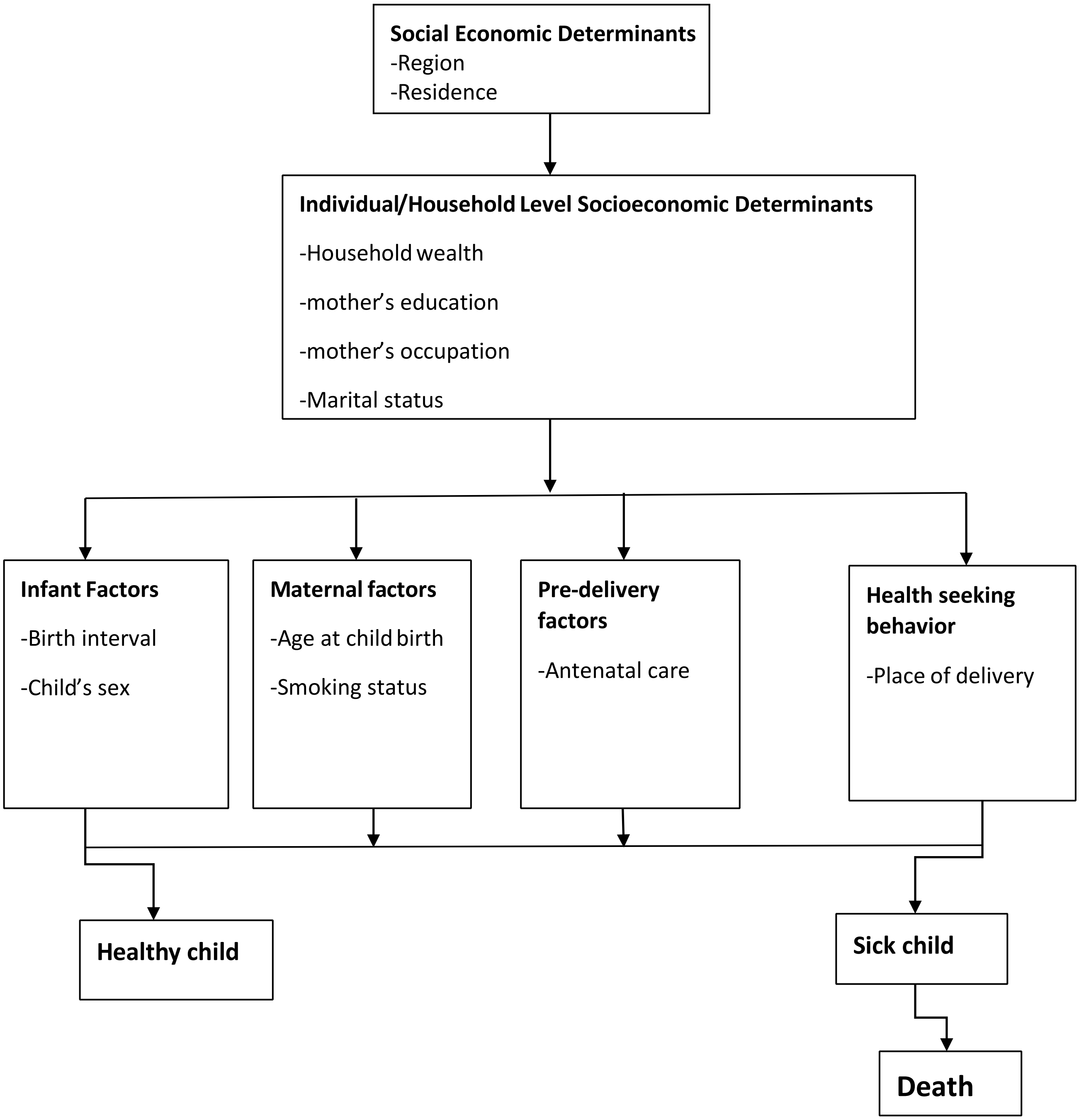
3.2 Justification of the problem

Globally, Uganda still contributes to the largest mortality rates and high number of children dying before celebrating their fifth birthday. Under-five mortality rate is often used as a proxy indicator of health and social economic development and thus a much concern for policy makers in the health system. Prevention of mortality and morbidity of infectious diseases benefits people of Uganda socially and economically.

Mortality rates generated using GEE methods will better explain the different factors that lead to the death of these children. These will hence help educate households about the benefits of healthcare, improve healthcare, enlight on lengthening birth intervals and help strengthen the health system thus proper measures applied to avert these preventable diseases. Proper measures could be carrying out immunisation programmes, promoting health seeking behaviours and advocating for child deliveries in health facilities. This will further reduce the under-five mortality in eastern Uganda (Ekholuenetale et al., 2020; UNICEF, 2021c).

3.3 Conceptual Framework

Figure 1: Conceptual framework



Adopted framework: The associated factors for under-five mortality were categorised into three categories based on Mosley and Chen's (1984) theoretical framework as social economic determinants, individual/household level socioeconomic determinants, infant factors, maternal factors, predelivery factors and health seeking behaviour (Chen, 1984)

CHAPTER FOUR: RESEARCH QUESTIONS, HYPOTHESES AND OBJECTIVES

4.1 Research questions

i. What is the difference between the model efficiency of GEE and GLMM while determining under-five mortality rates?
ii. What is the association between the predictor variables and under-five mortality rates?

4.2 Study hypotheses

Below are the study hypothesis

H_0: There is no association between the woman's age and under-five mortality

H_0: There is no association between marital status and under-five mortality

H_0: There is no association between wealth index and under-five mortality

H_0: There is no association between level of education and under-five mortality

H_0: There is no association between place of delivery and under-five mortality

H_0: There is no association between the gender of the child and under-five mortality

4.3 General Objective

To identify the determinants of under-five mortality rate in IMHDSS.

4.4 Specific Objectives

i. To compare the model efficiency of GEE and GLMM while determining under-five mortality rate.
ii. To assess the associations of predictor variables with under-five mortality rate using GEE.

CHAPTER FIVE: METHODOLOGY

5.1 Study site

This study used data from Iganga-Mayuge Health and Demographic Surveillance Site in eastern Uganda. The area consists of 65 villages in seven sub-counties spread over a 155km^2 area. The IMHDSS carries out household census surveys biannually. Since 2005, IMHDSS has collected longitudinal data on demographic events of births, deaths and migrations. In addition to these core indicators collected twice a year, information is collected on the educational status of all persons of school age also. Information on socio-economic status of households is updated every two years (Kajungu et al., 2020).

5.2 Study Population

This study included children below the age of five. These children were from all eligible women aged 13 to 49years living in the active households of IMHDSS.

5.3 Study Design

This study was a retrospective cohort utilizing secondary data from IMHDSS datasets for the period 2010 to 2015.

5.4 Sample Size

The sample was 1,156 children using the level of significance of 5% and power of 90%. Total sample size with attrition of 10% was 1,272 children

$$n = \frac{\left(Z_{1-a/2} + Z_{1-\beta}\right)^2 * [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

5.5 Sampling procedure

Data on randomly selected households containing women aged between 13years to 49years of the period 2010 to 2015 was considered.

The women selected had a record of ever given birth and whose birth outcome was a live birth.

The data on their children who died before their fifth birthdays were extracted from the routine household census surveys for the respective years 2010 to 2015.

5.6 Study Variables

5.6.1 Dependent variable

The dependent variable child living status defined as death before the fifth birthday was coded as a binary variable with being alive coded as 0 and dead coded as 1.

5.6.2 Independent variables

The variables considered include; place of residence, wealth index, highest level of education, occupation, mother's marital status, birth order, gender of the child, mother's age, attending antenatal care, place of delivery and year of collection.

Woman's age was be categorized, into two groups of 13 to 24 years as teenage mothers and 25 to 49 years described as non-teenage mothers.

Place of delivery is a place where the child was delivered from. It was classified as hospital/health facility, Home/on the way to clinic, TBA (Traditional birth attendants) and clinic.

Highest level of education attained was categorized as below primary level and above primary level of education, primary level is equivalent to seven years of education.

Marital status extracted from the mother's marital status as reported during the period of the study was classified as Married, Divorced/separated/widowed, cohabiting, never married and unknown. The married category included those that were currently married, another category included those that were divorced, separated and widowed, unknown were those whose marital status was not known at the time of data collection.

Each household was assigned a wealth quintile constructed using the principal component analysis (PCA). Using the resultant five household wealth quintile score of poorest, poorer, poor, less poor and least poor each household was categorised into two groups: poor and less poor whereby the poor included the poor, poorer and poorest while the less poor included both the less poor and the least poor.

Year of collection was the year at which the data was collected.

Random effect variable included: village.

5.7 Data Collection methods

A team of qualified Research Assistants (RAs) who collect data were selected and these went through a training course covering data collection tools. These were oriented and trained on the different tools.

The questionnaires were pre-tested and experiences were discussed prior to data collection. Each RA visited households daily and conducted face-to–face interviews in each of the assigned households.

They then sought for consent from each of the respondents before interviews were conducted.

From these surveys information about women's socio-demographic characteristics, as well as birth history and outcomes were obtained (Kajungu et al., 2020).

5.8 Data management methods

Data were extracted from common HDSS software Household Registration System (HRS) where it is stored.

Data was checked for missing whether MAR or MCAR.

Recoding and encoding of categorical variables was done where necessary.

Data management and statistical analyses were performed using Stata version 18 (StataCorp, 2023) and Microsoft Excel.

5.9 Data Analysis Methods

5.9.1 Descriptive statistics

For descriptive analysis, I used proportions to compare categorical variables among children who survived and those who died during the period of 2010 to 2015. Box plots, bar charts and line graphs were used for visualization.

To assess the differences, Wald chi-square tests for categorical variables and Wilcoxon rank sum tests for continuous variables were run. The explanatory variables too were checked for multicollinearity to ensure validity for further analysis.

5.9.3 GLMM

Parameters of the GLMM model were obtained using the maximum likelihood.

The covariance structure (autoregressive, compound symmetry, unstructured, independent, and exchangeable) with the lowest Akaike's information criteria (AIC) was considered for the final model.

Both unadjusted and adjusted odds ratios (including all variables) were estimated.

A multivariate model was constructed using stepwise selection of independent variables with probability values of <0.10.

5.9.2 GEE

GEE model that uses robust standard errors was employed to account for clustering due to repeated measurement of the women over time.

Residual plots based on the quantile–quantile (Q–Q) plots of a chi-square distribution were used to determine the goodness of fit for the GEE model.

The validity of the fitted model was determined through checking whether the linear systematic component of the model is correctly specified (by using plots of likelihood residuals or leverage) and whether the data contains some outliers or not.

The correlation structure (autoregressive, compound symmetry, unstructured, independent, and exchangeable) with the lowest quasi-AIC (QIC) was used to model the subject variation. The model with the lowest loss of information was considered. The QIC was used to test the goodness of fit.

The Liang-Zeger sandwich estimator used frequently in GEE was used to produce valid standard errors asymptotically for the resulting estimator β.

Both unadjusted and adjusted odds ratios (including all variables) were estimated.

A multivariate model was constructed using stepwise selection of independent variables with probability values of <0.10.

A two-tailed p-value of <0.05 was used to assess the statistical significance for all tests.

Adjusted odds ratios with 95% confidence intervals (CI) were calculated to measure the strength of associations between child living status and the response variables in the model.

For both the GLMM and GEE model, the coefficient estimates, standard errors and p-values produced with a binomial distribution and logit link were compared to check which model produces robust standard errors.

5.10 Ethical consideration

Approval to use the data from IMHDSS was sought from the School of Public Health (MakSPH) review board and a written letter, giving permission to use the data. Also, all data was anonymized upon receipt.

Ethical approval was sought from Makerere University School of Public Health Research and Ethics Committee (MakSPH IRB 042) and the Uganda National Council of Science and Technology (UNCST SS2002).

CHAPTER SIX: RESULTS

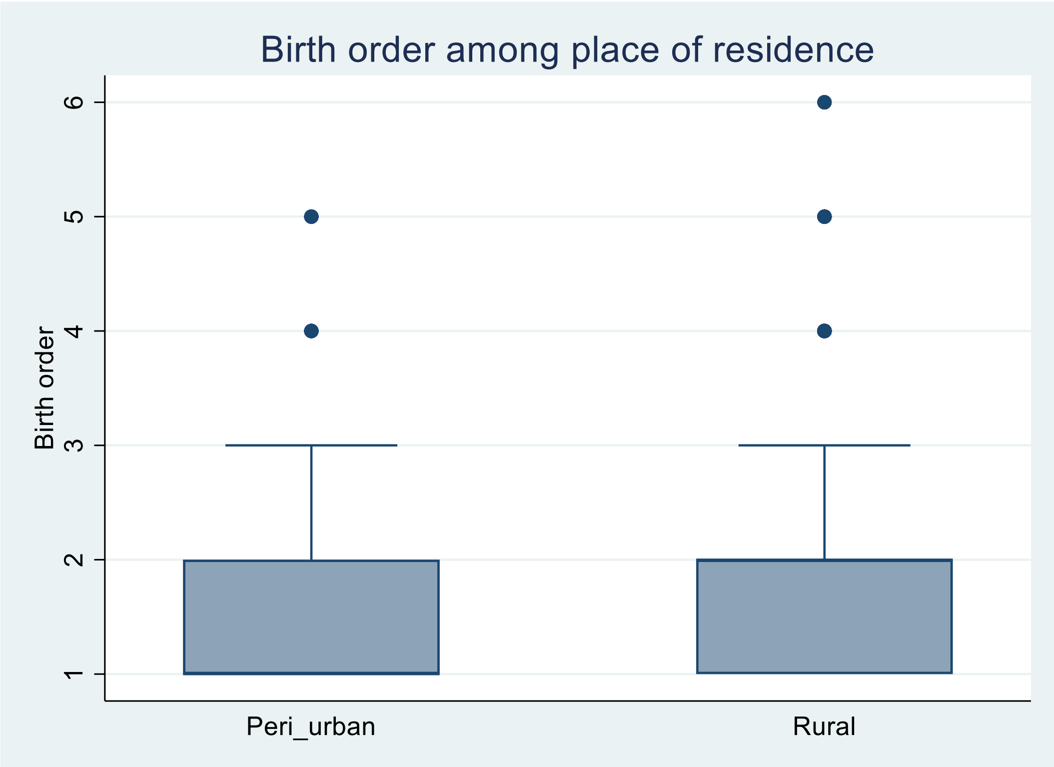
6.1 Descriptive analysis results of study participants and household characteristics

The dataset was first explored for preparation purposes. Before any analysis was conducted, the data were sorted, some variables recoded while other variables and some observations that were not of interest to the research problem were eliminated.

Figure 2: Birth order among the residence places.

According to the figure above, the average number of children between the age of one and five the mothers had from both places of residence was one.

The maximum number of children mothers from the Peri_urban areas had were only 5 and those from the rural areas had 6 children (figure 2).



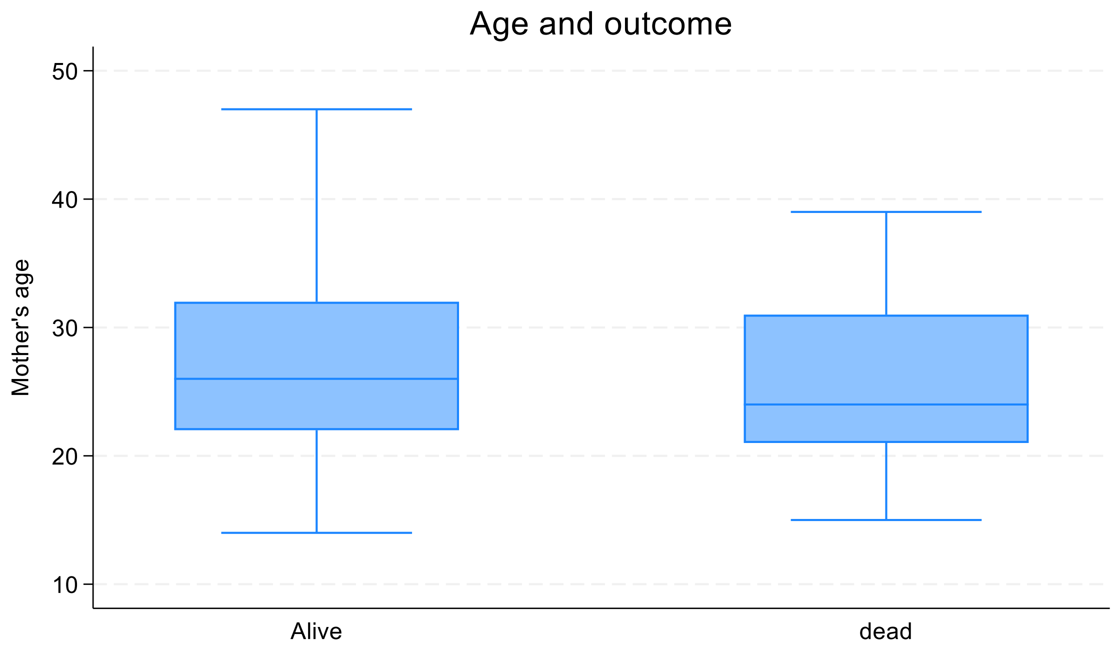
From table 1 below, children who died between the period of 2010-2015 were 31 and alive were 1241. The majority of the children were from rural areas (68.3%), were male (52.8%), were given birth from hospital/heath facility (64.1%) and their mothers' highest level of education were below primary (71.9%). The average number of children a mother had between the age of one and five was one. The number of children that died in this period were mostly among teenage mothers (58.6%), mothers whose highest level of education was below primary (64.5%) and the poor (71.0 %).

Table 1: Socio-demographic characteristics of children who died within the period (2010 – 2015)

	Dead n=31 n(%)	**Alive** n=1241 n(%)	**Total** N=1272 N(%)
Mean birth order (SD)	1.7(1.0)	1.7(0.9)	1.7(0.9)
Place of Residence			
Peri_urban	6(19.4)	397(32.0)	403(31.7)
Rural	25(80.6)	844(68.0)	869(68.3)
Gender			
Female	14(45.2)	587(47.3)	601(47.2)
Male	17(54.8)	654(42.7)	671(52.8)
Mother Age			
Non teenage mothers	14(45.2)	730(58.8)	744(58.5)
Teenage mothers	17(54.8)	511(41.2)	528(41.5)
Place of delivery			
Hospital/Health facility	20(64.5)	795(64.1)	815(64.1)
Home/way to clinic	2(6.5)	139(11.2)	141(11.1)
TBA	2(6.5)	76(6.1)	78(6.1)
Clinic	7(22.6)	231(18.6)	238(18.7)
Level of Education			
Above primary level	11(35.5)	346(27.9)	357(28.1)
Below primary level	20(64.5)	895(72.1)	915(71.9)
Marital Status			
Married	13(41.9)	698(56.2)	711(55.9)
Divorced/separated/widowed	3(9.7)	48(3.9)	51(4.0)
Cohabiting	5(16.1)	314(25.3)	319(25.1)
Never married	3(9.7)	58(4.7)	61(4.8)
Unknown	7(22.6)	123(9.9)	130(10.2)
Attend ANC			
Yes	29(93.5)	1166(94.0)	1195(93.9)
No	2(6.5)	75 (6.0)	77(6.1)
Wealth index			
Less poor	9(29.0)	380(30.6)	389(30.6)
Poor	22(71.0)	861(69.4)	883(69.4)
Year			
2010	6(19.4)	207(16.7)	213(16.7)
2011	7(22.6)	216(17.4)	223(17.5)
2012	5(16.1)	216(17.4)	221(17.4)
2013	4(12.9)	209(16.8)	213(16.8)
2014	4(12.9)	190(15.3)	194(15.3)
2015	5(16.1)	203(16.4)	208(16.4)

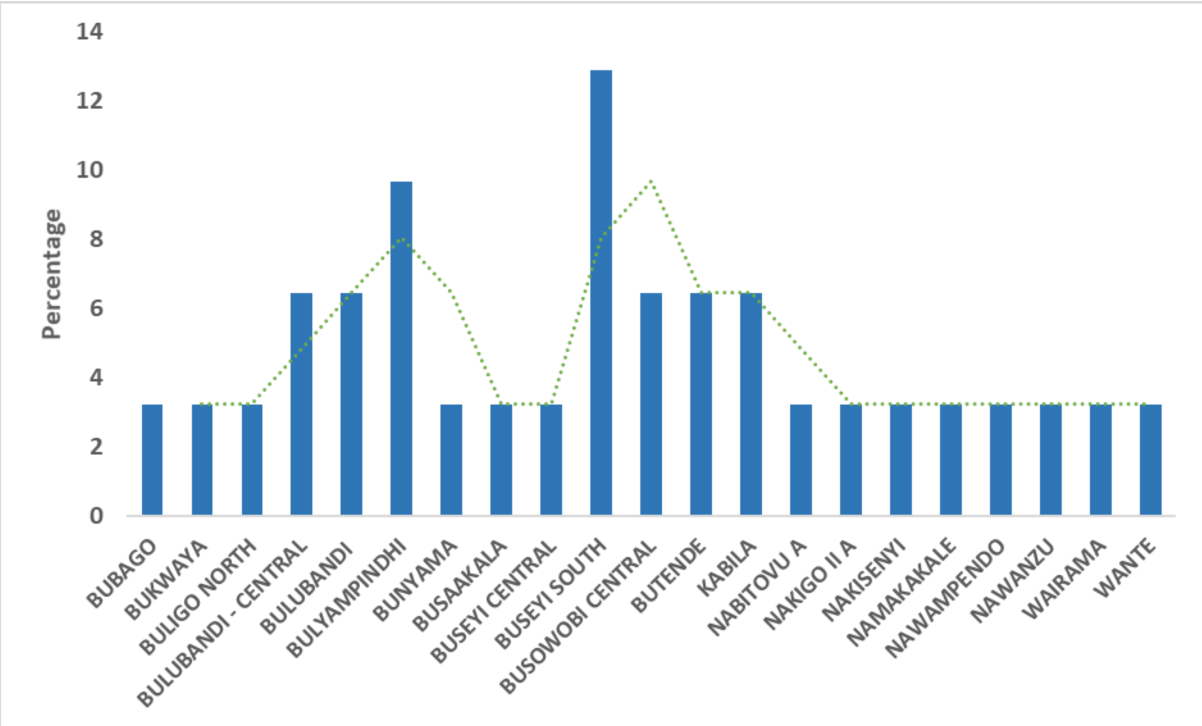
N = total number of study participants, n = frequency of study participants, %= percent

Figure 3: Mother's Age and outcome.



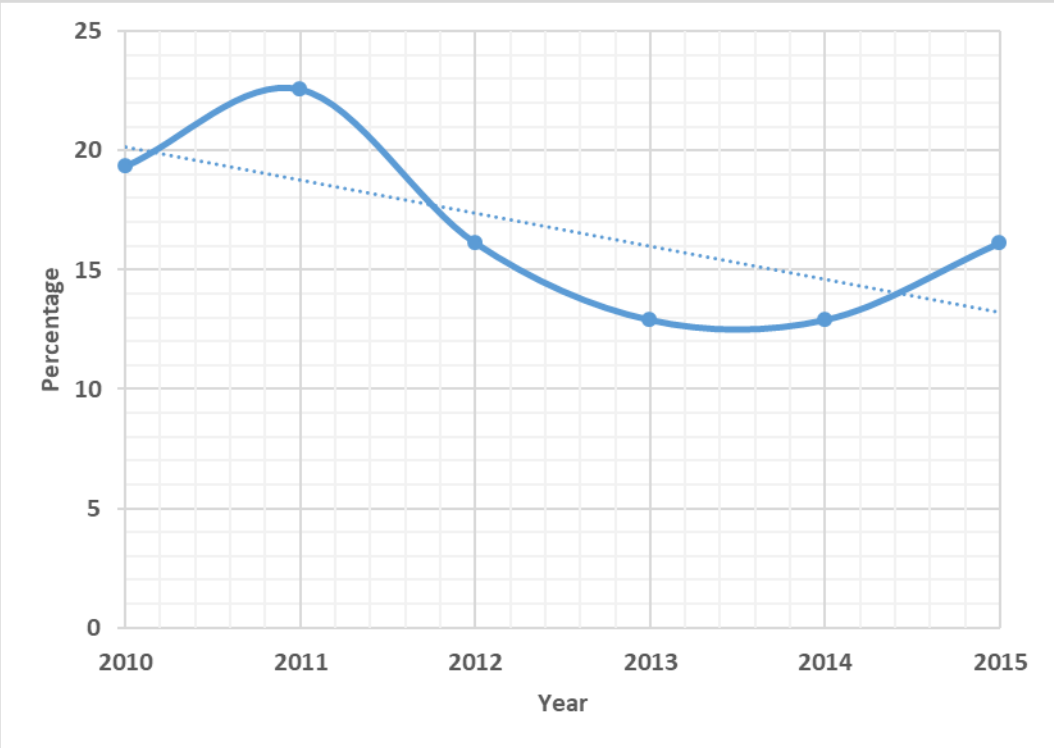
The age range of mothers whose children had died before their 60th birthday was between 15 years to 39 years and the average age being 25 years. While the age range of mothers whose children were alive was between 14 years to 47 years and the average age being 27 years (figure 3).

Figure 4: Proportion of deaths among children per Village.



Compared to the other villages in the figure above (figure 4), buseyi south had the highest proportion of children who died before the age of five between the year of 2010 and 2015 (12.9%) followed by Bulyampindhi village (9.68%).

Figure 5: Proportion of deaths



Over all from the period of analysis (2010 to 2015), the highest proportion of children that died before the age of five was in 2011 and after that there was a decline in the number of children that died. Children died the least in the years of 2013 and 2014 (figure 5).

6.2 Determining factors associated with under-five mortality using GLMM and GEE.

6.2.1 Selecting the best GLMM model using AIC and BIC.

The variance–covariance structure of the random effects was used to account for the within-subject correlation. Therefore different structures of the variance-covariance matrix were run in order to select the best GLMM model. The variance-covariance structures unstructured, independent, exchangeable and identity were used. The AIC index value of the model was determined using the maximum likelihood estimate and the number of parameters in the model (AIC = -2 * log-likelihood + 2 * number of parameters). The best-fit model according to AIC is the one that explains the greatest amount of variation using the fewest possible predictor variables.

From the table 2 below both the AIC and BIC were the same for all the variance–covariance structures and therefore a default correlation structure was used for the final model.

Table 2: Variance–covariance structures.

Model	N	ll(null)	ll(model)	df	AIC	BIC
unstructured	1,272	.	-137.498	15	304.997	382.222
independent	1,272	.	-137.498	15	304.997	382.222
exchangeable	1,272	.	-137.498	15	304.997	382.222
identity	1,272	.	-137.498	15	304.997	382.222

**ll (log likelihood)*

6.2.1.2 Using Generalized linear mixed models to determine the factors associated with under-five mortality.

From the table 3 below, after adjusting for other factors at the univariate analysis using the probability values of <0.10 independent variables like place of residence, gender, mother age, level of education, marital status and the Year of collection were found to be significant. Gender and education had a p-value>0.10 but were considered to be important variables.

The odds of children from the rural areas were two times higher likely to die before the age of five compared to the children from the Peri_urban areas (AOR=2.43; 95%CI: 0.87, 6.79).

Being a boy increased the odds dying before the age of five by 10% (AOR=1.10; 95%CI: 0.53, 2.30) compared to the girls.

Children whose mothers were teenagers had an increased odds of children dying before the age of five by 34% compared to children whose mothers were non-teenagers (AOR=1.34; 95%CI: 1.62, 2.90).

The odds of children dying before 60 months and whose mothers' highest level of education was below primary reduced by 55% (AOR=0.45; 95%CI: (0.19, 1.07) for every unit increase of death among children whose mothers' highest level of education was above primary.

The odds of dying before the 60th birthday among children whose mothers were divorced/separated/widowed were 3 times higher, never married were 2 times, marital status not known were 3 times higher than those children whose mothers were married. (AOR=3.01; 95%CI: (0.80, 11.35), (AOR=2.11; 95%CI: (0.52, 8.51), (AOR=3.91; 95%CI: (1.37, 11.15) respectively. Whereas children whose mothers were cohabiting reduced the odds of children dying before the age of five by 18% where those (AOR= 0.82; 95%CI: 0.27, 2.46).

Table 3: Factors associated with under-five mortality using Generalized linear mixed model in Iganga-Mayuge IMDHSS.

Factors	Univariate			Multivariate		
Social demographic characteristics	**OR**	**95% CI**	**pvalue**	**AOR**	**95% CI**	**pvalue**
Place of Residence						
Peri_urban	1.00			1.00		
Rural	1.96	0.80-4.82	**0.142***	2.43	0.87-6.79	**0.090***
Birth order						
First child	1.00					
Second child	0.53	0.20-1.44	0.213			
Third child	1.14	0.45-2.92	0.778			
Fourth child	1.41	0.32-6.26	0.648			
Fifth child	1					
Sixth child	1					
Gender						
Female	1.00			1.00		
Male	1.09	0.53-2.23	0.814	1.10	0.53-2.30	0.797
Mother Age						
Non teenage mothers	1.00			1.00		
Teenage mothers	1.73	0.85-3.55	**0.132***	1.34	1.62-2.90	0.462
Place of delivery						
hospital/health facility	1.00					
Home/way to clinic	0.57	0.13-2.47	0.455			
TBA	1.05	0.24-4.56	0.952			
Clinic	1.20	0.50-2.88	0.676			
Level of Education						
Above primary level	1.00			1.00		
Below primary level	0.70	0.33-1.48	0.354	0.45	0.19-1.07	**0.070***
Marital Status						
Married	1.00			1.00		
Divorced/separated/widowed	3.36	0.92-12.18	**0.066***	3.01	0.80-11.35	**0.104***
Cohabiting	0.85	0.30-2.42	0.768	0.82	0.27-2.46	0.725
Never married	2.78	0.77-10.02	**0.119***	2.11	0.52-8.51	0.296
unknown	3.06	1.20-7.81	**0.020****	3.91	1.37-11.15	**0.011****
Attend ANC						
Yes	1.00					
No	1.07	0.25-4.58	0.925			
Wealth index						
Less poor	1.00					
Poor	1.08	0.49-2.36	0.850			
Year						
2010	1.00			1.00		
2011	1.12	0.37-3.38	0.843	1.24	0.40-3.89	0.709
2012	0.80	0.24-2.66	0.714	1.45	0.26-3.08	0.867
2013	0.66	0.18-2.37	0.525	0.84	0.21-2.94	0.720
2014	0.73	0.20-2.61	0.624	0.47	0.22-3.06	0.772
2015	0.85	0.26-2.83	0.791	0.91	0.28-3.45	0.986

*OR: Unadjusted odds ratio, AOR: Adjusted odds ratio, CI: Confidence interval. *** p<0.01, ** p<0.05, * p<0.1*

$\text{logit}\{E(Y_{ij}|b_i)\} = \beta_0 + \beta_1 Place\ of\ residence + \beta_2 Gender + \beta_3 Mother\ age + \beta_4 Level\ of\ Education + \beta_5 Maritalstatus + \beta_6 Year + b_{Village}$

Both the odds of children dying before the age of five in the year of 2011 and 2012 increased by 24% and 45% compared to children in the year of 2010 (AOR= 1.24; 95%CI: 0.40, 3.89), (AOR= 1.45; 95%CI: 0.26, 3.08) respectively. Whereas the odds of children dying before the age of five in the year of in 2013 reduced by 16%, in 2014 by 53% and in 2015 by 9% compared to children in the year of 2010 (AOR= 0.84; 95%CI: 0.21, 2.94), (AOR= 0.47; 95%CI: 0.22, 3.06), (AOR= 0.91; 95%CI: 0.28, 3.45) respectively.

Village was used a random effect parameter which was used to give information about the error structure. Since the estimated coefficient and standard error was zero, this showed evidence that the random effect was not beneficial.

6.2.2 Selecting the best GEE model using the variance–covariance structure using the Quasi Likelihood Criterion

Table 4: Working Correlation structures

Correlation structure	QIC
Independent	294.850
Exchangeable	295.226
Unstructured	convergence not achieved
Stationary	observations not equally spaced
Autoregressive	observations not equally spaced
nonstationary	observations not equally spaced

In order to check the quality of the model or the goodness of fit of the model while including the correlation structure, the Quasi-likelihood Information Criterion (QIC) was used. Basing on the GEE regression approach, the slope and standard error estimates were obtained with the assumption of six working correlation structures that is independent which assumes that all measurements are independent/no correlation between any pairs of observations within the same cluster , exchangeable assumes identical correlation among all observations in a cluster/subject, unstructured assumes no specific pattern for correlations, stationary assumes the correlation between observations remains constant within clusters, autoregressive assumes that the correlation between observations decreases as the time gap between them increases and nonstationary assumes variations within clusters.

Based on the observed correlation structures in the above table (table5), an Independent correlation structure was found to be the most appropriate choice because it had the least QIC compared to the rest of the correlation structures and hence the best working correlation structure. Correlation structures unstructured, stationary, autoregressive and non-stationary did not produce any estimates because the observations were not equally spaced.

6.2.2.1 Using Generalized Estimating Equations to determine the factors associated with under-five mortality.

From the table 4 below, after adjusting for other factors at the univariate analysis using the probability values of <0.10 independent variables like place of residence, gender, mother age, level of education, marital status and the Year of collection were found to be significant. Gender and education had a p-value>0.10 but were considered to be important variables.

The logistic GEE model with a binomial distribution and logit link was used to provide both unadjusted odds ratios (OR) and adjusted odds ratios (AOR).

The ‘gender’ variable was maintained in the final model regardless of its statistical Insignificance and because of its biological plausibility. The variable ‘level of education’ was maintained because of its relational importance to the research problem according to previous research literature.

The odds of children from the rural areas were two times higher likely to die before the age of five compared to the children from the Peri_urban areas (AOR=2.28; 95%CI: 0.90, 5.80).

Being a boy increased the odds dying before the age of five by 10% (AOR=1.10; 95%CI: 0.52, 2.32) compared to the girls.

Children whose mothers were teenagers had an increased prevalence of children dying before the age of five by 36% compared to children whose mothers were non-teenagers (AOR=1.36; 95%CI: 0.68, 2.72).

The odds of children dying before their 60’s months and whose mothers’ highest level of education was below primary reduced by 54% (AOR=0.46; 95%CI: (0.19, 1.11) for every unit increase of death among children whose mothers’ highest level of education was above primary.

Table 5: Using Generalized Estimating Equations to determine the factors associated with under-five mortality in Iganga-Mayuge IMDHSS

Factors	Univariate			Multivariate		
Social demographic characteristics	**OR**	**95% CI**	**pvalue**	**AOR**	**95% CI**	**pvalue**
Place of Residence						
Peri_urban	1.00			1.00		
Rural	1.96	0.80-4.82	**0.142***	2.28	0.90-5.80	**0.084***
Birth order						
First child	1.00					
Second child	0.53	0.20-1.44	0.214			
Third child	1.15	0.45-2.93	0.776			
Fourth child	1.41	0.32-6.28	0.648			
Fifth child	1					
Sixth child	1					
Gender						
Female	1.00			1.00		
Male	1.09	0.53-2.23	0.813	1.10	0.52-2.31	0.808
Mother Age						
Non teenage mothers	1.00			1.00		
Teenage mothers	1.73	0.85-3.55	**0.132***	1.36	0.68-2.72	0.385
Place of delivery						
hospital/health facility	1.00					
Home/way to clinic	0.57	0.13-2.48	0.455			
TBA	1.05	0.24-4.57	0.952			
Clinic	1.20	0.50-2.89	0.676			
Level of Education						
Above primary level	1.00			1.00		
Below primary level	0.70	0.33-1.48	0.355	0.46	0.19-1.11	**0.085***
Marital Status						
Married	1.00					
Divorced/separated/widowed	3.36	0.92-12.25	0.939	3.06	0.91-10.33	**0.071***
Cohabiting	0.86	0.30-2.42	0.768	0.88	0.25-3.05	0.841
Never married	2.77	0.77-10.02	**0.120***	2.00	0.46-8.64	0.355
Unknown	3.05	1.19-7.81	**0.020****	3.77	1.14-12.49	**0.030****
Attend ANC						
Yes	1.00					
No	1.07	0.25-4.57	0.927			
Wealth index						
Less poor	1.00					
Poor	1.08	0.49-2.37	0.849			
Year						
2010	1.00			1.00		
2011	1.12	0.37-3.38	0.842	1.28	0.35-4.62	0.708
2012	0.80	0.24-2.65	0.714	0.90	0.28-2.85	0.856
2013	0.66	0.18-2.37	0.525	0.78	0.23-2.67	0.697
2014	0.73	0.20-2.61	0.625	0.82	0.20-3.30	0.778
2015	0.85	0.26-2.83	0.790	1.03	0.37-2.90	0.954

*OR: Unadjusted odds ratio, AOR: Adjusted odds ratio, CI: Confidence interval. *** p<0.01, ** p<0.05, * p<0.1*

$\text{logit}\{E(Y_{ij})\} = \beta_0 + \beta_1 Place\ of\ residence + \beta_2 Gender + \beta_3 Mother\ age + \beta_4 Level\ of\ Education + \beta_5 Maritalstatus + \beta_6 Year$

The odds of dying before 5years among children whose mothers were divorced/separated/widowed were 3 times higher, those whose mothers were never married, mothers whose marital status was not known were 2 times and 3 times higher than those children whose mothers were married (AOR=3.06; 95%CI: (0.91, 10.33), (AOR=2.00; 95%CI: (0.46, 8.64), (AOR=3.77; 95%CI: (1.14, 12.49) respectively. Whereas children whose mothers were cohabiting reduced the odds of children dying before the age of five by 12% where those (AOR= 0.88; 95%CI: 0.25, 3.05).

Both the odds of children dying before the age of five in the year of 2011 and 2015 increased by 28% and 3% compared to children in the year of 2010 (AOR= 1.28; 95%CI: 0.35, 4.62), (AOR= 1.03; 95%CI: 0.37, 2.90) respectively.

Whereas the odds of children dying before the age of five in the year of in 2012 reduced by 10%, in 2013 by 22% and in 2014 by 18% compared to children in the year of 2010 (AOR= 0.90; 95%CI: 0.28, 2.85), (AOR= 0.78; 95%CI: 0.23,2.67), (AOR= 0.82; 95%CI: 0.20, 3.30) respectively.

6.3 Comparing the model efficiency of both GLMM and GEE models in determining under-five mortality.

According to the two criteria (table 2 and 5), the model with the smallest statistic (AIC/QIC) was preferred. Village was included in both models as a random effect to account for the within-cluster similarity and between-cluster heterogeneity. Model coefficient estimates, standard errors and p-values were compared.

The coefficient estimates of the GEE model were slightly lower in magnitude compared to those of the GLMM model for most of the variables except for the mother's age.

The standard errors of the GEE model were all different and smaller than that of the GLMM model except those for the marital status variable.

Both models showed that place of residence, level of education and the marital status of the mother were statistically significant with pvalue<0.1.

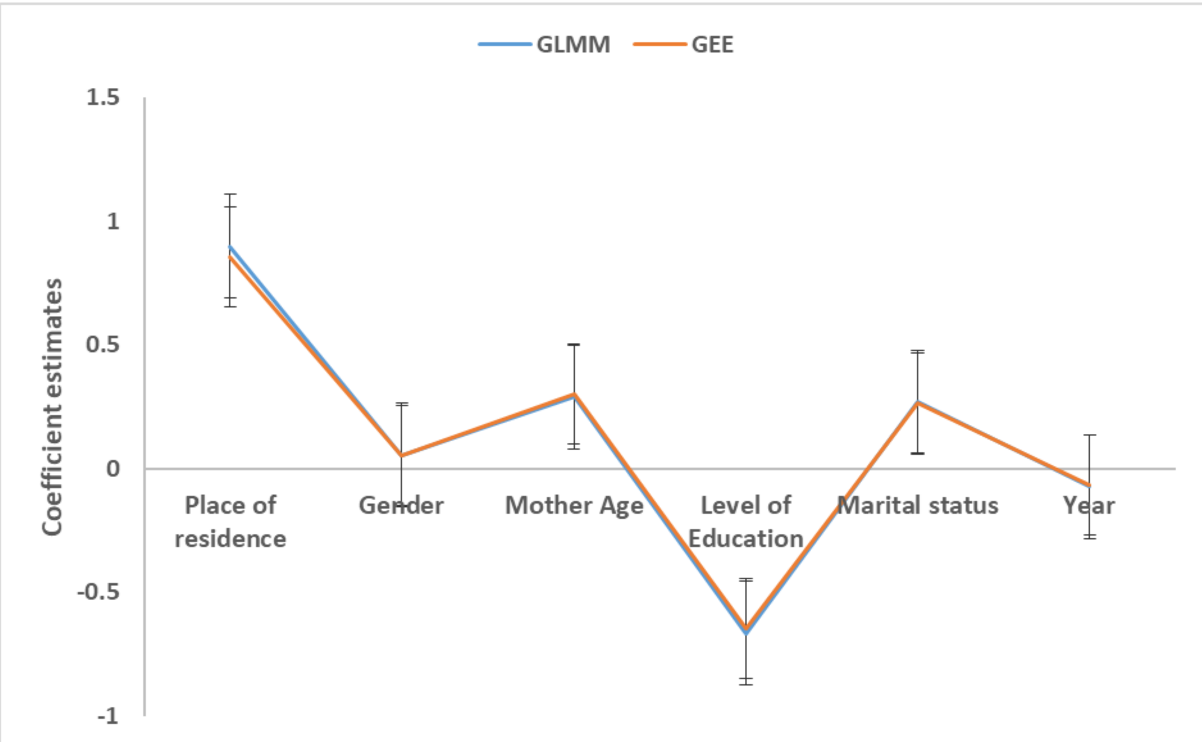
Table 6: Parameter Estimates, Standard Error (SE) Estimates and p-values with Different models (GLMM and GEE)

	GLMM			GEE		
	Estimate	SE	p-value	Estimate	SE	p-value
Place of residence(Rural/Peri_urban)	0.900	0.504	**0.075***	0.857	0.440	**0.052****
Gender(Male/Female)	0.055	0.370	0.881	0.051	0.386	0.897
Mother age(teenagers/non-teenagers)	0.290	0.384	0.451	0.303	0.332	0.363
Level of education(below primary/above primary)	-0.665	0.405	**0.101***	-0.648	0.361	**0.073***
Maritalstatus(Divorced,separated,widowed,cohabiting,nevermarried,uknown/Married)	0.297	0.129	**0.037****	0.264	0.144	**0.066***
Year(2011,2012,2013,2014,2015/2010)	-0.072	0.110	0.509	-0.068	0.090	0.448

**** p<0.01, ** p<0.05, * p<0.1*

Figure 6: Visualisation of estimates using GLMM and GEE

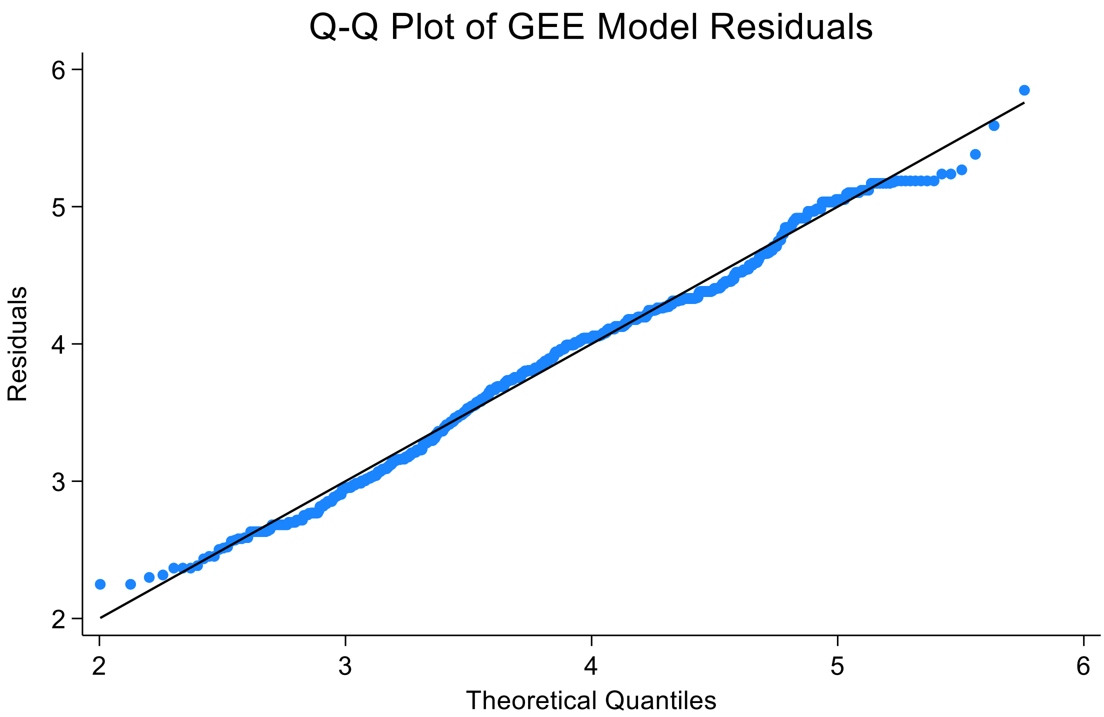
There was a slight difference in the coefficients predicted by both the GLMM and GEE models for the variables determining under-five mortality (figure 6).



6.2.4 Evaluation of model performance

Since the best model chosen was GEE that uses robust standard errors to correct model misspecification less of the diagnostics are needed therefore besides using the QIC to check for the overall goodness-of-fit of the GEE model a Q-Q plot was also fitted (*Generalized Linear Mixed Model and Generalized Estimating Equation for Binary Longitudinal Data*, 2014).

Figure 7: Q-Q plot showing the distribution



The Q-Q plot above shows that the residuals tend to deviate from the line especially on the tail ends indicating that the residuals are less likely to follow a normal distribution.

CHAPTER SEVEN: DISCUSSION

Under this chapter I discuss the specific objectives of the study and the results obtained from the analysis. This study investigated the difference between the model efficiency of determinants of under-five mortality rates while using GLMM and GEE and determined the association between the predictor variables and under-five mortality rates in the Iganga and Mayuge districts.

7.1 Comparison of model efficiency (analysis of objective 1)

Because of the nature of the data of repeated measurements where mothers have had more than one child both the GLMM and GEE were used to account for correlation and therefore avoid bias in the estimates (Astuti et al., 2023). Model efficiencies of GLMM and GEE models with a binomial distribution and logit link were used for comparison while determining the under-five mortality rates.

Both models included the same variables that were statistically associated with under-five mortality in the multivariate analysis.

There was no difference while using the AIC to choose the best working structure for GLMM whereas for GEE model while using QIC there were some differences in selecting the best working correlation structure.

Parameter estimates like the coefficient estimates, standard errors and p values were used to attain the comparison in the analysis. Most of the coefficient estimates from the GEE were lower in magnitude than corresponding estimates from the GLMM except for a few variables. This is consistent with the previous studies where the model estimates were similar (Koper & Manseau, 2009). The interpretation of the parameter estimates of both models were the same.

From the results obtained, this study therefore chose the GEE model to assess the factors associated with under-five mortality rates.

7.2 Factors associated with under-five mortality using GEE (analysis of objective 2)

Findings from the GEE model in this study found that amongst other factors place of residence, gender, mother's age, highest level of education, marital status and the year of collection had a significant influence on mortality among children below the age of five.

This study showed that children who reside in the rural areas were more likely to die compared to those that resided in the Peri_urban places. This could be attributed to the fact that rural areas have limited services in terms of health facilities, paediatric care among others (Khadka et al., 2015).

The boys had higher odds of dying before the age. This may be explained by biological reasons other than social and cultural factors. Biologically the boys are born weaker and are more prone to diseases and death compared to girls. This is in line with findings from other studies (Antehunegn & Worku, 2021).

Furthermore, this study found reduction in the under-five mortality among mothers who were cohabiting this would imply that male involvement while raising a child is key. Our findings are consistent with previous studies that assessed the impact of father's involvement on child development, functioning and quality of life (Pedersen & Liu, 2012).

Children whose mothers were teenager had higher odds of death compared to those whose mothers were non teenagers. Teenage mothers are disadvantageous compared to the non-teenage mothers. This is consistent with other findings by others. This could be associated with either social or economic factors like low income, higher rates of premature birth and low birth weight of babies, poor prenatal health behaviours, incomplete or delayed vaccination that could lead to preventable diseases (Pedersen & Liu, 2012; Y & S, 2016).

Children whose mothers had their highest level of education as below primary level had reduced odds of dying before the age of five compared to those whose education level was above primary level of education. Better maternal education increases the under-five survival. In Uganda, there is provision of free education in government institutions which is a good start and many of the mothers have been able to learn. Educated mothers are likely to have better income, and higher health literacy levels. And therefore achieve desirable maternal health outcomes (Akinyemi et al., 2015; Sartorius et al., 2011).

7.3 Strength and limitations of the study

This study used a hierarchical statistical model to identify factors associated with under-five mortality that accounts for correlation. The model don't require assumptions to be made and it produces robust standard errors. Therefore the findings of this study provide information that could help policy-makers and others working toward reducing child mortality to identify demographics that may be more vulnerable to dying before their fifth birthday.

However, the model has some limitations like; GEE does not use tests that use the likelihood function that can be used for model selection and also GEE treats the random effects as nuisance, GEE requires the data to be missing completely at random.

CHAPTER 8: CONCLUSIONS AND RECOMMENDATIONS

The conclusions and recommendations were discussed in line with the study specific objectives and based on the study findings.

8.1 Conclusions

The slight differences in the models could be due to the fact that;

i. GLMM incorporates random effects to model the variability within groups, while GEE focuses on population-averaged effects.
ii. GLMM accounts for individual variations within clusters more explicitly.
iii. GEE assumes that the correlations between observations within a cluster are the same for all clusters, while GLMM allows for varying correlations. If there is substantial variation in cluster correlations, GLMM might capture these nuances better, leading to different estimates.

Therefore a slight difference in the parameter estimates between GLMM and GEE models suggesting that the modelling approach does influence the results to some degree. The choice between these methods should be guided by the underlying data structure, assumptions, research objectives, and the practical significance of the differences observed. If the differences are minor and do not substantially alter the conclusions, it's important to acknowledge the differences of both models and the potential implications for your specific study context.

Improving socio-economic conditions, healthcare access, fostering community engagement, could potentially alleviate the identified risk factors. Therefore these interventions could include improve maternal and child healthcare services, enhance education and awareness among teenage mothers and rural communities, and addressing gender-specific health needs

These conclusions can guide future research, policy development, and interventions aimed at improving overall child well-being and reduce under-five mortality in Iganga and Mayuge districts.

8.2 Recommendations

Basing on the results from this study, the following are some of the recommendations:

To ensure the model is not under or over fit, one ought to look at the goodness-of-fit statistics, both the AIC and the QIC values.

GLMM model is suitable for accounting for individual-level variability since it incorporates in random effects that is not explained by fixed predictors and also it uses the likelihood estimation that make appropriate at simulations. GEE is a better model for population-averaged effects.

Policymakers should take these findings into account when designing and implementing health policies and programs like;

- Allocating resources to build and maintain healthcare facilities, especially in rural and underserved areas, ensuring access to quality maternal and child healthcare.
- Develop and implement public health campaigns to raise awareness about child health like immunization campaigns, hygiene, clean water, nutrition, family planning and the importance of seeking timely medical care that promote maternal and child health
- Train local healthcare workers and community health volunteers to deliver essential healthcare services, conduct health screenings, and promote healthy behaviours like emphasizing proper prenatal care, skilled birth attendance, and postnatal care.

Combatting under-five mortality requires tailored approaches that involve governments, Non-Governmental Organizations, healthcare providers and communities working together. By addressing factors above can significantly reduce under-five mortality rates and improve the overall health and well-being of children in Iganga- Mayuge districts.

REFERENCES

1. Abir, T., Ogbo, F. A., Stevens, G. J., Page, A. N., Milton, A. H., & Agho, K. E. (2017). *The impact of antenatal care, iron-folic acid supplementation and tetanus toxoid vaccination during pregnancy on child mortality in Bangladesh*. https://doi.org/10.1371/journal.pone.0187090

Agho, K. E., Ezeh, O. K., Ferdous, A. J., Mbugua, I., & Kamara, J. K. (2020). Factors associated with under-5 mortality in three disadvantaged East African districts. *International Health*, *12*(5), 417. https://doi.org/10.1093/INTHEALTH/IHZ103

Ahinkorah, B. O., Budu, E., Seiduid, A.-A., Agbaglo, E., Aduid, C., Oseiid, D., BankeThomas, A., & Yaya, S. (2022). *Socio-economic and proximate determinants of underfive mortality in Guinea*. https://doi.org/10.1371/journal.pone.0267700

Akhter, N., Fairbairn, R. S., Pearce, M., Warren, J., Kasim, A., Bambra, C., & Swartz, A. M. (2021). Local Inequalities in Health Behaviours: Longitudinal Findings from the Stockton-On-Tees Cohort Study. *International Journal of Environmental Research and Public Health Article Public Health*, *18*, 11018. https://doi.org/10.3390/ijerph182111018

Awor, P., Chola, L., Nsabagasani, X., Kagwire, F., Mpanga, F., & Mbonye, A. (2018). Improving Child Survival in Uganda: Trends, Opportunities and Policy Implications for Achieving Universal Health Coverage. *UNIVERSAL HEALTH COVERAGE IN UGANDA*, 374.

Ballinger, G. A. (2004). Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods*, *7*(2), 127–150. https://doi.org/10.1177/1094428104263672

Bauserman, M., Lokangaka, A., Thorsten, V., Tshefu, A., Goudar, S. S., Esamai, F., Garces, A., Saleem, S., Pasha, O., Patel, A., Manasyan, A., Berrueta, M., Kodkany, B., Chomba, E., Liechty, E. A., Michael Hambidge, K., Krebs, N. F., Derman, R. J., Hibberd, P. L., … Bose, C. L. (2015). Risk factors for maternal death and trends in maternal mortality in low- and middle-income countries: A prospective longitudinal cohort analysis. *Reproductive Health*, *12*(2), 1–9. https://doi.org/10.1186/1742-4755-12-S2S5/FIGURES/3

Cha, S., & Jin, Y. (2019). Have inequalities in all-cause and cause-specific child mortality between countries declined across the world? *International Journal for Equity in Health*, *19*(1), 1–13. https://doi.org/10.1186/S12939-019-1102-3/FIGURES/4

Chao, F., You, D., Pedersen, J., Hug, L., & Alkema, L. (2018). National and regional under-5 mortality rate by economic status for low-income and middle-income countries: a systematic assessment. *Lancet Glob Health*, *6*(5), e535–e547. https://doi.org/10.1016/s2214-109x(18)30059-7

Das, T., Holland, P., Ahmed, M., & Husain, L. (2021). Sustainable Development Goal 3: Good Health and Well-being. In *South-East Asia Eye Health* (pp. 61–78). https://doi.org/10.1007/978-981-16-3787-2_4

Ekholuenetale, M., Wegbom, A. I., Tudeme, G., & Onikan, A. (2020). Household factors associated with infant and under-five mortality in sub-Saharan Africa countries. *International Journal of Child Care and Education Policy*, *14*(1). https://doi.org/10.1186/s40723-020-00075-1

Exemplars. (2020). *WHY IS UNDER-FIVE MORTALITY IMPORTANT?* Https://Www.Exemplars.Health/. https://www.exemplars.health/topics/under-fivemortality/what-is-under-five-mortality

Filippi, V., Ronsmans, C., Campbell, O. M., Graham, W. J., Mills, A., Borghi, J., Koblinsky, M., & Osrin, D. (2006). Maternal health in poor countries: the broader context and a call for action. In *Lancet* (Vol. 368, Issue 9546). https://doi.org/10.1016/S01406736(06)69384-7

Gebremichael, S. G., & Fenta, S. M. (2020). *Under-Five Mortality and Associated Risk Factors in Rural Settings of Ethiopia: Evidences from 2016 Ethiopian Demographic and Health Survey*. https://doi.org/10.1155/2020/8430246

Ghisletta, P., & Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics*, *29*(4), 421–437. https://doi.org/10.3102/10769986029004421

Gonçalves, B. P., Huang, C.-Y., Morrison, R., Holte, S., Kabyemela, E., Rebecca, D., Fried, M., & Duffy, P. E. (2014). From the Laboratory of Malaria Immunol-ogy and Vaccinology. *N Engl j Med*, *19*(8), 1799–1808. https://doi.org/10.1056/NEJMoa1303944

Grantz, K. H., Rane, M. S., Salje, H., Glass, G. E., Schachterle, S. E., & Cummings, D. A. T. (2016). Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918. *Proceedings of the National Academy of Sciences*, *113*(48), 13839–13844. https://doi.org/10.1073/PNAS.1612838113

Hirose, A., Kajungu, D., Tusubira, V., Waiswa, P., Alfven, T., & Hanson, C. (2020). Postneonatal under-5 mortality in peri-urban and rural Eastern Uganda, 2005-2015. *BMJ Global Health*, *5*, 3762. https://doi.org/10.1136/bmjgh-2020-003762

Hubbard, A. E., Ahern, J., Fleischer, N. L., Laan, M. Van Der, Lippman, S. A., Jewell, N., Bruckner, T., & Satariano, W. A. (2010). To GEE or not to GEE: Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*, *21*(4), 467–474. https://doi.org/10.1097/EDE.0B013E3181CAEB90

Islam, M. Z., Billah, A., Islam, M. M., Rahman, M., & Khan, N. (2022). Negative effects of short birth interval on child mortality in low- and middle-income countries: A systematic review and meta-analysis. *Journal of Global Health*, *12*, 04070. https://doi.org/10.7189/JOGH.12.04070

Juul, S. E., Comstock, B. A., Wadhawan, R., Mayock, D. E., Courtney, S. E., Robinson, T., Ahmad, K. A., Bendel-Stenzel, E., Baserga, M., LaGamma, E. F., Downey, L. C., Rao, R., Fahim, N., Lampland, A., Frantz III, I. D., Khan, J. Y., Weiss, M., Gilmore, M. M., Ohls, R. K., … Heagerty, P. J. (2020). A Randomized Trial of Erythropoietin for Neuroprotection in Preterm Infants. *New England Journal of Medicine*, *382*(3), 233–243. https://doi.org/10.1056/nejmoa1907423

Kain, M. P., Bolker, B. M., & McCoy, M. W. (2015). A practical guide and power analysis for GLMMs: Detecting among treatment variation in randomeffects. *PeerJ*, *2015*(9). https://doi.org/10.7717/PEERJ.1226

Kajungu, D., Hirose, A., Rutebemberwa, E., Pariyo, G. W., Peterson, S., Guwatudde, D., Galiwango, E., Tusubira, V., Kaija, J., Nareeba, T., & Hanson, C. (2020). Cohort

Profile: The Iganga-Mayuge Health and Demographic Surveillance Site, Uganda (IMHDSS, Uganda). *International Journal of Epidemiology*, *49*(4), 1082-1082g. https://doi.org/10.1093/IJE/DYAA064

Kantorova, V., Kariuki, S., Kunju, S., Lay, K. K., Li, N., Riffe, T., Spoorenberg, T., & Williams, I. (2021). *Levels & Trends in*.

Khadka, K. B., Lieberman, L. S., Giedraitis, V., Bhatta, L., & Pandey, G. (2015). The socioeconomic determinants of infant mortality in Nepal: Analysis of Nepal Demographic Health Survey, 2011. *BMC Pediatrics*, *15*(1). https://doi.org/10.1186/s12887-015-04687

Kitui, J., Lewis, S., & Davey, G. (2013). Factors influencing place of delivery for women in Kenya: An analysis of the Kenya demographic and health survey, 2008/2009. *BMC Pregnancy and Childbirth*, *13*(1), 1–10. https://doi.org/10.1186/1471-2393-1340/TABLES/5

Klotzke, K., & Fox, J. P. (2019). Modeling Dependence Structures for Response Times in a Bayesian Framework. *Psychometrika*, *84*(3), 649–672. https://doi.org/10.1007/S11336019-09671-8/TABLES/4

Mani, K., Sada, ;, Dwivedi, N., Ravindra, ;, & Pandey, M. (2012). Determinants of UnderFive Mortality in Rural Empowered Action Group States in India: An Application of Cox Frailty Model. *International Journal of MCH and AIDS*, *1*(1), 61.

MoFPED. (2015). *Millennium Development Goals Report for Uganda 2015*. www.thenomadagency.com

Muff, S., Held, L., & Keller, L. F. (2016). Marginal or conditional regression models for correlated non-normal data? *Methods in Ecology and Evolution*, *7*(12), 1514–1524. https://doi.org/10.1111/2041-210X.12623

Nambuusi, B. B., Ssempiira, J., Makumbi, F. E., Utzinger, J., Kasasa, S., & Vounatsou, P. (2019). Geographical variations of the associations between health interventions and allcause under-five mortality in Uganda. *BMC Public Health*, *19*(1). https://doi.org/10.1186/S12889-019-7636-X

Nasejje, J. B., Mwambi, H. G., & Achia, T. N. O. (2015). Understanding the determinants of under-five child mortality in Uganda including the estimation of unobserved household and community effects using both frequentist and Bayesian survival analysis approaches Biostatistics and methods. *BMC Public Health*, *15*(1). https://doi.org/10.1186/s12889015-2332-y

Nisar, Y. Bin, & Dibley, M. J. (2014). Determinants of neonatal mortality in Pakistan: Secondary analysis of Pakistan Demographic and Health Survey 2006-07. *BMC Public Health*, *14*(1). https://doi.org/10.1186/1471-2458-14-663

Oduse, S., Zewotir, T., & North, D. (2021). The impact of antenatal care on under-five mortality in Ethiopia: a difference-in-differences analysis. *BMC Pregnancy and Childbirth*, *21*(1), 1–9. https://doi.org/10.1186/S12884-020-03531-5/TABLES/6

Pan, W. (2002). Goodness-of-fit tests for GEE with correlated binary data. *Scandinavian Journal of Statistics*, *29*(1), 101–110. https://doi.org/10.1111/1467-9469.00091

Panahi, F., Rashidi Fakari, F., Nazarpour, S., Lotfi, R., Rahimizadeh, M., Nasiri, M., & Simbar, M. (2022). Educating fathers to improve exclusive breastfeeding practices: a randomized controlled trial. *BMC Health Services Research*, *22*(1). https://doi.org/10.1186/S12913-022-07966-8

Pekár, S., & Brabec, M. (2018). Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology*, *124*(2), 86–93. https://doi.org/10.1111/ETH.12713

Ramirez, A. G., Byeong, ;, Choi, Y., Munoz, ; Edgar, Arely Perez, ;, Gallion, K. J., Moreno, P. I., & Penedo, F. J. (2020). Assessing the Effect of Patient Navigator Assistance for Psychosocial Support Services on Health-Related Quality of Life in a Randomized Clinical Trial in Latino Breast, Prostate, and Colorectal Cancer Survivors. *Cancer*. https://doi.org/10.1002/cncr.32626

Sartorius, B., Kahn, K., Collinson, M. A., Vounatsou, P., & Tollman, S. M. (2011). Survived infancy but still vulnerable: Spatial-temporal trends and risk factors for child mortality in rural South Africa (Agincourt), 1992-2007. *Geospatial Health*, *5*(2). https://doi.org/10.4081/gh.2011.181

UBOS. (2016). Uganda Demographic and Health Survey 2016. *Udhs*.

UNICEF. (2015). Under-five mortality rate. *Encyclopedia of Public Health*, *September*, 1425–1426. https://www.who.int/data/nutrition/nlis/info/under-five-mortality-rate

UNICEF. (2020). Child survival and the SDGs - UNICEF DATA. In *Unicef Data*. https://data.unicef.org/topic/child-survival/child-survival-sdgs/

UNICEF. (2021a). *Child and maternal health | UNICEF Uganda*. https://www.unicef.org/uganda/what-we-do/child-survival-development

UNICEF. (2021b). *Child Mortality - UNICEF DATA*. Unicef. https://data.unicef.org/topic/child-survival/under-five-mortality/

UNICEF. (2021c). United Nations Inter-agency Group for Child Mortality Estimation (UN IGME), 'Levels & Trends in Child Mortality: Report 2021. *Estimates Developed by the United Nations Inter-Agency Group for Child Mortality Estimation*.

Vaucher, Y. E., Peralta-Carcelen, M., Finer, N. N., Carlo, W. A., Gantz, M. G., Walsh, M. C., Laptook, A. R., Yoder, B. A., Faix, R. G., Das, A., Schibler, K., Rich, W., Newman, N. S., Vohr, B. R., Yolton, K., Heyne, R. J., Wilson-Costello, D. E., Evans, P. W., Goldstein, R. F., … Higgins, R. D. (2012). Neurodevelopmental Outcomes in the Early CPAP and Pulse Oximetry Trial. *N Engl J Med*, *367*, 2495–2504. https://doi.org/10.1056/NEJMoa1208506

Wang, M. (2014). *Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments*. https://doi.org/10.1155/2014/303728

Wang, M., Kong, L., Li, Z., & Zhang, L. (2016). Covariance estimators for Generalized Estimating Equations (GEE) in longitudinal analysis with small samples. *Statistics in Medicine*, *35*(10), 1706. https://doi.org/10.1002/SIM.6817

Woldeamanuel, B. T. (2019). *Socioeconomic, Demographic, and Environmental Determinants of Under-5 Mortality in Ethiopia: Evidence from Ethiopian Demographic and Health Survey, 2016*. https://doi.org/10.1155/2019/1073782

Worku, M. G., Teshale, A. B., & Tesema, G. A. (2021). Determinants of under-five mortality in the high mortality regions of Ethiopia: mixed-effect logistic regression analysis. *Archives of Public Health*, *79*(1), 1–9. https://doi.org/10.1186/S13690-021-005784/TABLES/5

World Health Organization. (2022). *Child mortality (under 5 years)*. Fact Sheets. https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-under-5mortality-in-2020

Zorn, C. J. W. (2001). Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *Source: American Journal of Political Science*, *45*(2), 470–490.