COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF STATISTICS

Modeling Determinants of Neonatal Mortality in Ethiopia

By: Bacha Ewunetu

A Thesis to be submitted to Department of Statistics, School of Graduate Studies, College of Natural Science, Jimma University, in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biostatistics

Jimma, Ethiopia

October, 2015

Modeling Determinants of Neonatal Mortality in Ethiopia

MSc Thesis

Bacha Ewunetu

Advisor: Yehenew Getachew (PhD)
Co-advisor: Akalu Banbeta (MSc)

October, 2015
Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES, JIMMA UNIVERSITY

As thesis research advisors, we her by certify that we have read the thesis prepared by Bacha Ewunetu under our guidance, which is entitled "Modeling Determinants of Neonatal Mortality in Ethiopia, in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Yehenew Getachew (PhD)	_____	_____
Advisor	Signature	Date
Akalu Banbeta(MSc)	_____	_____
Co-advisor	Signature	Date

As the members of the board of examiners of MSc thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____	_____	_____
Name of chairman	Signature	Date
_____	_____	_____
Name of major advisor	Signature	Date
_____	_____	_____
Name of Co-advisor	Signature	Date
_____	_____	_____
Name of internal examiner	Signature	Date
_____	_____	_____
Name of external examiner	Signature	Date

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the requirements of Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate. Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Bacha Ewunetu

Date_____

Signature _____

Jimma University, Jimma

ACKNOWLEDGEMENT

Firstly, I wish to express my profound gratitude to the Almighty God for his love, grace and protection throughout the course of my study in this university.

Secondly, I also wish to express my deepest appreciation to my advisor, Dr.Yehenew Getachew without whom accomplishment of this work would not have been made possible. His continuous supportive comments and constant guidance throughout the course of my thesis work get me out of my frustration. It is a real pleasure working with him. I can only hope that his cooperation will keep on going in the future.

Thirdly, I gratefully acknowledge my co-advisor Mr. Akalu Banbeta for his continuous and passionate support in every aspect I needed. His consistent concern and help made me concentrate on my thesis.

Finally am very grateful to my wonderful family for their prayers and given me the peace of mind to be able undertake this thesis work to a successful end. I cannot forget my loving and supportive wife, Sabone Abera: my deepest gratitude. Your encouragement when the times got rough are much appreciated and duly noted

ABSTRACT

Background: Although neonatal mortality estimates continue to decline in Ethiopia over the years, it is a matter of a great concern among stake holders as the decline is not enough to reduce NM.

Objective: The study aimed to investigate significant factors and appropriate model of neonatal mortality in Ethiopia and also to assess the effect region as a cluster.

Method: The data was obtained from the EDHS, 2011.The study sample (n = 2604) was based on infants (0–1 months old) during the survey period; extracted from the women data base. Two model families, generalized estimating equation and alternating logistic regression models from marginal model family, and generalized linear mixed model from cluster specific model family were used for the analysis. AIC and QIC were used for model selection.

Result: the result showed that among eligible children the proportion of NM was 14.78%. Alternating logistic regression model was best fits the data for population-averaged effects of the given factors on neonatal mortality than generalized estimating equation model and generalized linear mixed model with two random intercepts was the best model to evaluate within and between regional heterogeneity of neonatal mortality. From all the fitted model age of respondents (mothers), multiplicity of birth, birth interval, and birth order, age at first birth, residence, and birth size were found to be significant factors of neonatal mortality; whereas wealth, mothers educational level, sex of a child and place of delivery were non-significant factors.

Conclusion: in line with objectives of this study marginal models, GEE and ALR, have been compared for the analysis of marginal or average effects of covariates on the response variable and, we conclude that, ALR model with measure of association exhibited the best fit for this data than GEE models. For this study also GLMM, with two random intercept models was found to be appropriate for the analysis of within and between regional variations for neonatal mortality in Ethiopia. This concluded that there is heterogeneity of neonatal mortality between and within regions.

Keywords: Neonatal mortality; Generalized Estimating Equation; Alternating logistic regression; Generalized Linear Mixed Model

TABLE OF CONTENTS

STATEMENT OF AUTHOR I

ACKNOWLEDGEMENT II

ABSTRACT III

TABLE OF CONTENTS IV

LIST OF ABBREVIATIONS AND ACRONYMS VI

LIST OF TABLES VIII

LIST OF FIGURES IX

1. INTRODUCTION 1

1.1. BACKGROUND 1

1.2. STATEMENT OF THE PROBLEM 3

1.3. OBJECTIVE OF THE STUDY 5

1.3.1. General Objective: 5

1.3.2. Specific Objectives: 5

1.4. SIGNIFICANCE OF THE STUDY 5

2. LITERATURE REVIEW 6

2.1. INTRODUCTION 6

2.2. REVIEW OF LITERATURE ON STUDY METHODS 11

2.2.1. Generalized Estimating Equation (GEE) 11

2.2.2. Alternating Logistic Regression (ALR) 12

2.2.3. Generalized Linear Mixed Model 13

3. DATA AND METHODOLOGY 15

3.1. DATA SOURCE 15

3.1.1. Sampling design of 2011 EDHS data 15

3.2. VARIABLES DESCRIPTION 16

3.2.1. Response variable 16

3.2.2. Independent variables 16

3.3. METHODS FOR DATA ANALYSIS 17

3.3.1. Generalized Linear Model 17

3.3.2. Marginal models 18

3.3.2.1. Generalized Estimating Equation 19
3.3.2.1.1. Parameter Estimation for GEE 20
3.3.2.2. Alternating Logistic Regression (ALR) 20
3.3.2.2.1. Parameter Estimation of ALR 21
3.3.2.3. Model Building for Marginal Models 22
3.3.3. Variable Selection Technique 22
3.3.4. Model Comparison Technique 23
3.3.4.1. Quasi-Information Criterion (QIC) 23
3.3.5. Cluster Specific (Subject Specific) Models 24
3.3.5.1. Generalized Linear Mixed Model (GLMM) 24
3.3.5.2. Parameter Estimation for GLMM 25
3.3.5.3. Model Comparison in GLMM 26
3.3.5.4. Model Checking Technique 27
4. RESULT AND DISCUSSION 28
4.1. DESCRIPTIVE RESULT 28
4.2. COLLINEARITY DIAGNOSTIC TEST 30
4.3. STATISTICAL ANALYSIS OF MARGINAL MODELS 31
4.3.1. Analysis of Generalized Estimating Equations (GEE) 31
4.3.2. Analysis of Alternating Logistic Regression model (ALR) 34
4.3.3. Comparison of GEE and ALR Models 35
4.3.4. Parameter Interpretation of Marginal Models 37
4.4. ANALYSIS OF GENERALIZED LINEAR MIXED MODEL (GLMM) 39
4.4.1. Model Building in GLMM 39
4.4.2. Parameter Interpretation of GLMM 41
4.4.3. Model diagnostic for GLMM 43
4.5. DISCUSSION 45
5. CONCLUSIONS AND RECOMMENDATIONS 49
5.1. CONCLUSIONS 49
5.2. RECOMMENDATIONS 49
APPENDIX 55

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criterion
ALR	Alternating Logistic regression
AOR	Adjusted Odds Ratio
CSA	Central Statistical Agency
DHS	Demographic Health Survey
EDHS	Ethiopian Demographic Health Survey
GEE	Generalized Estimating Equations
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
LRT	Likelihood Ratio Test
MDG	Millennium Developmental Goal
MOH	Ministry of Health
NICU	Neonatal Intensive Care Unit
NMR	Neonatal Mortality Rate
ENMR	Early Neonatal Mortality Rate
LNMR	Late Neonatal Mortality Rate
OR	Odds Ratio
PHC	Population and Housing Census
PNM	Proportion of Neonatal mortality
QIC	Quassi Information Criterion
SGA	Small for Gestational Age
SNNPR	Southern Nation Nationalities and Peoples Representatives

UN	United Nations
UNICEF	United Nation Children's Fund
USAID	United States Agency International Development
VIF	Variance Inflation Factor
WHO	World Health Organization

LIST OF TABLES

Table 3.1: Description of variables and their coding used in the study.....16
Table 4.1: Neonatal Mortality of 2011 EDHS data 28
Table 4.2: Summary of descriptive statistics for neonatal mortality (EDHS, 2011).....29
Table 4.3: Collinearity test of independent variables 30
Table 4.4: Empirical and model based standard errors for two proposed working correlations (GEE) 32
Table 4.5: Parameter estimates (empirically corrected standard errors) for GEE 33
Table 4.6: Parameter estimates (empirically corrected standard errors) from ALR.....35
Table 4.7: model comparison of one and two random intercept model.....39
Table 4.8: Parameter estimates (standard errors) and corresponding P-value for GLMM. 41

LIST OF FIGURES

Figure 4.1: diagnostic plot for generalized linear mixed model (EDHS, 2011).....44

1. INTRODUCTION

1.1. Background

Neonatal mortality (NNM) is the probability of baby dying within the first month of life and is expressed as neonatal deaths per 1000 live births. In 2005, the World Health Organization reported that neonatal deaths accounted for 40% of deaths under the age of 5 worldwide; each year an estimated 8 million neonates died within the first 28 days of life (Lawn, 2005).

The first month of life, the neonatal period, carries one of the highest risks of death of any month in the human lifespan (Lancet, 2005). In high-income countries, neonates are now a major focus of child health for reducing both for mortality and morbidity. However, in low-income countries NNM rates, trends, and causes have attracted relatively little attention compared to maternal deaths or deaths among older children under-five. As far as international public health policy and programs are concerned neonatal deaths still do not receive attention commensurate with their burden (Lancet, 2010).

The Lancet report (The Lancet,2014a) indicated that globally, although the number of children under five who die has almost been halved since 1990, the progress in the reduction of neonatal deaths has been much slower than that of children over four weeks of age (average annual reduction of 2.1 % vs. 3.4 %) (5). The proportion of neonatal deaths among children under-five is now 44 percent, compared to 38 percent in 2005.For instance, in 2012, 2.9 million newborn babies died within 28 days after birth 1 million on the first and only day of life and there were an additional 1·2 million stillbirths shortly before or during labour.

In Ethiopia Neonatal mortality has decreased from 49 deaths per 1,000 live births according to EDHS 2000 to 39 deaths per 1,000 live births in the EDHS 2005; it remained stable at 37deaths per 1,000 as reported in EDHS 2011. This decline in NNM, as in other parts of the world, was lower than for infant and under-five mortality, which fell by 42% and 47%, respectively over the15-year period. As mentioned above the country is experiencing a high NNM rate (37per 1,000livebirths) which is comparable to the average rate of 35.9 per 1,000 for the African region overall (Lancet, 2014b).

There have been a lot of research works on neonatal mortality in the past years. Most of these investigations make use of the standard logistic regression as the method for

determining the factors that affect the probability a child dying in the first months of life due to binary nature of the outcome variable. There are also several other methods have been discussed in different literature to model neonatal mortality specially survival analysis when outcome variable is time dependent (time to event); however the authors contributed to modeling neonatal mortality have not reached a conclusion which model is appropriate if investigator wants to consider outcome variable as independent of time. In this thesis we used GEE, ALR, and GLMM.

Clustered data arise in a wide variety of applications including demographic surveys where samples are arranged in a hierarchy. The clustering can be expressed in terms of correlation among the measurements on units within the same cluster. EDHS 2011 data contains clusters in the various regions where respondents were selected and interviewed. Such clustered data are often correlated which violates the statistical assumption of independence of observations in the case of the logistic regression method. EDHS 2011 data describes respondents within the same region where there is the likelihood for these respondents within the same region to share similar characteristics which may affect the survival of their children.

Statistical models for clustered data must account for the intra cluster correlation at each level analysis else it could lead to misleading inferences. If the intra cluster correlation is not properly accounted for in the analyses, standard errors of the parameters may be biased (Ghisletta and Spini, 2004). Generalized linear model as described by Nelder and Wedderburn (1972) and McCullagh and Nelder (1989) are regression models to analyse continuous or discrete response variables. The association between response variable and covariates is given by the so-called link function. GLM assume that the observation are independent and does not consider any correlation between the outcomes of the n observation. Marginal and conditional models are extension of the GLM for correlated data.

In the marginal model, the primary interest of the analysis is to model the marginal expectation of response variable given the covariates. Here, the correlation or more general the association between the outcome variables modeled separately and is regarded as a nuisance parameter. The major goal is to investigate the effect of covariates in the population on the response variables. Including the correlation structure in estimating the effects mainly yields different variance estimation. Marginal models have been

introduced first by Zeger, Liang, and Self (1985), Liang and Zeger (1986) and Zeger and Liang (1986).

One method which allows researchers to analyze correlated data is the Generalized Estimating Equations (GEE). In the late 1980s, Generalized Estimating Equations (GEEs) were developed which allow for the analysis of non-normal, clustered data (e.g. repeated measures, Littel et al. 2002). These models are usually concerned with population-level inferences. The method is an extension of the Generalized Linear Models (GLMs) to accommodate correlated data with its main advantage in the unbiased estimation of regression coefficients despite a possible misspecification of the structure of the correlation (Ghisletta and Spini, 2004). The GEE is particularly effective for modeling clustered binary or count data (Wang, 2010). The two basic advantages for the GEE method when dealing with correlated data are as follows. Firstly, it could be used regardless of the nature of the response variable whether it is continuous, dichotomous, polychotomous, ordinal or an event-count. Secondly, the method allows for a variety of specifications of the correlation patterns within the clusters (Zorn, 2001). One attractive property of the Generalized Estimating Equations is that, one can use a working correlation structure that may be specified wrongly while the resulting regression coefficient estimate is still consistent and asymptotically normal (Pan and Connett, 2002). Finally, Generalized Linear Mixed Models (GLMMs) have been developed more recently and extend GLMs to include random-effects (Agresti, 2002). In contrast with GEEs, these models often have subject-specific interpretations.

1.2. Statement of the problem

Although neonatal mortality estimates continue to decline in Ethiopia over the years, it is a matter of a great concern among stake holders as the decline is not enough to reduce NM. Several studies have concluded a decline in NM in Ethiopia: (Negera W., Eshetu W., 2013), (Edward F., Fikre E., Peter B., 2009),(Bogale W,.Assaye K.,Amha M.,Birkineh T.,Alemayehu W.,2012) ,(Samuel M.,Eshetu W.,2012), but there is not much investigation into correlation of neonatal mortality among regions of Ethiopia.

Ethiopia is the second most populous country in Africa after Nigeria with a population estimated at nearly 87 million in 2010 (World data bank, 2013).The population grows at an annual rate of 2.6% which is slightly greater than that for sub Saharan African

countries average growth of 2.5%.The age structure suggests that nearly 45% of the population is under 15 years. High mortality, high fertility and low life expectancy characterize the demography, as in most sub-Saharan African countries (Ringheim K, Teller C, Sines E, 2009).

There is a rather limited research on NM in Ethiopia. Most of the information for any program planning and implementation has been based on Ethiopian demographic and health surveys conducted every five years There are two broad classes of models for clustered observations, conditional (or cluster specific) and marginal (or population averaged) models. Clustering occurs at the level of regions only and that each region contributes a variable number of observations (neonates in our case). Conditional models are models for intra-cluster responses that condition on the unmeasured characteristic(s). In these models we introduce a term to capture child-specific conditions that have a suspected influence on the mortality level of children in the household. The model for responses of children in a household is then conditional on the (unmeasured) child-specific covariate.

In our data set children are clustered within a locality (regions). It is recognized that individuals in the same region are more similar than individuals in different region. When there are clusters in the data, it is very appropriate to use a method that would cater for the correlation among variables within the same cluster. Zorn (2001) in" Generalized Estimating Equation Models for Correlated Data: A Review with Applications" reviewed the Generalized Estimating Equation (GEE) method for dealing with a correlated data. Staley (2013) further reported that, one technique which allows researchers to handle such forms of correlation is the GEE method. Furthermore, it was revealed in the research that GEEs allow for a range of correlation patterns within clusters and offer valuable insights into the dynamics of such correlation.

Moreover, this study apply GEE (using correlation) and ALR (using odds ratio), GLMM model on 2011 Ethiopia Demographic Health Survey (EDHS 2011) data. The focus is on the effect of important biological and social factors within (intra-cluster correlation) and between (inter-cluster correlation) the region as cluster in Ethiopia. The data clustered in to nine regional state and two administrative cities, with large variation in term of economic performance and standard of living. To have the correct estimates of our parameter we specify working correlation structure for GEE and we compare this model

with the Alternating Logistic regression. Therefore, given this fact this study is aimed to answer the following interesting research questions:

- Does regional variation have an effect on neonatal mortality?
- What are the variables that signifantly affect neonatal mortality?
- Which fitted model is well appropriate within and between the regional variations for neonatal mortality?

1.3. Objective of the study

1.3.1. General Objective:

The general objective of this study is modeling determinants of neonatal mortality in Ethiopia from 2011 EDHS dataset.

1.3.2. Specific Objectives:

Specifically, the study aims to:

- Identify the significant factors that affect Neonatal mortality
- Compare generalized estimating equation & alternating logistic regression models.
- To select an appropriate model for neonatal mortality, which fit the data well.

1.4. Significance of the Study

Findings obtained from this study will be useful in many ways:-

- Governmental and non-governmental organizations could take intervention measures and set appropriate plans to reduce neonatal mortality and giving priority for the areas which mostly affected in neonatal mortality in the country.
- The study would serve as a guide to stakeholders in making informed and intelligent policy decisions with regard to neonatal mortality and the management of the risk factors to avoid the death of neonates in the country
- To help individuals to have awareness about death of neonates in the country
- To help next researcher to give attention on the remaining issues of neonatal mortality using the alternative model.

2. LITERATURE REVIEW

2.1. Introduction

This section will review the general findings related to the effect of socio-economic, demographic and proximate (health related) determinants on neonatal mortality.

Derose and Kulkarni (2005) using multi-level logistic analysis found community HIV rates, women's education and immunization as significant determinants of child mortality in Zambia. In Egypt, Aly and Grabowski (1990) used logit analysis to model child death Probability using Egypt's World Fertility Survey in 1980. They concluded that source of drinking water and sanitation was significantly and negatively related to child mortality.

Study conducted in Ethiopia titled "Determinants of Neonatal Mortality in Ethiopia: a case control study, 2013" used multiple logistic regression to identify determinants of neonatal mortality. The study found that the major determinants of neonatal mortality were maternal education, birth interval, multiple birth, and age at birth, maternal and paternal employment and place of residence.

Desta Mekonnen (2011) also established several factors as influencing neonatal mortality. His study was undertaken for Ethiopia. His study examined and identified the important determinants of Infant and Child mortality in Ethiopia. The 2000 and 2005 Ethiopia Demographic and Health Survey (EDHS) data were used. The main aim of his study was to investigate the association between infant and child mortality and socio-economic and biodemographic factors in Ethiopia and distinguish which of these factors were more pronounced in the reduction of infant and child mortality between 2000 and 2005. The data consisted of a national representative sample of household level data. The results of the study established a strong link between birth order and neonatal mortality. Generally, those born from lower birth orders (e.g order 1) had higher chances of dying than those from higher orders. The increase in the preceding birth interval also reduced the risk of mortality.

Kumar and Gemechis (2010) used data from Ethiopia DHS survey (2005) and employed cross tabulation technique to examine the selected socioeconomic, bio-demographic and maternal health care factors that determine child mortality in Ethiopia. The results showed that among other variables, birth interval with preceding birth and mothers' education had a significant impact in lowering the risk of child mortality. The result

confirmed that the child mortality risk associated with children of less than 2 years of birth interval with previous child was highest (15 per cent) and lowest (4. 2 per cent) for the children whose birth interval was 4+ years. Birth order also was an important determinant of neonatal mortality in Ethiopia.

Christiana (2008) in her study of the determinants of neonatal mortality in Indonesia also found out that neonatal mortality was higher among infants with shorter birth intervals , males infants and smaller than average sized infants. Another study was carried out in Bangladesh and it used the 2007 Bangladesh Demographic and Health Survey data to investigate the effect of maternal education on neonatal mortality in Bangladesh. Among many factors, maternal age and birth order were important determinants of neonatal mortality in Bangladesh (Mostafa, 2012). Kamal (2012) in his study to investigate the socio-economic correlates of infant and child mortality in Bangladesh established a higher neonatal mortality among infants with mothers with a higher number of children ever born.

Deepak and others (2013) used Nepal 2001, 2006 and 2011 Demographic and Health Survey data to investigate the trends and determinants of neonatal mortality in Nepal. Findings revealed higher neonatal mortality for first or fourth or higher birth order than for second and third births. The study further revealed a higher mortality for neonates with less than two years birth interval than those with more than two years. Others including Arshad (2002) with the use of 1991 Pakistan Demographic and Health Survey data to investigate the determinants of neonatal and post-neonatal mortality in Pakistan revealed that preceding birth interval was a major factor determining neonatal mortality. Low birth weight was also found to be an important cause of neonatal mortality by Upadhyay (2011) in his study on the determinants of neonatal mortality in rural Haryana in India.

Ezra and Gurum (2002) used a logistic regression model to investigate the impact of birth interval on infant and child mortality in the context of communities characterized by high reproductivity, prolonged breast feeding practice, and poor living conditions in Ethiopia. The results revealed that short birth interval (less than 18 months) were significantly associated with neonatal mortality. They further observed that those born from younger mothers (15-19 years) and oldest mothers (35-49 years) had higher chances of death than those born from mothers aged 25-34 years. Okantey (2012) in his study of the

determinants of neonatal mortality in Ghana also found that maternal age and size of birth were significant contributors to neonatal mortality.

Boone and Zhan (2006) employed logistic regression for analyzing child mortality in a cross-section of countries. The study found mother's and father's education as significant determinants of child mortality in poor countries. Study on Bangladesh, Bairagi et.al (1999) using a duration model concluded that changes in mother's education, birth interval and birth order had little effect on mortality.

Using a logistic multilevel analysis, Curtis et al. (1993) found considerable variation between families in the risk of post-neonatal mortality in Brazil, even after controlling for a range of demographic, socioeconomic and locational variables at both the child and the maternal levels. Madise and Diamond (1995) found a strong correlation of neonatal and post-neonatal mortality risks for births from the same mother in Malawi even after controlling for a host of socio-economic, demographic and locational variables.

(Kojo, 2012) In his thesis titled "Modeling the Risk Factors of neonatal mortality in Ghana" sought to analyze the risk factors of neonatal deaths in Ghana and to suggest some interventions that can be used in order to improve the survival of newborn babies in Ghana. He developed three models each for mother level factors, child level factors and environmental level factors. The results of the research revealed that, for the mother level factors, it was found that the age of the mother and the wealth index were the main factors causing neonatal deaths in Ghana. Also, sex of the baby and whether the baby is a twin were not significant causes of neonatal deaths in Ghana. He added that, for the environmental level factors, only the region of delivery was a significant cause of neonatal mortality in Ghana. He however recommended for further research to be conducted using other factors of neonatal mortality.

(Hong, 2009) Conducted an analysis of infant and under-five mortality in Rwanda using the nearly two decades of Rwandan Demographic and Health Surveys data. They found that, in the case of neonatal mortality, the number of children ever born, birth interval, availability of professional antenatal and delivery care, full immunization of children, mother's education, and urban-rural residence were important determinants. The same study also revealed that, in the case of under-five mortality; multiplicity of births (i.e. number of births for each pregnancy), birth intervals, antenatal care and deliveries by

health professionals, full immunization of children, mother's education, and use of contraception and possession of mosquito nets were determinants.

Pandey et al (1998) examined infant and child mortality of India. This research found that sex of the child, mother's residence, mother's exposure to mass media, use of clean cooking fuel, mother's literacy status, access to a toilet facility, mother's religion and ethnicity, income of the household, birth order, mother's age at birth and mother's health care were important determinants of infant and child mortality in India.

Similar finding were reported in a study by Kumar and Gemechis (2010) using data from the 2005 Ethiopian Demographic and Health Survey. The study reported that birth interval, mother's literacy, household wealth, mother's age at birth, mother's exposure to mass media, sex of the child, religion, family size, birth order and residence were important predictors of infant and child mortality. Many studies reported that children born at short birth intervals are at higher risk of infant and child mortality (Rutstein 2005; Rutstein 2008; Hong et al, 2009; Saha and van Soest 2012). Using multi country DHS data collected between 2000 and 2005 Rutstein (2008) found that waiting 36 months or more to have another pregnancy substantially decreases risk to children of death and under nutrition.

Zenger (1993) found the family-level correlation of neonatal mortality in Bangladesh to be slight and not significant, after controlling for the length of the preceding birth interval. She also modeled the variation in residual correlation by difference in birth order, a type of model which may be readily estimated by the multilevel software package MLn, and found the association in neonatal mortality risks to be stronger for immediate pairs of siblings than for siblings which are further apart. She suggests the stronger association for immediate siblings may be because familial effects on neonatal mortality may change over time or with maternal age.

The most wide-ranging assessment of family effects on neonatal mortality has been that of Curtis and Steele (1996). Using logistic multilevel models, they found a substantial and significant family level effects on the risk of neonatal mortality in Bolivia, Kenya, Peru and Tanzania, even after controlling for the effects of a standard range of child-level and family-level variables, including the survival status of the preceding child. In all four countries they also found a significant reduction in the correlation of mortality risks for

the same mother with increases in the difference in birth order, but did not investigate how this changed after controlling for other variables. They conclude by suggesting biological sources of family-level correlations in neonatal mortality risks as a promising area for future research.

Child's Sex

A child's sex has been shown to affect the probability of infant and child mortality: Owing to biological factors, male infants have a higher risk of mortality during the first year of life, as highlighted for example in WHO (2003a). In addition, differential treatment of boys and girls, owing to cultural and socioeconomic factors, may also be expected to affect the chances of survival during childhood.

Birth Order

Birth order may also play a role in the probability of infant and child mortality, though the direction of the effect is a priori ambiguous. According to the hypothesis of intra household resource competition, first born children are more likely to capture vital resources such as food and care, thereby reducing their mortality risk (see e.g. Vos et al., 2004). On the other hand, it has been found that first born children, who are more likely to be born to mothers at younger reproduction ages, experience a higher mortality risk than children of a higher birth order. A number of studies indeed point to a U-shaped effect of birth order, with the probability of infant mortality declining after the first child and increasing again for children of birth order four and higher (see e.g. Titaley et al., 2008 and Uddin and Hossain, 2008). To account for this effect, we construct two dummies: one for first born children, and one for children with birth order 4 and higher.

Maternal age

Maternal age is an important risk factor for perinatal and neonatal mortality. In a systematic review by Carolan & Frankowska (2011) older maternal ages 35-39 years and over 40 years were at increased risk of neonatal mortality. Furthermore in a cohort study by Lisonkova (2010) older mothers were found to be at increased risk of pre-term birth and small for gestational age infants (adjusted Odds Ratio 1.5 (CI 1.4 to 1.7) for women aged 35 to 39 years; and aOR 1.6 (95% CI 1.3 to 2.0) for women aged 40 years.

Birth interval

A short preceding birth interval has also been found to increase the probability of infant mortality, resulting in a WHO recommendation of at least 24 month spacing between a preceding birth and a new pregnancy (WHO, 2006b). Assuming a full pregnancy of nine

months, this translates into an optimal spacing of 33 months between succeeding births. We use a dummy to control for preceding birth intervals shorter than that. Finally, we also include a dummy variable for different years of birth, to capture the changes in mortality over time. This dummy takes on the value 1 for all children born between 1971 and 1987, 2 for all children born between 1988 and 1997, and 3 for all children born between 1998 and 2007 (i.e. during the 10-year interval before the survey). In line with improving health conditions over time, we expect a higher value of the dummy to be associated with lower mortality rates.

2.2. Review of Literature on Study Methods

2.2.1. Generalized Estimating Equation (GEE)

According to Agresti, computationally simple alternative to maximum likelihood (ML) for clustered categorical data is a multivariate generalization of quasi likelihood. Rather than assuming a particular type of distribution for the response variable, this method only links each marginal mean to a linear predictor and provides a guess for the variance covariance structure of the response. The method uses the observed variability to help generate appropriate standard errors and called the GEE method because the estimates are solutions of generalized estimating equations. These equations are multivariate generalizations of the equations solved to find ML estimates for generalized linear models (Agresti, 2007). Generalized estimating equations (GEE) models are a direct extension of basic quasi-likelihood theory from cross-sectional to repeated or otherwise correlated measurements. They estimate the parameters associated with the expected value of an individual's vector of binary responses and phrase the working assumptions about the association between pairs of outcomes in terms of marginal correlations (Molenberghs & Verbeke, 2005).

When we are mainly interested in first-order marginal mean parameters and pair wise interactions, a full likelihood procedure can be replaced by quasi-likelihood based methods (McCullagh and Nelder, 1989). In quasi-likelihood, the mean response is expressed as a parametric function of covariates, and the variance is assumed function of the mean up to possibly unknown scale parameters.

Wedderburn first noted that likelihood and quasi-likelihood theories coincide for exponential families and that the quasi-likelihood estimating equations provide consistent estimates of the regression parameters in any generalized linear model, even for choices

of link and variance functions that do not correspond to exponential families (Wedderburn, 1974). Consequently, Liang and Zeger proposed the method of generalized estimating equations (GEE) as an extension of GLM to accommodate correlated data using quasi-likelihood approach. Rather than assuming a particular distribution for the response, GEE method requires a correct specification of the mean as well as how the variance depends on the mean. One of the desirable properties of the GEE method is that it yields consistent and asymptotically normal solutions even with the misspecification of the covariance structure (Liang and Zeger, 1986).

In the methodology of generalized estimating equations, the user may impart a correlation structure that is often called a working correlation matrix. One often does not know what the true correlation is, hence, the term working correlation. Common correlation structures include; Unspecified: all correlations are to be independently estimated from the data, Exchangeable: all correlations within subjects are equal, Independent: all correlations are assumed to be zero (Myers *et al*, 2010). Because GEE does not have likelihood function, likelihood-ratio methods are not available for checking fit, comparing models, and conducting inference about parameters.

2.2.2. Alternating Logistic Regression (ALR)

Generalized estimating equation (GEE), allows estimation of first and second moment parameters in regression models for multivariate binary data. When association among the observation is importance and is measured using marginal odds ratios, the computations required will exclude the applications in studies with large clusters. An alternative approach that overcomes the computational limitations encountered in many problems is proposed what is called alternative logistic regression (Zeger *et*, 1993). As explained by Zeger *et al*, alternating logistic regression is reasonably efficient relative to GEE. In ALR, we estimate the association parameters by modeling the conditional distribution of one response given another.

Molenberghs & Verbeke also expressed ALR as extension of classical GEE, in the sense that precision estimates follow for both the parameters. However, unlike with GEE, no working assumptions about the third- and fourth-order odds ratios are required. The clever combination of a marginal and a conditional specification, addressing the third and fourth moments is avoided all together, which is strictly different from setting them equal

to zero. This combination of marginal and conditional specification can be advantageous of ALR (Molenberghs & Verbeke, 2005).

2.2.3. Generalized Linear Mixed Model

Agresti explained that, generalized linear model (GLM) extend ordinary regression by allowing non-normal responses and a link function of the mean. The generalized linear mixed model is a further extension that permits random effects as well as fixed effects in the linear predictor (Agresti, 2007). Antonio & Beirlant defined GLMM as extend of GLM by allowing for random or cluster-specific effects in the linear predictor. These models are useful when the interest of the analyst lies in the individual response profiles rather than the marginal mean. The inclusion of random effects in the linear predictor reflects the idea that there is natural heterogeneity across subjects or clusters in some of their regression coefficients (Antonio & Beirlant, 2006).

According to McCulloch clarification, GLMM is very versatile in that they can handle non-normal data, nonlinear models, and a random effects covariance structure. This can be used to incorporate correlations in models, model the correlation structure, identify sensitive subjects and can be used to handle heterogeneous variances. The modeling process is relatively straightforward, requiring the following decisions: what is the distribution of the data, what is to be modeled, what are the factors, and are the factors fixed or random? This all makes GLMM attractive for use in modeling. Unfortunately, computing methods for much of the class of GLMM is an area of active research. No general-purpose software exists and, tests and confidence intervals are asymptotic and approximate (McCulloch, 1997).

Generalized the above explanation, GLMM is an extension to generalized linear model (GLM) that includes random effects in the linear predictor, giving an explicit probability model that explains the origin of the correlations. The resulting cluster-specific parameter estimates are suitable when the focus is on estimating the effect of changing one or more components of the predictor on a given individual.

The key problem in GLMM is maximization of the marginal likelihood, obtained by integrating out the random effects. In general, no analytic expressions are available for the integrals and numerical approximations are needed. There are large statistical literatures on various methods like approximation of the data, approximation of the Integral (Molenberghs & Verbeke, 2005). To summarize, this brief literature review has

shown the importance of a range of characteristics in determining maternal delivery service behavior.

3. DATA AND METHODOLOGY

3.1. Data Source

The data used in this study was obtained from the Ethiopian Demographic and Health Survey conducted in 2011, which is the third comprehensive survey conducted as part of the worldwide Demographic and Health Surveys project (2011, EDHS). The 2011 EDHS was carried out under the aegis of the Ministry of Health (MOH) and was implemented by the Central Statistical Agency (CSA). The survey interviewed a nationally representative population in about 18,500 households, and all women age 15-49 and all men age 15-59 in these households. The 2011 EDHS used three questionnaires: the Household Questionnaire, the Woman's Questionnaire, and the Man's Questionnaire. The Woman's Questionnaire was used to collect information from all women age 15-49 from the selected households.

3.1.1. Sampling design of 2011 EDHS data

The sample for the 2011 EDHS was designed to provide population and health indicators at national (urban and rural) and regional levels. Administratively, the regions of Ethiopia were divided into zones, and zones, into administrative units called weredas. Each wereda was subdivided into the lowest administrative units, called kebeles which are further subdivided into census enumeration areas. A representative sample of 17,817 households was selected for the 2011 EDHS. Of these households 16,702 were successfully interviewed in which 1 7,385 eligible women were identified for individual interview; full interviews were conducted with 16,515 women. The number of children at this level was 11,654 representing the number of live births born to the interviewed mothers in the period of five years preceding the date of the survey. After removal of missing values we obtained complete information about 2,605 neonates.

An estimate of Neonatal mortality was based on information from the birth history section of the questionnaire. The data for Neonatal mortality estimation will be extracted from the birth history section of the Woman's Questionnaire from 16,515 women age 15-49. The birth history section begins with questions about the respondent's experience with childbearing (i.e., the number of sons and daughters living with the mother, the number who live elsewhere, and the number who have died). These questions are followed by a retrospective birth history, in which each respondent is asked to list each of her births, starting with the first birth. For each birth, data were obtained on sex, month

and year of birth, and current age, or, if the child is dead, age at death. If the child's age at death is above 28 days, then it will be excluded from the analysis.

3.2. Variables Description

3.2.1. Response variable

The dependent variable is child survival status. One question from the EDHS used to examine the dependent variable, which is child alive at the time of interview "Yes (1) or no (0)". The response was binary: yes or no. As mentioned above, the dependent variables are dichotomous, coded as zero if death has not occurred and coded as 1 if death has occurred (alive =0 and dead =1).

3.2.2. Independent variables

The explanatory variables that would be included are explained as follows. The choice of these variables is guided by different literatures as the determinant factors of neonatal mortality. As various literatures supported, the major predictors of Neonatal mortality considered in this study were categorized as socioeconomic and demographic variables. These potential predictors included in the table 3.2.

Table 3.1: Description of variables and their coding used in the study

Variables	Description	Codes/categories
Age	Age of mother during survey	0=15-19 1=20-29 2=30-39 3=40-49
Sex	Sex of a child	0=Female 1=Male
Age at birth	Age at first birth of mothers	0=<20 1=20-29 2=>29
Wealth	Wealth of household	0=Poor 1=Middle 2=Rich
Multiple birth	Number of births during pregnancy	0=Single 1=Multiple
Residence	Place of residence for mother	0=Rural 1=Urban
Birth size	Size of a child at birth	1=Very large 2=Larger than average 3=Average 4=Smaller than average 5=Very small
Birth	Previous birth interval	0=<24 1=24-47 2=>47

interval		
Birth order	Birth order of a child	1=one 2=Two to four 3=Five and above
Place of delivery	The place mothers delivered during pregnancy	0=Home 1=Health facility
Maternal education	Highest level of education attained by mothers	0=No education 1=Primary 2=Secondary and above

3.3. Methods for data analysis

3.3.1. Generalized Linear Model

Generalized linear models (GLMs) represent a class of regression models that allow us to generalize the linear regression approach to accommodate many types of response variables including count, binary, proportions and positive valued continuous distributions. (Nelder and Wedderburn, 1972; Hilbe, 1994; Hoffman, 2004). Because of its flexibility in addressing a variety of statistical problems and the availability of software to fit the models, it is considered a valuable statistical tool and is widely used. In fact, the generalized linear model has been referred to as the most significant advance in regression analysis in the past twenty years (Hoffman 2004).

Generalized linear models include three components: 1) a random component which is the response and an associated probability distribution; 2) a systematic component, which includes explanatory variables and relationships among them (e.g., interaction terms): and 3) a link function, which specifies the relationship between the systematic component or linear predictor and the mean of the response. It is the link function that allows generalization of the linear models for count, binomial and percent data thus ensuring linearity and constraining the predictions to be within a range of possible values (Guisan, 2002). This ability to handle a larger class of error distributions and data types is a key improvement of GLMs over linear models. In general, GLM is a linear model for a transformed mean of a response variable that has distribution in the natural exponential family.

The Exponential Family

A random variable Y follows a distribution that belongs to the exponential family, if the density function is of the form.

$$\mathbf{f(y/\theta,\phi) = exp\{\phi^{-1}[y\theta - \psi(\theta)] + c(y,\phi)\}} \qquad (\mathbf{3.1})$$

For a specific set of unknown parameters θ and ϕ , and for known functions ψ (·) and c (·, ·). The parameter θ is called the canonical parameter and represents the location while, ϕ is called the dispersion parameter and represents the scale parameter and for the Poisson and binomial distribution it is fixed to be one (Faraway, 2006). An important property of the GLM is the functional relation between mean and variance.

Generalized linear model assumes that the response variables are independent. In clustered data however, observations are usually taken from the same unit, and thus this information forms a cluster of correlated observations. For instance, in the EDHS the dependent variable (survival status of neonate) was measured once for each eligible mothers nested within clusters from each region.

3.3.2. Marginal models

Marginal models also known as population averaged models has the primary scientific objective of analyzing the marginal expectation of responses for a given explanatory variables. The population averaged parameters represent the average effects of a unit change in the explanatory variables for the whole population rather than individual subjects. In other words, marginal models are used to model the population averaged expectations of the dependent variable as a function of the independent variables across the entire population but not individual observations as in conditional models

Fixed effects models, which assume that all observations are independent of each other, are not appropriate for analysis of several types of correlated data structures, in particular, for clustered and/or longitudinal data. In clustered design, subjects are observed nested within a larger unit. For instance, the neonates are observed nested within a region. Marginal models are among the most statistical models widely used to model clustered or repeated data. The primary objective of marginal model is to analyze the population-averaged effects of the given factors in the study on the binary response variable of interest. This means that the covariates are directly related to the marginal expectations (Molenberghs & Verbeke, 2005). The marginal models fitted in this study would be

included are Generalized Estimating Equations (GEE) and Alternating Logistic Regression (ALR).

3.3.2.1. Generalized Estimating Equation

The GEE is a method of parameter estimation which represents an extension of the generalized linear models to accommodate correlated data. The method models the marginal expectations of the outcome variable such as the probability of a child dying before the first months of life. For binary data, a GEE approach is used to account for the correlation between responses of interest for subjects from the same cluster (Diggle et al., 1994). GEE is non-likelihood method that uses correlation to capture the association within clusters or subjects in terms of marginal correlations (Molenberghs & Verbeke, 2005).

For clustered as well as repeated measured data, (Liang & Zeger, 1986) proposed GEE which require only the correct specification of the univariate marginal distributions provided one is willing to adopt "working" assumptions about the correlation structure. The "working" assumptions as proposed by Liang and Zeger, included independence, unstructured, exchangeable and autoregressive AR (1). Independence and exchangeable working assumptions can be used in virtually all applications, whether longitudinal, clustered, multivariate, or otherwise correlated. Auto regressive AR (1) and unstructured correlation structures are less relevant for clustered data, studies with unequally spaced measurements or sequences with differing lengths (MolenberghsandVerbeke, 2005). Let $y_j = \left(y_{J1}, y_{J2}, \cdots y_{Jn_J}\right)'$ be the response values of observations from jth cluster, for j=1, 2, $\cdots m$ follows a binomial distribution i.e. that belongs to the exponential family with the density function of the form (3.1). Then, to model the relation between the response and covariates, one can use a regression model similar to the generalized linear models given by:

$$g(\Pi_j) = logit(\Pi_j) = X'_j\beta \qquad (3.2)$$

Where, $g(\Pi_j)$= logit link function, X_j= (n_j x p) dimensional vector of known covariates, β= (1 x p) dimensional vector of unknown fixed regression parameter to be estimated and $E(\mathrm{Y}_j) = \Pi_j$ expected values of the response variable from j^{th} cluster.

3.3.2.1.1. Parameter Estimation for GEE

As previously expressed GEE is not likelihood approach, rather it is quasi-likelihood based and estimated by solving estimating equations which consist of the working covariance matrix The score equation used to estimate the marginal regression parameters while accounting for the correlation structure is given by:

$$s(\beta) = \sum_{j=1}^{m} \frac{\partial \Pi_j}{\partial \beta'} \left[{A_j}^{1/2} R_j {A_j}^{1/2} \right]^{-1} \left(Y_j - \Pi_j \right) = 0 \qquad (3.3)$$

Where R_j is working correlation matrix and the covariance matrix of $Y_j$is decomposed into ${A_j}^{1/2} R_j {A_j}^{1/2}$ with A_j the matrix with the marginal variances on the main diagonal and zeros elsewhere and $Y_j$is multivariate vector of asymptotically normal response variables with mean vector π_j i.e. $Y_j \cong N(X_j \beta, V_j)$. An advantage of the GEE approach is that it yields a consistent estimator of, even when the working correlation matrix R_j is misspecified. However, severe misspecification of working correlation may seriously affect the efficiency of the GEE estimators (Molenberghs & Verbeke, 2005).

3.3.2.2. Alternating Logistic Regression (ALR)

The alternating logistic regression (ALR) method combines GEE1 for the regression parameters β with a modified logistic regression for estimating the association parameter α. In the standard GEE, the association parameter is a nuisance parameter which is not the case in the ALR where the marginal models are fitted based on the odds ratio such that inferences can be made not only about marginal regression parameters but on the pairwise association of the responses as well (Molenberghs and Verbeke, 2005). The ALR method proposed by Carey et al. (1993) used another approach to estimate the regression coefficients β and the association parameter α which can be efficient for both sets of parameters and avoids the computational problems of the standard GEE methods.

This method is very similar to that of GEE, in that they are both quasi-likelihood based and they account for dependency in the data. However, unlike GEE which measures the association among the observed data through the correlation structure; Alternating logistic regression (ALR) measures this association using the odds ratio, which is interpretable and more applicable for binary data. ALR extends beyond classical GEE in the sense that precision estimates follow for both the regression parameters β and the association parameters α. Moreover, with ALR inferences can be made, not only about marginal

parameters but also about pair wise associations between subjects as well (Molenberghs & Verbeke, 2005).

For a response variable Y_{ij} denoting the probability of a child dying before the first month of life for j^{th} respondent from the i^{th} region, let X_{ij} be a known set of covariates with $\mu_{ij} = p(Y_{ij} = 1)$ for $\mu_{ij} = E(Y_{ij})$;

$$logit\left(p(Y_{ij} = 1) = {X_{ij}}^{T}\beta\right) \tag{3.4}$$

The logit model above shows that, the effects of the covariates X_{ij} on the response Y_{ij} are averaged over all the regions. The marginal model therefore resembles a multiple logistic regression except that the parameter estimate β as well as the estimated standard errors of the parameters are corrected for by clustering for respondents in the same region as found in the EDHS 2011 data.

For cluster j=1, 2,…m, let $y_j = \left(y_{J1}, y_{J2}, \cdots y_{Jn_J}\right)'$ be a $n_j \times 1$ response vector with mean $E(y_j) = \pi_j$ and let ψ_{jkl} be the odds ratio between responses Y_{jk} and Y_{jl} $\left(1 \le k \le l \le n_j\right)$

Defined by:

$$\psi_{jkl} = \frac{p(Y_{jk} = 1, Y_{jl} = 1)p(Y_{jk} = 0, Y_{jl} = 0)}{\left(p(Y_{jk} = 1, Y_{jl} = 0)p(Y_{jk} = 0, Y_{jl} = 1)\right)} \tag{3.5}$$

$j = 1,2, \ldots m, k, l = 1,2, \ldots n_j$, where Y_{jk} and Y_{jl} represent the response values for mothers k and l respectively from the same cluster. Let γ_{jkl} be the log odds ratio between the outcomes Y_{jk} and Y_{jl} , let $\pi_{jk} = p(Y_{jk} = 1)$ and $\gamma_{jkl} = p(Y_{jk} = 1, Y_{jl} = 1)$, then the association of the two responses (Zeger et al, 1993) is defined by:

$$logit\left(Y_{jk} = 1/Y_{jl} = y_{jl}\right) = \gamma_{jkl}y_{jl} + log\left(\frac{\pi_{jk} - \nu_{jkl}}{1 - \pi_{jk} - \pi_{jl} + \nu_{jkl}}\right) \tag{3.6}$$

Assume $\gamma_{jkl} = \alpha$. Then the pair wise log odds ratio α is the regression coefficient in logistic regression of Y_{jk} on Y_{jl} as long as the second term on the right hand side in (3.4) is used as an offset. Generally $log(\psi_{jkl}) = \gamma_{jkl} = z'_{jkl}\alpha$, where z_{jkl} is a $q \times 1$ vector of covariates which specifies the form of the association between Y_{jk} and Y_{jl} .

3.3.2.2.1. Parameter Estimation of ALR

Since ALR also not maximum likelihood approach like GEE, parameter estimation is based on the score equation of the approximate likelihood that is based on quasi

likelihood approximation. Let μ_j be a vector with elements $\mu_{jkl} = E(Y_{jk}/Y_{jl} = y_{jl})$ and let R_j be the vector of residual with elements $R_{jkl} = Y_{jk} - E(Y_{jk}/Y_{jl} = y_{jl}) = Y_{jk} - \mu_{jkl}$. Let S_j be a vector of diagonal matrix with diagonal elements $\mu_{jkl}(1 - \mu_{jkl})$ and let W_j denote matrix $\frac{\partial \mu_j}{\partial \beta}$. Finally, let$A_j = Y_{j-}\pi_j$,$B_j = cov(Y_j)$, $C_j = \frac{\partial \pi_j}{\partial \beta}$.

Then the ALR parameter $\delta = (\beta, \alpha)$ is the simultaneous solution of the following unbiased estimating equations (Zeger et al, 1993).

$$U_\beta = \sum_{j=1}^{m} C'_j B_j^{-1} A_j = 0 \quad (3.7)$$

$$U_\alpha = \sum_{j=1}^{m} W'_j S_j^{-1} R_j = 0 \quad (3.8)$$

Estimating equation 3.5 and 3.6 are solving for β and α by using Gauss-Seidel procedure algorithm. ALR is computationally feasible for very large cluster.

3.3.2.3. Model Building for Marginal Models

Model selection is an important issue in almost any practical data analysis. A common problem is variable selection in regression given a large group of covariates (including some higher order terms) one needs to select a subset to be included in the regression model. Model selection is data analysis strategy, which leads to a search of best model. With this, we mean selecting the best subset of the covariates from the available covariates in the data.

3.3.3. Variable Selection Technique

To select significant variables, firstly under the GEE, model building strategy started by fitting a model containing all possible covariates in the data. This was done by considering two working correlation assumptions (exchangeable and independence). In order to select the important factors related to the response variable, the backward selection procedure was used. The strategy is called backward because we are working backward from our largest starting model to a smaller final model. In this case, the procedure is used to remove covariates with non-significant p-values. This means that variables that did not contribute to the model based on the highest p-value was eliminated sequentially and each time a new model with the remaining covariates was refitted, until we remained with covariates necessary for answering our research question. Finally, the two models were compared using model comparison techniques. Additionally, using the

same procedures, an ALR model, which provides information about pair wise association of observations between two different individuals within the same cluster, was fitted. It turned out that the model with selected covariates is found to be the most parsimonious model.

3.3.4. Model Comparison Technique

3.3.4.1. Quasi-Information Criterion (QIC)

In a condition, when the likelihood function cannot be fully specified, such as in the GEE case, the Akaike's Information Criterion (AIC) cannot be directly applied to select either the optimal set of explanatory variables or correlation matrix. As an alternative, one can use the modified Akaike's Information Criterion called Quasi Information Criteria (QIC), which is based on the quasi-likelihood function (Pan, 2001). QIC is derived from the AIC and conceptually similar. The quasi-likelihood function takes the following form (McCullagh & Nelder, 1989).

$$Q(\pi) = \int_y^{\pi} \frac{y-t}{\phi v(t)} dt \qquad (3.9)$$

Where$\pi = E(y)$,$v(y) = \Phi v(\pi)$, and Φ being the dispersion parameter. An equation for the $\text{QIC} = -2\text{Q}(\hat{\pi}, \text{I}) + 2\text{trace}\left[\Omega_{\text{I}}^{-1}\hat{\text{v}}_{\text{R}}\right]$ where I represent the independent correlation structure (diagonal matrices) and R is the specified working correlation structure.

The P-dimensional structure Ω_{I}^{-1} and $\hat{\text{v}}_{\text{R}}$ are variance estimators of the regression coefficients under the correlation structure I and R respectively. The QIC value is computed based on the quasi-likelihood estimate and is used to select the candidate explanatory variables. The model with the smallest QIC value for all correlation structures is considered as the best candidate model.

The generalized Wald test: is used to compare models with different subsets of the regression parameters, i.e. to select the candidate covariates. That is, one can use the generalized Wald tests to test the joint null hypothesis that a set of regression parameters s are equal to zero. In general, for any matrix L a test for hypothesis can be written as follows, $H_0: L_\beta = 0 \; vs \; H_1: L_\beta \neq 0$,Where L is a p x q indicator matrix of ones and zeros. Here, p is equal to the number of parameters in the full model (including the intercept) and q equals the number of parameters in the generalized Wald test (that is, the difference in parameters between the full and reduced model). The Wald statistic is a quadratic form defined as follows:

$W^2{}_{stat} = \beta^t L^t (L\ var(\beta) L^t)^{-1} L\beta$. It is distributed as χ^2 with q degrees of freedom under the null hypothesis.

In addition to select the appropriate working correlation structure, the two models with exchangeable and independence working correlation were compared via their naïve (model based) and robust (empirical) standard error estimates and the one with the closest empirical and model based standard error estimates was preferred (Molenberghs & Verbeke, 2005). Moreover, unless one expects dramatic differences among the correlations, using the exchangeable working correlation structure is recommended (Agresti, 2007).

3.3.5. Cluster Specific (Subject Specific) Models

When interest is in the marginal or population-averaged models to analysis the relationships of the covariates to the dependent variable for an entire population, marginal models as discussed in previous section are preferred. However, in most biomedical and biological data problems, interest often lies in understanding the response of individual patient characteristics and how this response is influenced by a given set of possible covariates (Myers et al.,2010). This proves even to be essential when individual interventions may be necessary. Cluster specific models are useful in such cases. Cluster specific models differ from the marginal models by inclusion of parameters that are specific to clusters or subjects within a population. Consequently, random effects are directly used in modeling the random variation in the dependent variable at different levels of the data.

3.3.5.1. Generalized Linear Mixed Model (GLMM)

Generalized linear models (GLM) is one parts of subject specific models which extends ordinary regression by allowing non-normal responses and a link function of the mean. The generalized linear mixed model is a further extension that permits random effects as well as fixed effects in the linear predictor (Agresti, 2002). Let y_{ij} denote the response of i^{th} individual mother from j^{th}cluster where $i = 1,2, \dots n_j$ and $y_j\ the\ n_j$ dimensional vector of all measurements available for cluster j. Let $f(b_j/D)$ be the density of the $N(O, D)$ distribution for the random effects b_j. Assumed conditionally on q-dimensional random effects $b_j$to be drawn independently from $N(O, D)$, the outcomes y_{ij} of Y_j are independent with the density of the form

$$f_j(y_{ij}/b_j, \beta, \phi) = exp\{\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi)\} \quad (3.10)$$

Then the generalized linear mixed model (Molenberghs and Verbeke, 2005); with logit link is defined as

$$logit(\pi_{ij}) = X'_{ij}\beta + Z'_{ij}b_j, \qquad j = 1,2,\dots,m \tag{3.11}$$

Where $E(Y_{ij}/b_j) = \pi_{ij}$ is the mean response vector conditional on the random effects b_j, for mothers in cluster j and, X_{ij} andZ_{ij} are p-dimensional and q-dimensional vectors of known covariate values. The random effects b_j are assumed to follow a multivariate normal distribution with mean 0 and covariance matrix D.

3.3.5.2. Parameter Estimation for GLMM

Random-effects models can be fitted by maximization of the marginal likelihood, obtained by integrating out the random effects. Such likelihood may involve high-dimensional integrals that cannot be evaluated analytically. The likelihood of the data expressed as a function of unknown parameters is

$$L(\beta, D, \phi) = \prod_{j=1}^{m} f_j(y_j/\beta, D, \phi) = \prod_{j=1}^{m} \int \prod_{j=1}^{n_j} f_{ij}(y_{ij}/b_j, D, \phi) f(b_j/D)\, db_j \tag{3.12}$$

It is the integral over the unobserved random effects of the joint distribution of the data and random effects. The problem in maximizing (3.9) is the presence of m integrals over the q dimensional random effectsb_j. With Gaussian data, the integral has a closed form solution and relatively simple methods exist for maximizing the likelihood or restricted likelihood. With non-linear models, numerical techniques are needed. The Laplace method (Molenberghs & Verbeke, 2005) has been designed to approximate integrals of the form:

$$I = \int e^{Q(b)}\, d(b) \tag{3.13}$$

Where Q(b) is a known, unimodal, and bounded function of a q-dimensional variable b. Let($\hat{b}$) be the value of b for which Q is maximized. Then the second order Taylor expansion of $Q(b)$is the form

$$Q(b) \approx Q(\hat{b}) + \frac{1}{2}(b - \hat{b})' Q''(\hat{b})(b - \hat{b}) \tag{3.14}$$

Where, $Q''(\hat{b})$ is the matrix of the second order derivative of Q, evaluated at $\hat{b}$. Replacing $Q(b)$ in (3.10) by its approximation in (3.11), we obtain

$$\boldsymbol{I \approx 2\pi^{\frac{q}{2}}\left|-Q\left(\hat{b}\right)\right|^{\frac{-1}{2}} e^{Q(\hat{b})}} \quad (3.15)$$

Clearly, each integral in (3.9) is proportional to an integral of the form (3.10) for functions $Q(b)$ is given by:

$$\boldsymbol{Q(b) = \phi^{-1}\sum_{i=1}^{n_j}\left[y_{ij}\left(x'_{ij}\left(x'_{ij}\beta + z'_{ij}b\right) - \psi\left(x'_{ij}\beta + z'_{ij}b\right)\right)\right] - \frac{1}{2}b'D^{-1}} \quad (3.16)$$

This is called the Laplace's method or approximation of integrands. Note that the mode $\hat{\boldsymbol{b}}$ of Q depends on the unknown parameters β, $\boldsymbol{\phi}$ and D, such that in each iteration of the numerical maximization of the likelihood, $\boldsymbol{b}$ will be recalculated conditionally on the current values for the estimates for these parameter.

3.3.5.3. Model Comparison in GLMM

This study will be used Likelihood ratio test and Information criteria to select the best model based on the values of asymptotic estimations.

Likelihood Ratio Test:

The likelihood ratio is a function of log likelihood and used in significance testing. The likelihood ratio test is a test of the significance of the difference between the likelihood ratio (-2LL) for the researcher's model minus the likelihood ratio for a reduced model.

In order to decide on the best of the two random effects models, two models will be fitted, one with the two random intercepts (between and within regional variations) and another with one random intercept (within regional variation). One can use the approximate restricted maximum likelihood ratio test (LRT) to compare these two models (Myers et al., 2010). Let $\boldsymbol{LR_{full} = -2LL_f}$ and $\boldsymbol{LR_{redu} = -2LL_r}$. Then the likelihood ratio statistic is given by:

$\boxed{\boldsymbol{\lambda = LR_{full} - LR_{redu}}}$ Where LL_f log likelihood is value for full model and LL_r is log likelihood value for reduced model. The asymptotic null distribution of the likelihood ratio test statistic is a chi-square distribution with degrees of freedom equal to the difference between the numbers of parameters in the two models.

Akaike's Information Criterion (AIC)

AIC is a measure of goodness of fit of an estimated statistical model. It is not a test on the model in the sense of hypothesis testing; rather it is a tool for model selection. The AIC penalizes the likelihood by the number of covariance parameters in the model, therefore

$$\boldsymbol{AIC = -2\log(L) + 2p} \tag{3.17}$$

Where, L is the maximized value likelihood function for the estimated model and p is the number of parameters in the model. The model with the lowest AIC value is preferable

3.3.5.4. Model Checking Technique

In GLMM, it is assumed that the random effects are normally distributed and uncorrelated with the error term. Normality of the random effects is assessed using normal plot of each random effect. Normal Q-Q plot of estimated random effects is an important method for checking the normality (Myers et al., 2010).

4. RESULT AND DISCUSSION

4.1. Descriptive Result

The data extracted from the Ethiopia demographic health survey 2011 is represented in Table 4.2 below. Total of 2604 mothers from nine regional states and two city administrations in Ethiopia were eligible for this study. Among these eligible mothers 385 (14.78%) of them were mothers whose their baby died in the first months of life, whereas 2219 (85.22) of them were mothers whose their baby alive within the same period.

Table 4.1: Neonatal Mortality of 2011 EDHS data

Child Alive	Frequency (%)
Yes (0=for child alive)	2219 (85.22)
No (1=for child died)	385 (14.78)
Total	2604 (100)

Table 4.2 presents basic descriptive information that summarizes the associations between the risk factors and neonatal mortality. From the table the proportion of neonatal mortality (PNM) for the women aged 20 to 29 were slightly higher (7.60%) compared to women aged 40 to 49 (1.15%). Considering the wealth index PNM for the women of poor level were higher (7.80) compared to middle (4.57%) and rich level (2.42%) women. For the child level factors; the PNM for males (8.37%) were higher compared to females (6.41%). Also PNM for a single birth were higher (12.48%) compared to multiple birth (2.30%). The PNM for women deliver at home (12.52%), while PNM deliver at a health facility were (2.27%). The PNM for women living in rural areas were higher (12.83%) compared to women living in urban areas (1.96%). Also, the PNM for non-educated women were slightly higher (10.06%) compared to women attained primary school (4.22%) and secondary and above school (0.50%). Considering Age of mother at first birth the PNM for mother aged less than 20 were higher (9.91%) compared to mother aged 20 to 29 (4.72%) and above 29 (0.15%) respectively. The PNM for infants born in the order two to four were higher (6.61) compared to first birth (3.30%) and five and above order birth (4.88%). Also, NM for mother giving birth in the interval of 24 to 47 months were higher (7.24%) compared to mothers giving birth in the interval of less than 24 months (4.48%) and above 47 months (2.19%). Lastly, the PNM for infants of average size at birth were higher (4.22%) compared to infants of smaller than average size

(0.77%). Also, PNM for infants of very large, larger than average, and very small size were 3.84%, 2.19%, and 3.76%.

Table 4.2: Summary of descriptive statistics for neonatal mortality (EDHS, 2011)

Attributes	Alive (%)	Died (%)	Total (%)
Age			
15-19	34(1.31)	191(7.33)	225(8.64)
20-29	198(7.60)	1228(42.16)	1426(45.76)
30-39	123(4.72)	683(26.23)	806(30.95)
40-49	30(1.15)	117(4.49)	147(5.65)
Sex			
Male	1117(42.90)	218(8.37)	1335(51.27)
Female	1102(42.32)	167(6.41)	1269(48.73)
Birth order			
One	408(15.67)	86(3.30)	494(18.97)
Two to Four	1012(38.86)	172(6.61)	1184(45.47)
Five and above	799(30.68)	127(4.88)	926(35.56)
Birth size			
Very large	303(11.64)	100(3.84)	403(15.48)
> average	215(8.26)	57(2.19)	272(10.45)
Average	897(34.45)	110(4.22)	1007(38.67)
< average	202(7.76)	20(0.77)	222(8.53)
Very small	602(23.12)	98(3.76)	700(26.88)
Birth interval			
Less than 24 months	244(11.62)	94(4.48)	338(16.10)
24 to 47 months	1092(52.00)	152(7.24)	1244(59.24)
Above 47 months	472(22.48)	46(2.19)	518(24.67)
Mother's level of education	1461(56.11)	262(10.06)	1723(66.17)
No education	623(23.92)	110(4.22)	733(28.15)
Primary	135(5.18)	13(0.50)	148(5.68)
Secondary and above			
Residence			

Rural	1839(20.62)	334(12.83)	2173(83.45)
Urban	380(14.59)	51(1.96)	431(16.55)
Wealth			
Poor	1103(42.36)	203(7.80)	1306(50.15)
Middle	735(28.23)	119(4.57)	854(32.80)
Rich	381(14.63)	63(2.42)	444(17.05)
Multiple birth			
Single	2177(83.60)	325(12.48)	2502(96.08)
Multiple	42(1.61)	60(2.30)	102(3.92)
Place of delivery			
Home	1890(72.58)	326(12.52)	2216(85.10)
Health facility	329(12.63)	59(2.27)	388(14.90)
Total	2219(85.22)	385(14.78)	2604(100)

4.2. Collinearity Diagnostic Test

Before building the model for factors of neonatal mortality, the set of independent variables must be tested for collinearity. Table 4.2 below displays the Tolerance and the Variance Inflation Factor (VIF) of the independent variables to be used for predicting factors that affect neonatal mortality. The two statistics used for collinearity test are the Tolerance and the VIF. From Table 4.3, the Variance Inflation Factor (VIF) for each independent variable is less than 10 meaning that, there is no interaction (linear relationship) between the independent variables that might affect the results in the analysis. This implies that all the independent variables in table 4.2 are fit to be used in developing the model for neonatal mortality in Ethiopia.

Table 4.3: Collinearity test of independent variables (EDHS, 2011)

Independent variables	Collinearity statistics	
	Tolerance	VIF
Age of mother	0.51139	1.95544
Residence	0.59555	1.67912
Mother's level of education	0.80354	1.24450
Wealth	0.73010	1.3696

Age at birth	0.81399	1.22851
Birth order	0.55539	1.80054
Multiple birth	0.97922	1.02122
Sex	0.98768	1.01248
Birth interval	0.90773	1.10165
Place of delivery	0.67495	1.48160
Birth size	0.95239	1.04999

4.3. Statistical analysis of marginal models

In this section, NM has been analyzed using marginal models including generalized estimating equation and alternating logistic regression models.

4.3.1. Analysis of Generalized Estimating Equations (GEE)

In the methodology that is termed generalized estimating equations, the user may impart a correlation structure that is often called a working correlation matrix. Before selecting the correct correlation structure, consider the model building strategy (variable selection). Under the GEE, model building strategy is started by fitting a model containing all possible covariates in the data. This was done by considering two different working correlation assumptions (exchangeable and independence). In order to select the important factors related to neonatal mortality, the backward selection procedure was used. The full logit model for the probability of getting dying i^{th} child from j^{th} cluster (Region) was fitted as:

$$\begin{aligned} logit(\pi_{ij}) = {} & \beta_0 + \beta_1 Age_1 + \beta_2 Age_2 + \beta_3 Age_3 + \beta_4 Res_U + \beta_5 Mled_P + \beta_6 Mled_{S+} \\ & + \beta_7 Wealth_M + \beta_8 Wealth_{Pr} + \beta_9 BF_Y + \beta_{10} Bord_{5+} + \beta_{11} Mbirth_m \\ & + \beta_{12} Sex_{ma} + \beta_{13} Pbint_{\mathrm{fm}} + \beta_{14} Pbint_{\mathrm{nm}} + \beta_{15} Szbirth_{\mathrm{la}} + \beta_{16} Szbirth_{\mathrm{a}} \\ & + \beta_{17} Szbirth_{\mathrm{sa}} + \beta_{18} Szbirth_{\mathrm{vs}} + \beta_{19} Pdel_{\mathrm{H}} \end{aligned}$$

Where π_{ij} is the probability of a child dying before the age of twenty eight days. The subscripts in each covariates are defined as,1=20-29,2=30-39,3=40-49,U=urban, P=primary, S+=secondary and above=middle, Pr =poor,Y=yes, 5+=five and higher order,m=multiple,ma=male,fm=24-47 month,nm=above 47 month,la=larger than average, a=average, sa=smaller than average,vs=very small=Home.

After fitting the model, covariates with the largest p-value of Wald test is removed and refitted the model with the rest of the covariates sequentially. Then, place of residence, maternal education, wealth index of household and place of delivery were the covariates excluded from the model; with Wald test p-value for the given covariates are large (P-value > 0.05) which is found in the appendix. The QIC values of full model (which is found in appendix) and reduced models are 915.5822 and 911.5799 respectively. Then it turned out that the model with age, type of place of residence, mothers' education level, multiple births, birth order, birth size as covariates was the most parsimonious model.

Table 4.4: Empirical and model based standard errors for two proposed working correlations (GEE)

	Exchangeable			Independent		
coefficient	Estimates	Empirical (S.E)	Model based (S.E)	Estimates	Empirical (S.E)	Model based (S.E)
β_0	3.7954	(1.2043)	0.8721	3.7322	1.2023	0.8480
β_1	3.4673	(0.7640)	0.7139	3.4958	0.7705	0.7223
β_2	1.9447	(0.4377)	0.4271	1.9352	0.4381	0.4249
β_3	-1.8298	(0.36820	0.2787	-0.8318	0.3725	0.2832
β_4	-4.5616	(0.2336)	0.2279	-4.5648	0.2344	0.2278
β_5	-1.6099	(0.2795)	0.3027	-1.5981	0.2794	0.3022
β_6	3.2126	(0.3140)	0.3869	3.2086	0.3152	0.3869
β_7	-2.3758	(0.3095)	0.2967	-2.3680	0.3094	0.2961
β_8	-1.0222	(0.2717)	0.2489	-1.0139	0.2716	0.2477
β_9	-0.6866	(0.3067)	0.2610	-0.6702	0.3076	0.2613

Finally, as a customary, comparison of empirical and model based standard errors for the parameter estimates obtained based on the given working correlation assumptions (in this study exchangeable and independence) was performed using selected covariates. The correlation structure with the model based and empirical standard errors are closest to each other, is referred to be the best assumption correlation structure.

Moreover, since no dramatic differences among the correlations, using the exchangeable working correlation structure is recommended, as stated in agrestic book (Agrestic, 2007). In addition, the empirically corrected standard errors for exchangeable correlation structure are somewhat smaller than their counterpart under the independence assumptions.

Then, from table 4.5, exchangeable working correlation assumption was found to be plausible since the two standard errors were closer each other with (α = -0.0113).

Therefore, the final proposed generalized estimating equation model for neonatal mortality was given as:

$$logit(\pi_{ij}) = \beta_0 + \beta_1 Age_1 + \beta_2 Age_2 + \beta_3 Age_3 + \beta_4 Res_U + \beta_9 BF_Y + \beta_{10} Bord_{5+} + \beta_{11} Mbirth_m + \beta_{13} Pbint_{\text{fm}} + \beta_{14} Pbint_{\text{nm}} + \beta_{15} Szbirth_{\text{la}} + \beta_{16} Szbirth_{\text{a}} + \beta_{17} Szbirth_{\text{sa}} + \beta_{18} Szbirth_{\text{vs}}$$

Parameter estimates and their corresponding empirically corrected standard errors alongside the p-values from the final GEE model (model 4.2) are presented at table 4.4

Table 4.5: Parameter estimates (empirically corrected standard errors) for GEE

Effects	Coefficients	Standard errors	Confidence intervals Lower Upper
Intercept	3.7954	0.8721	2.086 5.5046
Age			
15-19			
20-29	3.4673	0.7139	2.0681 4.8666
30-39	1.9447	0.4271	1.1077 2.7817
40-49			
Multiple birth			
Single	3.2126	0.3027	2.4543 3.9709
Multiple			
Age at birth			
<20	-2.7818	0.6628	-4.2260 -1.4343
20-29	-1.7755	0.6880	-3.1538 -0.4659

>29			
Residence			
Urban	-0.8296	0.3359	-1.3760 -0.2836
Rural			
Birth size			
Vlarge	-0.4555	0.2714	-0.9874 0.0764
>Average	-0.6866	0.2439	-1.1982 -0.1750
Average	0.3747	0.2439	-0.1105 0.8600
<Average	0.4670	0.4239	-0.8638 1.2979
Vsmall			
Birth order			
2-4	-1.6099	0.2279	-2.2033 -1.0166
Birth interval			
<24	-2.3758	0.3869	-2.9574 -1.7942
24-47	-1.0222	0.2967	-1.5101 -0.5343
>47			
QIC value	911.5799		

Source: EDHS, 2011,* p<0.05 was statistically significant. CI=confidence interval, se=standard error, Lower=lower class limit, Upper=upper class limit

4.3.2. Analysis of Alternating Logistic Regression model (ALR)

Model building for ALR is follows the same procedure in GEE model building strategy. First ALR model was fitted using all proposed covariates. Then the covariate with the large pvalue is removed. Mothers' level of education, place of delivery, wealth and sex of a child are removed covariates with Wald test (p-value > 0.05). The QIC values of both saturated and reduced models are given by 915.5656 (found in appendix) and 911.5202 respectively.

Therefore, the reduced model with the rest of seven covariates was considered as the best candidate model. Using the selected covariates and the association parameter α, alternating logistic regression (ALR) model that provides information about pair wise association of observations between two different individuals within the same cluster was fitted. Therefore, the final proposed ALR model included the association parameter for neonatal mortality is given as follows:

$$logit(\pi_{ij}) = \beta_0 + \alpha + \beta_1 Age_1 + \beta_2 Age_2 + \beta_3 Age_3 + \beta_4 Res_U + \beta_9 BF_N + \beta_{10} Bord_{5+} + \beta_{11} Mbirth_m + \beta_{13} Pbint_{\text{fm}} + \beta_{14} Pbint_{\text{nm}} + \beta_{15} Szbirth_{\text{la}} + \beta_{16} Szbirth_{\text{a}} + \beta_{17} Szbirth_{\text{sa}} + \beta_{18} Szbirth_{\text{vs}}$$

Parameter estimates and their corresponding empirically corrected standard errors alongside the p-values from the final ALR model are presented in table 4.6.

4.3.3. Comparison of GEE and ALR Models

Since the likelihood function does not fully specified in marginal models, model comparison is based on quasi likelihood criteria (QIC) which is the modified AIC criteria. From table 4.3 and table 4.4, we found that the QIC values are 911.5799and 911.5202 for the GEE and ALR respectively. However, the empirically corrected standard errors for ALR model are somewhat smaller than their counterpart under the GEE model. This implies that the ALR fits the data with small disturbance than GEE. Moreover, ALR extends beyond classical GEE in the sense that precision estimates follow for both the regression parameters β and the association parameters α. We were also in a position to emphasize that the association is strongly significant (P < 0.0001), provided it has been correctly specified, a declaration we could not make in the corresponding exchangeable GEE analysis. Therefore, we can conclude that ALR is the better model for explaining the marginal association between survival status of neonates and the selected predictor variables. Thus, our interpretation of parameters is based on the final proposed ALR model. Overall, parameter estimates under ALR are slightly less than those of GEE. This difference in parameter estimates from the two models might be due to the fact that ALR takes the associations into account, where as GEE not consider the association parameter in the model.

Table 4.6: Parameter estimates (empirically corrected standard errors) from ALR

Effects	**Level**	**Parameter**	**Estimates (S.E)**	**95% conf.int**	**P-value**
Intercept		β_0	3.8197(0.8868)	(2.0816,5.5579)	<.0.0001*
Age	15-19	β_1	3.4392(0.7116)	(2.0446,2.8339)	<.0.0001*

	(ref)				
	20-29	β_2	1.9318(0.4273)	(1.0943,2.7692)	<.0.0001*
	30-39	β_3	0.6971(0.3361)	(0.0381,1.3558)	0.0359*
Residence	Urban	β_4	-0.8209(0.2794)	(-1.3685,-0.2733)	0.0033*
Age at first birth	<20	β_5	-2.7702(0.6918)	(-4.1260,-1.4143)	<0.0001*
	20 to 29	β_6	-1.7749(0.6780)	(-3.1038,-0.4459)	<0.0001*
Birth order	2-4	β_7	-1.6007(0.3024)	(-2.1933,-1.0081)	<.0.0001*
Multiple birth	Single	β_8	3.1926(0.3851)	(2.4378,3.9474)	<.0.0001*
Birth interval	<24	β_9	-2.3717(0.2963)	(-2.9525,-1.7909)	<.0.0001*
	24-47	β_{10}	-1.0185(0.2477)	(-1.5041,-0.5330)	<.0.0001*
Birth size	Very large	β_{11}	-0.4517(0.2707)	(-0.9822,0.0788)	0.0952
	> average	β_{12}	-0.6782(0.2609)	(-1.1894,-0.1669)	0.0093*
	Average	β_{13}	0.3803(0.2470)	(-0.1039,0.8644)	0.1237
	< average	β_{14}	0.4659(0.4229)	(-0.3630,1.2947)	0.2707
		α	-0.1317(0.2082)	(-0.5397,0.2764)	0.5271

QIC	911.5202

4.3.4. Parameter Interpretation of Marginal Models

Table 4.6 presents parameter estimates and their corresponding empirically corrected standard errors alongside the p-values from ALR model. Each parameter reflects the effect of factor on the log odds of the probability of neonatal mortality, statistically controlling all the other covariates in the model. Then, the odds ratio of variables is calculated as the exponent of β_j i.e odds ratio = exp (β_j). The ALR analysis from table 4.4 suggests that, age is significantly related to neonatal mortality. After controlling all other variables in the model, the odds ratio of neonatal mortality whose mothers aged 15-19 years is exp(β_1) = exp(3.43) = 31.2 (95% CI: 7.73, 125.7) times higher than compared to those aged between 40 and 49, and the odds ratio for mothers aged 20-29 is exp(1.93) = 6.9 (95% CI: 2.99, 15.95) times higher when compared with mothers aged 40-49 (the reference group),also the odds ratio for mothers aged 30-39 is exp(0.69)=2 (1.04,3.88) times higher when compared with those aged 40-49.

As we have seen from the result of the ALR model, type of place of residence is statistically significant on neonatal mortality. The odds ratio of neonatal mortality of mothers living in urban area is exp (-0.8209) = 0.44 (95% CI 0.25, 0.76) which means that when all other covariates are held constant.

This means that the probability of neonatal mortality of mothers who lives in urban area is around 44% times less likely than mothers who live in rural area. There is also a strong association between age at first birth and the neonatal mortality. This implies that, after adjusting all other predictor variables in the model, the estimated odds ratio of neonatal mortality for mother's aged less than 20 is given as exp (-2.77) =0.063 (95% CI: 0.016, 0.243) times and the odds ratio for age group of 40 to 49 is *exp*(-1.77)= 0.17 (95% CI: 0.045, 0.640) times lower when compared with the reference category age 15 to 19. This means that neonatal mortality is decreased by 6.3% and 17% for early and middle age mothers respectively compared to old age group mothers.

From the estimated ALR model in the Table 4.6 above birth order had also significant effect on neonatal mortality (negatively related). The odds ratio of neonatal mortality for

second to fourth order birth is exp(-1.60) =0.202 (95% CI: 0.112,0.365) times lower than first order birth(referent group). This implies that, after controlling other variables in the model the respondents of 2 to 4 order births less likely to lose an under-one month than respondents in the first order.

The logit of neonatal mortality positively related with multiplicity of birth. The odds ratio of neonatal death is extremely high, which is exp(3.1926) = 24.35 (95% CI: 11.45, 51.80) times higher for single birth mothers as compared with mothers belongs to greater than equal to two births . Similarly the birth interval is significantly affect neonatal mortality. In the same fashion estimated odds ratio of the neonatal mortality is exp (-2.3717) = 0.093 (95% CI: 0.052, 0.167) times lower than for birth interval less than 24 month as compared with reference group of above 47 month interval. This implies that the neonatal mortality is reduced by 9.3% for birth interval less than 24 month. The estimated odds ratio of neonatal mortality for mothers who get birth by 24 through 47 month birth interval is exp(-1.0185)=0.3611 (95% CI: 0.222,0.587) times lower as compared to mothers who get birth by 47 month interval. This implies that the neonates of mothers experienced birth in the interval of 24 to 47 is 36% less likely to die than above 47 month birth interval counterpart.

Birth size also another influential predictor variable for neonatal mortality. The odds ratio of neonatal mortality for mothers whose their neonates' birth size larger than average is exp (-0.6782) =0.51 (95% CI: 0.30, 0.85) times lower as compared to mothers of very small size neonate. This implies that the neonate who are very large in size when they born is less likely to die than the reference group (very small).

The ALR model also presents the estimated constant log odds ratio (alpha) which, provide information about the association between individual observations within the same cluster. Table 4.13 shows that, the association parameter (log odds ratio) was -0.1317 with a p-value of 0.5271. There is the indication that, the association within the households is not significant at 5% level of significance, hence household effect is not a contributing risk factor for neonatal mortality in the sample. This means that, neonatal mortality is not significantly influenced by respondents found in the same household in the EDHS 2011 data.

4.4. Analysis of Generalized Linear Mixed Model (GLMM)

4.4.1. Model Building in GLMM

Under the GLMM, model fitting began by adoption of the marginal model covariates. Additionally, the model also included the random effects in this case, random intercepts to address the between and within-regional correlations. First, all main effect covariates and the two random intercepts model were fitted and as usual, non-significant covariates were removed sequentially starting from variables with highest p-value for fixed effect covariates. Then the saturated models for GLMM were fitted as follows where, $v_j \; and \; u_{ij}$ two random intercepts.

$$\begin{aligned} logit(\pi_{ij}) = {} & \beta_0 + \beta_1 Age_1 + \beta_2 Age_2 + \beta_3 Age_3 + \beta_4 Res_U + \beta_5 Mled_P + \beta_6 Mled_{S+} \\ & + \beta_7 Wealth_M + \beta_8 Wealth_{Pr} + \beta_9 Agbirth_1 + \beta_{10} Agbirth_2 \\ & + \beta_{12} Bord_{\geq 5} + \beta_{13} Mbirth_m + \beta_{14} Sex_{ma} + \beta_{15} Pbint_1 + \beta_{16} Pbint_2 \\ & + \beta_{17} Szbirth_2 + \beta_{17} Szbirth_3 + \beta_{18} Szbirth_4 + \beta_{19} Szbirth_5 + \beta_{20} Pdel_1 \\ & + v_j + u_{ij} \end{aligned}$$

In order to decide on the better of the two random effects models, two models were fitted, one the saturated model above with two random intercepts to estimate between and within regional variations and the other with one random intercept model to estimate within regional variation. AIC and Likelihood ratio test (LRT) were used to compare the two models to select an appropriate models.

Table 4.7: model comparison of one and two random intercept model

Models	AIC	BIC	LogLik	Deviance	$\boldsymbol{\sigma_W}$	$\boldsymbol{\sigma_B}$	P
With one random intercept	916.16	1046.1	-435.08	870.16	3.186e-09		
With two random intercept	911.13	1046.7	-431.57	863.13	0.0001421	0.4064521	0.008017

Where, σ_W and σ_B are within and between regional standard deviation respectively, and P is the p-value of the log likelihood ratio test of the two models.

As we have seen from table 4.7, the AIC value of two random intercept model is reduced from 916.16 to 911.13, the -2loglikelihood is reduced from -435.08 to -431.57 & the deviance of the model is reduced from 870.16 to 863.13. The small p-value of the log likelihood ratio test (P < 0.008017) also indicates that the model with two random intercept is parsimonious model.

Also when considered a model without random effects (i.e. simply the generalized linear model), it gives AIC value of 914.16 which is large as compared to the above two models with random effects. In addition, the likelihood ratio test at the bottom panel of table 4.6 in GLMM parameter estimate output also shows that the comparison of random effect model versus the ordinary logistic model (GLM) without random effects. The resulting p-value (P < 0.008017) of this test supports that considering the random effect model is essential. Therefore, we conclude that, the model with two random intercepts should be used to address the between and within-regional heterogeneity in the given data.

Next, the covariates for the fixed effect were assessed and the candidate covariates were selected by removing covariates starting from with highest p-value sequentially. Then the first removable covariate is work of mothers with the highest p-value (P = 0.8799) and refitted the reduced model with the remaining covariates. The AIC is reduced from 3795 to 3793 and the p-value of log likelihood ratio test (P = 0.8811) supports the reduced model is preferable one. The next removable variable is sex of the household leader with p-value (P = 0.1181) and refitted the reduced model. For this model, AIC is similar with the previous one but the likelihood ratio test indicates that the reduced model is better with the p-value (P=0.1228). In addition, the model with small number of covariates is considered to be preferable. Therefore, the final proposed GLMM for neonatal mortality is given as:

$$logit(\hat{\pi}_{ij}) = \beta_0 + \beta_1 Age_1 + \beta_2 Age_2 + \beta_3 Age_3 + \beta_9 Agbirth_1 + \beta_{10} Agbirth_2 + \beta_{11} BF_Y + \beta_{12} Bord_{\geq 5} + \beta_{13} Mbirth_m + \beta_{15} Pbint_1 + \beta_{16} Pbint_2 + \beta_{17} Szbirth_3 + \beta_{18} Szbirth_4 + \hat{v}_j + \hat{u}_{ij}$$

$$logit(\hat{\pi}_{ij}) = -2.16 - 1.48 Age_1 - 2.68 Age_2 - 3.35 Age_3 + 0.93 Agbirth_1 + 2.67 Agbirth_2 + 1.59 Bord_{\geq 5} - 3.26 Mbirth_m + 1.45 Pbint_1 + 2.49 Pbint_2 + 0.85 Szbirth_3 + 1.06 Szbirth_4 + v_j + \hat{u}_{ij}$$

The parameter estimates and standard errors of the GLMM are presented in table 4.8.

4.4.2. Parameter Interpretation of GLMM

Unlike in the marginal models, (GEE and ALR) where parameters are treated as population averages, in the GLMM analysis, parameter interpretation is based on specific subjects or cluster. The parameter interpretation is conditional on the random effects, which is common for all individual children in the same cluster.

Table 4.8: Parameter estimates (standard errors) and corresponding P-value for GLMM

Effects	**Level**	**Parameter**	**Estimates (S.e)**	**P-value**	**95% Conf.Int**
Intercept		β_0	-2.161916 (0.744568)	0.004	(-3.6292425,0.71058848)
Age	15-19(ref)		.	.	.
	20-29	β_1	-1.489494 (0.651702)	0.022	(-2.7668061,0.21218121)
	30-39	β_1	-2.689861 (0.705433)	0.000	(-4.0724855,-1.30723706)
	40-49	β_2	-3.355781 (0.797953)	2.61e-05	(-4.9197405,-1.79182172)
Residence	Rural(ref)				
	Urban	β_2	0.7680 (0.3372)	0.022734	(0.10719038, 1.4288798)
Age at birth	<20 (ref)	.	.	.	.
	20-29	β_3	0.934621 (0.226624)	3.72e-05	(0.4904459,1.37879616)
	30-39	β_4	2.674870 (1.176227)	0.023	(0.3695069,4.98023382)
Birth order	One (ref)	.	.	.	.
	2-4	.	.	.	.
	>=5	β_6	1.595931 (0.283535)	1.82e-08	(1.0402130,2.15164849)
Multiple	Single	.	.	.	.

birth	(ref)				
	multiple	β_7	-3.264505 (0.324251)	2e-16	(-3.9000249 ,2.62898581)
Birth interval	<24	.	.	.	.
	24-47	β_8	1.455070 (0.225822)	1.17e-10	(1.0124678,1.89767279)
	>47	β_9	2.490112 (0.317552)	4.45e-15	(1.8677220,3.11250126)
Birth size	Very large	.	.	.	.
	Average	β_{10}	0.856464 (0.269861)	0.002	(0.3275454,1.38538304)
	< average	β_{11}	1.066275 (0.430011)	0.013	(0.2234695,1.90908068)

Ref=reference

Given the same random effects b_j, the estimated odds ratio of neonatal mortality is exp (-1.49) = 0.225 (95% CI: 0.0633, 1.234) times lower for age group 20-29 , exp (-2.68) = 0.068 (95% CI: 0.017,0.27) times lower for age group 30-39 and exp(-3.35)= 0.035 (0.0073,0.167) times lower for age group 40-49 compared to mothers with age group 15-19 in the same j^{th}cluster keeping constant the other fixed effect variables in the model. This implies that the probability of neonatal mortality is 22.5%, 6.8 % and 3.5 more likely for mothers whose age group is 20-29 , 30-39 $ 40-49 respectively than with mothers whose age group is 15-19 in the same cluster at the given random effects. In the same way, the estimated odds ratio of neonatal mortality was exp (0.93) = 2.53 (95% CI: 1.63, 3.97) times higher for age at first birth 20 to 29 and exp(2.67) = 14.44 (95% CI: 1.45, 145.5) times lower for age at first birth 30 to 39 respectively compared with age group 15-19 in the same j^{th} cluster with constant random effect in the given cluster and the other fixed effect covariates in the model are constant. The estimated odds ratio of neonatal mortality for the higher order birth (birth order five and greater) is exp (1.59) = 4.90 (95% CI: 2.83, 8.58) times higher than mothers with first birth order in the same cluster. Similarly, the odds ratio of neonatal mortality is exp(-3.26)= 0.038 (95% CI: 0.020 ,13.87) times lower for mothers who experienced two and above births (multiple) than for mothers experienced single birth. This implies 3.8% of neonatal mortality occurs due to multiple births less likely when compared to single birth mothers at the same cluster with the same random effect.

Birth interval was also another a key factor for neonatal mortality. According to estimated odds ratio, considering less than two year (24 months) birth interval as reference group the estimated OR is exp (1.45) =4.26 (95% CI: 2.75, 6.67) times higher for mothers experienced birth by interval from 24 to 47 month (two to four year) when compared to mothers experienced birth less than two year interval keeping the same cluster with the same random effect as it is. In the same fashion the estimated odds ratio is exp(2.49)= 12.061 (95% CI: 6.488,22.488) times higher for mothers experienced birth in above 47 month interval as compared to mothers experienced birth less than 24 month, at the same cluster with the same random effect. The estimated odds ratio for neonatal mortality was also set up for neonates who were larger than average, average, very small and smaller than average in size during birth time. Accordingly, the estimated OR is exp(0.85) = 2.34 (95% CI: 1.39, 3.99) and exp(1.06)= 2.89 (95% CI: 1.25,6.75) times higher for mothers whose neonate was average and smaller than average in birth size respectively as compared to neonates who were very large in size (reference group) keeping the same cluster with the same random effect. The other category of birth size such as larger than average and very small size were non-significant. This means neonatal mortality was not affected by these factors.

Except the variable place of residence, the interpretation of other predictor variables can be done in a similar manner. Since clustering for 2011 EDHS was considered urban and rural area, parameter interpretation of the covariate, type of place of residence is at regional level random effects. Then, the odds ratio of neonatal mortality of mothers who lives in urban place is exp (0.7680) = 2.155 (95% CI: 1.113, 4.174) times higher than mothers who lives in rural area in the same region keeping constant other covariates and regional level random effects. This implies that the probability of neonatal mortality for urban area is around two folds more likely than rural mothers in the given region.

4.4.3. Model diagnostic for GLMM

The Q-Q plot from the following figure in first panel verifies that the residuals are close to normally distributed and symmetric around zero. Thus, it meets the assumption of the distribution of error terms. As well, to the above, the non-linearity of the Q-Q plot confirms the model is not linear. Residuals versus observation CLID number plot panel two, also suggested that the residuals are symmetric around zero (i.e. positive and

negative residuals are almost equal). Q-Q plots for normality of random effects at regional and cluster levels are also given in the figure at panel three and four, and illustrates that the random effects are normally distributed with mean zero and variance covariance matrix D. Thus, the fitted GLMM model is fine for the given data.

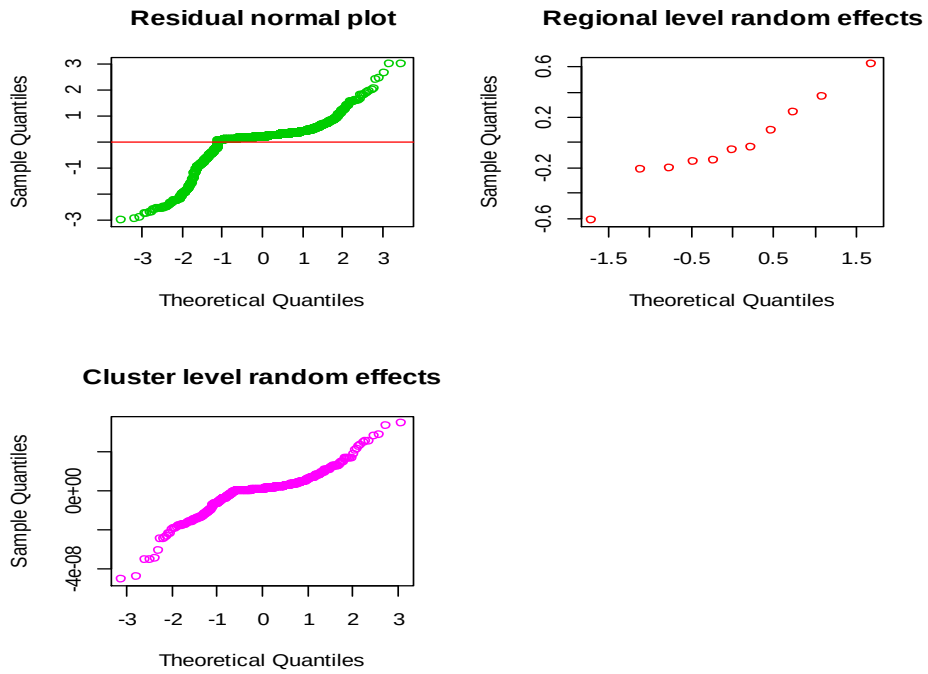

Figure 4.1: diagnostic plot for generalized linear mixed model (EDHS, 2011)

4.5. Discussion

The main aim of this study was to identify significant factors of Neonatal mortality in Ethiopia using the nationally representative 2011 EDHS data. In the present study the variables significantly affecting NNM were multiplicity of birth, age at first birth, breastfeeding, birth order, birth interval, birth size and mother's age at birth, whereas place of delivery,sex of a child, wealth index and mother's level of education seen as non-significant factors.

This study was aimed at modeling the determinants of neonatal mortality in Ethiopia. As a preliminary analysis, assortments of summary statistics were employed to explore the association between the response variable of interest and available covariates. It should be well-known that there is inconsistency in the conclusion from the analysis of various summary statistics, which might be due to the fact that they make use of varying amount of information, which determines the power of their inferences. Thus, the analysis was extended to other statistical methods to account for the clustered nature of correlated observations. The data were then analyzed using two model families one with marginal models (GEE and ALR), and the other is random effects model (Generalized linear mixed model).

Two proposed working correlation structures, exchangeable and independence correlation assumptions were taken for the comparison, in GEE model-building strategy. The model with exchangeable working correlation structure was found to be better fits the data than independence. This supports that considered the clustering nature of the data was essential for the analysis and the dependency of individuals for the given data. In addition, ALR was fitted for simultaneously regress the response variable on explanatory variables as well as association among responses in terms of pair wise odds ratio.

Two models from marginal model families were compared in order to assess which model is efficiently explain the relations between response and explanatory variables as well as to evaluate that whether considering pair wise association is important. After then, ALR model was selected as best model and the model shows that there is a positive pair wise association between responses. This is supported the idea explained by Zeger *et al*, alternating logistic regression is reasonably efficient relative to GEE (Zeger *et al*, 1993).

The purpose of GLMM was to evaluate within and between regional variations of NM in Ethiopia. Two models was fitted one with only one random intercept model to assess only

within regional variation and other with two random intercepts model, in order to account within and between regional variations. Additionally, generalized linear model was fitted as the sake of comparison whether including random effects in the analysis is important or not. The three models were compared using the AIC value followed by likelihood ratio test and we got a model with two random intercept was favorable. This demonstrates that, accounting within and between regional variations for the analysis of NM should be vital and, indicates within and between regional heterogeneity in NM. This finding is supported by the explanation or suggestion of Antonio & Beirlant (2011). Even though the two model families are different and their comparability may not be meaningful as they have different parameter interpretations and estimations, parameter estimates obtained from GLMM are generally bigger in absolute values than those from marginal models (GEE and ALR) similar with Agresti (2007).

From all the fitted model age of respondents (mothers), multiplicity of birth, birth interval, and birth order, age at first birth, residence, and birth size were found to be significant factors of neonatal mortality; whereas wealth, mothers educational level, sex of a child and place of delivery were non-significant factors.

A study in Ghana titled "Modeling the risk factors of neonatal mortality in Ghana using logistic regression "found that for the mother level factors two factors out of the three factors namely age of mother and wealth index were significant factors as contributing to neonatal mortality in Ghana (Kojo,2012). For the child level factors which included size of child, sex of child and whether the child was a twin or not, none of these factors seen significant as causing neonatal mortality. For the environmental level factors it was found that only the region (site of delivery) of the respondent was significant the findings of this study revealed that the risk of neonatal death was higher among twin or multiple births than among single births, which was comparable to another study (Asefa, Drewett, Tessema, 2002) in southwest Ethiopia suggested that twins were much more likely to die than singletons, even after taking their birth weight into account.

A study in Brazil (Araujo, Bozzetti, and Tanaka, 2000) also found that multiplicity of birth was significantly associated with NNM. One possible explanation for this observed association could be that multi-foetal pregnancy and multiple births including twins and higher order multiples such as triplets and quadruplets were at high-risk during both

pregnancy and birth. These high-risk births were frequently accompanied by a number of associated foetal and neonatal complications that required special.

In this study two important marginal models (GEE and ALR) are compared by using QIC value to select appropriate model, exchangeable working correlation structure also used throughout and we found that ALR give the best fit to the data and we interpreted our model based on ALR model, based on this which is also consistent with a study in Ghana (Gyabaah, 2014).

In several studies birth order of a neonate brought to light controversial results about NNM. Some studies showed that first order births were at higher risk of NNM while others showed that higher order births were at higher risk. For instance, a study about risk factors of NNM in Bangladesh showed that infants of first birth had a higher risk of NNM (Kamal, Ashrafuzzaman, Nasreen, 2012). A study in the Empowered Action Group States of India (Arokiasamy, Gautam, 2008) showed that neonates with first and higher order births experienced a high risk of death. In rural Iran, neonates with four or higher order births were at increased risk of NNM (Chaman, Naieni, Golestan, Nabavizadeh, and Yunesian, 2009). A study in Kenya (Mustafa, Odimegwu, 2008) found that increased risk of NNM was associated with first born children.

The present study showed that neonates with first order birth experienced the lowest NNM compared with 2-4 births, concurring with those researches which came to the conclusion that higher birth orders were at increased risk. The arguments put forward supporting these opposing conclusions are that first birth is associated with very young mothers with little or no experience of taking care of an infant (health care, feeding, and the like). On the other hand, with higher birth order (many children) comes scarcity of food, lack of attention by mothers, and so on.

The present study revealed that sex of a neonate is non-significant factor of neonatal mortality; this finding is opposed with results of other studies elsewhere. It was found that in Indonesia (Titaley, Dibley, Agho K., Roberts, Hall, 2008) and in Brazil (Araújo, Bozzetti, Tanaka, 2000) the risk of neonatal death was higher for male infants. It is not apparent why more male neonates than females died on the basis of the information provided by the EDH surveys.

The results of this study suggested that the risk of neonatal death was higher for neonates with preceding birth interval less than two years. Another study in Ethiopia showed that birth intervals shorter than two years led to higher NNM rates than in higher birth intervals (Susman, 2012), while analysis of pooled data on birth history from 52 countries (Rutstein, 2008), to see the effect of preceding birth intervals on NNM, showed that the risk was higher for birth intervals shorter than 24 months as well as for the periods longer than 47 months, compared to the referent category (24-47 months). Similar to the findings of the present study, the three studies (Arokiasamy, Gautam ,2008),(Chaman R, Naieni KH, Golestan B, Nabavizadeh H,Yunesian M,2009), and (Mustafa E. Odimegwu C,2008) provided evidence that the risk of dying was higher for neonates with birth spacing less than 24 months. This could probably be attributed to biological factors: giving a second birth within such a short span of time affects the health of both child and mother.

In this study physical size at birth came out as one of the statistically significant predictors of neonatal death despite the limitation associated with it because of the way the EDH survey gathered data about weight. This has been highlighted earlier – physical size is actually not weight, but rather mothers' perceived size of their children. The risk of death for neonates with very large and very small physical size at birth was higher than neonates with average size at birth whereas the risk of death for neonates with smaller-than-average and larger-than average size was statistically not significant compared to the reference category (average). Neonatal death was higher for very small infants in Indonesia (Titaley CR., Dibley MJ., Agho K., Roberts CL, Hall J., 2008).

The findings of this study showed that neonates born to mothers who were under 20 years of age and to mothers 30 years and older were at a higher risk of death compared with those born to mothers in the age bracket 20 -29 years. Another study in Ethiopia showed that finding in Bangladesh (Mondal NI., Hossain K., Korban A., 2009) showed that mothers' age at birth (age under 20 years) was the most significant predictor of NNM. The studies (Arokiasamy P., Gautam A., 2008), (Seedhom AE., Kamal NN., 2008), and (Kamal SMM., 2012) provided similar evidence: the risk of dying was higher for neonates whose mothers' age was below 20 years

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. CONCLUSIONS

The study aimed at investigating determinants of neonatal mortality in Ethiopia from 2011 EDHS data set. In line with the objectives of the study, it is clear that in 2011 EDHS, age of mother during pregnancy, preceding birth interval, birth order of child, residence of mother, age of mother at first birth, multiple birth (birth type) and size of child at birth significantly contributed to neonatal mortality. Place of delivery, wealth of household, sex of a child and mother's level of education had no significant influence on neonatal mortality. Furthermore, age of mother and multiple births are negatively related with neonatal mortality; while ages of mother at first birth, birth order, previous birth interval and size of child at birth are positively related with neonatal mortality. For this study two marginal models, GEE and ALR, have been compared for the analysis of marginal or average effects of covariates on the response variable and, we conclude that, ALR model with measure of association exhibited the best fit for this data than GEE models. For this study also GLMM, with two random intercept models was found to be appropriate for the analysis of within and between regional variations for neonatal mortality in Ethiopia. This concluded that there is heterogeneity of neonatal mortality between and within regions.

5.2. RECOMMENDATIONS

Based on the findings of the study different factors were identified for cause of neonatal mortality either directly or not in Ethiopia. In summary the key recommendations emerging from this study for policy makers, clinicians and the public at large are:

- Education and health intervention policies should be designed to reach the various ethnic groups on the issues of infant mortality in Ethiopia
- Very old women should be discourage from giving birth as they stand a greater risk of losing their infants under-one month
- Since rural infants are more exposed to neonatal mortality special attention is expected from concerned body in order to discourage neonatal mortality in rural areas.
- Further research should be conducted in subsequent Ethiopian Demographic and Health Surveys in order to monitor the causes and progress of neonatal mortality.

BIBLIOGRAPHY

Abuqamar, M.D., et al., 2011. The Impact of Parental Education on Infant Mortality in Gaza Strip, Palestine. *Journal of Public Health and Epidemiology*, **3**: 28-33.

Agresti, A., (2002).*Categorical Data Analysis*, Second Edition

Agresti, A., (2007).*An Introduction to Categorical Data Analysis*, 2nd Ed, Wiley Inc

Ananth CV, Joseph KS, Demissie K, Vintzileos AM,(2005).Trends in twin preterm birth subtypes in the United States, 1989 through 2000: Impact on perinatal mortality. Am J Obstet Gynecol; **193**:1076-1082.

Anderson, J., Verkuilen, J., Johnson, R., (2012).Applied Generalized Linear Mixed Models:, Continuous and Discrete Data For the Social and Behavioral Sciences.

Antonio, K. & Beirlant, J., (2006).Actuarial Statistics with Generalized Linear Mixed Models, University Center for Statistics, Belgium.

Araújo BF, Bozzetti MC, Tanaka CAA. Early neonatal mortality in Caxias do Sul: a cohort study. *Jornal de Pediatria* 2000; **76(3)**: 200-206

Araújo BF., Bozzetti MC., Tanaka CAA.. "Early neonatal mortality in Caxias do Sul: a cohort study". *Journal de Pediatria*; 2000; **76(3)**: 200-206.

Arokiasamy P., Gautam A., 2008. Neonatal mortality in the Empowered Action Group States of India: Trends and determinants. *Journal of Biosocial Science*; **40**:183-201.

Assefa N., Gebeyehu A., Terefe B., Tesfayi G., Roger P.,(2013) An Analysis of the Trends, Differentials and Key Proximate Determinants of Infant and Under-five Mortality in Ethiopia; *Ethiop J Health Dev* 2002; **16**:13-20

Awour IE., Abed Y., Ashour M., 2012. "Determinants and risk factors of neonatal mortality in the Gaza Strip, occupied Palestinian territory: A case-control study". Published Online (cited 2013); Available at: URL:http://www.thelancet.com/

Bashir, A.O., Ibrahim, G.H., Bashier, I.A. and Adam, I. (2013) Neonatal Mortality in Sudan: Analysis of the Sudan Household Survey, 2010; *BMC Public Health*, 13,287. http://dx.doi.org/10.1186/1471-2458-13-287

Black RE., Cousens S., Johnson HL., Lawn JE., Rudan I., Bassani DG., et al,2008. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 2010; **375**: 1969–87.

Bogale W., Assaye K., Amha M., Birkineh T., Alemayehu W, 2012. "Predictors of early neonatal mortality at a neonatal intensive care unit of a specialized referral teaching hospital in Ethiopia". *Ethiop. J. Health Dev*; **26(3)**:200-207

Bryce, J. and Requejo, J.H. (2008) Countdown to 2015. Tracking Progression on Maternal Newborn & Child Survival. The 2008 Report. UNICEF, New York. (2013) http://www.unicef.org/specialsession/accessed

Chaman R., Naieni KH., Golestan B., Nabavizadeh H., Yunesian M., 2009. Neonatal mortality risk factors in a rural part of Iran: A nested case-control study. *Iranian J Publ Health;* **38(1)**:48-52.

Chaman, R., Naieni, K.H., Golestan, B., Nabavizadeh, H. and Yunesian, M. (2009). Neonatal Mortality Risk Factors in a Rural Part of Iran: A Nested Case-Control Study. *Iranian Journal of Public Health*, **38**: 48-52.

Edward F., Fikre E., Peter B." the distribution and effects of child mortality risk factors in Ethiopia: A comparison of estimates from DSS and DHS".2009, **23(2)**:163-168.

Ethiopia Mini Demographic and Health Survey: Central Statistical Agency, Addis Ababa, Ethiopia, 2014.

Ethiopian Demographic and Health Survey, 2011. Addis Ababa, Ethiopia and Calverton, Maryland, USA: Central Statistical Agency; 2012

Factors influencing infant and child mortality: A case study of Ragshahi District, Bangladesh

Hong, R., Ayad M., Rutstein S., and Ren R., (2009). Childhood Mortality in Rwanda: Levels, Trends, and Differentials. Further Analysis of the Rwanda Demographic and Health Surveys 1992-2007/08. 66, 27.

Kamal SMM. Maternal Education as a determinant of neonatal mortality in Bangladesh, 2012. *Journal of Health Management*; **14(3)**:269–281.

Kamal SMM. Maternal Education as a determinant of neonatal mortality in Bangladesh. *Journal of Health Management* 2012; **14(3):**269–281

Kamal SMM., Ashrafuzzaman M., Nasreen SA, 2012. Risk factors of neonatal mortality in Bangladesh. *J Nepal Paediatr Soc*; **32(1)**:37-46.

Kamal SMM., Ashrafuzzaman M., Nasreen SA, 2012. Risk factors of neonatal mortality in Bangladesh. *J Nepal Paediatr Soc*; **32(1)**:37-46.

Kumar, P. and Gemechis. (2010): Infant and child mortality in Ethiopia: As statistical analysis approach.

Kwara K., 2012. Modeling the risk factors of neonatal mortality in Ghana using Logistic Regression, Master's thesis, Kwame Nkrumah University, Ghana.

Lawn JE., Cousens S., Zupan J. 4 million neonatal deaths: when? Where? Why? Lancet 2005; **365**:891 -900

Lawn JE., Kerber K., Enweronu-Laryea C., Massee Bateman O. Newborn survival in low resource settings – are we delivering? BJOG 2009; **116**: 49–59.

Logistic regression available at: http: // faculty. class. ncsu. edu/ garson /PA765 /logistic. htm #sigtests

Logistic regression using the SAS system: Theory and Applications, Paul D. Allison

MacFarlane, A.J., Johnson A., and Mugford, M. (2006) Epidemiology of Neonatal Mortality. In: Rennie, J.M. and Roberton, N.R.C., Eds., Textbook of Neonatology, Churchill Livingstone, Edinburgh, 3-33.

Mason, E. (2013) Newborns in Sub-Saharan Africa: How to Save These Fragile Lives. UN Chronicle. Dec 2007.

Masuy-Stroobat G. The Determinants of infant mortality: How far are conceptual frame works really.

Mekonnen Desta (2011): Infant and Child Mortality in Ethiopia: The role of Socioeconomic, Demographic and Biological factors in the previous five years period of 2000 and 2005.

Mendes, K.G., Olinto, M.T.A. and da Costa, J.S.D. (2006) Case-Control Study on Infant MortalityinSouthernBrazil.RevistadeSaúdePública,**40**:240-248. http://dx.doi.org/10.1590/S0034-89102006000200009

Millennium Development Goal indicators. United Nations, 2009. Available from: URL://Unstat.Un.org/Unsd/mdg/metadata.aspx? Indicator Id=0 & Series Id=562.

Millennium Development Goals, 2014: assessing progress in Africa towards the MDGs; Analysis of the common African positions on the post-2015 development Agenda.

Mondal NI., Hossain K., Korban A., 2009. Factors influencing infant and child mortality: A case study of Rajshahi District, Bangladesh. *Journal of Human Ecology*; **26**: 31 -39.

Mortality. In: Rennie, J.M. and Roberton, N.R.C., Eds., Textbook of Neonatology, Churchill Livingstone, Edinburgh, 3-33.Development Review, **41**:723-744.

Mostafa Kamal (2012): The effect of maternal education on Neonatal Mortality in Bangladesh, Islamic University, Bangladesh.

Mustafa E. Odimegwu C. Socioeconomic determinants of neonatal and post- neonatal mortality in Kenya: Analysis of Kenya DHS 2003. *Journal of Humanities and Social Sciences* 2008; Vol. 2, Issue 2.

Negera W., Eshetu W. Risk factors of neonatal mortality in Ethiopia, 2013; *J. Health Dev.*; **27(3)**:192-199

Oestergoard MZ, Counsens S, Lawn JE, Mathers CD. Neonatal mortality levels for 193 countries in 2009 with trends since 1990: A systematic analysis of progress, Projections, and priorities.PLOS Med 2011; **8(8)**:e1001080.

Pan, W., (2001).*Akaike's information criterion in generalized estimating equations*, Biometrics, 57, 120-125

Rajaratnam, J.K., Marcus, J.R., Flaxman, A.D., Wang, H., Levin-Rector, A., Dwyer, L., et al. (2010) Neonatal, Post Neonatal, Childhood, and under-5 Mortality for 187 Countries, 1970-2010:

Ringheim K., Teller C., Sines E., 2009. Ethiopia at across road: Demography, and Development. Washington, D.C: Population Reference Bureau.

Rutherford, M.E., Dockerty, J.D., Jasseh, M., Howie, S.R.C., Herbison, P., Jeffries, D.J., et al. (2009) Access to Health Care and Mortality of Children under 5 Years of Age in Gambia: A Case-Control Study. Bulletin of World Health Organization, **87**:216-224. http://dx.doi.org/10.2471/BLT.08.052175

Rutstein SO. Further evidence of the effects of preceding birth intervals on neonatal, infant, and

Samuel M., Eshetu W. Determinants of infant mortality in Ethiopia: A study based on the 2005 EDHS data. Ethiopia, Journal of Health Development 2012; **26(2)**:72-77

Seedhom AE., Kamal NN. Some determinants of neonatal mortality in a rural area, El-Minia governorate, Egypt, 2008.Egyptian J Comm Med 2008; **28**:63-72

Seedhom AE., Kamal NN. Some determinants of neonatal mortality in a rural area, El-Minia governorate, Egypt, 2008. *Egyptian J Comm Med* 2008; **28**:63-72.

Sharma, V., Katz, J., Mullany, L.C., Khatry, S.K., LeClerq, S.C., Shrestha, S.R., et al., 2008 Young Maternal Age and the Risk of Neonatal Mortality in Rural Nepal. Archives of Pediatrics & Adolescent Medicine, 1**62**:828-835.

Susman AS. Child mortality rate in Ethiopia, 2012. *Iranian J Publ Health*; **41(3)**:9-19.

Tachiweyika, E., Gombe, N., Shambira, G., Chadambuka, A., Mufuta, T. and Zizhou, S. (2009) Determinants of Perinatal Mortality in Marondera District, Mashonaland

East Province of Zimbabwe: A Case Control Study. Pan African Medical Journal, 8, 7.

The 2013 UN Inter-agency Group to Child mortality estimation, September 2013

The causes of neonatal mortality in Afghanistan

Titaley CR., Dibley MJ., Agho K., Roberts CL., Hall J. Determinants of neonatal mortality in Indonesia. BMC Public Health 2008; 8:232

Titaley CR., Dibley MJ., Agho K., Roberts CL., Hall J. Determinants of neonatal mortality in Indonesia. BMC Public Health 2008; **8**:232.

Titaley, C.R., et al. (2008) Determinants of Neonatal Mortality in Indonesia. BMC Public Health, 8,232. http://dx.doi.org/10.1186/1471-2458-8-232

under-five-years mortality and nutritional status in developing countries: Evidence from the Demographic and Health Surveys. DHS Working Paper No.41, 2008.

UNICEF. State of world's children, 2007. Newyork: UNICEF: 2006

Upadhyay, R.P., Dwivedi, P.R., Rai, S.K., Misra, P., Kalaivani, M. and Krishnan, A. (2012) Determinants of Neonatal Mortality in Rural Haryana: Nested Case-Control Study. Indian Pediatrics, 49, 291-294. http://dx.doi.org/10.1007/s13312-012-0044-2

Upadhyay, R.P., Dwivedi, P.R., Rai, S.K., Misra, P., Kalaivani, M. and Krishnan, A. (2012) Determinants of Neonatal Mortality in Rural Haryana: Nested Case-Control Study. Indian Pediatrics, 49, 291-294. http://dx.doi.org/10.1007/s13312-012-0044-2

WHO World Health Report: Make every mother and child count. Geneva: 2005

World Data Bank: World Development Indicators. The World Bank Group; 2013 (cited 2013);Availableat:URL:http://www.worldbank.org/en/country/Ethiopia

World Health Organization (2006) Neonatal and Perinatal Mortality: Country, Regional and Global Estimates. Geneva.

World Health Organization (2006). Neonatal and Perinatal Mortality: Country, Regional and Global Estimates. Geneva.

World Health Organization (WHO). (2006). Neonatal and Perinatal Mortality; Country, Regional and Global Estimates

APPENDIX

SAS program for fitting GEE and ALR

```
/* Fitting GEE model*/
data research;
infile 'c:\\Users\\User\\Desktop\\research.sav' ;
input CLID Age Reg Res Mled Wealth Agbirth BF Bord Mbirth Sex Survstats Pbint Pdel
Szbirth;
run;
proc print data=research;
run;
proc genmod data=research descending;
class CLID Reg Age(ref="15-19")Res(ref="R") Mled(ref="No
education") Wealth(ref="Rich") Agbirth(ref="<20")BF(ref="N0") Bord(ref="first birth")
Mbirth(ref="Single") Sex(ref="female")
Pbint(ref="<24") Pdel(ref="Home") Szbirth(ref="very large")
/ param=ref ref=first;
model Survstats=Age Res Mled Sex Wealth Agbirth BF Bord Mbirth Pbint Pdel /
dist=bin link=logit;
repeated subject=CLID /type=exch modelse;
run;
/*Fitting ALR model: */
proc genmod data=research descending;
class CLID Reg Age(ref="15-19") Res(ref="Rural") Mled(ref="No
education")Wealth(ref="Rich") Agbirth(ref="<20")BF(ref="N0")Bord(ref="first
birth")Mbirth(ref="Single") Sex(ref="female")
Pbint(ref="<24") Pdel(ref="Home") Szbirth(ref="very large")
/ param=ref ref=first;
model Survstats=Age Res Mled Sex Wealth Agbirth BF Bord Mbirth Pbint Pdel /
dist=bin link=logit;
repeated subject=CLID /logor=exch modelse;
run;
```

The full model Wald test for variable selection in GEE

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
Region	10	27.36	0.0023
Age	3	27.77	<.0001
Residence	1	2.71	0.0999
Mother's level of education	2	6.65	0.0360
Sex	1	2.83	0.0926
Wealth	2	1.84	0.3979
Age at first birth	2	20.40	<.0001
Birth order	1	24.39	<.0001
Multiple birth	1	24.67	<.0001
Birth interval	2	45.49	<.0001
Place of delivery	1	0.01	0.9296
Birth size	4	18.76	0.0009

The full model Wald test for variable selection in ALR

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
Region	10	27.53	0.0021
Age of mother	3	27.24	<.0001
Residence	1	2.68	0.1014
Mother's level of education	2	6.25	0.0438
Sex	1	2.77	0.0959
Wealth	2	1.89	0.3883
Age at first birth	2	20.02	<.0001
Birth order	1	24.52	<.0001
Multiple birth	1	24.80	<.0001

Birth interval	2	45.16	<.0001
Place of delivery	1	40.23	0.0093
Birth size	4	18.03	0.0012

R code for fitting generalized linear mixed model

```
library(Mass)

library(foreign)

library(lme4)

research=read.spss("C:\\Users\\User\\Desktop\\research.sav")

research=as.data.frame(research)

attach(research)

names(research)

View(research)

res.glmm1<-
lmer(Survstats~Age+Res+Mled+Wealth+Agbirth+BF+Bord+Mbirth+Sex+Pbint+Pdel+S
zbirth+(1|CLID) + (1|Reg), family="binomial",link="logit",data=research)

print(res.glmm1, corr=F)

summary(res.glmm1)

res.glmm2<- lmer(Survstats ~

Age+Res+Mled+Wealth+Agbirth+BF+Bord+Mbirth+Sex+Pbint+Pdel+Szbirth+

(1|CLID), family="binomial",link="logit",data=research)

print(res.glmm2, corr=F)

summary(res.glmm2)
```