ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

DEPARTMENT OF STATISTICS

DETERMINANTS OF INFANT MORTALITY IN ETHIOPIA: A MULTILEVEL COUNT REGRESSION MODELS

By: Jiregna Olani

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN STATISTICS (BIOSTATISTICS)

June, 2016

Addis Ababa, Ethiopia

DETERMINANTS OF INFANT MORTALITY IN ETHIOPIA: A MULTILEVEL COUNT REGRESSION MODELS

A thesis submitted to the school of Graduate Studies, Department of Statistics, College of Computational and Natural Sciences, Addis Ababa University in partial fulfillment of the requirements for the Degree of science in Biostatistics

Jiregna Olani

Advisor: Dejen Tesfaw (PhD)

Addis Ababa University

School of Graduate Studies

This is to certify that the thesis prepared by Jiregna Olani, entitled Determinants of Infant Mortality in Ethiopia: a Multilevel Count Regression Models and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Biostatistics complies with the regulation of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name	Signature	Date
Examiner___________________________________	________________	______________
Examiner___________________________________	______________	____________
Advisor: Dejen Tesfaw (PhD)	________________	_____________

Chair of Department or Graduate Program Coordinator

DECLARATION

I hereby declare that the thesis is my original work, to the best of my knowledge, it has not been presented for degrees in any other University and all sources of materials used for the thesis have been duly acknowledged with proper citation.

Name: Jiregna Olani

Signature: _____________

Place of submission: Department of Statistics, College of Natural and computational Sciences, Addis Ababa University

Date of submission: June, 2016

This thesis has been submitted for examination with my approval as a University advisor.

Dr. Dejen Tesfaw ---------------------------------

Advisor's Name signature

TABLE OF CONTENTS

ABSTRACT

Infant mortality rate is used as an indicator of a nation's economic welfare. Many researchers analyzed determinants of infant deaths in Ethiopia using a limited set of variables. The aim of this study was to identify factors that affect infant mortality based on 2014 Ethiopia mini Demographic and health survey (EMDHS) dataset using multilevel count regression models. The descriptive result showed that 4905 out of 5534 of the respondents, have no infant deaths per women, indicating excess zeroes. Among families of count models the ZIP model was found to be a better fit to the dataset than the others. The results of ZIP regression model showed that region, household size, birth order, and birth interval were identified as significant factors. At the stage of multilevel, ZIP model *showed that age, household size, birth order and birth interval were significantly affect infant death. The study also showed that there is a significant regional variations of infant mortality* $({\sigma_{u_0}}^2 = 3.6920\, p - value = 0.0001)$*.Further, the age* $({\sigma_{u_2}}^2 = 0.1181,$ $pvalue = 0.0001)$ *and household size* $({\sigma_{u_4}}^2 = 0.0501,\ p - value = 0.0002$*) effects on infant mortalities varies among regions of Ethiopia.*

Key words: EMDHS, Vuong test, ZIP, multilevel count regression models

ACKNOWLEDGEMENTS

Thanks to almighty God! Firstly, I extend my heartfelt appreciation to my Advisor Dejen Tesfaw (PhD) for the guidance rendered to me in the development of this thesis into what it is today. I also thank for taking your time to advice through my work, as well as the criticisms and the suggestions. I am grateful to the entire staff of Addis Ababa University and my working place Mekelle University Department of Statistics for their support by various inputs. My thanks also go to, Fikru Olani, my classmates, and my families for your interest for this thesis from the beginning up to this stage by giving a great support by financial and materials; I have said may God bless you.

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

AIC	Akaike Information Criteria
ANOVA	Analysis of Variance
BIC	Bayesian Information Criteria
CSA	Central Statistical Agency
EDHS	Ethiopian Demographic and Health Survey
GLMs	Generalized Linear Models
IMR	Infant Mortality Ratio
LL	Log likelihood
LRT	Likelihood Ratio Test
MDGs	Millennium Development Goals
NB	Negative Binomial
p.m.f	Probability Mass Function
UNICEF	United Nations Children's Fund
WHA	World Health Assembly
WHO	World Health Organization
ZINB	Zero-inflated negative binomial
ZIP	Zero-inflated Poisson

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Infant mortality is defined as the deaths under one year of age of babies who were live born (i.e. stillbirths are not included) and infant mortality rate (IMR) is defined as the total number of infant deaths per 1,000 live births in a specified period of time (Siegel and Swanson, 2004). Infant deaths are divided into two groups: those occurring at less than 28 days after birth, referred to as neonatal deaths; and those occurring at ages 28 days and over but under one year, referred to as post-neonatal deaths. Infant mortality rates have for a long time been used as country or regional level proxy indicators of socio-economic position. Infant mortality has also long been regarded as an important indicator of population health (Brocklehurst et al., 2011).

In nearly all populations, deaths before age one comprise the majority of deaths before age five. The infant mortality rate has been widely used as an indicator of population health (UNICEF, 2013). Ministry of health of federal democratic republic of Ethiopia claimed a target for infant mortality, to reduce infant mortality rate from 2013 level of 44/1000 to 20/1000 by 2019/20. By 2035, every child in Ethiopia enjoys the highest attainable standard of health and development with an end to all preventable child deaths (FMOH/MCH, 2015). The other less developed countries on average reached the ICPD target, with an average infant mortality of 33 per 1,000 in 2010-2015. Meanwhile, the more developed regions, which had already reached a low infant mortality level of 11 per 1,000 in 1990-1995, saw a further decline to 6 per 1,000 by 2010-2015. The relative levels and changes in infant mortality amongst the world's major areas are on the whole similar to those for under-five mortality. Africa has the highest level of infant mortality in 2010-2015 at 64 per 1,000, and has some distance to go to reach the ICPD target of 35 per 1,000. On the other hand, on average each of the world's other major areas met the ICPD target by 2010-2015. In Ethiopia, under-five mortality rate has declined by two thirds from the 1990 figure of 204/1,000 live births to 68/1,000 live births in 2012, thus meeting the target for Millennium Development Goal 4 (MDG 4) on child survival three years ahead of time. However, the mortality reduction was not uniform across the different childhood age groups (neonatal, infant, and under-five), geographic and socio-demographic population groups. Disaggregation of the mortality data by age indicated that the decline in neonatal mortality is not as relevant as the infant and child mortality figures (FMOH/MCH, 2015). In Ethiopia, results from the 2011 EDHS data show a remarkable decline in all levels of childhood mortality. Mortality trends can also be examined by comparing data from DHS surveys conducted in 2000, 2005, and 2011. Infant mortality rates obtained by these surveys evidence a continuous declining trend in mortality. Infant mortality decreased from 97 deaths per 1,000 live births in the 2000 survey to 59 in the 2011 survey. On the other hand, even though neonatal mortality rate decreased from 49 deaths per 1,000 live births in 2000 to 39 deaths per 1,000 live births in 2005, it has since remained stable at 37 deaths per 1,000, as reported in the 2011 EDHS. Mortality rates in urban areas are consistently lower than in rural areas, although the difference is quite small for neonatal mortality. Infant mortality is 29 percent higher in rural areas (76 deaths per 1,000 live births) than in urban areas (59 deaths per 1,000 live births) (CSA, 2012).

Despite the tremendous reduction since 1900s IMR is still high for developing countries. Though impressive infant mortality reduction was done in 1900s but IMR of a child born in a specified year dying before reaching the age of one is still high especially in less developed countries UNICEF (2013). One in every 17 Ethiopian children dies before the first birthday (CSA, 2012). Infant mortality declined by 39 percent over the 15-year period between the 2000 EDHS and the 2011 EDHS, from 97 deaths per 1,000 live births to 59 deaths per 1,000 live births. Childhood mortality is higher in rural areas than in urban areas. These rates were highest in Benishangul-Gumuz and lowest in Addis Ababa. The neonatal mortality rate was 37 deaths per 1,000 live births, the post neonatal mortality rate was 22 deaths per 1,000 live births. Programmes to increase the proportion of births attended by skilled health personnel, to increase immunizations against the vaccine-preventable diseases, to provide early care and treatment to sick children, and to upgrade the status of women through education and enhanced participation in the labor force can all help to improve the probability of survival of young children. Results from the 2011 EDHS are timely in evaluating the impact on the achievement of some major national policies, such as the National Population Policy, the National Policy on Ethiopian Women, and the National Health Policy (UNICEF, 2015).

In biomedical or health research, outcomes of interest often consist of count variables. For such count data, the standard framework for explaining the relationship between the outcome variable and a set of explanatory variables includes the Poisson and negative binomial regression models. However, the basic Poisson regression model forces the conditional variance of the outcome to equal the conditional mean, which is of limited use in real life. The negative binomial regression can be written as an extension of Poisson regression and it enables the model to have greater flexibility in modeling the relationship between the conditional variance and the conditional mean compared to the Poisson model. Also, an often encountered characteristic of count data, is the number of zeros in the sample can exceed the number of zeros predicted by either Poisson or negative binomial model, and this is of interest because zero counts frequently have special status (FANG, 2008).

A popular approach to count data with excess zeros relative to a Poisson distribution is to use a zero-inflated Poisson (ZIP) regression model. Zero-inflated models attempt to account for excess zeros. In other words, two kinds of zeros are thought to exist in the data, "true zeros" and "excess zeros". Zero-inflated models estimate two equations simultaneously, one for the count model and one for the excess zeros. Often, because of the hierarchical study design or the data collection procedure, zero-inflation and lack of independence may occur simultaneously, which render the standard ZIP model inadequate (Zhu et al., 2015). To account for the majority number of zero counts and the inherent correlation of observations, a class of multilevel ZIP regression model with random effects can be possibly used.

1.2 Statement of the Problem

IMR is used as an indicator of a nation’s economic welfare. Many researchers do recommend the need to conduct in-depth studies on the various aspects of infant and child health status in a different demographic, economic, and social-cultural setting. Mortality rate tends to cluster by area and high in most parts of the sub-Saharan African and Southern Asia countries. Therefore, understanding the geographic distribution of mortality of children under five is important to policy interventions. The aim of this study was to examine factors related to the mortality among children under the age one in Ethiopia. Moreover, some explanatory variables were used to observe the differentiated effects of factors among people from different socioeconomic groups and there is disparity of infant mortality rate from region to region in Ethiopia. In this study, the regional and individual level factors related with infant mortality in Ethiopia were assessed. For this purpose multilevel count regression model was thought to be appropriate method. This is because multilevel count regression model allows us to analyze the variations of infant mortality among regions of Ethiopia and the predictors of infant deaths that vary across regions.

Researchers studied the determinants of infant deaths in Ethiopia using some set of variables and statistical methods such as logistic regression, survival analysis and the like. Most of those researches are done on small-scale survey data which were came from only certain regions of the country. On the other hand, in the study of number of infant deaths count regression models are more appropriate to use than other methods. Many researches prefer to categorize count variable as binary and do the analysis. Particularly, in logistic regression which is widely used in analyzing infant mortality dataset, by categorizing the count variables into binary variable, we are losing some information. Another advantage of count regression model is, it takes it in to account many zero counts in the data. In this case, two states may be assumed to better reflect the situation. One of the states is the structural zero (or zero count) state where the only counts are zeros. The other state is the sampling zero state where the counts could be zeros or values greater than zero. Therefore, this study which attempts to analyze the infant mortality in Ethiopia using a varied set of explanatory variables makes a contribution to the literature in this context using dataset based on Ethiopian 2014 infant mortality mini EDHS data. Therefore, this study was performed to fill the problem addressed using Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), multilevel Poisson, multilevel NB, and multilevel zero-inflated Poisson (multilevel ZIP) regression models.

1.3 Objective of the Study

1.3.1 General objective of the study

The general objective of this study was to identify factors that affect the infant mortality in Ethiopia based on EDHS 2014 mini dataset using multilevel count regression models.

1.3.2 Specific objective of the study

The specific objectives of this study were:

1. To apply single and multilevel count regression models for analyzing determinants of infant mortality in Ethiopia.
2. To handle over-dispersion and zero-inflation problem for the number of infant deaths.
3. To identify the best multi-level count regression models to analyze infant death dataset.

1.4 Significance of the Study

This study was aimed to identify the predictors of infant mortality in Ethiopia and concerns deaths that happen from birth through infant, it must be noted that stillbirths are a significant problem and many of the causes and underlying determinants are similar to those of deaths that happen later in childhood. This has a great importance in providing timely registering, awareness of risks and early seeking to care and birth preparedness. Information on which factor that determines the infant death in the area could be helpful for policy makers, program implementers, monitoring and evaluation activities to develop maternal care programs and set appropriate plans to tackle the existing health and antenatal care problems. Therefore, this study will try to provide related statistical methods to analyze and interpret the dataset. Those statistical methods constituted by the study were Poisson, NB, ZIP, ZINB, and multilevel count regression models (multilevel-Poisson, multilevel-NB, and multilevel zero-inflated Poisson (ZIP). On the other hand, this study might support researchers in determining the appropriate model to use a given multilevel count regression models.

1.5 Limitations of the Study

In this study there were some challenges that we faced. Some important variables like infant weight at birth that might affect infant mortality were not available in the 2014 EDHS dataset. SAS PROC NLMIXED took longer time for multilevel count regression models. In addition to this, multilevel zero-inflated negative binomial regression model was not considered under this study due to convergence issue.

CHAPTER TWO

LITERATURE REVIEW

2.1 Trends in Infant Mortality

With respect to infant and child mortality, the ICPD Programme of Action embraced the absolute targets of the World Health Assembly (WHA) Health for all by 2000 resolution, aiming all countries to achieve an infant mortality rate less than 50 deaths per 1,000 live births and an under-five mortality rate less than 70 deaths per 1,000 live births by 2000. For countries that already had comparatively low levels of infant and child mortality in 1994, the Programme of Action offered relative targets: one-third reductions in infant and child mortality rates by 2000. For horizons beyond 2000, the Programme of Action again identified absolute targets, urging countries to achieve an infant mortality rate below 35 deaths per 1,000 live births and an under-five mortality rate below 45 deaths per 1,000 live births by 2015. In nearly all populations, deaths before age one comprise the majority of deaths before age five. The infant mortality rate has been used widely as an indicator of population health. Between 1990-1995 and 2010-2015, global infant mortality fell from 59 per 1,000 to 37 per 1,000. Infant mortality in the least developed countries is 63 per 1,000 in 2010-2015, not quite double the ICPD target. The other less developed country including Ethiopia on average have reached the ICPD target, with an average infant mortality of 33 per 1,000 in 2010-2015. Meanwhile, the more developed regions, which had already reached a low infant mortality level of 11 per 1,000 in 1990-1995, saw a further decline to 6 per 1,000 by 2010-2015. The relative levels and changes in infant mortality amongst the world's major areas are on the whole similar to those for under-five mortality. Africa has the highest level of infant mortality in 2010-2015 at 64 per 1,000, and has some distance to go to reach the ICPD target of 35 per 1,000. Each of the world's other major areas has, on average, met the ICPD target by 2010-2015 (UNICEF, 2013).

2.2 International Literature on Infant Mortality

Many studies have investigated the determinants of childhood mortality in different socio economic settings. Seckin (2009) conducted an analysis of infant mortality in Turkey using survival analysis. The results of the logistic regression showed that birth interval was associated with infant mortality at lower levels of wealth index. Children from poorer families with preceding birth interval shorter than 14 months or children whose mothers experience a subsequent birth fare badly. Breastfeeding is important for the survival chance of the infants under the age 3 months. Place of delivery and source of water were also found to be statistically significant effect with infant mortality risk.

In another study, examined infant and child mortality in India using survival analysis. This study 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 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 (Pandey et al., 1998).

2.3 Literature on Infant Mortality in Ethiopia

In Ethiopia, Ezra and Gurmu, 2002, employed a logistic regression model to investigate the effect of birth interval on infant and child mortality in the context of communities characterized by high reproductively, prolonged breast feeding practice and poor living conditions. It was found that, short birth interval (less than 18 months) significantly related with infant and child mortality as compared with long birth interval (more than 24 months), this implies the influence of short birth interval are more prominent on infant mortality while weaker effect on child mortality. On the other hand, they obtained the children born to young mothers' age (between 15 and 19) and oldest mothers' age (between 35 and 49) had found to be a significant effect on infant and child mortality as compared with children born to mothers in the age category (between 25 and 34). Education was also found to be a significant determinant of infant and child mortality.

A study was performed to identify determinants of under-five child mortality in Ethiopia (Senayit, 2012). For this purpose survival analysis was employed. The results from cox hazard's proportional model shows that mother's education, mother's age, marital status, birth order and place of residence are the significant factors that contribute to under-five child mortality. However, sex of a child, family size, wealth index, source of drinking water and toilet facility are not found to be significant.

Samuel and Eshetu (2012) used survival analysis model to identify determinant factors of infant mortality in Ethiopia. The study investigated the relationship between infant mortality as measured in months (birth to completion of 12 months) and predictor variables: birth spacing, mother's age at birth of a child, sex of infant, breastfeeding status, family size, marital status, mother's education, father's education, wealth index, area of residence, and source of drinking water. The Cox regression analysis showed that the covariates: breastfeeding status, mother's education, mother's age, birth spacing, source of drinking water, and the gender of an infant were found to be statistically significant predictors of infant mortality.

Tibebu (2011) used single and multilevel count model to explore the major risk factors and regional differentials in under-five mortality in Ethiopia. The study was used predictor variables: place of residence, region, education level of mother, source of water supply, employment status of water, economic status of the house hold, religion, availability of toilet facility, and age of mother at first birth. The relationship between under-five mortality and the explanatory variables age at first birth, mother education level, religion, employment status of mother and economic status of mother are significant. The multi-level Poisson regression model analysis further showed that there are substantial under-five death variations per mother among regions in Ethiopia.

Esayas (2003) used survival analysis to analyze factor that affect infant mortality and which combination of explanatory variables affect the form of the hazard function. Despite its many advantages, the work of women in economic activities in Ethiopia has been associated with increased and decreased mortality of infants. Thus, this study examines whether these conclusions are upheld at the level of the typical Ethiopian mothers. In general, this study finds that working type, status of work (high, medium, low), literacy, birth order and previous interval has a significant effect on infant mortality. Multivariate analysis assesses the strength of apparent association between work status of the mother and infant mortality by controlling other characteristics likely to influence the outcomes. The relative risks associated with several other variables are statistically significant and in the expected direction. Among these factors the length of the preceding birth interval for infants exerted an expected beneficial effect on the hazard ratio of infants. There is also evidence of a strong negative effect of the birth order on the hazard ratio of infant death and being first born significantly increases the probability of dying at infant stage.

The risk of infant death is higher in those who are not married than married; among other variables the probability of infant death will be lower if the mother is educated.

2.4 Literature on count regression models

Models for count data have been prominent in many branches of the recent applied literature, for example, in health economics (e.g., in numbers of visits to health facilities) management (e.g., numbers of patents) and industrial organization (e.g., numbers of entrants to markets). The foundational building block in this modeling framework is the Poisson regression model. But, because of its implicit restriction on the distribution of observed counts – in the Poisson model, the variance of the random variable is constrained to equal the mean – researchers routinely employ more general specifications, usually the negative binomial (NB) model which is the standard choice for a basic count data model. There are also many applications that extend the Poisson and NB models to accommodate special features of the data generating process, such as zero inflation. The basic models for fixed and random effects, have also been extended to the Poisson and NB models for counts (Greene, 2007).

There have, however, been scores of further refinements and extensions that are documented in a huge literature and several book length treatments such as (Cameron A.C. and Trivedi P.K., 2005). The multilevel regression model has become known in the research literature under a variety of names, such as 'random coefficient model', 'variance component model', and 'hierarchical linear model'. Statistically oriented publications tend to refer to the model as a mixed-effects or mixed model. The multilevel count regression models assume that there is a hierarchical data set, with one single outcome or response variable that is measured at the lowest level, and explanatory variables at all existing levels. Conceptually, it is useful to view the multilevel regression model as a hierarchical system of regression equations (Joop, 2010).

In this study, count regression models with two levels are used in the analysis to the dataset. The relative importance of these predictive variables of infant mortality may vary depending on the prevailing socio-economic conditions in a community. Although it is true that the problem of persisting high levels of infant and child mortality concerns most developing countries, it is also clear that not only the levels of mortality, the characteristics of the phenomenon differ, often widely. In fact, the mechanisms of infant and child mortality are influenced by geographical, climatic, social, cultural and economic characteristics that differ from one country to another and, very often, even among different regions of the same country. Therefore, careful judgment should be used when extending the results of a study to other populations, especially if there are substantial differences among the populations considered.

CHAPTER THREE

DATA AND METHODOLOGY

3.1 Source of data

The data in this study is based on 2014 infant mortality mini EDHS which was obtained from the Central Statistical Authority (CSA), Ethiopia. The 2014 EMDHS was undertaken on a representative sample of 8,070 women in the reproductive ages of between 15 and 49. Its specific objective was to collect information and estimate of some of the MDG indicators such as childhood mortality, knowledge and use of family planning methods, maternal and child health, nutrition, knowledge of HIV/AIDS were provided for the nine regional states and two city administrations. Therefore, this study was based on information gathered on infant mortality found from the birth history of children's born to women who were included in the survey. Since the interest of this study was about infants under age one a dataset consisting of 5,534 infants was used.

3.2 Variables included in the study

Depending on the demonstrated related literature reviews the variables included in this study are listed as follows.

3.2.1 Response variable

The response variable of the study was the number of infant deaths per mother of reproductive age.

Explanatory variables

The predictor factors assessed as the main determinants of infant mortality in this study were described as follows.

Child Characteristics

- Sex of the child refers to whether the child is male or female and is included in the study because probability of death is not equal between the two sexes.

- Birth order refers to the number of siblings the child has at birth. Birth order has a U-shaped pattern with children who are born first or last having the highest probability of death. This variable is divided into four categories as 1, 2 and 3, 4-6 and 7 and above.
- Previous birth interval, measures the time between the current birth and the preceding birth. It is assessed in this study for its contribution to infant. It is categorized into three categories as less than 2, 3-4, 5 and above.

Mother's Characteristics

- Mother's education is an important factor in infant mortality reduction; more education is related with a lower risk of infant death. In this study mothers were categorized as illiterate and literate categories.
- Marital status refers to whether the mother is currently in married or not. It is included in this study since children from mothers not married have more risk of death than children born to married mothers. The variable is categorized into two: mothers who are married and currently in union and mothers who are not.

Household Characteristics

- Region is potential variable that influences childhood mortality in the country. This variable is classified in to nine administrative regions and two city administrations.
- Wealth index is important to determine household capacity for health and education services and then infant mortality. It is categorized as poor, middle and rich.
- Place of residence is whether the household is in a rural or urban area is a factor that always affects infant mortality because living in urban areas is advantageous in terms of easy access to quality health and educational services facilities as compared with rural areas.
- Age of mother at her first birth give another covariate considered to affect the death of infant.
- On the other hand, household size is the variable which takes the total number of all members of the individual mother's household. Similarly, we considered this variable as a covariate (continuous).
- Source of drinking water is also another factor that thought to be affecting the infant mortality. It is categorized as piped water, improved drinking water sources, and unimproved drinking water sources (UNICEF and WHO (2011); WHO, 2006).

3.2.2 Descriptions and coding of the study variables

The detail description of independent variables (covariates) of the study is presented according to the following table.

Table 3.1: Descriptions of and categories of study variables

No	Variables	label of variables	Category
1.	Region	Region of the household	0 Tigray 1 Afar 2 Amhara 3 Oromia 4 Somali 5 Benishangul-Gumuz 6 SNNP 7 Gambela 8 Harari 9 Dire Dawa 10 Addis Ababa
2.	Place of residence	Place of residence of the household	0 Rural 1 Urban
3.	Age	Age of mother at her first birth	Continuous variable
4.	Marital status	Current marital status of mother	0. Married and currently in union 1. Others
5.	Wealth Index	Household's level of income	0 Lower 1 Middle 2 Higher
6.	House hold size	House hold size	Continuous variable
7.	Education	Mother's level of education	0 Illiterate 1 Literate
8.	Sex	Sex of the infant	0 Female 1 Male
9.	Birth interval	Previous birth interval	0 less than 24 months 1 between 24-48 months 2 49 months and above
10.	Birth order	Birth order of the child	0 if 1 1 2 and 3 2 between 4 and 6 3 7 and above
11.	Source of water	Type of drinking water used	0 Piped water 1 Improved drinking water sources 2 Unimproved drinking water sources

3.3 Methodology

Poisson distribution is the most common probability model for discrete data with observations assumed to have a constant rate of occurrence amongst individual units with the property of equal mean and variance. However, in many applications the variance is greater than the mean and over-dispersion is said to be present. The application of the Poisson distribution to data exhibiting over-dispersion can lead to incorrect inferences and/or inefficient analyses. The most commonly used extension of the Poisson distribution is the negative binomial distribution which allows for unequal mean and variance but may still be inadequate to model datasets with long tails and/or value-inflation (Wondewosen et al., 2014; Ayati and Abbasi, 2014; Sileshi, 2007 and Loquiha et al., 2013).

Models for continuous data such as linear regression and Analysis of Variance (ANOVA) should not be directly applied to discrete response variables due to the underlying distributional assumptions required by these models for their correct application. Generalized linear models (GLMs) use a regression procedure to fit relationships between predictor and dependent variables. Unlike classical regression model where the random component (i.e., the error term) is assumed to follow a normal distribution, the random component in a GLM is assumed to follow an exponential family of distributions. In this section, several common features of skewed discrete random variables related to over-dispersion and zero-inflation are included. Overdispersion occurs where there is greater variability in a dataset than expected under a standard statistical model (normally Poisson), i.e. the variance in a dataset is greater than the mean (McElduff, 2012; Akbarzadeh et al., 2013). The presence of overdispersion in discrete data causes summary statistics resulting from a simple statistical model to be larger than anticipated and can lead to incorrect inferences under such a simple hypothesis (Gupta et al., 2013).

Even though there are several statistical models, some models may not be appropriate to deal with some specific types of data. Their use is solely depending on the types and nature of the data. In this study, the form of response variable is a count data, which is most often characterized as non-normal distribution. Thus, to deal with the data and methodological issues related with modelling the number of infant deaths, a wide variety of statistical methods can be used. There are count regression models which had been developed to analyze data with count response variables.

In this study, count regression models such as Poisson, negative binomial, zero-inflated Poisson regression, and zero-inflated negative binomial regression models were applied. Further multilevel count regression models like multilevel Poisson, multilevel NB, and multilevel ZIP were used to check the variation among the groups.

3.3.1 Single level count regression models

3.3.1.1 Poisson regression model

The Poisson distribution is the most common probability distribution for count data. The Poisson probability model is appropriate for events that occur randomly over time and/or space. The probability function for Y is given by

$$\Pr(Y = y_i|\mu_i) = \frac{e^{-\mu_i}\mu^{y_i}}{y_i!}, \quad \mu_i \geq 0, \tag{1}$$

where y_i = 0, 1, 2 … are discrete counts and μ is the mean of the Poisson distribution. The mean and variance of Poisson distribution is given as

$$E(y_i) = Var(y_i) = \mu_i$$

One specification that is mostly used for the mean parameter μ_i is the exponential specification. This specification ensures that μ_i is non-negative and it is given as

$$E(y_i) = \mu_i = \exp(x_i{}'\beta)$$

where $x_i{}' = (1, x_{i1}, \ldots, x_{ip})$, is a vector of explanatory variables and β is a $(p+1)$-dimensional column vector of unknown parameters to be estimated. The estimation is undertaken by using maximum likelihood method. There are two basic criteria commonly used to check the presence of overdispersion: the deviance, $D(y, \mu_i)$ or the Pearson χ^2 statistic be greater than its degrees of freedom. For the Poisson regression, $D(y, \mu_i)$ and χ^2 are defined in expression below respectively.

$$D(y, \mu_i) = 2\sum_{i=1}^{n}\left\{y_i ln\frac{y_i}{\widehat{\mu_i}} - (y_i - \widehat{\mu_i})\right\} \tag{2}$$

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \widehat{\mu_i})^2}{\widehat{\mu_i}} \tag{3}$$

However, these two rules of thumb can yield misleading inference from a direct likelihood point of view. Therefore, selecting between Poisson regression and an overdispersed Poisson model should be performed using some appropriate modeling procedure (Dejen et al., 2015; Dobson, 2002).

3.3.1.2 Negative binomial regression model

The negative binomial model is an extension of the Poisson model to overcome possible over-dispersion in the data (Lord, 2010). If a Poisson regression model doesn't fit the data and it appears that the variance of Y is increasing faster than the Poisson model allows (i.e. if a plot of the residuals versus linear predictors appears to fan out), then a simple scale-factor adjustment is not appropriate. One way to handle this situation is to fit a parametric model that is more dispersed than the Poisson. A natural choice is the negative binomial (John and Pamela, 2010). The probability mass function for the negative binomial distribution is:

$$f(y_i; \mu_i; \delta) = \frac{\Gamma\left(y_i + {}^{1}/_{\delta}\right)}{y_i!\ \Gamma\left(\frac{1}{\delta}\right)} (1 + \delta\mu_i)^{\frac{-1}{\delta}} \left(1 + \frac{1}{\delta\mu_i}\right)^{-y_i}, \quad y_i \geq 0;\ \delta > 0 \tag{4}$$

with mean and variance are expressed as:

$$E(y_i) = \mu_i = \exp(x_i{}'\beta), var(y_i) = \mu_i(1 + \delta\mu_i)$$

where

- The term δ is the dispersion factor and it is a constant.
- We relate parameters μ_i to covariates $x_i \in R^p$ through the log-link function so that

$$\log\mu_i = x_i{}'\beta \tag{5}$$

where $x_i{}' = (x_{i1}, x_{i2}, \dots, x_{ip})$, $1xp$ row vector of covariates is the number of covariates in the model. $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ is the corresponding $(p + 1)\ x\ 1$ column vector of unknown regression parameters. The maximum likelihood estimation method is used to estimate the parameter vector

$$\xi = (\delta, \beta^T)^T$$

The likelihood function of the NB model based on a sample of n independent observations is given by

$$L(\mu, \delta; y_i) = \prod_{i=1}^{n} \left\{ \frac{\Gamma\left(y_i + {}^{1}/_{\delta}\right)}{y_i!\ \Gamma\left(\frac{1}{\delta}\right)} (1 + \delta\mu_i)^{\frac{-1}{\delta}} \left(1 + \frac{1}{\delta\mu_i}\right)^{-y_i} \right\} \tag{6}$$

Then the log-likelihood function is expressed as follows

$$l = \log\mathrm{L}(\mu_i, \delta; y_i)$$

$$= \sum_{i=1}^{n} \left\{ -\log(y_i!) + \sum_{k=1}^{y_i} \log(\delta y_i - \delta k + 1) - \left(y_i + {}^{1}/_{\delta}\right) \log(1 + \delta\mu_i) + y_i \log(\mu_i) \right\} \tag{7}$$

The likelihood equations for estimating μ_i and δ are obtained by taking the partial derivations of the log-likelihood function and setting them equal to zero. Thus, we obtain the first derivatives of $l = \log\mathrm{L}(\mu, \delta; y_i)$ with respect to the underlying parameters as follows:

$$\frac{\partial l}{\partial \beta} = \frac{\partial l}{\partial \mu}\frac{\partial \mu}{\partial \beta} = \sum_{i=1}^{n} \left(\frac{y_i - \mu_i}{1 + \delta\mu_i}\right) x_i\ ,$$

$$\frac{\partial l}{\partial \delta} = \sum_{i=1}^{n} \left\{ \left(\sum_{k=1}^{y_i} \frac{y_i - k}{\delta y_i - \delta k + 1} \right) + \frac{\log(1 + \delta\mu_i)}{\delta^2} - \frac{\left(y_i + \frac{1}{\delta}\right)\mu_i}{1 + \delta\mu_i} \right\} \tag{8}$$

3.3.1.3 Zero-inflated count regression models

There are situations where a major source of over-dispersion is a relatively large number of zero counts, and the resulting over-dispersion cannot be modeled accurately with negative binomial model. In such cases, one can use zero-inflated Poisson or zero-inflated negative binomial model to fit the data. Zero-inflated distributions can be formed from a component mixture of two distributions. They allow for zero-inflated data and involve a mixture of two distributions where the zeros are modelled separately from the counts. Let $f(y_i; \mu)$ be a distribution function for count data, such as the Poisson and negative binomial distribution, with unknown parameters μ. Then, a zero-inflated distribution, denoted as ZI $f(y_i; \mu)$, is given by (Agarwal et al., 2002).

$$P(y_i|\omega, \mu) = \begin{cases} \omega + (1 - \omega) f(y_i = 0; \mu), & y_i = 0 \\ (1 - \omega) f(Y_i = y_i; \mu), & y_i = 1, 2, \ldots \end{cases} \tag{9}$$

The mean and variance of the ZI $f(y_i; \mu)$ distribution are given by

$$E_{zi}f(y_i; \omega, \mu) = (1-\omega)E_f(y_i; \mu) \text{ and}$$

$$Var_{Zi}f(y_i; \omega, \mu) = (1-\omega)[E^2{}_f(y_i; \mu)] - [(1-\omega)E_f(y_i; \mu)]^2$$
$$= (1-\omega)\{Var_f(y_i; \mu) + \omega E^2{}_f(y_i; \mu)\}.$$

3.3.1.4 Zero-inflated Poisson (ZIP) regression model

The Zero-inflated Poisson regression model is expressed as

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\mu}, & y_i = 0 \\ (1-\omega_i)\dfrac{e^{-\mu_i}\mu_i{}^{y_i}}{y_i!}, & y_i = 1, 2, \qquad 0 \le \omega_i \le 1 \end{cases} \tag{10}$$

The mean and variance of Zero-inflated (ZIP) distribution is given as

$$E_{ZIP}(y_i|\omega_i, \mu_i) = (1-\omega_i)\mu_i\,, \qquad \text{and}$$

$$var_{ZIP}(y_i|\omega_i, \mu_i) = E_{ZIP}(y_i|\omega_i, \mu_i)(1 + \omega_i\mu_i).$$

To apply the ZIP model in practical modeling situations, (Agarwal et al., 2002; Afifi et al., 2007) suggested the following joint models for μ_i and ω_i

$$\log(\mu_i) = x_i{}^T\beta \text{ and } \log\left(\frac{\omega_i}{1-\omega_i}\right) = z_i{}^T\gamma,\ i = 1, 2, \dots, n \tag{11}$$

where x_i and Z_i are covariate matrices. β and γ are (p+1) x 1 and (q+1) x 1 vector of unknown parameters, respectively. The vector of covariates x_i and z_i can be the same or different. For a random sample of observations $y_1, y_2, \dots\ , y_n$ the log-likelihood function $l(\mu, \omega; y)$ is given by

$$l = \sum_{i=1}^{n}\{\ln[\omega_i + (1-\omega_i)e^{-\mu_i}]I_{(y_i=0)}\} + [\ln(1-\omega_i) - \mu_i + y_i ln\mu_i - \ln(y_i!)]I_{(y_i=0)} \tag{12}$$

where I (.) is the indicator function for the specified event, i.e. equal to 1 if the event is true and 0 otherwise. The first and the second derivatives of $l = l(\mu, \omega; y)$ with respect to β and γ are as follows

$$\frac{\partial l}{\partial \beta_j} = \frac{\partial l(\mu,\omega)}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j} = \sum_{i=1}^{n}\left\{I_{(y_i=0)}\left[\frac{-(1-\omega_i)\mu_i e^{-\mu_i}}{\omega_i + (1-\omega_i)e^{-\mu_i}}\right] + I_{(y_i>0)}[y_i - \mu_i]\right\}x_{ij}\,, \quad j = 1, 2, \dots\ , p$$

$$\frac{\partial l}{\partial \gamma_r} = \frac{\partial l(\mu,\omega)}{\partial \omega_i}\frac{\partial \omega_i}{\partial \gamma_r}$$

$$= \sum_{i=1}^{n} \left\{ I_{(y_i=0)} \left[\frac{1 - e^{-\mu_i}}{\omega_i + (1 - \omega_i) e^{-\mu_i}} \right] - I_{(y_i>0)} \left[\frac{1}{1 - \omega_i} \right] \right\} Z_{ir} \ , \quad r = 1,2, \dots \ q$$

3.3.1.5 Zero-inflated negative binomial (ZINB) regression model

The zero-inflated negative binomial (ZINB) model is a general model for counts which nests the Zero-inflated Poisson (ZIP), negative binomial (NB), and Poisson models. A Zero-inflated negative binomial ZINB model for the response y_i can be written

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1 - \omega_i)(1 + \delta\mu_i)^{-1/\delta}, & y_i = 0 \\ (1 - \omega_i) \dfrac{\Gamma\left(y_i + \frac{1}{\delta}\right)}{y_i! \, \Gamma\left(\frac{1}{\delta}\right)} (1 + \delta\mu_i)^{-1/\delta} \left(1 + \dfrac{1}{\delta\mu_i}\right)^{-y_i}, & y_i > 0 \end{cases} \quad (13)$$

where $\delta > 0$ is a dispersion parameter and is assumed not to depend on covariates. The mean and variance of the ZINB model are given by

$$E(Y_i) = (1 - \omega_i)\mu_i \ \text{and}$$

$$Var(Y_i = (1 - \omega_i)(1 + \omega_i\mu_i + \delta\mu_i)\mu_i$$

The parameters μ_i and ω_i depend on vectors of covariates x_i and z_i, respectively. The zero-inflated negative binomial (ZINB) distribution is not a standard generalized linear model (GLM) type, even when the over-dispersion parameter δ is known, and standard GLM fitting methods are not applied. To obtain the parameter estimates of ZINB regression models $\hat{\delta}, \ \hat{\beta}$ and $\hat{\gamma}$, the Newton-Raphson method or the method of Fisher scoring will be used. However, the method of Fisher scoring is more appropriate for ZINB regression because the second derivative $l = l(\delta, \mu_i, \ \omega_i, \ ; y_i)$, is simplified by taking expectations (Agarwal et al., 2002; Agresti, 2003).

3.3.2 Multilevel Count Regression Models

3.3.2.1 Multilevel Poisson regression model

Many kinds of data, including observational data collected in the human and biological sciences, have a hierarchical or clustered structure. For example, animal and human studies of inheritance deal with a natural hierarchy where offspring are grouped within families. Offspring from the same parents tend to be more alike in their physical and mental characteristics than individuals chosen at random from the population at large. We prefer to a hierarchy as consisting of units grouped at different levels (Goldstein, 1999).

The multilevel Poisson model deals with certain kinds of dependence. The model can be further extended by including a varying exposure rate *m*. The multilevel Poisson regression model for a count Y_{ij} for i^{th} individual in the j^{th} group is expressed as follows (Joop, 2010).

$$Y_{ij}|\lambda_{ij} = Poisson(m_{ij}, \lambda_{ij}).$$

The link function for maximum likelihood Poisson distribution is given as;

$$log(\lambda_{ij}) = \eta_{ij} \quad (14)$$

$$\text{where, } \eta_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij} \quad (15)$$

letting

$$\beta_{0j} = \beta_0 + U_{0j} \quad \text{and}$$

$$\beta_{hj} = \beta_h + U_{hij}$$

$$\text{from (15) } \log(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h X_{hij} + U_{0j} + \sum_{h=1}^{k} U_{hj} X_{hij} \quad h = 1, 2, \ldots, k \quad (16)$$

The first part of equation (16) $\beta_0 + \sum_{h=1}^{k} \beta_h X_{hij}$, is called the fixed part of the model. The second part, $U_{0j} + \sum_{h=1}^{k} U_{hj} X_{hij}$ is called the random part. The groups are characterized by $k + 1$ random coefficients $U_{0j}, U_{1j}, U_{2j}, \ldots, U_{hj}$. The random coefficients are independent between groups, but may be correlated within groups. It is assumed that the vectors $(U_{0j}, U_{1j}, U_{2j}, \ldots, U_{hj})$ is distributed with means zero and has a multivariate normal distribution with a constant variance matrix. The variances and covariances of the level two random effects are

$$Var(U_{hj}) = \sigma_{hh} = \sigma_h^{\,2}, \qquad h = 0, 1, 2, \ldots, k$$

$$Cov(U_{hj}, U_{pj}) = \sigma_{hp}, \qquad p = 0, 1, 2, \ldots, k, \qquad for\ h \neq p$$

3.3.2.2 Multilevel negative binomial regression model

Multilevel negative binomial regression model is given by;

$$Pr\ (Y_{ij} = y_{ij}) = \frac{\Gamma(y_{ij} + v)}{y_{ij}!\ \Gamma(v)} \frac{v^v \mu_{ij}^{\ *y_{ij}}}{(v + \mu_{ij}^{\ *})^{v + y_{ij}}} \quad y_{ij} = 0, 1, \ldots, \quad (17)$$

with mean and variance given, respectively, as follows:

$$E(Y_{ij}) = \lambda_{ij}^{\ *} = log(\eta_{ij}), \text{ and } Var(Y_{ij}) = \lambda_{ij}^{\ *} + \alpha(\lambda_{ij}^{\ *})^2.$$

where $\eta_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij}$.

3.3.2.3 Multilevel zero-inflated Poisson regression model

Count data with excess zeros relative to a Poisson distribution are common in many biomedical applications. A popular approach to the analysis of such data is to use a zero-inflated Poisson (ZIP) regression model. Often, because of the hierarchical study design or the data collection procedure, zero-inflation and lack of independence may occur simultaneously, which render the standard ZIP model in adequate. To account for the preponderance of zero counts and the inherent correlation of observations, a class of multilevel ZIP regression model with random effects is presented. The multilevel ZIP model is then generalized to cope with a more complex correlation structure (Andy et al., 2006). Suppose a discrete count response variable Y follows a ZIP distribution:

$$P(Y_i = y_i) = \begin{cases} \phi + (1-\phi)e^{-\mu_i}, & y_i = 0 \\ (1-\phi)\dfrac{e^{-\mu_i}{\mu_i}^{y_i}}{y_i!}, & y_i = 1, 2, \end{cases} \tag{18}$$

where, $0 \leq \phi_i \leq 1$ so that it incorporates more zeros than those permitted under the Poisson assumption ($\phi = 0$), whereas $\phi < 0$ corresponds to the zero-deflated situation. The ZIP distribution may be regarded as a mixture of a Poisson (λ) and a degenerate component placing all its mass at zero. Recently, the ZIP regression model has been extended to the random effects setting, whereby random components s_i and v_i are incorporated within the logistic and Poisson linear predictors to account for the dependence of observations within clusters. These random effects ZIP models are household-specific in the sense that the random effects s_i and v_i so introduced are specific to the i^{th} region. In the following, a multilevel ZIP regression model is developed to handle correlated count data with extra zeros.

Without loss of generality, consider the two-level hierarchical situation where Y_{ij} represents the j^{th} observation of infant death in the i^{th} individual region $(i = 1,2, \ldots ,m \; and \; j = 1,2, \ldots ,n_i). Let\; m$ be the total number of individuals in each region and $N = \sum_{i=1}^{m}\sum_{j=1}^{n_i} n_i$ gives the total number of observations. The observations may be taken to be independent between regions, but certain within-household and within-individual correlations are anticipated, which can be modelled explicitly through random effects attached to the linear predictors:

$$\log(\lambda_{ij}) = \beta_0 + \sum_{h=1}^{k} \beta_h X_{hij} + U_{0j} + \sum_{h=1}^{k} U_{hj} X_{hij}$$
$$= X_{ij}'\beta + X_{ij}'U$$
$$\log\left[\frac{\phi_{ij}}{(1-\phi_{ij})}\right] = \xi_{ij} = Z_{ij}^{T}\gamma + V_{0j} + \sum_{h=1}^{k} V_{hj} Z_{hij} \qquad (19)$$

Here, the covariates X_{ij} and Z_{ij} appearing in the respective Poisson and logistic components are not necessarily the same, and β and γ are the corresponding vectors of regression coefficients (Moghimbeigi et al., 2008; Meng, 1997). The EM algorithm was also used for over-dispersed count data (McLachlan and Krishnan, 2008).

3.3.2.4 Assessing Model Adequacy

Assume that estimation is by the method of maximum likelihood. Tests for the validity for the null hypothesis can be based on any one of the following three principles:

3.3.2.5 Wald Test

Suppose we are testing $H_0: \beta = \beta_0$ then with non-null standard error of $\widehat{\beta}$, the test statistic is

$$Z = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})} \qquad (20$$

has an approximate standard normal distribution. The multivariable extension for the Wald test of $H_0: \beta = \beta_0$ has test statistic

$$W = (\hat{\beta} - \beta_0)^T [cov(\hat{\beta})]^{-1} (\hat{\beta} - \beta_0) \qquad (21)$$

where $cov(\hat{\beta})$ denote the asymptotic covariate matrix of $\hat{\beta}$ and is the inverse of the information matrix. The (j, k) element of the information matrix is

$$-E\left(\frac{\partial^2 l(\beta)}{\partial\beta_j \partial\beta_k}\right)$$

The asymptotic multivariable normal distribution for $\hat{\beta}$ implies an asymptotic distribution for W. The degrees of freedom equal the rank of $cov(\hat{\beta})$, which is the number of non-redundant parameters in β (Joop, 2010).

3.3.2.6 Likelihood Ratio Test (LRT)

The negative binomial (NB) regression model reduce to the Poisson regression model as $a \to 0$. The test for over-dispersion in NB regression model, $H_0: a = 0 \;\; vs \;\; H_1; a > 0$, can be performed using likelihood ratio test (LRT), $LRT = 2(lnL_1 - lnL_0)$, where L_1 and L_0 are the models' log likelihood under alternative and null hypothesis. Since the null hypothesis is on the boundary of parameter space, the LRT is asymptotically distributed as half of probability mass at zero and half of chi-square with one degree of freedom (Lawless, 1987). In other words, to test the null hypothesis at significance level α, the critical value of chi-square distribution with significance level 2α is used, or reject H_0 if $T > \chi_{1-2\alpha}(1)$.

3.3.2.7 Vuong Test

For non-nested models, a comparison between models with p.m.f. $p_1(.)$ and $p_2(.)$ can be performed using Vuong test, (Vuong, 1989; Greene, 2007).

$$v = \frac{\bar{m}\sqrt{n}}{sd(m)} \tag{22}$$

where *m* is the mean of m_i , $sd(m)$ is the standard deviation of m_i , *n* is the sample size and

$$m_i = ln\left[\frac{p_{1i}(y_i)}{p_{2i}(y_i)}\right]$$

The Vuong test statistic follows a standard normal distribution. As an example, for 0.05 significance level, the first model is "closer" to the actual model if *V* is larger than 1.96. In the other hand, the second model is "closer" to the actual model if *V* is smaller than -1.96, otherwise, neither model is "closer" to the actual model and there is no difference between using the first or the second model. For models with unequal number of parameters, the equation for m_i in Vuong test is slightly modified to account for the difference in the number of parameters,

$$m_i = ln\left[\frac{p_{1i}(y_i)}{p_{2i}(y_i)}\right] - \frac{k_1 - k_2}{2}\ln(n), \tag{23}$$

where k_1 and k_2 are the number of parameters in model 1 and model 2, respectively.

3.3.2.8 Information Criteria (AIC and BIC)

When several models are available, one can compare the models' performance based on several likelihood measures which have been proposed in statistical literatures. Two of the most regularly used measures are Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The AIC penalizes a model with larger number of parameters, and is defined as $= -2lnL + 2p$, where lnL denotes the fitted log likelihood and p the number of parameters. The BIC penalizes a model with larger number of parameters and larger sample size, and is defined as $BIC = -2lnL + pln(n)$, where lnL denotes the fitted log likelihood, p the number of parameters and n the sample size (Ismail and Zamani, 2013).

3.3.2.9 Statistical software packages

In this study we used SAS version 9.2, STATA 13.0, R version 3.2.3 and also ML WIN 2.36 statistical software packages to the analysis of the data.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Descriptive Statistics

This chapter provides the results of the analyses using the methods outlined in chapter three. As shown in Table 4.1, the variance of dependent variable is greater than the mean, implying there is a possibility of overdispersion. A further screening on the data also showed that more than 88% of the respondents, 4905 out of 5534, have no infant deaths per women, indicating an excess zeroes in the dataset.

Table 4.1: Summary statistics for the number of infant death

Variable	Obs	Mean	var	Min	Max
Infant death	5534	0.143	0.192	0	4

Table 4.2 showed the descriptive results of infant deaths per mother. It indicated that more than 88% of the number of infants did not face death, which indicates excess zero and less percentage of non-zero counts.

Table 4.2: Frequency distribution of Number of infant mortalities per women

N	Frequency	Percent
0	4903	88.6
1	500	9.0
2	106	1.9
3	21	0.4
4	4	0.1
Total	5534	100.0

As Figure 1: showed there is massive counts of zero outcomes, the histograms are highly picked at the beginning (zero values). However large observations (i.e. large number of infant deaths per women) are less frequently observed. This leads to have a positively (or right) skewed distribution. Count data models such as zero-inflated models provide better fit which takes into account excess zeros.

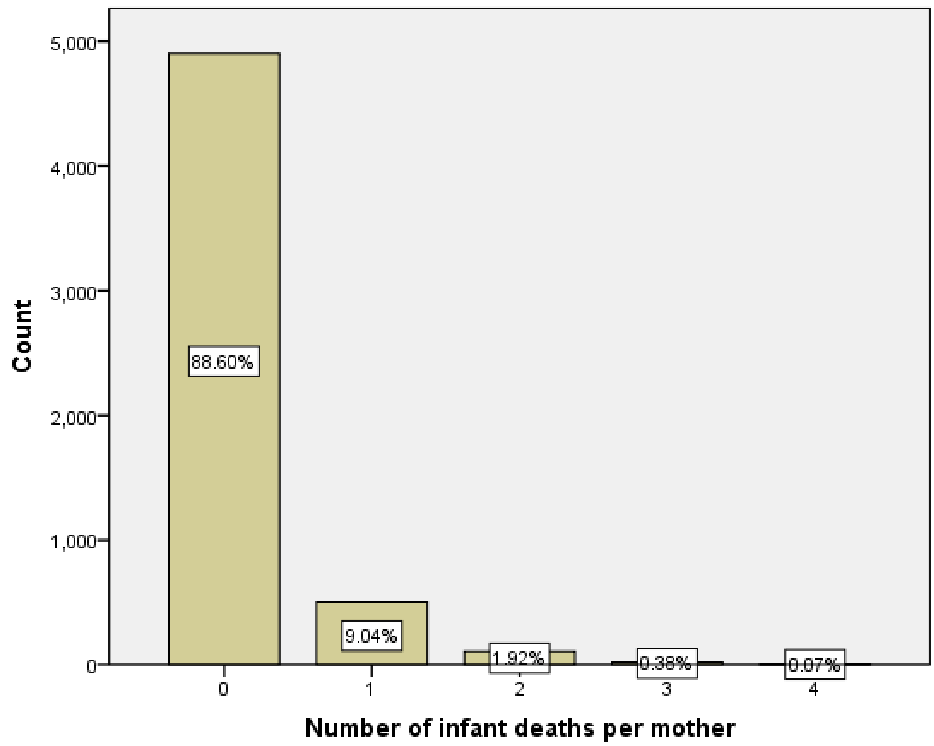

Figure 1: Distribution of response variable, count of infant mortalities per mother

Table 4.3: Descriptive statistics of predictor variables of infant mortalities

Independent Variables	Categories of the variables	N	Mean	Std.D
Region	Tigray	459	0.13	0.402
	Affar	627	0.23	0.617
	Amhara	521	0.10	0.321
	Oromia	684	0.10	0.361
	Somali	655	0.20	0.491
	Benishangul-Gumuz	471	0.18	0.497
	SNNP	722	0.15	0.454
	Gambela	418	0.13	0.457
	Harari	357	0.12	0.343
	Addis Ababa	216	0.04	0.200
	Dire Dawa	404	0.09	0.299
Place of residence	Urban	970	0.09	0.332
	Rural	4564	0.15	0.457
Wealth index	Poor	2929	0.15	0.463
	Middle	878	0.19	0.501
	Rich	1727	0.10	0.350
Education level	Illiterate	3888	0.20	0.505
	Literate	1646	0.02	0.149
Marital status	Married and currently in union	5198	0.14	0.434
	Others	336	0.17	0.503
Sex	Male	2887	0.15	0.455
	Female	2647	0.13	0.420
Birth order	1	1029	0.09	0.333
	2 and 3	1699	0.12	0.392
	4-6	1412	0.13	0.427
	7 and above	1394	0.22	0.550
Birth interval	less than 2	1247	0.28	0.608
	3-4	2246	0.12	0.407
	5 and above	2041	0.08	0.311
Sources of drinking water	Piped water	516	0.06	0.250
	Improved drinking water sources	2176	0.13	0.395
	Unimproved drinking water sources	2842	0.17	0.492

Table 4.3 showed that the lowest mean number of infant mortality was occurred in Addis Ababa and Dire Dawa 0.04 and 0.09, respectively, whereas the highest mean number of infant mortality occurred in Affar, Somali, Benshangul-Gumuz, and SNNP. Regardless of place of residence, large mean number of infant mortalities was occurred in rural areas (mean = 0.15) than urban areas (mean = 0.09).

Table 4.3 revealed that the mean number of infant death for poor and middle economy level were 0.15 and 0.19, respectively. Therefore, a women's living in better and standard economic level experienced to have less mean number of infant deaths as compared with a women living in high income economic level. Literate mothers have lower mean number of infant mortalities than illiterate mothers.

According to the variable, marital status, women's who are not married and not currently in union showed less mean number of infant death. On the other hand women's who are married and currently who live with their husband's encountered less mean number of infant deaths. The mean number of deaths of infant's for male and female were 0.15 and 0.13, respectively, and indicated that the mean number of infant death is more for male infants as compared with female infants.

With regard to birth order, the highest mean number of infant death was observed for infant of birth order of seven and above; and the second highest mean number of infant deaths was observed for infant of birth order of between four and five. The lowest mean number of infant deaths was observed for mothers of first birth order. According to preceding birth interval the observed mean number of infant death is 0.28 and 0.08 for children born less than 2, and 5 and above previous birth interval which mean that the mean death is high for infants of born within short interval of birth. It is also showed that the highest mean number of infant's death occurred with families who use unimproved drinking water sources, which is around 0.17.

4.1.1 Single-level Analysis

4.1.1.1 Variable Selection method

For the identification of potential predictors of infant mortality at the first glance uni-variable analysis was performed using Poisson regression model and all the explanatory variables included in the model are chosen in advance and then multiple regression model with backward selection method was used to select variables before applying different count model. The results identified that: region, place of residence, age, marital status, household size, sex, birth order, and birth interval were significant and the other variables are found to be non-significant and thus excluded from analysis. However, the explanatory variable place of residence was included in the model though it was dropped by backward variable selection method because it was thought purposively to be included in the part of the analysis. After Poisson regression model, the analysis using other count regression models (NB, ZIP, and ZINB) are used with variables selected using backward variable selection method under Poisson and NB.

4.1.1.2 Test of overdispersion and Goodness-of-fit tests

As shown in Table 4.4, the formal test of overdispersion (test of α), in the dataset was performed using the likelihood ratio test with the null hypothesis H_0 of that there is no overdispersion in the dataset. Since, the likelihood ratio test statistic value equal to 58.95 with $p-value=0.000$ was highly significant and thus we conclude that there is overdipersion in the dataset which favor NB regression model. The result of likelihood ratio test statistics was given in the Table 4.4.

Table 4.4: Goodness-fit-of criteria's

Criterion	Model	Value	$p-value$
Likelihood ratio test	NB	58.95	0.000

4.1.2 Statistical Model Selection

4.1.2.1 Information criteria's and LRT values

In this study we considered different count regression models like Poisson, NB, ZIP, and ZINB: likelihood ratio test (LR), Akaike information criterion (AIC), and Vuong test were used to compare and identify the most appropriate count regression model that can provide better fit to the dataset than others.

Table 4.5: Model selection criteria's

	Models			
Criteria	Poisson	NB	ZIP	ZINB
LRT	511.67	389.36	74.85	76.63
AIC	4494.5	4380.701	4352.9334	4404.50

Table 4.5 provides the computed values of AIC and likelihood ratio test for each models and we can use it to compare the goodness-of –fit of model. Further it indicates that NB and ZIP regression models were better fitted than Poisson and NB, respectively, based on their corresponding LRT as well as information criterion's. It was found that the model with the smallest AIC was considered as the best fit of the dataset. Since the LRT value for ZIP model and ZINB model (LRT = 74.85 and 76.63 respectively) were highly significant (p-value < 0.0001) which is in turn supported by the Akaike information criteria obtained under each of the models. According to all these model selection criteria's, the ZIP model is identified as the best model which gives appropriate fit to the dataset than the other models.

4.1.2.2 Vuong test

On the other hand, we used Vuong test to indicate whether zero-inflated model provided an improvement over its counterpart count model. We can determine this by running the corresponding Vuong test of the two models, in our case a zero-inflated model and its non-zero inflated analog. The values of the Vuong test were provided in the Table 4.6.

Table 4.6: Vuong test statistic

	ZIP	ZINB
P	Z = 4.75, Pr > z = 0.0001	_
NB	_	Z = 4.20, Pr > Z = 0.000

We can see from Table 4.6 that the zero-inflated models are better than their non zero-inflated analogs in both versions of the zero-inflated models. The Vuong test shows us that all the candidate models: NB, ZIP, and ZINB performed better than the standard Poisson model. Zero-inflated Poisson performed better than Poisson (V = 4.75, *P-value* = 0.000), zero-inflated negative binomial performed better than negative binomial (Z = 4.20, *P-value* = 0.000). Therefore, again the Vuong test supports the improvement of the zero-inflated Poisson model over Poisson model.

4.1.2.3 Plots of differences between observed and predicted values

The plots of difference between predicted and observed values from each model against the observed value of the response was used to visualize how the model adequately expresses the response variable. In the following table, the values for observed and predicted probabilities for each model was presented. It indicated that, the values are very close to the observed values for both ZIP and ZINB in predicting each count of infant death per mother.

Table 4.7: Values of observed and predicted probabilities

Number of infant death	Observed probabilities	Values of predicted probabilities			
		Poisson model	NB model	ZIP model	ZINB model
0	0.8859	0.8723	0.8848	0.8859	0.8859
1	0.0904	0.1142	0.0948	0.0905	0.0905
2	0.0192	0.0122	0.0157	0.0197	0.0197
3	0.0038	0.0013	0.0034	0.0033	0.0033
4	0.0005	0.0001	0.0009	0.0005	0.0005

Figure 2 provides the fit of Poisson, NB, ZIP, and ZINB models expressed by different colors. It showed that Poisson regression model gives poor fit to predict count of infant deaths per mother. On the other hand though negative binomial regression model predicted 0's as strong as ZIP and ZINB regression models but it showed less predictions of other counts. The graph of ZIP and ZINB regression models for the differences between predicted and observed values looks overlaid which mean that both regression models efficiently predicted the count of infant deaths per mother.

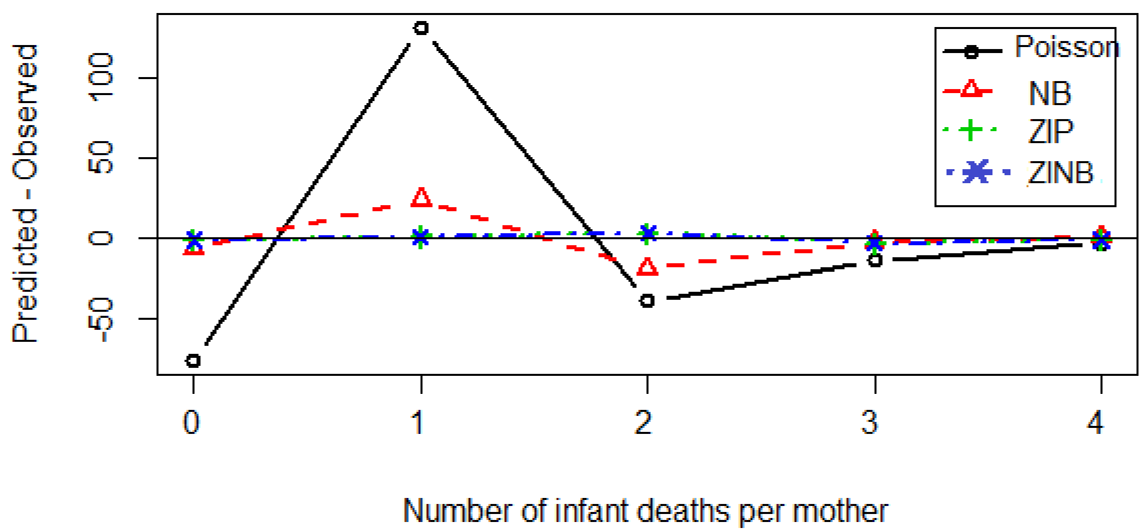

Figure 2: Graph of differences between observed and predicted values for number of infant deaths

It also showed that the ZIP and ZINB regression models account for the excess zeros quite well and both reasonably capture the shape of the distribution of the relative frequencies. Clearly, a zero-inflated model can account for the excess zeros and thus zero-inflated Poisson (ZIP) might be a solution because it can account for the excess zeros and it provides a more flexible estimator for the variance of the response variable.

4.1.3 Discussion of the results from ZIP regression model

It turned out that the ZIP model with region, place of residence, age, marital status, household size, sex, birth order, and birth interval as covariates was the most parsimonious model. Based on the above mentioned criteria for model selection and evaluation, especially, Vuong test, AIC and log-likelihood, we selected zero-inflated Poisson model for fitting the number deaths of infants per women dataset. The cumulative evidence suggested that the zero-inflated model provided an adequate fit to the data than ZINB model for these dataset. Based on the results of all models, it is reasonable to assume that the standard errors of the ZIP model's parameter estimates are unbiased and that the model's estimates are suitable for statistical inference.

Table 4.8: Estimates of model odds ratios and their standard errors of ZIP regression

Parameters	OR	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Region						
Tigray	0.8574	0.1850	0.69	0.4058	0.5968	1 .2322
Dire Dawa	0.5884	0.2183	5.90	0.0151	0.3836	0.9027
Amhara	0.6491	0.1908	5.13	0.0235	0.4466	0.9435
Oromia	0.6473	0.1748	6.19	0.0128	0.4596	0.9117
Somali	0.9444	0.1542	0.14	0.7108	0.6981	2.2776
Affar	1.4531	0.1540	5.89	0.0153	1.0744	1.9652
SNNP	0.9318	0.1585	0.20	0.6559	0.6829	1.2714
Gambela	0.8397	0.1900	0.85	0.3579	0.5786	1.2186
Harari	0.9054	0.2072	0.23	0.6313	0.6031	1.3589
Addis Ababa	0.5789	0.3988	1.88	0.1704	0.2649	1 .2648
Benishangul-Gumuz (Ref.)						
Place of residence						
Rural	1.1465	0.1375	0.99	0.3203	0.8756	1.5013
Urban (Ref.)						
Age	1.0228	0.0226	1.01	0.3141	0.9787	1.0692
Marital status						

Others	1.3557	0.1586	3.68	0.0550	0.9935	1.8498
Married (Ref.)						
House hold size	0.8470	0.0316	27.57	< 0.0001	0.7961	0.9011
Sex						
Male	1.1530	0.0784	3.30	0.0694	0.9887	1.3446
Female (Ref.)						
Birth order						
7 and above	0.6504	0.2064	4.34	0.0372	0.4341	1.2292
4-6	0.3827	0.1204	63.70	< 0.0001	0.3023	0.4845
2 and 3	0.4861	0.1076	44.97	< 0.0001	0.3936	0.6001
1 (Ref.)						
Birth interval						
3 - 4	1.5065	0.2531	2.62	0.1055	0.9172	2.4742
Less than 2	2.1819	0.2392	10.64	0.0011	1.3653	3.4869
Greater than 5 (Ref.)						
cons	0.0403	0.7650	17.62	< 0.0001	0.0090	0.1806
inflate						
Age	1.2058	0.0371	25.45	< 0.0001	1.1212	1.2965
House hold size	1.1220	0.0448	6.62	< 0.0001	0.0278	1.2249
Birth interval						
Between 3 and 4	0.5800	0.3375	2.61	0.1065	0.2993	1.1238
Less than 2	0.2275	0.3484	18.06	< 0.0001	0.1149	0.4504
Greater than 5 (Ref.)						
Birth interval						
cons	-3.7553	0.7491	-5.01	0.000	-5.2236	-2.2871

In order to study the covariates related with infant mortality we fitted the ZIP regression model to predict the count of death of infant's per mother. The predictors related to infant mortality among those in Poisson part of the model such as region, household size, birth order, and birth interval were identified as statistically significant.

The risk of infant mortality for infants born to mothers from Affar, Amhara, Oromia and Dire Dhowa was higher as compared to that of infants born to mother in Benishangul-Gumuz region. The estimated odds of infant death is 1.45, 0.65, 0.65, and 0.59, respectively, for Affar, Amhara, Oromia and Dire Dhowa being Benishangul-Gumuz region as a reference category. Furthermore, household size is found to be the most important factor affecting infant mortality (OR = 0.85). As the size of members of the household increases then it increases the likelihood of infant deaths. Also, the birth order number increases the likelihood of infant mortality, infants of birth order seven and above do have experienced highest likelihood of death and of order two and three do have also relatively high chance of death compared to the death of infants to order one (OR = 0.65 and OR = 0.49, respectively), infants born to first birth number was taken as a reference category.

The results of the study also showed us that the risk of infant mortality was 2.18 times more for infants born less than two years birth interval as compared to infants born after five year birth interval, and the death of infant's increased more for those of between three and four years birth intervals compared to death of infants of birth interval 5 and above but less than the death of infant's born less than two birth interval (OR = 1.51), infant's born 5 and above year preceding birth interval was taken as a reference category.

4.2 Multilevel Analysis of the data

In this study, the dependent variable is the count of infant deaths per mother in her life time. We considered two levels: individual mothers (as level-1 units) and region (as level-2 units) in order to identify significant explanatory variables with the number of infant deaths. The model estimation for the multilevel Poisson, multilevel NB and multilevel ZIP models were done using the SAS NLMIXED procedure with Adaptive Gaussian Quadrature. The number of quadrature points varied from 1 to 25, according to model complexity and computational time. For simpler models like those with independence correlation assumption for random effects, we used 10 quadrature points. Initial values were obtained from the corresponding ZIP and ZINB models.

4.2.1 Model selection and heterogeneity test

4.2.1.1 Goodness of fit and criteria for model selection

The AIC, BIC and deviance are presented in Table 4.9 for model comparison and fit criteria. A smaller value of these criteria indicates a better fit. Therefore, for multilevel ZIP regression model the values of deviance, AIC and BIC are 4094.4, 4144.4 and 4235.7, respectively. The values were small as compared to values of the criteria's for multilevel Poisson and multilevel NB regression models. In general, all criteria indicates that the multilevel zero-inflated Poisson model predicted each count outcome very close to the observed counts and suggests better fit than multilevel Poisson and multilevel zero-inflated negative binomial models. Based on the above mentioned criteria's for model selection and evaluation, we opted for the zero-inflated Poisson model to number of infant death per women data among the candidate model provided.

Table 4.9: The values of model fit comparison criteria's

Criteria	**Models**		
	Poisson model	NB model	ZIP model
Deviance	4320.3	4302.1	4094.4
AIC	4338.8	4322.1	4144.7
BIC	4371.5	4358.5	4235.7

4.2.1.2 Test of heterogeneity

It is important to deal with testing of whether presence of heterogeneity in the dataset is captured by the count regression models. With respect to model performance of count models to account for the heterogeneity in the dataset, we computed the likelihood ratio test as presented in Table 4.10 was used as a measure of each model's performance.

Table 4.10: Likelihood ratio test value

Criteria	**Multilevel Models**		
	Poisson model	NB model	ZIP model
LRT	152.8	99.0	269.1558

As shown in Table 4.10, to compare the heterogeneous multilevel Poisson, multilevel NB and multilevel ZIP models with their standard counterparts, we obtained LRT statistic of 152.8, 99.0 and 269.1558, respectively The values of LRT's for each model is larger than the critical value ${\chi^2}_{1-\alpha}(2) = 5.99$ which indicates an existence of heterogeneous random effect models.

4.2.2 Multilevel zero-inflated Poisson regression model

4.2.2.1 Model comparisons and assessment

Most of the criteria are useful only for comparing the model fit among given alternative models. The Table 4.11 given provides the values of -2* log likelihood and information criteria's for multilevel zero-inflated Poisson (ZIP) regression model for random intercept-only, random intercept and random intercept and coefficient models.

Table 4.11: Model comparison criteria's

Criteria	**Multilevel ZIP regression model**		
	Random intercept-only model	Random intercept model	Random intercept and coefficient model
-2LL	4569.9	4246.1	4049.7
AIC	4575.9	4280.1	4099.7
BIC	4586.9	4342.0	4190.7

The result in the above Table 4.11 showed the values of -2*log likelihood, AIC and BIC were large under random intercept model and it was relatively small value compared values for the other two models for random intercept model and random intercept and coefficient model. The -2*LL, AIC, and BIC value for random intercept and coefficient model were small (4049.7, 4099.7, and 4190.7 respectively), revealed that the model is the most parsimonious. Therefore it suggested that the random intercept and coefficient multilevel ZIP model can efficiently predict the frequency of infant deaths per mother better than random intercept-only and random intercept models can.

4.2.2.2 Random intercept-only model for multilevel zero-inflated Poisson model

The random intercept-only model would try to identify how much variation in between mothers' is due to differences between regions after we control for all our independent variables. The results from the model of random intercept-only model is given in the Table: 4.12.

Table 4.12: Results of random intercepts-only at regional variations model

Parameter	Estimate	Std.Err	DF	t Value	Pr > |t|	Exp (β)
Logistic part						
β_0	-1.4390	0.1085	280	-13.26	< 0.0001	0.2372
Poisson part						
α_0	0.2863	0.1189	280	2.41	0.0167	1.3312
Random effect						
Random intercept-only (σ_{u_0})	0.6982	0.1271	280	5.49	<0.0001	

The result of the random intercept-only model showed that, there is significant variations at regional levels. On the other hand the Poisson part of the result indicates that at the national level, on average, the expected number of under one death per mother in the regions is about 1.3312.

4.2.2.3 Random intercept model for multilevel zero-inflated Poisson regression

Firstly, zero-inflated Poisson model with random effects (mixed or multilevel ZIP) was carried out to account for both clustering and excessive zeros. The model is expected that it would explain the heterogeneity effects due to regional variations (level-2 units). As can be seen from the Table 4.13, infant mortalities per mother varies among the regions since the fact that variance of the random intercept, at region level (i.e. ${\sigma_{u_0}}^2$) was found to be significant (P-value = < 0.0001), Which revealed the number of under one death per mother varies among regions of the country.

Table 4.13: Parameter estimates and standard errors for random intercept multilevel ZIP model

Parameter	Estimate	Std.Err	df.	t-value	Pr > |t|	Exp (β)
ZIP part						
β_0 (const.)	0.2341	0.5769	1753	0.41	0.6849	1.26
Place of residence						
Rural	-0.0522	0.6617	1753	-0.7139	0.0821	0.95
Urban(ref.)						
Age	-0.0506	0.0179	1753	-2.82	0.0049	0.95
Marital status						
Others	-0.5686	0.2383	1753	-2.39	0.0171	0.57
Married(ref.)						
Household size	-0.1922	0.0269	1753	-7.15	< 0.0000	0.83

Sex						
Male	0.1364	0.0897	1573	-1.52	0.1287	1.15
Female (ref.)						
Birth order						
7 and above	-0.1098	0.2503	1753	-0.44	0.6609	0.89
4-6	0.7295	0.1377	1753	5.30	<0.0001	2.07
2 and 3	0.4298	0.1232	1753	3.49	0.0005	1.54
First birth (ref.)						
Birth interval						
Less than 2	-1.2067	0.1329	1753	-9.08	<0.0001	0.29
3-4	-0.7038	0.1262	17532	-5.58	< 0.0001	0.49
5 and above (ref.)						
Logistic part						
α_0 (const.)	-8.9787	4.6695	1753	-1.92	0.0547	0.00
Place of residence						
Rural	1.7011	0.2308	1753	7.37	0.1275	5.48
Urban(ref.)						
Age	0.3771	0.0553	1753	6.83	<0.0001	1.46
Marital status						
Others	-1.0374	0.9795	1753	-1.06	0.2897	0.35
Married(ref.)						
Household size	0.0745	0.0818	1753	0.91	0.3622	1.08
Sex						
Male	0.3542	0.5570	1753	0.61	0. 5393	1.43
Female (ref.)						
Birth order						
7 and above	0.4493	0.1549	1753	2.90	0.0040	1.57
4-6	0.8821	0.2565	1753	6.49	<0.0001	2.42
2 and 3	-1.9969	1.0701	1753	-1.87	0.0622	0.14
First birth (ref.)						
Birth interval						
Less than 2	2.0069	1.5659	1753	1.28	0.202	7.44
3-4	-0.5098	0.0571	1753	-8.94	<0.0001	0.60
5 and above (ref.)						
Random part						
Region effect (σ_{u_0})	2.3137	0.2762	279	8.38	<0.0001	

The results from the random intercept model are given in Table 4.13. Variance components of random effects are observed in both between regions (region-level, ${\sigma_{u_0}}^2$) and between mothers (mother-level, ${\sigma_{u_1}}^2$) and it is significant (p-value = < 0.0001). The results of multilevel ZIP regression of random intercepts was identified that the covariates age, marital status, household size, birth order, and birth interval were found to be related with the infant mortality and was showed to vary among regions of the country.

On the other hand, the predictor variable age of mother at her first birth giving was indicated that it was related with child mortality and it mean that infant death was observed to decreases as age of mother increases. There was also significant relationship between the death of infant and household size. Accordingly, the infant deaths was higher for infants' born from mother of who live with household size of large number compared to infants' born from mother of who live with small number of household size.

On the other hand, it indicated the likelihood deaths of infant was high for infants born of higher birth order compared with infants born of first birth order. According to the results, the infant deaths was higher for infant's born in less birth interval compared with infants born of wide enough of birth interval time.

4.2.3 Random intercept and slope model for multilevel zero-inflated Poisson regression model

In this session we are going to fit the same model, but include a random coefficient and covariance between the intercept and coefficient. According to the results of both models, the model with random intercept and fixed explanatory variables, the model with random intercept and random coefficients was found to be a best fit (i.e. LRT = 4246.1 - 4094.4 = 151.7 which is large enough) in explaining regional differences in infant death per mother. The results of fitted random intercept and random coefficient model are given in Table 4.14.

Table 4.14: The results of random intercept and coefficients multi-level ZIP model

Parameter	**Estimate**	**Std.Err**	**df.**	**t-value**	**Pr > |t|**	**Exp (β)**
ZIP part						
β_0 (const.)	-0.8484	0.7808	279	-1.09	0.2782	0.43
Place of residence						
Rural						
Urban(ref.)	-0.3842	0.2009	279	-1.91	0.0569	0.68

Age	-0.0747	0.0190	279	-3.92	0.0001	0.93
Marital status						
Others	-0.3887	0.2319	279	-1.68	0.0949	0.68
Married(ref.)						
Household size	-0.3339	0.0452	279	-7.39	< 0.0001	0.72
Sex						
Male	0.0504	0.1011	279	0.50	0.6181	1.05
Female (ref.)						
Birth order						
7 and above	1.4301	0.3836	279	3.73	< 0.0002	4.17
4-6	1.0822	0.1503	279	7.20	0.0001	2.95
2 and 3	0.6751	0.1308	279	5.16	0.0001	1.96
First birth (ref.)						
Birth interval						
Less than 2	0.0593	0.1329	279	-9.08	<0.0001	0.94
3-4	0.2550	0.1262	279	0.78	< 0.0001	1.29
5 and above(ref.)						
Logistic part						
α_0 (const.)	-5.8023	1.6884	279	-3.44	0.0007	0.003
Place of residence						
Rural						
Urban(ref.)	0.4130	0.4169	279	0.99	0.3227	1.51
Age	0.0835	0.0510	279	1.64	0.1029	1.09
Marital status						
Others	-0.1326	0.4916	279	-0.27	0.7875	0.88
Married(ref.)						
Household size	0.0658	0.0518	279	1.27	0.2052	1.07
Sex						
Male	0.6758	0.2864	279	2.36	0. 0190	1.97
Female (ref.)						
Birth order						
7 and above	1.2110	0.5528	279	2.19	0.0293	3.36
4-6	-0.0446	0.3997	279	-0.11	0.9113	0.96
2 and 3	-0.1988	0.3891	279	-0.51	0.6097	0.82
First birth (ref.)						
Birth interval						
Less than 2	1.2850	1.9710	279	0.05	0. 9592	3.62
3-4	1.9008	0.4781	279	3.98	< 0.0001	6.69
5 and above(ref.)						
Random part						
Intercept (σ^2_{u0})	3.6920	0.7351	279	5.02	< 0.0001	
Age (σ^2_{u2})	0.1181	0.0223	279	5.30	< 0.0001	
Household size (σ^2_{u4})	0.0501	0.0133	279	3.77	0.0002	

4.2.4 Discussion of intercept and slope model for multilevel zero-inflated (ZIP) regression model results

ZIP model with random effects (mixed ZIP) was carried out to account for both hierarchical structure of the dataset and excessive zeros as presented in the Table 4.14, then two parameters of the random effects were estimated. The first estimate of random effect model is random intercept term ($\sigma_{u_0}{}^2 = 3.6920$) indicate that there is fixed effects of each of covariates incorporated in the model, which is significant. Also this random effect shows that the model with random intercept was found to have a better fit in explaining regional differences in under one death per mother. Overall regional variance constant term and the random variability for age, and household size were found to be significant and suggest that there exists difference in the number of infant death per mother among all regions in Ethiopia. On the other hand it indicates the effect of age, household size on infant mortalities per mother differs across regions significantly.

The results of the multilevel zero-inflated Poisson regression model provided that age, household size, birth order and birth interval were the significant independent variables that were identified to affecting infant death. The impact of maternal age at birth draws attention of many researchers studying determinants of infant and child mortality. General notion is that; unfavorable effect of teen maternity on infant mortality is related to reproductive immaturity of mother. Most of the studies investigate a curvilinear relationship between the maternal age at birth and infant mortality, high risks having infant mortality at very young and old ages, Bhalotra and van Soest (2008).

In this study, the variable age was found to be increase the likelihood death of infant born to teenage mothers and it then there was a tendency to have less infant deaths as they become young (β = -0.0747, p-value = 0.0001) and these finding is consistent with (Tibebu, 2011). As household size, infant's born of mother who live with large size of household were more likely to die than those infants of mother with small household size (β = -0.3339, p-value = 0.0001). According to the results from the model the infant mortality is 1.40 times higher as number of household members increases.

In addition, Birth order was another variable which can significantly affect with the infant to die as per the results of the multilevel ZIP model regression to the dataset. Accordingly, infant's birth to order 7 and above were more likely to die than infant of birth order one (ref. category). The result showed also the risk of infant mortality for infants belonging to birth order 7 and above is 1.43 times more than those of infant born to first birth order. The risk of infant mortality for infants belonging to birth order between 4 and 6 is 1.08 times more than those of infant born to first birth. The risk of infant mortality for infants belonging to birth order 2 and 3 is 0.67 times more than those of infant born to first birth.

According to the variable birth interval (we considered previous birth interval in this study); there is a general thought that longer birth intervals improve the survival chance of the following infant. Short preceding birth interval influences infant mortality through three mechanisms. First, closely spaced births cause depletion of the mother. Second mechanism is through sibling competition and the third is transition of infectious diseases between the closely spaced children (Majumder et al. 1997). The first one is the biological and the other two are behavioral effects of short preceding birth interval (Koenig et al. 1990). Maternal depletion occurs as a result of repeated and closely spaced pregnancies. Closely spaced pregnancies do not give the mother enough time to recover from the adverse physiologic and nutritional demands related to pregnancy (Koenig et al. 1990). The child who is born in such an environment suffers from low birth weight, short duration of gestation and growth retardation (Majumder et al. 1997). The result of the study revealed that, the risk of infant mortality for infants born to birth interval between 3 and 4 were 0.26 times more than those of reference group. On the other hand it is 0.06 times more for birth interval less than 2 as compared with reference group.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The main objective of this study was to apply the applications of single level and multilevel count regression models whenever there is an excess number of zeros and additional sources of heterogeneity in the dataset. From the results of single level count regression models, zero-inflated Poisson model was selected as the best fit to predict current infant death. Accordingly, the variables like region, household size, birth order, and birth interval were found to have significant effect with infant mortality.

The study showed that when there is heterogeneity across count data, modelling multilevel count regression model with random effects as covariates provides an unbiased and accurate estimates of parameters. The multilevel zero-inflated Poisson analysis identified that a large degree of heterogeneity in infant mortalities per mother is different among regions of Ethiopia. This study was often interested in understanding how explanatory variables of infant mortalities such as region, place of residence, age, marital status, household size, sex of infant, birth order and birth interval time would affect the risk of infant deaths per mother in Ethiopia. The study also help to examine and address the function of explanatory variables with number of infant deaths per mother.

Using fixed part of multilevel ZIP model the variables like age, household size, birth order, and birth interval were found to be statistically significant effect with infant mortality. The random part of multilevel ZIP model also revealed that infant deaths per mother differs among regions of the country in terms of both age and household size. Most importantly, the covariate age of mother was found to be significantly increase the risk of death of infant born at early age of the mother. Furthermore, the covariate household size was also identified statistically that infants born to mother live with large family members have more likely to die. In addition to that, birth order also was identified as significant variable affecting infant mortality. It indicates that as birth order increases the likelihood death of infant death become wide large. On the other hand, there is a significant effect between the variable birth interval and number of infant deaths.

5.2 Recommendations

Based on the findings that we have obtained we recommend the following issues:

- Different efforts should be made by the society and the government to control the death of infants.
- It is useful to give attention that a lot of effort needs to be made in family planning programs to give awareness on early marriage, spacing birth interval to reduce infant mortality.
- We recommend also health institutions made effort to give awareness for mothers regarding birth control methods.
- Further researchers can extend this study by using multilevel ZINB model.

REFERENCES

Afifi A.A., Kotlerman J.B., Ettner S.L., Cowan M., 2007. Methods for Improving Regression Analysis for Skewed Continuous or Counted Responses. Annual Review of Public Health 28, 95–111.

Agarwal D., Gelfand A., Citron P., 2002. Zero-infated models with application to spatial count data. 9, 341–355.

Agresti A., 2002. Agresti - Categorical Data Analysis, 2nd ed.

Akbarzadeh B.A., Pourhoseingholi A., Zayeri F., Akbar A.J., Moayed S.A., 2013. Application of Zero-Inflated Poisson Mixed Models in Prognostic Factors of Hepatitis C. Bio Med Research International 2013, 1–5.

Andy H., Wang K., Jane A., Kelvin K.W., Geoffrey J.M., 2006. Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research* 15, 47–61.

Ayati E., Abbasi E., 2014. Modeling Accidents on Mashhad Urban Highways. *Open Journal of Safety Science and Technology* 2014, 22–35.

Bhalotra S., van Soest A., 2008. Birth-spacing, fertility and neonatal mortality in India: Dynamics, frailty, and fecundity. *Journal of Econometrics* 143, 274–290.

Brocklehurst, P., Kurinczuk J., Hollowell J., Juszczak E., Knight M., Gray R., 2011. National Perinatal Epidemiology Unit Final Report to the Policy Research Programme, Department of Health, National Perinatal Epidemiology Unit, University of Oxford.

Cameron A.C., Trivedi P.K., 2005. Micro Econometrics Methods and Applications.

CSA, 2014. Ethiopia Mini Demographic and Health Survey 2014.

Dejen T, Muniswamy B., Konda N.R., 2015. Comparison of Test Statistic for Zero-Inflated Negative Binomial against Zero-Inflated Poisson Model. Indian Journal of Science and Technology 8, 349–357.

Dobson A.J., 2002. An Introduction to Generalized Linear Models, 2nd ed.

Esayas M., 2003. Mother's Work Status and Infant Mortality in Ethiopia: a study based on demographic and health survey data. Addis Ababa.

Ezra M., Gurmu E., 2002. Breastfeeding, birth intervals and child survival analysis of the 1997 community and family survey data in southern Ethiopia. *Ethiopian Journal of Health 16*, 41–51.

FANG R., 2008. Zero-Inflated Negative Binomial Regression Model for Over-dispersed Count Data with Excess Zeros and Repeated Measures.

Goldstein H., 1999. Multilevel Statistical Models.

Greene W., 2007a. Functional Form and Heterogeneity in Models for Count Data.

Greene W., 2007b. Functional forms for the negative binomial model for count data. Economics Letters 99, 585–590.

Gupta R., Marino B.S., Cnota J.F., Ittenbach R.F., 2013. Finding the right distribution for highly skewed zero-inflated clinical data. *Epidemiology Biostatistics and Public Health* 10, 1–15.

Ismail N., Zamani H., 2013. Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models. Casualty Actuar. Soc. E-Forum.

John M.M., Pamela K.L., 2010. Count Models in Criminology.

Joop J., 2010. Multilevel Analysis Techniques and Applications, 2nd ed.

Koenig M.A., Phillips J.F., Campbell O.M., Dsouza S., 1990. Birth intervals and Childhood Mortality in Rural Bangladesh 27, 251–265.

Lawless J., 1987. Negative binomial and mixed Poisson regression. *Negative binomial regression models* 15, 209–225.

Loquiha O., Hens N., Chavane L., Temmerman M., Aerts M., 2013. Modeling heterogeneity for count data a study of maternal mortality in health facilities in Mozambique 55, 647–660.

Lord D., 2010. The Statistical Analysis of Crash-Frequency Data A Review and Assessment of Methodological Alternatives.

Majumder A.K., May M., Pant P.D., 1997. Infant and child mortality determinants in Bangladesh: Are they changing? Journal of Biosocial Science 29, 274–290.

McElduff F., 2012. Models for Discrete Epidemiological and Clinical Data.

McLachlan G.J., Krishnan T., 2008. The EM Algorithm and Extensions. pp. 1–40.

Meng X.L., 1997. The EM Algorithm an Old Folk-song to a Fast New Tune. *Journal of Royal Statistical Society* 59, 511–567.

Moghimbeigi A., Reza M., Mohammad K., Mcardle B., 2008. Multilevel Zero-inflated Negative Binomial Regression Modeling for Over-dispersed Count Data with Extra Zeros 35, 1193–1202.

FMOH/MCH, 2015. National Newborn and Child Survival Strategy Document.

Pandey A., Choe MK., Luther NY., Sahu D., Chand J., 1998. Infant and Child Mortality in India. Mumbai, India.

Samuel M., Eshetu W., 2012. Determinants of Infant Mortality in Ethiopia: a study based on the 2005 EDHS data 26, 1–6.

Seckin N., 2009. Determinants of Infant Mortality in Turkey.

Senayit S., 2012. Determinants of mortality among one to five years old children in Ethiopia. Addis Ababa.

Siegel J.S., Swanson D.A., 2004. The Methods and Materials of Demography, 2nd ed.

Sileshi G., 2007. The excess-zero problem in soil animal count data and choice of appropriate models for statistical inference 52, 1–17.

Tibebu G., 2011. Identification of Risk Factors and Regional Differentials in Under-five Mortality in Ethiopia Using Multilevel Count Model. Addis Ababa.

UNICEF, 2013. World Mortality Report 2013.

UNICEF, 2015. The Millennium Development Goals Report.

UNICEF, WHO, 2011. Drinking Water Equity, safety and sustainability, JMP Thematic Report on Drinking Water 2011.

Vuong Q.H., 1989. Likelihood Ratio Tests for Model Selection and Non-Nested hypothesis. *The Econometric Society*, 307–333.

WHO, 2006. Guidelines for Drinking-water Quality, 3rd ed.

Wondewosen K., Neyens T., Faes C., Molenberghs G., Verbeke G., 2014. A Zero-Inflated Overdispersed Hierarchical Poisson Model.

Zhu, H., Luo S., DeSantis S.M., 2015. Zero-inflated count models for longitudinal measurements with heterogeneous random effects. *Statistical Methods in Medical Research*, 1–16.

Appendix

A. Results of Poisson, NB, ZIP and ZINB regression models

Table 5.1: Estimates and standard errors for Poisson, NB, ZIP and ZINB regression models

Parameters	Poisson		ZIP		NB		ZINB	
	OR	Std.Err	OR	Std.Err	OR	Std.Err	OR	Std.Err
Region								
Tigray	0.5822**	0.0907	0.5899**	0.1013	0.5453**	0.0950	0.5940**	0.1017
Dire Dawa	0.4373**	0.0836	0.4049**	0.0840	0.4320**	0.0898	0.4099**	0.0851
Amhara	0.4530**	0.0746	0.4465**	0.0797	0.4455**	0.0805	0.4533**	0.0809
Oromia	0.4535**	0.0668	0.4454**	0.0716	0.4497**	0.0732	0.4501**	0.0721
Somali	0.7021**	0.0863	0.6498**	0.0888	0.6948**	0.0981	0.6638**	0.0911
Affar	0.7229**	0.1003	0.6881**	0.1060	0.7091**	0.1127	0.6995**	0.1084
SNNP	0.6355**	0.0826	0.6411**	0.0925	0.6370**	0.0943	0.6478**	0.0934
Gambela	0.6032**	0.0977	0.5776**	0.1022	0.5773**	0.1049	0.5809**	0.1029
Harari	0.6499**	0.1163	0.6229**	0.1219	0.6355**	0.1252	0.6371**	0.1254
Addis Ababa	0.4048**	0.1480	0.3983**	0.1563	0.4041**	0.1541	0.4199**	0.1652
Benishangul-Gumuz (reference)								
Place of residence								
Rural	1.2299	0.1553	1.1467	0.1577	1.2285	0.1691	1.1393	0.1571
Urban (reference)								
Age	0.9232**	0.0106	1.0228	0.0231	0.9188**	0.0117	1.0199	0.0238
Marital status								
Others	1.3574**	0.1932	1.3573	0.2151	1.3722	0.2218	1.3735**	0.2216
Married (reference)								
House hold size	0.7846**	0.0175	0.8470**	0.0268	0.7979**	0.0193	0.8034**	0.0189
Sex								
Male	1.1484	0.0825	1.1534	0.0905	1.1741**	0.0941	1.1533	0.0909
Female (reference)								
Birth order								
7 and above	0.6129**	0.1155	0.6490**	0.1339	0.6378**	0.1282	0.6847	0.1398

4-6	0.3597**	0.0401	0.3823**	0.0460	0.3782**	0.0472	0.3765**	0.0459
2 and 3	0.4625**	0.0457	0.4858**	0.0523	0.4776**	0.0530	0.4786**	0.0518
First birth (reference)								
Birth interval								
Less than 2	4.6465**	0.6562	0.6671	0.1690	4.5919**	0.6874	2.5580**	0.5947
3-4	2.1142**	0.2977	1.4517**	0.2285	2.0586**	0.3049	1.6823**	0.4007
5 and above (reference)								
constant	2.6703**	0.8612	1.4262	0.6718	2.6002**	0.9319	0.9902	0.4966
/ ln alpha					0.0354**	0.1817	-1.6254	1.1472
Alpha					1.0360**	0.1882	0.1968	0.2258
Inflate								
Age			0.1866**	0.0371				
House hold size			0.1157**	0.0448				
Birth interval								
Less than 2			0.5532	0.3376			-	0.4665
3-4			-0.9307**	0.3120			1.5609**	0.3899
5 and above (reference)							-0.4397	
constant			3.7553**	0.7491			3.7039**	1.1470

B. Results of Multilevel Poisson regression

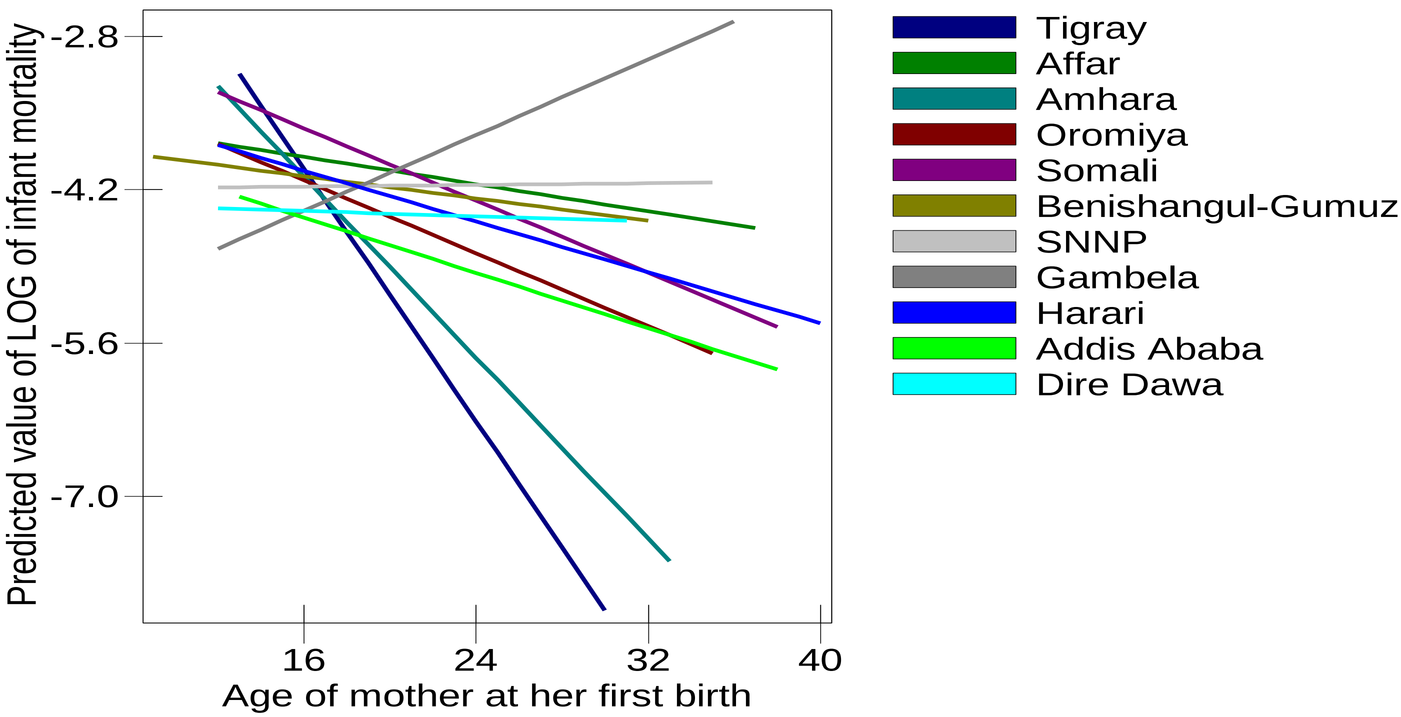

Figure 3: Plot for the predicted value of infant mortality per mother when coefficients of mothers age at her first birth varies

Table 5.2: Estimates and standard errors for multilevel Poisson regression model

Parameters	OR	Std.Err	z	P>|z|	[95% C. I]	
Place of residence						
Urban (Reference)						
Rural	1.4797	0.1838	3.15	0.002	1.1599	1.8877
Age	0.9272	0.0106	-6.62	0.000	0.9067	0.9482
Marital status						
Married (reference)						
Others	1.2887	0.1831	1.79	0.074	0 .9754	1.7025
House hold size	0.7762	0.0175	-11.25	0.000	0.7427	0.8112
Sex						
Female (reference)						

Male	1.1132	0.0801	1.49	0.136	0.9667	1.2818
Birth order						
First birth (reference)						
7 and above	0.5318	0.1007	-3.33	0.001	0.3669	0.7709
4-6	0.3311	0.0372	-9.83	0.000	0.2656	0.4127
2 and 3	0.4397	0.0439	-8.24	0.000	0.3618	0.5347
Birth interval						
5 and above (reference)						
less than 2	4.5737	0.6427	10.82	0.000	3.4727	6.0240
3-4	2.0629	0.2903	5.15	0.000	1.5657	2.7181
cons	0.0161	0.0052	12.71	0.000	0.0085	0.0304
ln(MOTHERS) 1						
(exposure)						
REGION						
var (cons)	0.1178	0.0575			0.0452	0.3067

C. Results of Multilevel negative binomial regression model

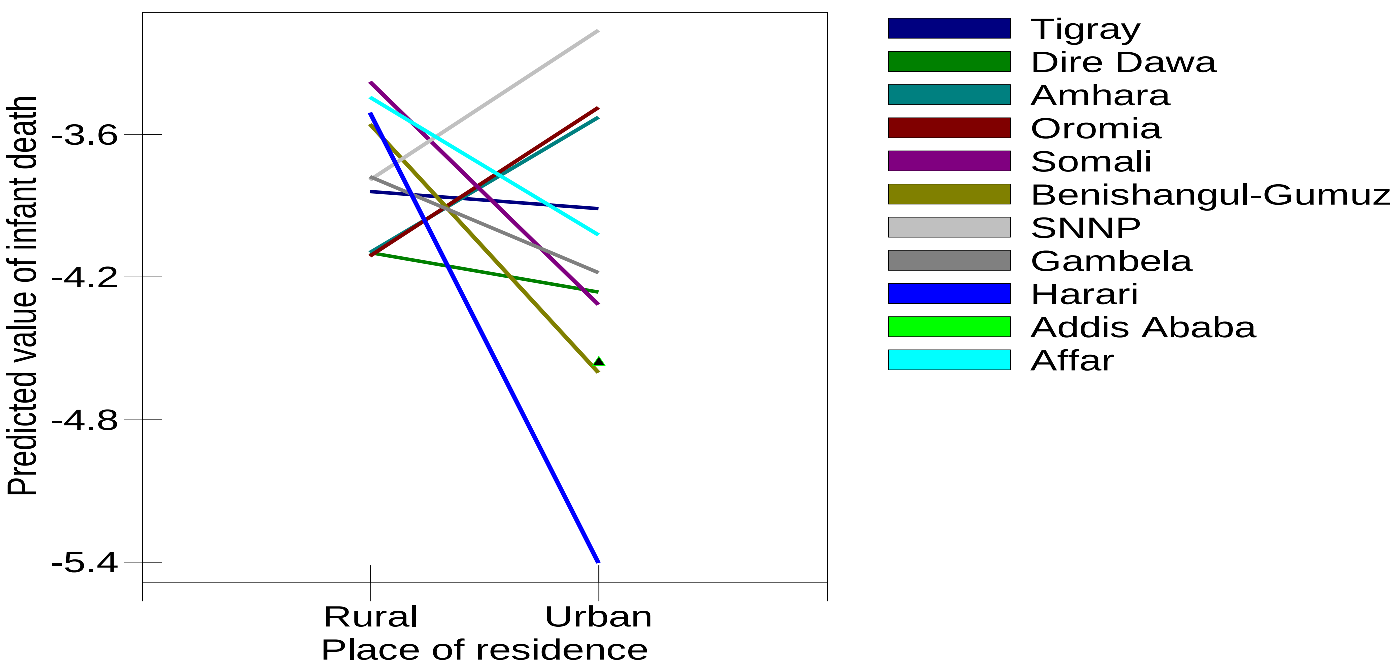

Figure 4: Plot for the predicted value of infant mortality per mother when coefficients or place of residence varies

The following equation identifies the predicted value of infant mortality per mother when coefficients of house hold size varies,

$$\text{RESPO}_{ij} \sim \text{-ve Binomial}(\pi_{ij})$$

$$\begin{aligned}\log(\pi_{ij}) = {} & \text{logEXP}_{ij} + \beta_{0j}\text{cons} + -0.674(0.678)\text{Tigray}_j + 0.121(0.719)\text{Dire Dawa}_j + -0.072(0.680)\text{Amhara}_j + -0.831(0.630)\text{Oromia}_j + -0.004(0.594)\text{Somali}_j + \\ & 0.173(0.615)\text{Benishangul-Gumuz}_j + -0.253(0.605)\text{SNNP}_j + -0.261(0.672)\text{Gambela}_j + -0.521(0.661)\text{Harari}_j + -1.907(0.965)\text{Addis Ababa}_j + 0.000(0.000)\text{Affar}_j + \\ & -0.017(0.071)\text{Tigray.HH\_SIZE}_{ij} + -0.209(0.089)\text{Dire Dawa.HH\_SIZE}_{ij} + -0.166(0.082)\text{Amhara.HH\_SIZE}_{ij} + -0.022(0.059)\text{Oromia.HH\_SIZE}_{ij} + \\ & -0.067(0.048)\text{Somali.HH\_SIZE}_{ij} + -0.123(0.059)\text{Benishangul-Gumuz.HH\_SIZE}_{ij} + -0.072(0.055)\text{SNNP.HH\_SIZE}_{ij} + -0.079(0.067)\text{Gambela.HH\_SIZE}_{ij} + \\ & -0.046(0.074)\text{Harari.HH\_SIZE}_{ij} + 0.074(0.128)\text{Addis Ababa.HH\_SIZE}_{ij} + -0.058(0.036)\text{Affar.HH\_SIZE}_{ij}\end{aligned}$$

$$\beta_{0j} = -3.067(0.401) + u_{0j}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u) : \Omega_u = \left[0.087(0.047)\right]$$

$$\text{var}(\text{RESPO}_{ij}|\pi_{ij}) = \pi_{ij} + \pi_{ij}^2 * 1.971(0.167)$$

(5534 of 5534 cases in use)

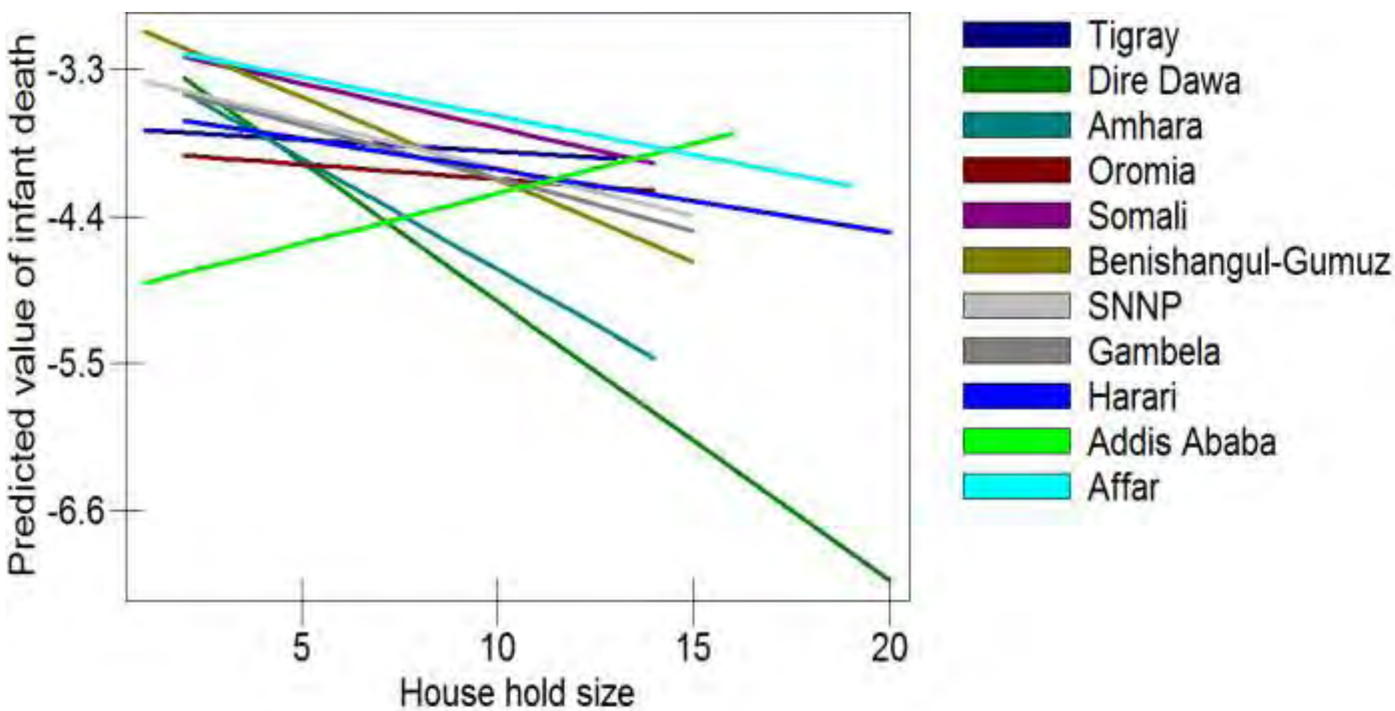

Figure 4: Plot for the predicted value of infant mortality per mother when coefficients of house hold size varies

Table 5.3: Estimates and standard errors for multilevel NB regression model

Parameters	OR	Std.Err	z	P>|z|	[95% C. I]	
Place of residence						
Urban (reference)						
Rural	1.4543	0.2113	2.58	0.010	1.0938	1.9334
Age	0.9178	0.0126	-6.26	0.000	0.8934	0.9427
Marital status						
Married (reference)						
Others	1.3309	0.2352	1.62	0.106	0.9414	1.8819
House hold size	0.7825	0.0204	-9.41	0.000	0.7435	0.8235
Sex						
Female (reference)						
Male	1.1461	0.1007	1.55	0.121	0.9648	1.3613
Birth order						
First birth (reference)						
7 and above	0.5706	0.1219	-2.63	0.009	0.3754	0.8675
4-6	0.3461	0.0474	-7.74	0.000	0.2646	0.4527
2 and 3	0.4475	0.0550	-6.54	0.000	0.3517	0.5695
Birth interval						
5 and above (reference)						
less than 2	5.0309	0.8031	10.12	0.000	3.6794	6.8789
3-4	2.0771	0.3234	4.70	0.000	1.5309	2.8183
cons	0.0187	0.0072	-10.37	0.000	0.0088	0.0397
ln(MOTHERS)1	0.4993	0.1497	3.33	0.001	0.2058	0.7928
lnalpha						
REGION						
var (cons)	0.1227	0.0619			0.0456	0.3298