



**DEBRE BERHAN UNIVERSITY**  
**SCHOOL OF POST GRADUATE STUDIES**  
**COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES**  
**DEPARTMENT OF STATISTICS**

**DETERMINANTS OF UNDER-FIVE DEATHS IN ETHIOPIA,  
WITH AN APPLICATION OF COUNT REGRESSION MODELS**  
**BY**  
**BUZUNEH TASFA**

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, COLLEGE OF NATURAL SCIENCES OF DEBRE BERHAN UNIVERSITY IN THE PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN BIostatISTICS

**June, 2019**  
**Debre Berhan,**  
**Ethiopia**

**DEBRE BERHAN UNIVERSITY**  
**SCHOOL OF POST GRADUATE STUDIES**  
**DEPARTMENT OF STATISTICS**

**DETERMINANTS OF UNDER-FIVE DEATHS IN ETHIOPIA,  
WITH AN APPLICATION OF COUNT REGRESSION MODELS**

**BY**

**BUZUNEH TASFA**

**ADVISOR BEZAREDE MEKONNEN (ASSISTANT PROF)**

**CO-ADVISOR MELAKU FEKADU (MSc)**

**A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, COLLEGE OF  
NATURAL SCIENCES OF DEBRE BERHAN UNIVERSITY IN THE PARTIAL  
FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF  
MASTER OF SCIENCE IN BIOSTATISTICS**

**JUNE 2019**  
**Debre Berhan,**  
**Ethiopia**

## DECLARATION

I, the undersigned, declare that the thesis is my original work, has not been presented for degrees in any other University and all sources of materials used for the thesis have been duly acknowledged.

**Name: Buzuneh Tasfa**

\_\_\_\_\_  
**Signature**

\_\_\_\_\_  
**Date**

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

**Bezarede Mekonnen (ass .prof)**

\_\_\_\_\_  
**Signature**

\_\_\_\_\_  
**Date**

**Date of submission: June, 2019**

**DEBRE BERHAN UNIVERSITY**  
**SCHOOL OF GRADUATE STUDIES**  
**DEPARTMENT OF STATISTICS**

As **Thesis Advisor**, We Have Read The Research Paper prepared by **Buzuneh Tasfa**. Entitled With, **DETERMINANTS OF UNDER-FIVE DEATHS IN ETHIOPIA, WITH AN APPLICATION COUNT REGRESSION MODELS**. The research paper fulfills the requirements of the university and department paper writing standard. So we have recommended that the research can be submitted for defense.

<b><u>Bezarede Mekonnen (ass .prof)</u></b>	_____	_____
<b>Advisor</b>	<b>Signature</b>	<b>Date</b>
<b><u>Melaku Fekadu (MSc)</u></b>	_____	_____
<b>Co-advisor</b>	<b>Signature</b>	<b>Date</b>

As member of the board examiners of MSc. Thesis open defense Examination of Buzuneh Tasfa, we have read and evaluated the thesis and examined the candidate. Hence we recommend that the thesis can be accepted as it fulfills the requirements for the Master of Science in Biostatistics.

<b><u>Mesfin (PhD)</u></b>	_____	_____
<b>Name of Chairman</b>	<b>Signature</b>	<b>Date</b>
<b><u>Bezarede Mekonnen (Ass .prof)</u></b>	_____	_____
<b>Name of Advisor</b>	<b>Signature</b>	<b>Date</b>
<b><u>Melaku Fekadu (MSc)</u></b>	_____	_____
<b>Name of Co-Advisor</b>	<b>Signature</b>	<b>Date</b>
<b><u>Dejen Tesfaw (PhD)</u></b>	_____	_____
<b>Name of Internal Examiner</b>	<b>Signature</b>	<b>Date</b>
<b><u>Nathnael Mamuye (Ass.prof)</u></b>	_____	_____
<b>Name of External Examiner</b>	<b>Signature</b>	<b>Date</b>

---

**Chair of Department or Graduate Program Coordinator**

## **ACKNOWLEDGMENT**

It gives me a great pleasure to come to this final point to express my deepest gratitude to all individuals and institutions who contributed in one way or another to my study and the realizing of this thesis paper.

First and foremost, I would like to thank my almighty God with his mother, Virgin Mary (**ቅድስት ድንግል ማርያም**) for being with me in all aspects during my life and giving the opportunity to pursue my graduate study. Thanks too **ቅድስት ድንግል ማርያም!!!!**

I would like to gratefully and sincerely thank, my thesis advisor Bezarede Mekonnen (Assistant Prof), for his supervision, kind help, invaluable advice, professional guidance, useful comments and suggestions and my deep appreciation and thanks goes to my Co-advisor Melaku Fekadu (MSc), for his invaluable advice, professional guidance, useful comments and suggestions.

My deepest thanks go to all members of Debre Berhan University, instructors and all individuals in the Department of Statistics for their moral support and positive attitude when I needed them during my course study.

I am also highly grateful to all those who helped me in my academic career. I owe profound thanks to Andualem Fentie, Aleka Getachew, Anberbir Abebe and all my friends who have provided me help and moral support in one way or another to accomplish this study.

Finally, I am very much indebted to members of my family, relatives and friends, who have given me, support and encouragement. My special appreciation goes to my brothers and sisters, Yirgu Abebe, Iyerus Alemu, Shola Tesfa, Aster Tesfa, Takele Tesfa and Getaw Sintayew who were eager to see my achievements.

# **Table of Contents**

<b>ACKNOWLEDGMENT .....</b>	<b>i</b>
<b>LIST OF TABLES.....</b>	<b>v</b>
<b>LIST OF FIGURES.....</b>	<b>vi</b>
<b>LIST OF ABBREVIATIONS.....</b>	<b>vii</b>
<b>ABSTRACT.....</b>	<b>viii</b>
<b>1. INTRODUCTION.....</b>	<b>1</b>
1.1 Background of the study.....	1
1.2 Statement of the problem .....	4
1.3. Objectives of the study .....	6
1.3.1 General objectives.....	6
1.3.2. Specific objective .....	6
1.4. Significance of the Study.....	6
1.5. Limitation of study.....	7
<b>2. LITERATURE REVIEW.....</b>	<b>8</b>
2.1. Conceptual Framework.....	8
2.2. Causes of under-five death .....	9
2.3. Socio economic factors.....	10
2.4. Environmental factors.....	13
2.5. Empirical studies on the determinants of under-five death.....	14
<b>3. METHODOLOGY.....</b>	<b>19</b>
3.1. Source of the Data .....	19
3.2. Variables in the Study.....	20
3.2.1. Response variable .....	20
3.2.2. Explanatory variables.....	20

<b>3.3 METHODS OF DATA ANALYSIS .....</b>	<b>22</b>
3.3.1 Introduction on count data model .....	22
<b>3.3.2. STATISTICAL MODELS .....</b>	<b>23</b>
3.4 Poisson Regression Model .....	24
3.4.1 Poisson Parameter estimation .....	25
3.5 Negative Binomial Regression Model.....	27
3.5.1 Parameter estimation of NB model .....	28
<b>3.6. Zero-inflated Regression Models .....</b>	<b>28</b>
3.6.1. Zero-inflated Poisson Regression Model.....	29
3.6 .1.1 Parameter estimation of zero inflated Poisson model.....	30
3.6.2. Zero-inflated negative binomial regression model .....	30
3.6.2.1. Parameter Estimation of ZINB Model.....	32
3.6.3. Zero-inflated Generalized Poisson (ZIGP) regression.....	33
3.6.4. Hurdle Regression Models .....	34
3.6.4.1 Hurdle Poisson regression model .....	36
3.6.4.2 Hurdle negative binomial regression model.....	37
<b>3.7. Model Fitting Test (Goodness of fit tests).....</b>	<b>38</b>
3.7.1 Tests for the Comparison of the Models .....	39
3.7.1.1 Tests for comparison of nested models .....	39
3.7.1.2 Test for comparison of Non-nested models .....	39
3.7.3 Test for individual predictors.....	41
3.8 Statistical software packages.....	41
<b>4. STATISTICAL DATA ANALYSIS AND RESULTS.....</b>	<b>42</b>
4.1 Result of descriptive statistics.....	42
4.1.1. Number of under-five mortality per mother in Ethiopia .....	42

4.1.2. Summary Statistics for explanatory Variables.....	44
<b>4.2. Count Regression Model Results For U5M .....</b>	<b>50</b>
4.2.1. Variable Selection .....	50
4.2.2. Goodness of Fit and Test of over dispersion .....	50
4.2.3. Parameter Estimation.....	51
4.3. Comparison of Models.....	54
4.3.1. Vuong Test for non-nested models.....	55
4.3.2. Predicted value and Probability .....	57
4.4. Parameter Estimation of HNB regression Model for under-five mortality .....	58
4.5. Interpretations of HNB Model result for under-five deaths .....	60
4.5.1. Interpretation of HNB model for covariates of zero parts .....	63
<b>5. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS.....</b>	<b>65</b>
5.1. Discussion of the Results.....	65
5.2. Conclusions.....	69
5.3. Recommendations .....	70
<b>REFERENCE.....</b>	<b>71</b>
<b>APPENDICES .....</b>	<b>76</b>

## LIST OF TABLES

<b>Table 3. 1.</b> Independent variables with their label and category.....	20
<b>Table 4. 1.</b> Frequency distribution of number of under-five deaths in Ethiopia.....	42
<b>Table 4. 2.</b> Summary data for the dependent variable, number of under-5 deaths .....	43
<b>Table 4. 3.</b> Presents summary statistics of the categorical variables that are assumed to affect the number of under-five mortality among Women ages 15-49.....	44
<b>Table 4. 4.</b> The results of over-dispersion test after fitting a Poisson regression.....	50
<b>Table 4. 5.</b> Estimated Coefficients and Std.err of the Poisson, NB, ZIP and ZINB models.....	51
<b>Table 4. 6.</b> Estimated Coefficients and Std.Err of the ZIGP, HP and HNB count models .....	53
<b>Table 4. 7.</b> Model selection criteria for PR, NB,GP, ZIP ZINB, ZIGP, HP and HNB models for the number of under-five death dataset.....	55
<b>Table 4. 8.</b> Vuong non- nested tests results.....	56
<b>Table 4. 9.</b> Zero count capturing in count models for under-five mortality.....	57
<b>Table 4. 10.</b> Observed and predicted probability from Poisson, NB, ZIP, ZINB, ZIGP, HP and HNB model for under-five deaths.....	57
<b>Table 4. 11.</b> Estimates and standard errors for HNB model for under five death in Ethiopia .....	58
<b>Table A 1.</b> Estimates and standard errors for NB model .....	75
<b>Table A 2.</b> Estimates and standard errors for ZIP model.....	78
<b>Table A 3.</b> Estimates and standard errors for ZINB model.....	81
<b>Table A 4.</b> Estimates and standard errors for HP model.....	84

## LIST OF FIGURES

<b>Figure 4. 1.</b> Histogram of the number of under- five mortality in Ethiopia.....	44
<b>Figure 4. 2.</b> The Observed and predicted probability of Poisson, NB, ZIP, ZINB, ZIGP, HP and HNB models.....	58

## **LIST OF ABBREVIATIONS**

WHO	World Health Organization
UNICEF	United Nations Children Fund
CSA	Central Statistical Agency
EDHS	Ethiopia Demographic and Health Survey
U5M	Under-five Mortality
SDG	Sustainable Development Goals
CMR	Child mortality rate
UNMDG	United Nations Millennium Development Goals
EHFP	Ethiopia health and family planning
GTP	Growth and transformation plan
AIC	Akaike Information Criteria
BIC	Bayesian Information Criteria
EA	Enumeration Area
LRT	Likelihood Ratio Tests
PRM	Poisson Regression Model
ZIRM	Zero Inflated count Regression Model
NB	Negative Binomial
ZIP	Zero-Inflated Poisson
ZINB	Zero-Inflated Negative Binomial
ZIGP	Zero-inflated Generalized Poisson
HRM	Hurdle Regression Model

## **ABSTRACT**

*Under-five mortality is the probability of dying between birth and the age of five of life. The first five years of life are the most crucial to the physical and intellectual development of children with strong reflections on the future of a country. Globally the number of deaths among children under age five has reduced from 12.4 million in 1990 to 5.6 million in 2016. About half of under-five mortality occurs in only six South-East Asia and Africa countries including Ethiopia. About 194,000 children dies before the fifth birthday every year in Ethiopia. Therefore, main objective of this study was to identify and analysis the most determinants of under-five mortality in Ethiopia based on Ethiopian Demographic and Health Survey the data which is conducted in 2016. To achieve the objective of this study count regression models were used to identify the most important significant independent variables such as socio-economic, demographic, biological and environmental related variables that affects the number of under-five deaths per women as the response variables. In this study, the most appropriate fitted model was selected from different seven count regression models: Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-inflated generalized Poisson (ZIGP), hurdle Poisson (HP) and hurdle negative binomial (HNB) using different comparison techniques. The result revealed that HNB model was found to be the most appropriate model to predict the number of under-five mortality in Ethiopia. This study captured predictor variables that had significant factors influencing the number of under-five deaths have been identified. For selected ordinary HNB model, the truncated negative binomial part, predictor variables like women education level, women age at first birth, place deliver, family size, breastfeeding status, wealth index, marital status, types birth, anemia status, source of water drinking, region, toilet facilitate, preceding birth interval and vaccination status were statistically significant factors influencing the number of under-five mortality. While for logistic part, women education level, family size, breastfeeding status, source of water drinking, toilet facilitate, wealth index, marital status and type's birth were found to be statistically significant determinants of the number of under-five deaths in Ethiopia. Concerned effort should be made to extend educational programmers aimed at educating mothers on the benefits of, age of first birth, spacing birth interval, vaccination of child and place of delivery in order to reducing under-five mortality.*

**Key words:** *under-five mortality, hurdle negative binomial, count regression models*

# **1. INTRODUCTION**

## **1.1 Background of the study**

Under-five mortality rate is defined as the probability (expressed as a rate per 1 000 live births) of a child born in a specified year dying before reaching the age of five years subject to current age-specific mortality rates (UNICEF, 2017). Improving child survival has been a priority for both policy makers and health advocates worldwide (Amiri and Gerdtham, 2013). Under-five child mortality rate is widely accepted as an indicator of socio-economic development and a reflection of a country's health care system and quality of life. In order to prevent child deaths and ensure healthy child survival, reducing under-five mortality to at least as low as 25 per 1,000 live births by 2030 are referred as the third Sustainable Development Goals (SDG) (Kumar et al., 2016). It is one of the serious issues facing not only Sub-Saharan Africa regions, but also many countries throughout the world (UNICEF, 2010). This shows the importance of interventions to reduce the rates of under-five mortality. Due to increasing under-five mortality rates, there have been efforts made to solve this issue facing the world (Getachew and Bekele, 2016).

Under-five death in the world has been a subject of academic interest, political advocacy and speculation. This is because, child mortality rate (CMR) is considered to be one of the strongest indicators of a country's wellbeing, as it reflects social, economic and environmental conditions in which children (and others in the society) live, including their health care. The first five years of life are the most crucial to the physical and intellectual development of children with strong reflections on the future of a country (Amponsah, 2017). Health and life actuaries have always had an especial interest in the development and construction of mortality rates as actuarial mortality reflect a country's socio-economic level of development and quality of life and are used for monitoring and evaluating population and healthcare intervention programs and policies.

According to (WHO, 2016), every year millions of children under-five years of age die. Approximately 5.6 million children under-five years die each year with large variation across region and countries. Globally the number of deaths among children under age five has reduced from 12.4 million in 1990 to 5.6 million in 2016 (UNICEF, 2017). Under-five mortality rates dropped by more than half from 1990 to 2016, decreasing from an estimated 93 deaths per 1000 live births to 41 deaths per 1000 live births, respectively. Much of that reduction has been achieved in recent years, with the rate of decline increased from 1.9 percent in 1990-2000 to 4.0 percent in 2000-2016 (Hug et al., 2017).

Despite the progress made, Sub-Saharan and South-East Asia still carry the burden of 79 percent of all global under-five deaths (UNICEF et al, 2017). These two regions share nearly the same part: 39 percent of all deaths occurred in southern Asia whereas 38 percent occurred in sub-Saharan African region (Nyinawajambo, 2018). Sub-Saharan Africa had an average rate of under-five mortality of 79 deaths per 1000 live births which is translated to 1 death for every 13 living children. This rate is almost 15 times as likely compared to countries high countries whose ratio of 1 death for every 180 living children (UNICEF et al., 2017). The astounding thing is that among five countries whose half of the global under-five deaths three of them is part of sub-Saharan countries namely: Nigeria, Ethiopia, and Congo Democratic Republic (UNICEF et al., 2017). Moreover by that time among 6 million estimated under-five deaths half of them will be newborns (Kyei, 2012). This persistence of high mortality rate slowed the growth of the economies in countries with higher mortality for instance sub-Saharan Africa (Amiri & Gerdtham, 2013)

In the past few decades there is decline in under-five mortality in almost all countries of the world, regardless of initial levels, socioeconomic circumstances and development strategies. In spite of this substantial drop in global child mortality rate, about 5.6 million children still die every year before their fifth birthday worldwide which implies 15000 under-five children die each day (WHO, 2017). Ethiopia, one of the developing and second most populous countries in sub-Saharan Africa countries located in Eastern Africa faces high U5M rates. According to Central Statistical Agency (Ethiopia) and ICF International (2016), Ethiopia is one of the few African countries that have made considerable progress in reducing under-five mortality, however the problem still needs due consideration (Organization, 2016). More than one third of

the deaths are largely due to communicable diseases that could have been easily prevented and treated using low-technology interventions (WHO, 2016). Though U5M is one of the public health problem in Ethiopia, evidence revealed that U5M rate has declined by 60% percent from the 1990 figure of 204/1,000 live births to 67/1,000 live births in 2016 (CSA, 2016).

In recent years, child mortality in Ethiopia is reported to show a decrement. For example, neonatal mortality decreased by 41 percent over the 16-year period between 2000 and 2016 EDHS, from 49 deaths per 1,000 live births to 29 deaths per 1,000 live births. Infant mortality declined by 50 percent during the same period (declined from 97 deaths per 1,000 live births to 48 deaths per 1,000 live births). Under-five mortality also decreased by 60% during the same period (declined from 166 deaths per 1,000 live births to 67 deaths per 1,000 live births). The decline reported being continued in the recent EDHS with neonatal, infant and under-five mortality of 29, 48 and 67 deaths per 1,000 live births, respectively. Even if the mortality rate is decreased from time to time, it remains to be among the highest in the world (Hug et al., 2017). As a result, about one in every 35 Ethiopian children dies before his/her first month, one in every 21 children dies before the first birthday, and one in every 15 children dies within the fifth birthday. The major causes of under-five mortality in Ethiopia were ARI/Pneumonia, diarrhea, complications of prematurity, intrapartum related events (birth asphyxia), meningitis, and measles. Other causes of death (including deaths due to severe malnutrition) of under-five mortality (WHO and Mathers, 2016).

In recent years, Ethiopia health and family planning (EHFP) has successfully implemented in a wide array of fertility and mortality reduction interventions. Besides, the growth and transformation plan (GTP) has been developed and under implementation starting from 2016 to improve access and quality of health services. However, despite all of these efforts, health care facilities in Ethiopia are limited and inadequate. Moreover, lack of health personnel, medicines and other facilities are not uniformly available. To expand our understanding about the most common and consistent factors on the risk of under-five child mortality, we have considered possible determinants of under-five mortality by using count regression model. Therefore, this study explores the demographic factors; biological, environmental, socioeconomic characteristics of under-five mortality in Ethiopia based 2016 EDHS data set.

## **1.2 Statement of the problem**

The study of under-five mortality becomes one of the most important researches in developing countries because there is high level of under-five mortality. The current levels of child mortality in Ethiopia are higher than the target of the Sustainable Millennium Development Goals (MDGs) in 2030. Nowadays, mortality is expected to decline very rapidly after birth like in most industrialized countries, with under-five mortality rates being relatively low. In these settings the highest proportion of under-five deaths occur in the first few days of life and are related to genetic, maternal, and perinatal factors. However, in less developed countries with Ethiopia being included, environmental and nutritional risks to survival persist through infancy and afterwards, and those risks will be reflected in an attenuated age pattern and an overall higher level of mortality under age five (Lawn et al.2011). In spite of this substantial drop in global child mortality rate, about 5.6 million children still die every year before their fifth birthday worldwide which implies 15000 under-five children die each day (WHO, 2017).

According to (WHO, 2016), a child's risk of dying is highest in the neonatal period, the first 28 days of life. 44% of children deaths under the age of five take place during the neonatal period. Nearly 3,000,000 babies die every year in their first month of life and a similar number are still born. Within the first month, up to one half of all deaths occur within the first 24 hours of life, and 75% occur in the first week. The 48 hours immediately following birth is the most crucial period for new born survival. This is when the mother and child should receive follow-up care to prevent and treat illness.

Ethiopia is one of the countries with better progress in achieving the MDG target of reducing under-five mortality by two thirds in the past two decades. However, the problem in the country is among the highest in the world. The country is among the 79 countries with under-five mortality rates above the SDG target of 25 under-five deaths per 1000 live births. About 194,000 under-five deaths every year in Ethiopia (UNICEF, 2016) where one in every 15 children dies before celebrating their fifth birthday. Moreover, under-five mortality in Ethiopia is an important concern, where it is essential for monitoring the current health programs and formulating policies for improving the current situation (WHO, 2017). Mostly in Ethiopia, there are limited studies that focused on the determinant factors of under-five mortality using count regression models. To understand the disparity in under-five mortality, this study therefore explores the demographic,

biological, maternal, environmental socioeconomic factors and their effect on under-five mortality in Ethiopia.

Several studies also investigated determinants of under-five mortality in Ethiopia with few variables and statistical methods such as multiple logistic regression, logistic regression, and survival Cox-regression survival analysis. Many of the previous studies done in Ethiopia the number of under-five deaths by linear regression models including the above studies. However, the number of under-five deaths count data with a positively-skewed distribution may not fit well in the linear regression model. Because, linear regression model produces negative values, in other words, linear regression does not account for data being truncated at zero; thus, it could predict negative values which are meaningless (King, 1988; Sturman, 1999). Due to the reasons mentioned above, using linear regression to analyze number of under-five death data may lead wrong conclusions, impossible mean predicted values, inefficient and the least square estimates of dependent variable suffer from these problems and are biased and inconsistent as well. Therefore, to study the number under-five mortality using count regression models is more appropriate than the previous above models.

Also some the previous studies, tried to predict the number of under-five deaths in Ethiopia by using Poisson and negative binomial regression models. However, the nature of count data often involves over-dispersion not only due to unobserved heterogeneity but also due to the excess number of zero (Asada and Kephart, 2007). In this case Poisson or the negative binomial models may not satisfactorily account for excess variability if this variability is indeed due to excess zeros. In such situations, use of these models may likely underestimate the response variable, and may provide misleading results. However, hurdle regression models can be used to increase predictability in situations with excess zeros in the dependent variable (Lambert, 1992).

Moreover, this study involved count regression models like, Poisson regression model, negative binomial regression models, zero-inflated regression models were considered in order to analyze the number of under-five death. In this study, hurdle regression model in addition to the existing models, we were considering different count regression model. Therefore, we were consider different count regression models for the number of under-five deaths and discuss how they can enhance our understanding of the risk factors of under-five mortality based 2016 Ethiopia demography health survey (EDHS) data set.

## **1.3. Objectives of the study**

### **1.3.1 General objectives**

The general objective of the study was to identify the most important determinants factors that affects the number of under-five mortality based on 2016 Ethiopia demography health survey (2016 EDHS) by using the count regression models.

### **1.3.2. Specific objective**

- ❖ To examine the socio-economic, demographic, environmental and biological related factors associated with the under-five motility in Ethiopia.
- ❖ To examine over-dispersion and zero-inflation problem for the number of under-five deaths.
- ❖ To identify appropriate count regression models in order to analyze the number under-five mortality data in Ethiopia.

## **1.4. Significance of the Study**

The findings from this study were useful in many ways. The findings are believed to be useful for policy making, monitoring and evaluation activities of the government and different concerned agencies.

Finally, this study has the following significant

- In order to give a better understanding of the factors affecting the under-five mortality and pronounced impact for the reduction of under-five child mortality in Ethiopia.
- The results can provide an important input for any possible intervention in this area for the future.
- The study would serve as a guide to stakeholders in making informed and intelligent policy decisions with regard to under-five deaths and the management of the risk factors to avoid the death of children in the country.
- This study can be a baseline for the other interested finders (researchers) for further investigation on the under-five mortality in Ethiopia.

## **1.5. Limitation of study**

In this study there were some challenges that we faced. Some important variables that might affect infant mortality were not available in the 2016 dataset. The study used data from national surveys that have inherent gaps such as lack of data on children for women who had died though attempts were made to address them arising from the fact that only surviving women aged 15-49 years were interviewed and Some important variables are not included because of large number of missing values like child weight at birth, and size of child at birth. Child specific some biological variables are not considered in this study as we considered mothers as our response subject.

## **2. LITERATURE REVIEW**

### **2.1. Conceptual Framework**

As noted by the World Bank (2005), child mortality has broad sets of determinants which interconnect with many factors such as fertility, maternal factors, nutrition, education, infrastructure development, safe water supply, electrification, disaster prevention and relief, agricultural output, and public policies and private acquisitions that aim for income generation. Several analytical frameworks through which to view the effects of different determinants on childhood mortality and morbidity have been developed.

Researchers used a number of different conceptual frameworks to analyze the impact of different factors on child survivals. Among these researchers (Mosley and Chen, 1984), classified the determinants of infant and child mortality as exogenous (socioeconomic or extrinsic) such as cultural, socioeconomic, community and regional determinants and endogenous (bio-medical or intrinsic) such as maternal, environmental, nutrition, injuries and personal illness. Socioeconomic factors affect indirectly infant and child mortality, they operate through the proximate factors while proximate determinants affect infant and child mortality directly (Mosey and Chen 1984; Schultz, 1984).

The study presented here, based on the conceptual framework of child survival for developing countries proposed by (Mosley and Chen, 1984) for the analysis of the impact socioeconomic factors and bio-demographic (bio-medical) factors on infant and child mortality (You et al., 2015). They identify clearly the proximate and socioeconomic determinants of infant and child mortality and they categorize fourteen proximate determinants of infant and child mortality into five general groups. (Mosley and Chen, 1984) set the framework of child survival based on the assumption of all socioeconomic factors of child mortality necessarily operate through a common set of intermediate factors. The assumptions are:-

1. In an optimal setting, over 97 percent of children born can be expected to survive until the fifth birthday.
2. The proximate determinants through the socioeconomic factors operate to influence the infant and child mortality.

3. Socioeconomic, biological and environmental factors are the driving force behind the reduction of infant and child mortality.

As mentioned above (Mosley and Chen, 1984) categorized a set of proximate determinants into five general groups that directly affect infant and child mortality. Socioeconomic factors affect infant and child mortality that must operate through these proximate determinates or indirectly affect infant and child mortality.

The five grouped proximate determinants that directly affect infant and child mortality are:-

- I. Maternal factors: age, parity and birth intervals: - each of these factors has an impact on infant and child mortality through affecting maternal health. Synergism may also be considered as maternal factors for example, birth interval categories (example short, medium, long) with mothers age categories (example five year age group).
- II. Environmental contaminations: hygiene factors, water and sanitation: - each factor spread or transmits infectious disease to mothers or children.
- III. Nutrient deficiency (calories, protein and micro nutrient deficiency):- the deficiency of nutrient decrease the probability of child survival and also an impact to the mothers through which affect infant and child mortality for instance, nutrition during pregnancy and lactation have an impact on new born children weight and quality of breast milk.
- IV. Injury: - this related to physical, burn and poisoning injury. Injuries have an impact on the infant and child mortality and more pronounced in the infanticide period.
- V. Personal illness control (Immunization, bed net, malaria prophylaxis etc.): - this includes both traditional and modern preventative measures to avoid disease during pregnancy and child births and the quality of preventative measure are important. Personal illness control factors influence on pregnancy outcome and child survival through its effect on both mothers and children (Mosley & Chen, 1984) On the other hand Mosley and Chen (1984) categorized the socioeconomic determinants of infant and child mortality into individual, household and community level variables.

## **2.2. Causes of under-five death**

Leading causes of death in children under-5 years are preterm birth complications, pneumonia, birth asphyxia, diarrhea and malaria. Children in sub-Saharan Africa are more than 15 times

more likely to die before the age of 5 than children in high income countries. The leading causes of death among children under five in 2017 were preterm birth complications, acute respiratory infections, intrapartum-related complications, congenital anomalies and diarrhea (UNICEF, 2016). Neonatal deaths accounted for 47% of under-five deaths in 2017. Ending preventable child deaths can be achieved by providing immediate and exclusive breastfeeding, improving access to skilled health professionals for antenatal, birth, and postnatal care, improving access to nutrition and micronutrients, promoting knowledge of danger signs among family members, improving access to water, sanitation, and hygiene and providing immunizations (Blencowe et al., 2010). Many of these lifesaving interventions are beyond the reach of the world's poorest communities. According to the World Health Organization, 9.2 million children under the age of 5 die every year, many from preventable conditions that could be treated with simple healthcare interventions. The majority of these deaths occur in Sub-Saharan Africa and South Asia, where the child mortality rate is 1 75 per 1 000 (compared to 6 per 1 000 in industrialized countries) (Beise et al., 2017).

### **2.3. Socio economic factors**

Socioeconomic factors are the independent variables that act through proximate determinants to influence the level of morbidity and mortality. They can be grouped into individual level, household level and community variables. Socio-economic factors may affect, directly and indirectly, environmental, behavioral, nutritional and demographic risk factors with the exception of age and sex (Mondal et al., 2009) and (Victoria et al., 1997).

#### **a. Mother's educational level**

Mother's education has frequently been used as a proxy indicator of socio-economic status in international surveys and studies. However, mother's education is also thought to be associated with hygiene, care seeking, and treatment of illness behaviors pertaining to early childhood morbidities (Stalling, 2004). Some authors argue that socio-economic factors including education may have a greater effect on infections mortality than on morbidity (Root, 2001). According to them, well-educated mothers may be unable to reduce risk of exposure due to factors beyond their control, such as a contaminated community environment, or lack of water. However, their

knowledge and wealth may allow them to use healthcare services more effectively than uneducated women (Root, 2001).

The mother's level of education is strongly linked to child survival. Higher levels of educational attainment are generally associated with lower mortality rates, since education exposes mothers to information about better nutrition, use of contraceptives to space births, and knowledge about childhood illnesses and treatment. Larger differences have been found to exist between the mortality of children of women who have attained secondary education and above and those with primary level of education or less (McGuire, 2006).

(Goro, 2007) used data from 1993, 1998, and 2003 DHS surveys in Ghana to examine the determinants of infant and child mortality in three northern regions by using multivariate logistic regression model found that education of mothers, birth order of child and marital status of mothers are powerful significant determinants for infant mortality, while only mothers education have a significant impact for child mortality. Similarly, (Twum-Baah et al, 1994) indicated that children born to mothers with higher educational level associated with lower risk of infant and child mortality as compared to children born to mothers with primary education level or non-educated.

### **b. Family size**

Family size has been found to influence infections in many studies. When many people live together, the chance of contact with pathogens increases, and hygiene may deteriorate; (Woldemicael, 2001, Manunebo et al., 1994). A large number of children in a household increase the likelihood of having disease like infections because of crowding and competition for mother's time and attention and other resources (Teklemariam et al., 2000). In Eritrea, the probability of having diarrhea is about 60% higher if there are six or more children living in the house than if the number of children is less than three. In Ethiopia, the odds of having infections associated with the number of children remained significant even after controlling for all environmental, behavioral and other socio-economic variables considered in a study conducted in study (Goulding et al., 2017).

### **c. Birth Interval**

Birth intervals of short duration are associated with adverse pregnancy outcomes, increased morbidity during pregnancy, and increased infant and child mortality (Rutstein, 2005). (Kumar & Gemechis, 2010) uses data from Ethiopia DHS survey 2005 and employs cross tabulation technique to examine the selected socioeconomic, bio-demographic and maternal health care factors that determine child mortality in Ethiopia. The result shows that among socioeconomic variables, birth intervals with preceding birth have significant impact to lowering the risk of child mortality. The result conformed that the child mortality risk associated with children of less than 2 years of birth interval with previous child was highest 15% and lowest 4.2% for the children whose birth interval was 4+ years.

(Mutunga, 2004) used data from 2003 DHS in Kenya to investigate the impact of socioeconomic and environmental variables of infant and child mortality in urban areas of Kenya. The results show that the child mortality was lower for children with birth interval of more than 2 years. A 2008 study by Rutstein using DHS data found that the risk of mortality rapidly decreases as the birth interval increases up to 24-29 months and then decreases more slowly with longer birth intervals, but increases again for intervals of 96 or more months (Hill et al., 2001).

### **d. Place of delivery**

Differentials in mortality by urban/rural residence have commonly been observed, with urban areas having more advantages and therefore better child survival prospects. The place of residence is one of the predictors of child health in general and mortality in particular. In developing countries, socio-economic status, access to health services and environmental conditions all affect the health of children of the rural areas. Children in urban areas where proper sanitation and water are available, and where modern treatment is more frequent will have lower incidences of under-5 mortality (Stallings, 2004). In Ethiopia, rural children exhibited more than five higher odds of experiencing mortality than their urban counterparts. In Eritrea, children living in urban areas are 46% less likely to experience mortality before the age of five. The variation in prevalence between urban and rural persisted even after adjusting for environmental, behavioral and other socio-economic variables (Espo, 2002).

## **2.4. Environmental factors**

The effect of the environment on health is complex and is conditioned by a wide range of characteristics and behaviors. For example, the effect of improved water and toilet facilities on child health may vary depending on parental education, child feeding practices or income (Manunebo et al., 1994) and (Timaheus & Lush, 1995). Environmental factors include water sources, availability of toilet facilities and method of excreta disposal (Espo, 2002). Most environmental factors are usually associated with socio-economic status and place of residence

### **a. Source of water**

The health benefits of improved water supply have been established in several studies. Improved sources of drinking water are less likely to be contaminated and likely to prevent the spread of water-related diseases, such as infections and cholera. Using DHS data from eight Sub-Saharan African countries (Fayehun, 2010) establishes that in countries with low under-five mortality such as Namibia and Lesotho, the proportion of children living in households with an improved source of drinking water is greater than in countries in the high under-five mortality group. Seventy one percent of households in Namibia and 55% of households in Lesotho access their drinking water from improved sources. Conversely, in countries with high under-five mortality at least 40% of children live in households with an open well or surface water as their source of drinking water (Fayehun, 2010)

(Mutunga, 2007) observes that even if water is from a clean tap, fetching it with unclean containers and improper storage can facilitate contamination with infection causing organisms. WHO reiterates that the most common and widespread danger associated drinking water is contamination either directly or indirectly by sewage, other waste including human and animal excrement. If drinking water is contaminated and then used in preparing foods, it may result to further cases of infection (Kumar et al., 2016). Faecal contamination of any other source of pollution may introduce a variety of intestinal pathogens, parasites and bacteria (Griggs et al., 2013) and (Teklemariam et al., 2000) holds that the organisms in water may cause severity from mild to sometimes fatal dysentery, cholera and typhoid. Using clean and treated water can prevent infections. Families who have plentiful supply of safe piped water and use it properly have reduced correlation of under-5 mortality.

## **b. Type of Toilet facility**

Researchers have shown that children living in households with some kind of toilet facility are less likely to be sick than children in households which do not have toilet facilities (Mohammed et al., 2013). In Ghana, the risk of having infections was found to be significantly associated with toilet facility, where children living in houses with toilet facilities are about 50% less likely to contract infections than children living in houses with no such facilities (Buttenheim, 2008). A similar finding was reported in Rural Zimbabwe where a cohort study on childhood infections was conducted for 45 weeks in two neighboring semi-arid communities with similar characteristics in terms of healthcare provision, water supply and socio-economic characteristics but different in terms of sanitation and population density. One community was densely populated and had improved sanitation facilities while the other was less dense but openly defecated (Mekonnen, 2011). Children from the less dense community that openly defecated had three times as many episodes of infections as the community which was densely populated but used an improved sanitation (Root, 2001). Similar results have also been obtained from studies in Uganda, Brazil, Bangladesh and Ethiopia (Buttenheim, 2008) and (Ezra and Gurum, 2002). As highlighted by (Klaauw & Wang, 2004) access to toilet facility can reduce under-five mortality rate significantly in rural areas of India as a whole. In urban Kenya, access to modern sanitation facilities, flush toilets, reduces diarrhea prevalence in urban areas and ultimately reduce the child mortality, (Musyoka et al., 2018, Van der Klaauw and Wang, 2004).

## **2.5. Empirical studies on the determinants of under-five death**

Several studies have been conducted both locally and globally on the determinants of under-five mortality. Most of the studies have shown significant effect between socioeconomic, demographic and environmental and biological factors and causes of under-five mortality.

Mamun, (2014) used data from DHS from Bangladesh to examine determinants of under-five child deaths per mother using zero-inflated regression models. He found that mother's level of education and wealth index are significant factors for under-five mortality.

Bello and Joseph, (2014) in their study examined the determinants of child mortality in Oyo State, Nigeria using logistic regression model. Their findings revealed that Breastfeeding status

of mothers was a significant factor for child mortality in Atiba Local government Area of Oyo state while age of mother at first birth was not significant determinant.

Logistic regression model is applied by Sampson, (2014), using the 2011 Ghana DHS to examine the determinants of under- five mortality. His findings suggested that breastfeeding was the main determinants of under-five mortality in the Tano South district(Kumar and File, 2010). On the other hand, other factors including mothers' education, mothers' occupation, household income and marital status did not show significant effect on under-five mortality.

Lemani, (2013), used data from 2004 and 2010 DHS in Malawi to examine determinants of Infant and Child Mortality in Malawi using logistic regression and survival analysis. In this study, the 2010 data indicated that age of mothers and mother's education were found to be statistically significant with child mortality. Using 2010 DHS in survival analysis, infants from poor and middle income households were more likely to die as compared to those from rich households (OR=2.063 and 1.616, respectively).

Iram & Butt, (2008) using the results of Pakistan Integrated Household Data (PIHD) to examine socioeconomic factors of childhood mortality in Pakistan. They found that the higher level of maternal education, the higher the infant and child mortality. They also found that maternal age at the first birth and working mother has significant and positive effect on childhood mortality. Mesike & Mojekwu, (2012), using data from Demographic and Health Survey (DHS) 2008 and the annual abstract statistics of the National Bureau of Statistics (NBS) 2009 in Nigeria examined Environmental Determinants of Child Mortality by using principal component analysis and simultaneous multiple regression. As for the socio-economic variables, better survival prospect were found to exist in homes with high income. As expected, environmental characteristics of the household has been found to be significantly related to child mortality.

Mondal et al., (2009) used the logistic regression model to investigated factors influencing infant and child mortality in Rajshahi District of Bangladesh. Findings revealed that the most significant predictors of neonatal, post-neonatal and child mortality levels are mother's age at birth. Similarly, Uddin et al.,(2009) in their study investigated child mortality in Bangladesh also using the logistic regression. Results of analysis showed that occupation of mother, economic

status of households and breastfeeding status were significant determinants of child mortality in Bangladesh.

Abimbola et al., (2012,) and Kittur, (2014) in their study examined the determinants of child mortality in rural Nigeria employing the in Nigeria Demographic and Health Survey (NDHS) data. Data were analyzed using the Logit regression model. The result of analysis showed that education of mother and age of mother at first birth were among the significant factors influencing child mortality in rural Nigeria (Antai, 2011).

Gideon, (2012) using the Poisson regression model, investigated under-five mortality differentials in urban East Africa: a study of three capital cities (Nairobi, Dares-Salaam & Kampala). The findings from these analyses underscore the fact that differences exist in under-five mortality levels in the three East African cities. Mothers age at the first birth, mother's level of education, marital status, and household living conditions (water and toilet facility) were significant determinants of infant and child mortality in all three cities.

Mohammed et al., (2013) the data used from Cairo University Specialized Hospital and Benha University Hospital to examine supportive strategies regarding deaths Prevention for mothers of children under five years old using chi-square test. This finding showed that mothers age at the first birth, mother level of education and mother currently working were significant factors for under-five deaths.(Antai, 2011), from world health center, analyzed the Nigerian National Demographic and Health Survey (2003) data by using Multilevel Cox proportional hazards model to investigate regional inequalities in under-5 mortality in Nigeria. This study showed that under-5 mortality was found to be statistically significant with region of residence, with higher risks of under-5 deaths for children of mothers resident in the South region after adjusting only for individual-level risk factors and only community-level risk factors, as well as simultaneously adjusting for individual and community-level risk factors.

Berhie and Yirtaw, (2017), a study was performed to statistical analysis on the determinants of under-five mortality using negative binomial showed that family size, mother's educational level, mother's age at first birth, child's, employment status of mother, type of birth, birth order and vaccination of child were found to be significant determinants of under-five mortality. Father's educational level, place of residence, sex of the child, the occupation of father, the

source of drinking water, place of delivery, availability of toilet facility, wealth index and religion were not statistically significant

Alemu, (2015) used data from 2011 EDHS to study determinants of under-five death using count regression models and found revealed that mother's age at the first birth, breastfeeding status, wealth index, current mother working, region and mother's level of education had statistically significant on the number of under-five deaths in rural parts of Ethiopia. Similarly, mother's level of education, age of mothers at the first birth, toilet facility and work/employment status of mothers were found to be statistically significant with the number of under-five deaths per mothers in urban parts of Ethiopia. Also region, age of mothers at the first birth, mother's level of education, breastfeeding status of mothers, wealth index and employment status of mothers were found to be statistically significant effect with the number of under-five deaths in Ethiopia.

Seyoum and Wencheke, (2013) used data from 2011 EDHS to study determinants of mortality among children aged one to five years using Cox proportional model and found that place of residence, mother's age at first birth, mother's education and marital status were found to be statistically significant with child mortality. The study showed that children living in rural areas face higher risk of mortality than children living in urban area. This study suggested that children born to unmarried mothers are expected to experience a higher risk of dying than children born to married mothers.

Amare et al., (2007), using the data of structural questionnaire from Jimma town, Ethiopia, examined the determinants of under-five mortality of children in Gilgel Gibe field research Center, Southwest Ethiopia. They found that mother education and mother age were found to be statistically significant with under-five mortality. Under-five mortality was doubled in mother less than 20 years of age compared to above 20 years.

Desta, (2011) used data from 2000 and 2005 DHS to examine the socioeconomic, demographic and biological factors of infant and child mortality in Ethiopia using logistic regression analysis and found that marital status and mother's educational levels are significant determinants of infant and child mortality. His finding showed that children born to mothers with more than 20 years of age at first birth are less likely to die before five years as compared to children born to

mothers of age less than 15 at first birth. This study also suggested that infant and child mortality widely varied between regions in Ethiopia. However, most variation between regions does not have any significant effect on infant and child mortality.

Getiye, (2011) used data from 2005 EDHS to study risk factors and regional differentials in under-five mortality in Ethiopia by using multilevel count model. He found that mother education level, employment status of mother and economic status of mother were found to be statistically significant with under-five mortality. He also found that under-five child mortality differentials per mother among regions are significant.

Tarihu & Eshetu, (2013) used survival analysis to estimate social determinants of under-five mortality in Ethiopia. The results showed that being teenage mother at birth and child births to mothers residing in poor households were identified risk factors for increased under-five mortality in Ethiopia.

Tesfaye, (2011), used data from EDHS 2005 to investigate the determinants of under-five mortality per mother in Ethiopia using Bayesian approach. He found that mother age at the first birth, mother level of education, current breastfeeding status of mother and source of drinking water were found to be statistically significant with under-five mortality. However, married women and mothers from rich and medium households were higher risk factors of under-five mortality than unmarried women and mothers from poor households, respectively.

Ermias , (2013), used data from the demographic and health survey (DHS) conducted in Ethiopia 2005. The analysis was conducted using Cox proportional hazards model which analyses the effects of covariates on child mortality in rural Ethiopia. The study showed that source of drinking water, wealth index of household, mother's education have significant contribution on child mortality.

On the application of cox hazard's proportional model by (Gebretsadik and Gabreyohannes, 2016) under-five mortality significantly associated with preceding birth interval, family size, birth type, breastfeeding status, the source of drinking water, and income of mother.

### **3. METHODOLOGY**

#### **3.1. Source of the Data**

The data used in this study is based on the 2016 Ethiopia Demographic and Health Survey (EDHS) which was implemented by the Central Statistical Agency (CSA). The 2016 EDHS is the 4th Demographic and Health Survey conducted in Ethiopia since 2000 as part of the DHS Program. The sampling frame used for EDHS 2016 is the Population and Housing Census (PHC) conducted in 2007 by the Central Statistical Agency (CSA). CSA has an electronic file consisting of 81,654 Enumeration Areas (EA) created for the 2007 census for 10 of its 11 geographic regions. Ethiopia is divided into 11 geographical regions. Each region is sub-divided into zones, each zone into Woredas, each Woreda into towns, and each town into Kebeles. Among the 81,654 EAs, The sample design for EDHS 2016 used a two stage stratified cluster sampling design. Stratification was achieved by separating each region into urban and rural areas. The main objective of the 2016 EDHS is to provide current estimates on fertility levels, marriage, sexual activity, family planning, breastfeeding practices, nutrition, childhood and maternal mortality, maternal and child health, domestic violence, FGC/M, and HIV/AIDS and other STIs. This information is essential for programmer's managers and policymakers to evaluate and design programmers and strategies for improving the health of Ethiopia.

The sample points were selected independently in each sampling stratum, by a two stage selection. Among the first stage, 645 selected EAs, 202 were in urban areas and 443 were in rural areas and also the second stage, overall, 18,008 households were selected of which 17,067 were occupied. Selected households were visited and interviewed. All women age 15-49 and all men age 15-59 in the selected households were interviewed. Anthropometry and biomarkers were collected in all households. The analysis presented in this study on under-five deaths was based on the 16650 eligible numbers of women age between 15 and 49.

## 3.2. Variables 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,  $Y_i$  is a count variable, the total number of under-five deaths per women of reproductive age (15-49) in Ethiopia,  $Y_i = 0, 1, 2, \dots$  where refers to the  $i^{\text{th}}$  individual women.

### 3.2.2. Explanatory variables

The predictor factors assess as the main determinants of under-five mortality in this studies were described as follows.

**Table 3. 1.** Independent variables with their label and category

No_	variables	Codes (category)	Descriptions of variables
1	MEL	0=No education 1=Primary 2=Secondary 3=Higher	Mother's Educational level
2	MAGEB	15-19 20-24 25-29 30-34 35-39 Above 40	Mother's age at the firth birth
3	ESM	0=Non-Working 1=Working	Employment status of mother's
4	MBFS	0=No 1=Yes	Mother's Currently Breastfeeding status
5	MotMS	0=single 1=married 2= widowed 3=divorced	Mother's Marital status
6	FELS	0=No Education 1=Primary 2=Secondary 3= Higher	Father's educational level

7	WEALTH	0=Poor 1=Medium 2=Rich	Wealth index (household income status)
8	FAMSIZE	0=1-3 1=4-6 2 ≥ 7	Family size
9	PRESIDENCE	0=Rural 1=Urban	Place of residence
10	RELIGION	0=Orthodox 1=Muslim 2= Protestant 3= Other	Religion
11	PDELIVERY	0=HOME 1=Health center	Place of delivery
12	SEX	1=Male 2=Female	Sex of child
13	TBIRTH	0=Single Birth 1=Multiple Birth	Type of birth
15	(CANEMIA)	0=Non-Anemia 1=Mild 2=Sever	Child's anemia status
16	REG	1=Tigray 2=Affar 3=Amhara 4=Oromiya 5=Somali 6=Benishangul-Gumuz 7=SNNP 8=Gambela 9=Harari 10=Addis Ababa 11=Dire Dawa	Region
17	(CVACCINATION)	0=No 1=Yes	Vaccination of child
18	BirthO	1=1, 2=2-3, 3=4-6, 4 ≥ 7	Birth order
19	SDW	0=protected 1=unprotected	Source of drinking water
20	ToilFa	0=has toilet facility 1=no toilet facility	Toilet facility status
21	PBI	0 ≤ 2 years 1= 2-3 years 2 ≥ 3 years	Preceding birth interval

## **3.3 METHODS OF DATA ANALYSIS**

### **3.3.1 Introduction on count data model**

Counts are non-negative integers. The observations can take only non-negative integer values; 0, 1, 2, 3,... These integers arise from counting rather than ranking. The foundation for the development of count data models is the Poisson distribution. For count data the appropriate models are divided into three parts: count model for equal dispersion; Poisson Regression model, count model common case of over dispersion includes; Negative Binomial Regression model (over-dispersion), Zero-Inflated Count Models (excess zeroes); Zero inflated Poisson model, Zero-inflated Negative Binomial model, Zero-Truncated Count Models; Zero-truncated Poisson model, Zero-truncated Negative Binomial model and Hurdle Models and Common models that provide for under dispersion include: the generalized event count (GEC(k)) model, double Poisson, Poisson polynomial expansion, and the generalized Poisson models. In the oldest time count variables are treated as continuous and linear regression model is applied. However, count data with a positively-skewed distribution may not fit well in the OLS linear regression model. There are four reasons. First, the OLS linear regression model produces negative values, but count data are always larger than or equal to zero. In other words, OLS linear regression does not account for data being truncated at zero; thus, it could predict negative values which are meaningless (King, 1988 & Sturman, 1999). Second, one of the assumptions for validating statistical tests from OLS linear regression is the normality of residuals. Count data with a positively-skewed distribution are unlikely to satisfy this assumption. Third, the validity of hypothesis tests in the OLS linear regression model depends on assumptions about the homogeneity of variance of residuals that are unlikely to be met in count data (Gardner, & Shaw, 1995).

Fourth, OLS linear regression is mainly for continuous dependent variables, not discrete variables, like count data. Due to the reasons mentioned above, using OLS regression to analyze count data may lead to conclusions that do not make sense for the data, such as impossible mean predicted values, and incorrect standard errors for significance tests and p-values. Using linear regression models for count data is very inefficient. It has inconsistent standard errors and may

produce negative predictions for the number of events. The least square estimates with a logged dependent variable suffer from these problems and are biased and inconsistent as well.

Unlike in the case of a classical regression model, the response variable is a discrete with a distribution that places the probability mass at non-negative integer values only. Regression models for counts, like other limited or discrete variable models are nonlinear with many properties and special features intimately connected to discrete-ness and non-linearity (Cameron and Trivedi, 2013). Despite the fact that count data regression modeling techniques have rather recent origin, the statistical analysis of count data has a long history. Most of the early statistical count analyses concerned univariate independent and identically distributed random variables within the framework of discrete parametric distributions (Johnson et al., 2005). With these statistical models for handling count data, it is difficult to know which one to choose by just someone's intuitive feelings. proposed a comparative approach for handling count data by comparing different count regression models on how they fitted their count data using the Akaike Information Criterion (AIC) , and Bayesian information criterion (BIC) (Johansson, 2014) .

### **3.3.2. STATISTICAL MODELS**

In this study, the variable of interest is a count variable. When the response or dependent variable (number of under-five mortality in this study) is a count (which can take on non-negative integer values (0, 1, 2 ...), it is appropriate to use non-linear models based on non-normal distribution to describe the relationship between the dependent variable and a set of predictor variables. For 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. Unlike linear regression, count data regression models have counts as the response variable that can take only nonnegative integer values (Cameron and Trivedi, 2013).

Numerous models have been developed specifically for count data (Long & Freese, 2006; Sano & Zvonkovic, 2005). These models can handle non-normality on the dependent variable and do not require the researcher to either dichotomize or transform the dependent variable. We shall focus on six of these models (Atkins & Gallop, 2007; Long & Freese, 2006; Sano et al., 2005): Poisson, Negative Binomial, Zero-inflated Poisson (ZIP), Zero-inflated Negative Binomial

(ZINB). Zero-inflated Generalized Poisson (ZIGP), Hurdle Poisson regression model and Hurdle Negative Binomial regression Model (Harris et al., 2014)

### 3.4 Poisson Regression Model

The Poisson distribution was first published by Siméon-Denis Poisson in 1838. The Poisson distribution is usually used as a standard model for count data and was derived as a limiting case of the binomial distribution by Poisson. It was the first model specifically used to model counts and it still stands at the base of many types of count models available to analysts (Cameron et al, 1998). The Poisson distribution becomes increasingly positively skewed as the mean of the dependent variable decreases (Lord et al., 2005) and (aLong & Freese, 2006), reflecting a common property of count data. According to Sturman (1999), the apparent simplicity of Poisson comes with two restrictive assumptions. First, the variance and mean of the count variable are assumed to be equal. This makes it unsatisfactory to use Poisson model on real study data. The other restrictive assumption of Poisson models is that occurrences of the event are assumed to be independent of each other.

In real data, many count variables have a variance greater than the mean, which is called over dispersion. The skewness of the Poisson distribution is  $(\frac{1}{\sqrt{\lambda_i}})$  and the kurtosis is  $(3 + \frac{1}{\sqrt{\lambda_i}})$ , so that for large  $\lambda_i$ , the distribution approaches the Normal with skewness of zero and kurtosis of three.

A Poisson random variable has the probability distribution function (pdf)

$$P(Y_i=y_i, \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

$$\lambda_i > 0, i=1, 2, \dots, n \text{ and } y_i = 0, 1, 2, \dots$$

Where,  $y_i$  denotes the value of an event count outcome variable occurring in a given time or exposure periods with mean parameter  $\lambda_i$ . In Poisson model, the conditional variance is equal to conditional mean

$$E(y_i) \text{ and } \text{var}(y_i) = \lambda_i$$

This property of the Poisson distribution is known as equal dispersion.

Let  $x$  be an  $n \times (p+1)$  covariate matrix. The relationship between  $Y_i$  and  $i^{\text{th}}$  row vector of  $x$ ,  $x_i$  linked by  $g(\lambda_i)$  is

This model is known as the Poisson regression or log-linear model. Where,  $x_i = (1, x_{i1}, \dots, x_{ip})^T$  is the vector of explanatory variables and  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  is the vector of the unknown regression parameters.

### 3.4.1 Poisson Parameter estimation

The parameters of Poisson model are estimated by maximum likelihood approach using an numerical iteratively reweighted least squares algorithm. The likelihood function of the Poisson model based on a sample of  $n$  independent observations is given by

$$\ell(y, \beta) = \prod_{i=1}^n \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

The log-likelihood function of Poisson is

$$\ell = \text{Log}(L(\beta)) = \sum_{i=1}^n [y_i \ln(\lambda_i) - (\lambda_i) - \ln(y_i!)]$$

The likelihood equations for estimating the parameter is obtained by taking the partial derivations of the log-likelihood function and solve them equal to zero.

Thus, we obtain the first derivatives of  $\ell$  with respect to the parameters  $\beta$  as follows:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^n (y_i - \lambda_i) x_{ij}$$

If  $E(y_i) < \text{var}(y_i)$  then we speak about over- dispersion, and when  $E(y_i) > \text{var}(y_i)$  we say we have under-dispersion. Next, we employed two tests of over dispersion where the Null Hypothesis ( $H_0$ ) is: mean and variance of the response variable are equal against the Alternative Hypothesis ( $H_1$ ): variance exceeds the mean. There are two basic criteria commonly used to check the presence of over dispersion:

#### 1. Deviance, $D(y_i, \hat{\lambda}_i)$ is given by

$$D(y_i, \hat{\lambda}_i) = 2 * \sum_{i=1}^n \{ y_i \ln \left( \frac{y_i}{\hat{\lambda}_i} \right) - (y_i - \hat{\lambda}_i) \}.$$

Where,  $y$  is the number of events,  $n$  is the number of observations and  $\hat{\lambda}_i$  is the fitted Poisson mean.

#### 2. Pearson chi-square test, $X^2$ is also given by

$$X^2 = \sum_{i=1}^n \left( \frac{(y_i - \hat{\lambda}_i)^2}{\hat{\lambda}_i} \right)$$

Over-dispersion may be a result of higher occurrence of zero counts and subject heterogeneity.

If the model fits the data, both deviance and Pearson Chi-square statistics divided by the degrees of freedom are approximately equal to one. Values greater than one indicate the variance is an over dispersion, while values smaller than one indicate an under-dispersion. It is possible to account for over-dispersion with respect to the Poisson model by introducing a scale (dispersion) parameter into the relationship between the variance and the mean (Pedan, 2001).

Another way of checking the presence of over-dispersion is a statistical test of the hypothesis:

$$H_0: \alpha = 0 \text{ VS } H_1: \alpha < 0$$

If P-value of  $LRT\alpha < (\text{level of significance})$ , then there is over-dispersion and the Negative Binomial model is preferred.

The Negative Binomial Regression Model is more appropriate for over-dispersed data because it relaxes the constraints of equal mean and variance

In the general Poisson regression model, we think of  $\mu_i$  as the expected number of under five child death from the  $i^{\text{th}}$  mother and the total number children ever born from the  $i^{\text{th}}$  mother is  $N_i$ . This means parameter will depend on the population size and the total number of children ever born from the individual mother. Thus the distribution of  $Y_i$  can be written as:

$$Y_i \sim \text{Poisson}(N_i, \mu_i)$$

where  $N_i$  is the total fertility rate of  $i^{\text{th}}$  mother and  $\mu_i = \exp(\mathbf{X}_i^T \boldsymbol{\beta})$ . The logarithm of the children ever born is introduced in the regression model as an offset Variable. By including [children ever born] as offset in the equation, it is differentiated from other coefficients in the regression model by being carried through as a constant and forced to have a coefficient of one (Gideon,2000). Thus, the GLM with an offset is given by

$$\log \mu_i = \log N_i + \mathbf{X}_i^T \boldsymbol{\beta}$$

The link between the expectation of the dependent variable and the linear predictor is a logarithmic function and the linear predictor contains a known part or offset. This allows for estimation of maximum likelihood, standard errors and the likelihood ratio goodness of fit chi-square statistics. The model suggests that both set of the parameters are dependent on the covariates. Furthermore, the number of children born will be equal to the observed deaths if the coefficients of the independent variables, denoted by  $\boldsymbol{\beta}$ , are all equal to zero. Since  $\log N_i$  is a constant, any variation in the coefficients of the independent variables will show up affecting the dependent variable and not the number of children born. The procedure therefore allows us to

obtain the maximum likelihood regression coefficients that can be easily interpreted in terms of differentials in the dependent variables. Using the negative binomial regression procedure, several regression equations are estimated to the relationship between under-five mortality changes when control variables earlier mentioned are introduced. Results from the negative binomial models are sometimes better expressed on more convenient scale.

### **3.5 Negative Binomial Regression Model**

The negative binomial regression model is more flexible than the Poisson model and is frequently used to study count data with over dispersion (Hilbe, 2007; Hoffman, 2004). Negative binomial distribution is used for modeling over dispersed count data and is a standard generalization of the Poisson. The limitation of the Poisson regression model is that the mean and variance of the outcome variable is not identical. Another alternative solution to overcome this problem is negative binomial regression model (Hilbe, 2011).

In fact, the negative binomial regression model is in many ways equivalent to the Poisson regression model because the negative binomial model could be obtained from the mixture of Poisson and Gamma distribution called Poisson-Gamma distribution (Hilbe, 2007). Over dispersion, caused by heterogeneity or an excess number of zeros (or both) to some degree is inherent to most Poisson data. By introducing a random component into the conditional mean, the Negative Binomial Regression Model addresses the issue of over-dispersion. However, it equally models both zero and nonzero counts, which might result in a poor fit for data with excessive number of zeros. Therefore, it is always necessary to check the proportion of zero counts before developing a Negative Binomial Regression Model.

(Hilbe, 2011) used Negative Binomial Regression to Model over dispersed Poisson data. In the negative binomial regression model, a random term reflecting unexplained between-subject differences is included (Gardner et al., 1995), that is, the negative binomial regression adds an over dispersion parameter to estimate the possible deviation of the variance from the expected value under Poisson regression. Therefore, using the negative binomial regression to model count data with a Poisson distribution has the consequence of generating more conservative estimates of standard errors and may modify parameter estimates (Hilbe, 2011)

A random variable  $y_i, i= 1, 2, 3 \dots\dots$  is called a negative binomial distributed count with parameter  $\lambda$  and  $\alpha$  the probability density function is expressed as follows

$$P (Y_i = y_i, \lambda_i, \alpha) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i! \Gamma(\frac{1}{\alpha})} (1 + \alpha\lambda_i)^{-1/\alpha} \alpha (1 + \frac{1}{\alpha\lambda_i})^{-y_i}$$

Where  $Y_i \geq 0$  and  $\alpha > 0$ . With mean and variance are given by

$$E (y_i) = \lambda_i = \exp (X^T \beta) \text{ and } \text{var} (y_i) = \lambda_i (1 + \alpha \lambda_i)$$

Where,  $\alpha$  shows the level of over-dispersion and  $\Gamma (.)$  is the gamma function.

If  $\alpha = 0$ , NB Regression Model will reduce to Poisson Regression Model. This Model adds unobserved heterogeneity by specifying

$$E (y_i) = \lambda_i = \exp (X^T \beta)$$

$X_i^T$  is row  $1 \times p$  vector of covariate (including an intercepts),  $p$  is the number of covariate in Where, the model and  $p \times 1$  column vector of unknown regression parameters.

### 3.5.1 Parameter estimation of NB model

The parameters of negative binomial model are estimated by maximum likelihood approach by using numerical iteratively algorithm commonly used is either Newton–Raphson or Fisher Scoring (McCullagh & Nelder, 1989). The likelihood function of the negative binomial model based on a sample of  $n$  independent observations is given by

$$\ell (y_i, \lambda_i, \alpha) = \prod_{i=1}^n \left\{ \frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i! \Gamma(\frac{1}{\alpha})} (1 + \alpha \lambda_i)^{-1/\alpha} \alpha (1 + \frac{1}{\alpha\lambda_i})^{-y_i} \right\}$$

The log-likelihood function  $\ell$  of NB regression model is

$$\ell = \sum_{i=1}^n \left\{ -\log y_i! + \sum_{k=1}^{y_i} (\alpha y_i - \alpha k + 1) - (y_i + 1/\alpha) \log (1 + \alpha\lambda_i) - y_i \log (\lambda_i) \right\}$$

$$\text{Where } \frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i! \Gamma(\frac{1}{\alpha})} = \prod_{k=1}^{y_i} (y_i + 1/\alpha - k) = \alpha^{-y_i} \prod_{k=1}^{y_i} (\alpha y_i - \alpha k + 1)$$

For estimating regression coefficients  $\beta$  and dispersion parameter  $\alpha$  the Newton-Raphson iteration procedure is applied like in the Poisson model.

### 3.6. Zero-inflated Regression Models

In some cases, excess zeros exist in count data and considered as a result of over dispersion. In such a case, the NB model cannot be used to handle the over-dispersion which is due to the high

amount of zeros. To do this, zero-inflation (ZI) can be alternatively used. Real-life count data are frequently characterized by over-dispersion and excess zeros (Gurmu and Trivedi, 1996). Zero inflated count models provide a parsimonious yet powerful way to model this type of situation. Such models assume that the data are a mixture of two separate data generation processes: one generates only zeros, and the other is either a Poisson or negative binomial data-generating process. Count data that have an incidence of zeros greater than expected for the underlying probability distribution can be modeled with a zero inflated distribution. The following five regression models, zero-inflated Poisson (ZIP) regression, zero-inflated negative binomial (ZINB) regression, zero-inflated generalized Poisson (ZIGP) regression, hurdle Poisson (HP) and hurdle negative binomial (HNB) regression are frequently used to model zero-inflated count data.

### 3.6.1. Zero-inflated Poisson Regression Model

This model was proposed by (Lambert, 1992), with an application to defects in a manufacturing process. The Zero-inflated Poisson regression study the relationship between dependent and independent variable(s) when there are many zeros value in the dependent variable, where the relationship is the mixture between Poisson model and Logistic model. Zero-inflated Poisson Regression also provides a flexible way of modeling zero counts and an attractive interpretation. The Zero-inflated Poisson, model is another model that one can use when the zeros in a dataset are argued to be caused by both chance and systematic factors (Min and Agresti, 2005). In ZIP regression, the responses  $Y = (Y_1, Y_2, \dots, Y_n)$  are independent. One assumption of this model is that with probability  $p$  the only possible observation is 0, and with probability  $(1 - p)$ , a Poisson ( $\lambda$ ) random variable is observed in  $Y$ . In order to explain the occurrence of extra zeros in the variable  $Y_i$ , The ZIP regression model is (Lambert, 1992),

$$P(y_i) = \begin{cases} \omega_i (1 + \omega_i) e^{-\lambda_i}, & y_i = 0 \\ (1 - \omega_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, & y_i = 1, 2, \dots \end{cases}, 0 \leq \omega_i \leq 1,$$

Where,  $Y_i \sim \text{ZIP}(\lambda_i, \omega_i)$ . The mean and variance of ZIP are given by

$$E(Y_i) = (1 - \omega_i) \lambda_i \text{ and } \text{Var}(Y_i) = (1 - \omega_i) \lambda_i (1 + \omega_i \lambda_i)$$

### 3.6 .1.1 Parameter estimation of zero inflated Poisson model

The parameters  $\lambda_i$  and  $\omega_i$  can be obtained by using the link functions,

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

Where,  $x_i^T$  and  $Z_i^T$  are covariate matrices,  $\beta$  and  $\gamma$  are the  $(p+1) \times 1$  and  $(q+1) \times 1$  unknown parameter vectors, respectively. The log-likelihood function of ZIP model is given by

$$\ell(\lambda, \gamma) = \sum_{i=1}^n \{ \ln [\omega_i + (1-\omega_i) e^{-\lambda_i}] I_{(y_i=0)} + [ \ln(1-\omega_i) - \lambda_i + y_i \ln(\lambda_i) - \ln y_i! ] I_{(y_i>0)} \}$$

Where,  $I(\cdot)$  is the indicator function for the specified event, i.e. equal to 1 if the event is true and 0 otherwise. To obtain the parameter estimates of ZIP regression models,  $\hat{\beta}$  and  $\hat{\gamma}$ , the Newton-Raphson method can be used. The first derivatives of with respect to  $\beta$  and  $\gamma$  are

$$\frac{\partial \ell}{\partial \beta_j} = \frac{\partial \ell}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial \beta_j} = \sum_{i=1}^n \{ I_{(y_i=0)} \left[ \frac{-(1-\omega_i) \lambda_i e^{-\lambda_i}}{\omega_i + (1-\omega_i) e^{-\lambda_i}} \right] + I_{(y_i>0)} (y_i - \lambda_i) \}, \quad j=0, 1, 2, \dots, p;$$

$$\frac{\partial \ell}{\partial \gamma_r} = \frac{\partial \ell}{\partial \omega_i} \frac{\partial \omega_i}{\partial \gamma_r} = \{ I_{(y_i=0)} \left[ \frac{(1-\omega_i) \lambda_i e^{-\lambda_i}}{\omega_i + (1-\omega_i) e^{-\lambda_i}} \right] - I_{(y_i>0)} \left[ \frac{1}{(1-\omega_i)} \right] \} Z_{ir}, \quad r=0, 1, 2, \dots, q;$$

Newton-Raphson iteration procedure can be used for estimating the parameter of ZIP regression model.

### 3.6.2. Zero-inflated negative binomial regression model

Zero-inflated negative binomial (ZINB) regressions have been used by researchers for handling both zero-inflation and over dispersion in count data. Zero-Inflated Negative Binomial (ZINB) regression is one of the methods used in troubleshooting over dispersion due to excessive zero values in the response variable (excess zeros). This model provides a way of modeling the excess number of zeros (with respect to a Poisson distribution or negative binomial distribution) in addition to allow for count data that are skewed and over dispersed (Harris et al., 2014). The ZINB distribution is a mixture distribution, similar to ZIP distribution, where the probability  $p$  for excess zeros and with probability  $(1-p)$  the rest of the counts followed negative binomial distribution. Note that the negative binomial distribution is a mixture of Poisson distributions, which allows the Poisson, mean  $\lambda$  to be distributed as Gamma, and in this way over dispersion is modeled (Hilbe, 2011).

We used the vuong test, likelihood ratio based test, to compare the zero inflated negative binomial model with negative binomial regression model. The Zero-inflated negative binomial (ZINB) regression model assumes there are two distinct data generation processes. The result of a Bernoulli trial is used to determine which of the two processes is used. For observation  $i$ , with probability  $\omega_i$  the only possible response of the first process is zero counts, and with probability of  $(1-\omega_i)$  the response of the second process is governed by a negative binomial with mean  $\lambda_i$ . The zero counts are generated from the first and second processes, where a probability is estimated for whether zero counts are from the first or the second process. The overall probability of zero counts is the combined probability of zeros from the two processes. ZINB also arises in Bernoulli trials with non-equal success probabilities (Gurmu, 1998). The ZINB distribution is a general model for counts which nests the ZIP, NB, and Poisson models. We consider  $Y_i$  as a ZINB distribution. Specifically, we consider the distribution. (Gurmu and Trivedi, 1996) used the zero-inflated negative binomial (ZINB) regression to model over dispersed data with an excess of zeros. This regression model was given by

$$P(y_i | \omega, \alpha, \lambda) = \begin{cases} \omega_i + (1 - \omega_i)(1 + \alpha\lambda_i)^{-\frac{1}{\alpha}} & , y_i = 0 \\ (1 - \omega_i) \frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i! \Gamma(\frac{1}{\alpha})} (1 + \alpha\lambda_i) - 1/\alpha (1 + \frac{1}{\alpha\lambda_i})^{-y_i} & , y_i > 0, \end{cases}$$

Where,  $\lambda_i$  is the mean of the underlying negative binomial distribution,  $\alpha > 0$  is the over dispersion parameter and is assumed not to depend on covariates and  $0 \leq \omega_i \leq 1$ . Also the parameters  $\lambda_i$  and  $\omega_i$  depend on vectors of covariates  $x_i$  and  $z_i$ , respectively. The formulations for  $\lambda_i$  and  $\omega_i$  are the same as those used in the zero-inflated Poisson regression model. In this case, the mean and variance of the  $Y_i$  are

$$E(Y_i) = (1 - \omega_i) \lambda_i \quad \text{and} \quad \text{Var}(Y_i) = (1 - \omega_i) \lambda_i (1 + \omega_i \lambda_i + \alpha \lambda_i)$$

ZINB approaches ZIP and NB as  $\alpha = 0$  and  $\omega_i = 0$ , respectively. If both  $\alpha$  and  $\omega_i = 0$ , then ZINB reduces to Poisson.

The parameter  $\lambda_i$  is modeled as a function of a linear predictor, that is,

$$\lambda_i = \exp(x_i^T \beta)$$

Where,  $\beta$  is the  $(p+1) \times 1$  vector of unknown parameters associated with the known covariate vector  $x_i^T = (1, x_{i1}, \dots, x_{ip})$ ,  $p$  is the number of covariates not including the intercept. The parameter  $\omega_i$ , which is often referred as the zero-inflation factor, is the probability of zero counts

from the binary process. For common choice and simplicity,  $\omega_i$  is characterized in terms of a logistic regression model by writing as

$$\text{logit}(\omega_i) = \log\left(\frac{\omega_i}{1-\omega_i}\right) = Z_i^T \gamma$$

where,  $\gamma$  is the  $(q+1) \times 1$  vector of zero-inflated coefficients to be estimated, associated with the known zero-inflation covariate vector  $Z_i^T = (1, z_{i1}, \dots, z_{iq})$ , where  $q$  is the number of the covariates  $Z$ 's not including the intercept. ZINB is also used to analyze exploratory data. When all the covariates are included in the log link model, as in the case of ZIP, the estimate of the inflated parameter was found to be zero.

### 3.6.2.1. Parameter Estimation of ZINB Model

The probability of the observed data, expressed as a function of the parameter is called the likelihood function. The maximum likelihood estimate of a parameter is the parameter value for which the probability of the observed data takes its greatest value. The ZINB distribution is not a standard GLM type exponential family distribution, even when the over dispersion parameter is known, and standard GLM fitting methods are not applied. To obtain the parameter estimates of ZINB regression models,  $\hat{\alpha}$ ,  $\hat{\beta}$  and  $\hat{\gamma}$ , the Newton-Raphson method can be used. The log-likelihood function  $\ell = \ell(\alpha, \lambda_i, \omega_i; y)$ , for the ZINB model is given below

$$\begin{aligned} \ell = \ell(\alpha, \lambda_i, \omega_i; y) &= \sum_{i=1}^n \{ I_{(y_i=0)} \log(\omega_i + (1-\omega_i)(1 + \alpha \lambda_i)^{-1/\alpha}) \\ &+ I_{(y_i>0)} \log[(1-\omega_i) \frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i! \Gamma(\frac{1}{\alpha})} (1 + \alpha \lambda_i) - 1/\alpha (1 + \frac{1}{\alpha \lambda_i})^{-y_i}] \} \end{aligned}$$

Since  $\frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i! \Gamma(\frac{1}{\alpha})} = \prod_{k=1}^{y_i} (y_i + 1/\alpha - k) = \alpha^{-y_i} \prod_{k=1}^{y_i} (\alpha y_i - \alpha k + 1)$ .

Furthermore,  $\ell$  can be written as

$$\begin{aligned} \ell &= \sum_{i=1}^n \{ I_{(y_i=0)} \log(\omega_i + (1-\omega_i)(1 + \alpha \lambda_i)^{-1/\alpha}) + I_{(y_i>0)} \log[(1-\omega_i) - \log y_i! + \\ &\sum_{k=1}^{y_i} (\alpha y_i - \alpha k + 1) - (y_i + 1/\alpha) \log(1 + \alpha \mu_i) + \log y_i! + y_i \log \mu_i] \} \end{aligned}$$

Newton-Raphson iteration procedure can be used for estimating the parameter of ZINB regression models.

### 3.6.3. Zero-inflated Generalized Poisson (ZIGP) regression.

(Famoye & Singh, 2003) introduced a zero-inflated generalized Poisson regression model. The generalized Poisson distribution has first been introduced by (Consul and Jain, 1970). ZIG models have recently been found useful for the analysis of count data with a large amount of zeros and proposed by (Famoye & Singh, 2006) with an application to domestic violence data. In the case of over dispersed data with too many zeros, ZINB model may fit the data well than ZIP model. But the iterative technique to estimate the parameters of ZINB model sometimes fails to converge; (Lambert, 1992) also pointed out this difficulty. (Famoye & Singh, 2006, (Dalrymple et al., 2003) advocated ZIGP model over ZINB while ZINB fails to converge in estimating parameters for the domestic violence data.(Dalrymple et al., 2003) (Later Czado et.al, 2007) proposed the unrestrictive dispersion parameter for ZIGP model as (Famoye & Singh, 2006) considered constant dispersion parameter for a given data. The generalized Poisson regression model for mean  $\lambda_i$  and for fixed dispersion parameter  $\alpha$  can be written as:

$$f(y_i; \lambda_i; \alpha) = \left( \frac{\lambda_i}{1+\alpha\lambda_i} \right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \exp \frac{(-\lambda_i(1+\alpha y_i))}{\alpha+\alpha y_i}$$

Here,  $y_i = 0, 1, 2, \dots$  for  $i = 1, 2, \dots$ ,  $\lambda_i = e^{x_i\beta}$ ;  $x$  is the covariate matrix and  $\beta$  are unknown regression parameters. Mean and variance of generalized Poisson regression (GPR) model are  $E(y_i / x_i) = \lambda_i$  and  $\text{Var}(y_i / x_i) = \lambda_i (1 + \alpha\lambda_i)$ , respectively.

Note that when  $\alpha = 0$  the probability model reduces to the Poisson regression model. When  $\alpha > 0$ , the GPR model in represents count data with over dispersion and when  $\alpha < 0$  then the model represents underdispersion.

Famoye and Singh (2006) proposed the ZIGP model as:

$$P(Y = y_i / x_i; z_i) = \begin{cases} \varphi_i(1 - \varphi_i)f(y_i; \lambda_i; \alpha), & y_i = 0 \\ (1 - \varphi_i)f(y_i; \lambda_i; \alpha) & y_i > 0 \end{cases}$$

Still  $\lambda_i = e^{x_i\beta}$ , and the new parameter  $\varphi_i$  is the probability of  $y_i = 0$  for given  $z_i$ .

Here  $z_i$  are the covariate matrix which are responsible for zeros in  $y_i$ .  $x$  and  $z$  can be identical or may have some common covariates. The link function for  $\varphi_i$  is

$$\text{Logit}(\varphi_i) = \log \left( \frac{\varphi}{1-\varphi} \right) = z\gamma.$$

Here is the unknown regression parameter for  $Z\gamma$ .

The mean and variance of the ZIGP model are given,

$$E(y_i / x_i) = (1 - \phi_i)\lambda_i$$

$$\text{Var}(y_i \setminus x_i) = (1 - \phi_i)[\lambda_i^2 + \lambda_i(1 + \lambda_i\alpha)^2] - (1 - \phi_i)^2\lambda_i$$

Note that the model reduces to GPR model if  $\alpha = 0$ . It reduces to ZIP if  $\alpha = 0$ .

### **3.6.4. Hurdle Regression Models**

The hurdle models are based on Poisson regression and negative binomial regression respectively but is used for modeling excess zeros (Hilbe, 2011), Hilbe, 2014). It was developed separately by (Mullahy, 1986) in economics and (King, 1989) in political science, although the term itself was most likely coined by (Cragg, 1971). (Welsh, Cunningham, Donnelly, & Linden Mayer, 1996) refer to it as a 'conditional Poisson model'. "The idea underlying the hurdle formulations is that a binomial probability model governs the binary outcome whether a count variety has a zero or a positive realization. If the realization is positive the 'hurdle' is crossed, and the conditional distribution of the positives is governed by a truncated-at-zero count data model (Mullahy, 1986) such as a truncated Poisson or truncated negative binomial distribution (Min & Agrestic, 2005, (Agresti, 2010).

The hurdle model may be defined as a two-part model where the first part is a binary outcome model, and the second part is a truncated count model. As per (Cameron & Trivedi, 1998, p. 123) such a partition permits the interpretation that positive observations arise from crossing the zero hurdle or the zero threshold. The first part models the probability that the threshold is crossed. In principle, the threshold need not be at zero; it could be any value. Further, it need not be treated as known. The zero value has special appeal because in many situations it partitions the population into subpopulations in a meaningful way." So, a data set is split into zero and non-zero (positive) values to fit two different models with associated covariates in regression. A variety of probability distribution can be considered for zero counts, and frequent use in real-life data are binomial distribution, Poisson distribution, and negative binomial distribution. Also, for the positive count data frequent use of probability distributions are Poisson distribution, negative binomial distribution and geometric distribution. Zero-inflated models and hurdle models provide a way of modeling the excessive proportion of zero values and allow for over dispersion.

Especially when there is a large number of zeros, these techniques are much better able to provide a good fit than Poisson or negative binomial models (Cameron & Trivedi, 1998).

Suppose that  $g_1(0)$  is the probability value when the value for response variable is zero and that  $g_2(k)$ ,  $k=1, 2, \dots$  is a probability function when the response variable is a positive integer. Therefore, the probability function of the hurdle-at-zero model is given by:

$$P(Y_i=k) = \begin{cases} g_1(0) & \text{for } k = 0 \\ (1 - g_1(0)) g_2(k) & \text{for } k = 1, 2, \dots \end{cases}$$

(Mullahy, 1986) discussed the hurdle-at-zero model and he assumes that both parts of the hurdle model are based on probability functions for non-negative integers such as  $f_1$  and  $f_2$ . In terms of the general model above,

Let  $g_1(0) = f_1(0)$  and  $g_2(k) = \frac{f_2(k)}{(1-f_2(0))}$ . In the case of  $g_2$ , normalization is required because  $f_2$  is defined over the non-negative integers ( $k=0, 1, 2, \dots$ ) whereas the support of  $g_2$  must be over the positive integers ( $k = 1, 2, \dots$ ). This means that we need to truncate the probability function  $f_2$ . However, this is a theoretical concept, i.e., truncation on  $f_2$  does not mean that there is truncation of the population here. All we need to do is to work with a distribution with positive support, and the second part of a hurdle model can use a displaced distribution or any distribution with positive support as well. Under the (Mullahy, 1986) assumptions, the probability distribution of the hurdle-at zero models is given by

$$f(Y=0) = f_1(0)$$

$$f(Y=k) = \frac{1-f_1(0)}{(1-f_2(0))} f_2(k) = \theta f_2(k), \quad k = 1, 2, \dots$$

Where  $f_2$  is referred to as parent-process. The numerator of  $\theta$  presents the probability of crossing the hurdle and the denominator gives a normalization that accounts for the (purely technical) truncation of  $f_2$ . It follows that if  $f_1 = f_2$  or, equivalently,  $\theta = 1$  then the hurdle model collapses to the parent model. The expected value of the hurdle model is given by

$$E(Y) = \theta \sum_{k=1}^{\infty} k f_2(k)$$

The difference between this expected value and the expected value of the parent model is the factor  $\theta$ . In addition, the variance value of the hurdle model is given by

$$\text{Var}(Y) = \theta \sum_{k=1}^{\infty} k^2 f_2(k) - [\theta \sum_{k=1}^{\infty} k f_2(k)]^2$$

If  $\theta$  exceeds 1, it means that the probability of crossing the hurdle is greater than the sum of the probabilities of positive outcomes in the parent model. Therefore, increasing the expected value

of the hurdle model is related to the expected value of the parent model. Alternatively, if  $\theta$  is less than 1 (that is the usual case in an application with excess zeros), it means that the probability of not crossing the hurdle is greater than the probability of a zero in the parent model, thus decreasing the expected value of the hurdle model relatively to the expected value of the parent model. Therefore, this model gives a new explanation of excess zeros as being a characteristic of the mean function rather than a characteristic of the variance function (Dalrymple et al., 2003).

### 3.6.4.1 Hurdle Poisson regression model

A Poisson model typically is assumed for count data. In many cases because of many zeros in the response variable, the mean is not equal to the variance value of the dependent variable. Therefore, the Poisson model is no longer suitable for this kind of data. Thus, we suggest using a hurdle Poisson regression model to overcome the problem of over dispersion.

We start with the binomial process, which determines whether the dependent variable takes on the value zero or a positive value. The probability mass function is

$$P(Y = y) = \begin{cases} \pi, & y = 0 \\ 1 - \pi, & y = 1, 2, 3, \dots \end{cases}$$

**The zero-truncated Poisson process has probability mass function**

$$P(Y = y | Y = 0) = \begin{cases} \frac{\lambda^y}{(e^\lambda - 1)y!} & y = 1, 2, 3, \dots \\ 0 & \text{otherwise} \end{cases}$$

Thus, the unconditional probability mass function for Y is

$$P(Y = y) = \begin{cases} \pi, & y = 0 \\ (1 - \pi) \frac{\lambda^y}{(e^\lambda - 1)y!}, & y = 1, 2, 3, \dots \end{cases}$$

And the log likelihood for the  $i^{\text{th}}$  observation, assuming the observations are independently and identically distributed, is

$$\ln L(\pi_i, \lambda_i, y_i) = \begin{cases} \ln \pi_i, & y = 0 \\ \ln \left\{ (1 - \pi) \frac{\lambda^y}{(e^\lambda - 1)y!} \right\}, & y = 1, 2, 3, \dots \end{cases}$$

If we model  $\pi_i$  using the complementary log-log link and  $\lambda_i$  using the log link, with a little algebra we have,  $\pi_i = e^{-e^{x_i\beta_1}}$  and  $\lambda_i = e^{x_i\beta_2}$

### 3.6.4.2 Hurdle negative binomial regression model

In many cases because of many zeros in the response variable, the mean is not equal to the variance value of the dependent variable. In this case we suggest using a hurdle negative binomial regression model to overcome the problem of over dispersion. Let  $Y_i$  ( $i = 1, 2, \dots, n$ ) be a nonnegative integer-valued random variable and suppose  $Y_i = 0$  is observed with a frequency significantly higher than can be modeled by the usual model. We consider a hurdle negative binomial regression model in which the response variable  $Y_i$  ( $i = 1, \dots, n$ ) has the distribution

$$P(Y_i = y_i) = \begin{cases} w_0, & y_i = 0, \\ (1 - w_0) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \frac{(1 + \alpha\mu_i)\alpha^{-1 - y_i} \alpha^{y_i} \mu_i^{y_i}}{1 - (1 + \alpha\mu_i)^{-\alpha^{-1}}}, & y_i > 0, \end{cases}$$

$$\text{Or } P(Y_i = y_i) = \begin{cases} w_0, & y_i = 0 \\ (1 - w_0) \frac{g}{1 - (1 + \alpha\mu_i)^{-\alpha^{-1}}}, & y_i > 0 \end{cases}$$

$$\text{Where } g = g(y_i; \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} (1 + \alpha\mu_i)\alpha^{-1 - y_i} \alpha^{y_i} \mu_i^{y_i}$$

Where  $\alpha$  ( $\geq 0$ ) is a dispersion parameter that is assumed not to depend on covariates. In addition, we suppose  $0 < w_0 < 1$  and  $w_0 = w_0(z_i)$  satisfy

$$\text{Logit}(w_0) = \log\left(\frac{w_0}{1 - w_0}\right) = \sum_{j=1}^k z_i j \delta_j$$

Where  $(z_{i1} = 1, z_{i2}, \dots, z_{im})$  is the  $i^{\text{th}}$  row of covariate matrix  $Z$  and  $(\delta_1, \delta_2, \dots, \delta_m)$  is an unknown  $m$ -dimensional column vector of parameters. In this set up, the nonnegative function  $w_0$  is modeled via logit link function. This function is linear and other appropriate link functions that allow  $w_0$  being negative may be used. In addition, there is interest in capturing any systematic variation in  $\mu_i$ , the value of  $\mu_i$  is most commonly placed within a log linear model

$$\log(\mu_i) = \sum_{j=1}^k x_{ij} \beta_j$$

Where  $\beta_j$ 's are the independent variables in the regression model and  $m$  is the number of these independent variables.

### 3.7. Model Fitting Test (Goodness of fit tests)

There are different count regression models to be compared in order to select the appropriate fitted model, which fits the data well. This was done using likelihood-ratio test (LRT), Akaike information criteria (AIC) and Bayesian information criteria (BIC). AIC is the most common means of identifying the model which fits well by comparing two or more than two models. The formula is given as:

$$AIC = -2\ell + 2k$$

Where  $\ell$  is the log-likelihood of a model that will compare with the other models and  $k$  is the number of parameter in the model including the intercept.

Bayesian information criteria (BIC) takes into account the size of the data under consideration. BIC is given by,

$$BIC = -2\ell + k \log(n)$$

Where  $\ell$  is the log-likelihood of a model that will compare with the other models,  $n$  is the sample size of the data and  $k$  is the number of parameters in the model including the intercept. The comparison will start from the model without any independent variable with the model with adding the independent variable one by one through the full model. The model which has the minimum value of AIC and BIC is the most appropriate fitted model to the dataset.

### 3.7.1 Tests for the Comparison of the Models

#### 3.7.1.1 Tests for comparison of nested models

##### Likelihood Ratio test (LRT)

The likelihood-ratio test is used to assess the adequacy of two or more than two nested models. It compares the maximized log-likelihood value of the full model and reduced model. For instance, the null hypothesis can be stated as the over dispersion parameter is equal to zero (i.e. the Poisson model can be fitted well the data) versus the alternative hypothesis can be stated as the over dispersion parameter is different from zero (i.e. the data would be better fitted by the negative binomial regression).

The likelihood-ratio test is given by:

$$\text{Likelihood ratio test} = G^2 = -2(\ell_{null} - \ell_k) \sim \chi^2_{p-1}$$

This statistic is called the likelihood-ratio test statistic.

Where:  $\ell_{null}$  is the log-likelihood of the null model and  $\ell_k$  is the log-likelihood of the model comprising  $k$  predictors,  $p$  is number of parameters and  $\chi^2_{p-1}$  is a chi-square distribution with  $p-1$  degree of freedom. If the test statistics exceeds the critical value, the null hypothesis is rejected. That means the overall model is significant. In this study, to compare different modern count regression models, we used significance of dispersion parameter and likelihood ratio (LR) test as criterions. The statistic of likelihood ratio test for  $\alpha$  is given by the following equation:

$$\text{LRT}\alpha = -2(\text{LL1} - \text{LL2})$$

This statistic has a Chi-squared distribution with 1 degrees of freedom and LL is log-likelihood. If the statistic is greater than the critical value then, the model 2 is better than the model

#### 3.7.1.2 Test for comparison of Non-nested models

##### Vuong Test

The Vuong test is a non-nested test that is based on a comparison of the predicted probabilities of two models that do not nest (Vuong, 1989). That means vuong test statistics are needed to provide the appropriateness of zero-inflated models against the standard count models. For

instance, comparisons between zero-inflated count models with ordinary Poisson, or zero-inflated negative binomial against ordinary negative binomial model can be done using Vuong test. This test is used for model comparison. For testing the relevance of using zero-inflated models versus Poisson and NB regression models, the Vuong statistic is used. Let's define

$$m_i = \log \left( \frac{P_1(Y_i / X_i)}{P_2(Y_i / X_i)} \right)$$

Where,  $P_1(Y_i / X_i)$  and  $P_2(Y_i / X_i)$  are probability mass functions of zero-inflated and Poisson or NB models, respectively. In general,  $(Y_i / X_i)$  is the predicted probabilities of observed count for case  $i$  from model  $N$ , then the Vuong test statistic is simply the average log-likelihood ratio suitably normalized.

The test statistic is

$$V = \sqrt{n} \frac{\frac{\sum_{i=1}^n m_i}{n}}{\sqrt{\frac{\sum_{i=1}^n (m_i - \bar{m})^2}{n-1}}} = \frac{\sqrt{n}}{s_m} (\bar{m})$$

Where,  $\bar{m}$  is mean of  $m_i$ ,  $s_m$  standard deviation and  $n$  sample size.

The hypotheses of the Vuong test are:

$$H_0: E[m_i] = 0 \quad \text{vs} \quad H_1: E[m_i] \neq 0$$

The null hypothesis of the test is that the two models are equivalent. Vuong showed that asymptotically,  $V$  has a standard normal distribution. As Vuong notes, the test is directional (vuong, 1989).

- If  $V > Z\alpha/2$ , the first model is preferred.
- If  $V < -Z\alpha/2$ , the second model is preferred.
- If  $|V| < Z\alpha/2$ , none of the models are preferred

### 3.7.2 Information Criteria (AIC and BIC)

AIC and BIC are goodness of fit criteria used for model selection. The likelihood ratio test was used to compare the Poisson model and NB model. Many Monte-Carlo simulations indicate that the BIC and AIC selection criteria need to be used together (Dalrymple et al., 2003) & (Wang et

al, 1996). The model with smallest value of AIC or BIC is preferable. Selecting an appropriate model is often based on a standard likelihood information criteria, for example, Akaike information criteria (Akaike, 1973) or Bayesian information criteria (Raftery, 1986) abbreviated by AIC and BIC, respectively, where

$$\text{AIC} = -2 \log \text{likelihood} + 2k$$

$$\text{BIC} = -2 \log \text{likelihood} + k \ln(n)$$

Where,  $k$  = number of parameters and  $n$  = number of observations.

### 3.7.3 Test for individual predictors

Let  $\beta$  denote an arbitrary parameter. Consider a significance test of  $H_0: \beta = 0$ . The simplest test statistic uses the large-sample normality of the ML estimator  $\hat{\beta}$ , let  $\text{SE}(\hat{\beta})$  denote the standard error of  $\hat{\beta}$ , evaluated by substituting the ML estimate for the unknown parameter in the expression for the true standard error. When  $H_0$  is true, the test statistics

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

Has approximately a standard normal distribution. Equivalently, has approximately a chi-squared distribution with  $df = 1$ . This type of statistic, which uses the standard error evaluated at the ML estimate, is called a Wald statistic.

The Wald statistic is  $Z^2 = \left( \frac{\hat{\beta} - \beta_0}{\text{SE}(\hat{\beta})} \right)^2$

Under  $H_0$  true,  $Z^2$  is a chi-square distribution with 1 degree of freedom. Wald statistics are for small samples. Likelihood-ratio tests are generally considered to be superior (Agresti, 2007).

### 3.8 Statistical software packages

In this study we used STATA 14, SPSS version 21 and R versions 3.5.2 were used for statistical analysis and graphics. For statistical tests, at 5% level of significance was used.

## 4. STATISTICAL DATA ANALYSIS AND RESULTS

This chapter discusses results of the study showing how selected socio economic, environmental, biological, demographic factors and the most important variable that affects the number of under-five mortality among women at reproductive age in Ethiopia. The Poisson, negative binomial, zero-inflated Poisson, zero-inflated negative binomial regression, zero-inflated generalized Poisson, hurdle Poisson regression and hurdle negative binomial regression models were used to analyze the data. The statistical analyses were performed using Statistical Package for Social Science (SPSS) version 21, South Texas Art Therapy Association (STATA) version 14 and R version 3.5.2.

### 4.1 Result of descriptive statistics

Before proceeding to fit an appropriate count models, we make a descriptive analysis of the data in order to have an overall picture of the distribution of the number of under-five mortality. Thus we start with the description of the response variables.

#### 4.1.1 Number of under-five mortality per mother in Ethiopia

The data to be analyzed for this study were obtained from Ethiopian Demographic and Health survey (EDHS) 2016. This study includes 16650 women under-fifty years of age. The result showed that among the 16650 women, 8584 (51.56 %) never experienced under-five deaths of their children while 8066 (48.44%) of the women experienced death of their under-five children.

**Table 4. 1.** Frequency distribution of number of under-five deaths in Ethiopia.

Number of deaths per mother	Frequency	Percent
0	8584	51.56
1	4416	26.5
2	2810	16.9
3	745	4.5
4	81	.5
5	10	.1
6	4	.0001
Total number of deaths	8066	48.44

As shown in Table 4.1, 26.5%, 16.9%, 4.5%, 0.5%, 0.1%, and 0.0011%, of the women lost 1, 2, 3, 4, 5, and 6 of their under-five children, respectively.

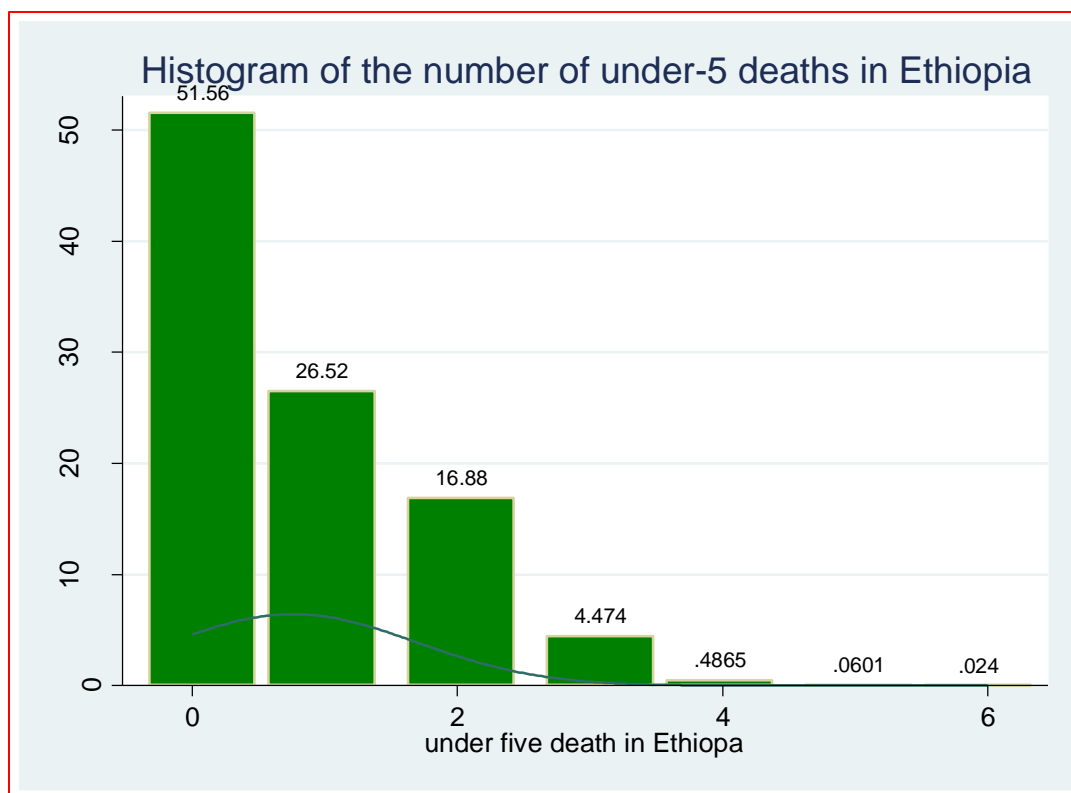
Summary statistics for the dependent variable used in this study are presented in Table 4.2. The 16650 number of women values corresponding to each variable were used in the study. Among the 16650 women, considered in the study, 8066 women were died there children before the age of five. While the smallest value for number of under-five deaths was 0, the highest child deaths were 6. Out of 16650 women, 8584 (51.56%) have no under-five deaths there child in Ethiopia, this indicating an excess zero in the response variable.

**Table 4. 2.**Summary data for the dependent variable, number of under-5 deaths

	Minimum	Maximum	Mean	Variance	Skewness
<b>Number of under-5 deaths</b>	0	6	.76	.88	1.050

**Table 4.2** showed that the variance of the outcome variable (number of under-five death) was .88, the sample mean of the response variable, the number of under-five deaths was 0.76 the fact that the mean is smaller than the variance, and the **ratio**  $\frac{.88}{.76} = 1.16 > 1$  the results suggesting over dispersion in the dataset. Moreover, the data has excess zeros and thus one might expect that the zero-inflated models and hurdle regression models would be appropriate to predict the number of under-five deaths in Ethiopia.

**As Figure 4.1:** showed there is many counts of zero in the outcomes, the histograms are highly picked at the beginning (zero values). However large observations (i.e. large number of under-five deaths per women) are less frequently observed. This leads to have a positively (or right) skewed distribution. This is an indication that the data could be fitted better by count data models which takes into account excess zeros models



**Figure 4. 1.** Histogram of the number of under- five mortality in Ethiopia

#### 4.1.2. Summary Statistics for explanatory Variables

Summary statistics of categorical variable for the number of children mortality by women aged 15- 49 years are presented in Table 4.3.

**Table 4. 3.** Presents summary statistics of the categorical variables that are assumed to affect the number of under-five mortality among Women ages 15-49.

Variable	The number of under-five mortality		
	Number of death	proportion	Percentage
<b>Region</b>			
Tigray	697	0.0864	8.64%
Affar	973	0.1206	12.06%
Amhara	797	0.0988	9.88%
Oromia	835	0.11036	11.04%
Somali	972	0.1205	12.05%
Benishan-Gumuz	1036	0.1284	12.84%

SNNPR	850	0.1054	10.54%
Gambela	700	0.0868	7.38%
Harari	482	0.0598	6.7%
Addis Ababa	291	0.0361	6.4%
Dire Dawa	432	0.0536	6.9%
<b>Sex</b>			
Male	3997	.4955	49.55%
Female	4069	.5045	50.45%
<b>Women education level</b>			
No education	4862	.6028	60.28%
Primary	2150	.2666	26.66%
Secondary	762	.0945	9.45%
Higher	292	.0362	3.62%
<b>Mother's age at the fifth birth</b>			
15-19	2482	.3077	30.77%
20-24	824	.1022	10.22%
25-29	521	.0646	6.46%
30-34	1012	.1255	12.55%
35-39	1314	.1629	16.29%
≥ 40	1883	.2334	23.34%
<b>Employment status of mother's</b>			
Yes	5317	.6592	65.92%
No	2749	.3408	34.08%
<b>Mother's Currently Breastfeeding status</b>			
No	4751	.5890	58.9%
Yes	3315	.4110	41.1%
<b>Mother's Marital status</b>			
Single	2365	.2932	29.32%
Married	1787	.2215	22.15%
Widowed	1863	.2310	23.1%
Divorced	2051	.2498	24.98%
<b>Wealth index (household income status)</b>			
Poor	4027	.4993	49.93%
Medium	2875	.3564	35.6%

Rich	1164	.1443	14.4%
<b>Family size of household</b>			
1-3	2285	.2833	28.3%
4-6	2735	.3391	33.9%
≥ 7	3046	.3776	37.76%
<b>Place of residence</b>			
Urban	3020	.03744	37.44%
Rural	5046	.6771	67.71%
<b>Religion</b>			
Orthodox	2940	.3645	36.45%
Muslim	3491	.4328	43.28%
Protestant	1594	.1976	19.76%
Other	41	.0051	0.51%
<b>Type of birth</b>			
Single birth	3319	.4115	41.15%
Multiple	4747	.5885	58.85%
<b>Place of delivery</b>			
Home	5044	.6253	62.53%
Health center	3022	.3746	37.46%
<b>Child's anemia status</b>			
Non-Anemia	1091	.1353	13.53%
Mild	2491	.3088	30.88%
Sever	4484	.5559	55.59%
<b>Source of drinking water</b>			
Unprotected	4939	.6123	61.23%
Protected	3127	.3877	38.77%
<b>Toilet facility status</b>			
No toilet facility	4892	.6064	60.64%
has toilet facility	3174	.3936	39.36%
<b>Preceding birth interval</b>			
≤ 2 years	3592	.4453	44.5%
2-3 years	2967	.3678	36.78%
≥ 4 years	1507	.1868	18.86%
<b>Vaccination of child</b>			
No	4984	.6179	61.79%
Yes	3080	.382	38.2%
<b>Father's educational level</b>			

No Education	4024	.4989	49.89%
Primary	2265	.2808	28.08%
Secondary	939	.1164	11.64%
Higher	838	.1039	10.39%
<b>Birth order</b>			
first birth	2045	.2535	25.35%
2-3	2872	.3561	35.6%
4-6	2286	.2834	28.34%
≥ 7	863	.1070	10.7%

**Table 4.3** the result of the socioeconomic, biological, environmental, and demographic, related factors that affected the number of under-five mortality are summarized as follows.

The result in table 4.3 showed that the lowest proportion of under-five mortality was occurred in Addis Ababa, Dire Dawa and Harari, 0.0361, 0.0536 and 0.0598, respectively, whereas the highest proportion of under-five mortality occurred in, Benishan-Gumuz region (0.1284) and Affar (0.1206)

**Women education level:** - There is difference in the number of under-five mortality has been observed between women with no education and those with education. Women with some education exhibited lower proportion of under-five mortality than those without education. That is, difference in the proportion of under-five deaths was observed between women who have never been to school and those with primary, secondary and higher education. The proportion of under-five deaths of women with no education was 0.6028. Conversely, the proportion of under-five deaths of women with primary education, secondary education and higher education were 0.2666, 0.0945 and 0.362 respectively.

**Mother's age at the first birth:-** The mothers age at the first birth the age between 15 years and 19 years have the highest proportion of under-five deaths (0.3077) and the second highest proportion of under-five deaths was observed for mothers age at the first birth above 40 years (0.2334). The lowest proportion of under-five deaths was observed for mothers' age at first birth the age between 25 years 29 years (.0646) and the age between 20 years and 24 years (.1022). The proportion of under-five deaths of Childs for male and female were 0.4955 and 0.5045, respectively. The descriptive result indicated that the proportion of under-five deaths higher for female as compared with male Childs.

**Mother's Currently Breastfeeding status:-** Table 4.3 showed that non- breastfeeding mothers have a higher proportion of under-five child deaths (.5890) as compared to breastfeeding mothers (.4110). In addition, working mothers have a higher number of under-five child deaths (.6592) as compared to non-working mothers (.3408). Similarly, the proportion of under-five deaths for multiple births was (.5885) higher as compared to single birth (.4115).

**Place of residence:** There was a difference in the proportion of under-five mortality by place of residence. The proportion of under-five mortality was higher in rural areas (.6771) as compared to urban areas (0.3744). The results we can also observe that, the proportion of under-five death that children who are delivered at home (.6253) had higher as compared with children delivered at health center (.3746). In the same way, the proportion of under-five deaths was lower among parents who vaccinated their children (.382), than does who did not (.6179).

**Wealth index (household income status):** Among the women included in the study, revealed that the proportion of under-five mortality for poor and middle economy level were (.4993) and (.3564), higher as compared with a women living in high income economic level (.1443). Because, a women's living in better and standard economic level experienced to have less proportion of under-five deaths as compared with a women living in low income economic level.

**Religion:-**The proportion of under-five mortality for other category religion (.0051) and protestant women (.1976) is lower as compared with Orthodox women (.3645) and Muslim (.4328) religion followers of women. Similarly, the proportion of under-five deaths for no educated father (.4989) is higher than fathers with primary, secondary, and higher education were (0.2808, 0.1164 and 0.1039), respectively.

**Mother's Marital status:** the result in table 4.3 showed that, marital status, women's who are married and divorced showed less proportion of under-five deaths (.2215, .2310) respectively. On the other hand women's who are single and widowed marital status higher proportion of under-five mortality (.2932, .2498) respectively. It is also showed that the highest proportion of under-five deaths occurred with families who use unprotected drinking water sources (.613) as compared with families who use protected drinking water (.387). With regards to type of toilet facility, the higher proportion of under-five child deaths occurred, the mothers without toilet facility (.6064) as compared to mothers used with toilet facility (.3936).

**Preceding birth interval:** According to preceding birth interval, the highest proportion of under-five deaths is occurred children born less than or equal to two years (.4453) and the lowest proportion of under-five deaths is occurred children born greater than equal to four years (.1868).

**With regard to birth order,** the highest proportion of under-five deaths observed for children birth order between second and third birth order (.3561) and the second highest proportion of under-five deaths was observed for children birth order of between fourth and sixth birth order (.2834). The lowest proportion of under-five deaths was observed for mothers greater than or equal to seventh birth order (.107).

**Family size of household:** There was a difference in the proportion of under-five mortality by number of family size in the household. The proportion of under-five mortality was higher number of family greater than or equal to seven (.3776) and the smallest proportion of under-five mortality was occurred between one and three number of family (.2833).

## 4.2. Count Regression Model Results For U5M

### 4.2.1. Variable Selection

In this study we have used Stepwise variable selection which is a combination of backward elimination and forward selection to identify the predictors in the model. This was done on Poisson regression model as it is the bench mark for other count regression models. Stepwise selection method addresses where variables were added or removed with respect to the p-value in the process. The result recognized that: predictor variables like women education level, women age at first birth, place deliver, family size, breastfeeding status, wealth index, marital status, types birth, anemia status, source of water drinking, region, toilet, preceding birth interval and vaccination status were statistically significant factors for the number of under-five mortality.

### 4.2.2. Goodness of Fit and Test of over dispersion

The result in Table 4.4, show that the over-dispersion parameter alpha is significantly different from zero indicating over-dispersion of the data. Hence, there was an over dispersion problem in the data. As a result of this the standard error of the standard Poisson regression model is smaller than that of the other models. Particularly, when we compare the negative binomial with the Poisson regression model, the standard error of the Poisson regression model is smaller than negative binomial regression model and this leads to wrong interpretation (Cox, 1983; Aklilu, 2011). When the assumption of the equality of variance and mean in the Poisson regression model is violated, over dispersion occurs and the standard error estimates will be biased which leads to incorrect value of the test statistic. Consequently, the covariates may wrongly interpreted (Yaacob et. al., 2010). Moreover, the ratio of the Deviance and Pearson Chi-square statistic to their corresponding degrees of freedom are greater than one, indicating over dispersion in the data and the Negative Binomial (NB) regression model is preferred over the Poisson model.

**Table 4. 4.**The results of over-dispersion test after fitting a Poisson regression

Statistics	Value	Df	Value/Df	P-value
Deviance test statistics	18495.19	14988	1.234	0.0000
Pearson Chi-square statistic	17147.9	14988	1.15	0.0000

### 4.2.3. Parameter Estimation

The various models obtained when the number of under-five mortality per women was regressed on different factors such as socioeconomics factor, demography factor, biological factors and environmental factor were involved in the under-five deaths per women model. Sometimes when analyzing a count response variable, the number of zeros may be excessive. When analyzing a dataset with an excessive number of zero outcomes, a zero-inflated model and hurdle models should be considered. In this study we apply seven count regression approaches to model the number of under-five mortality. Those models are the standard Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-inflated generalized Poisson (ZIGP), hurdle Poisson (HP) and hurdle negative binomial (HNB) models. Table 4.5 presented parameter estimates with their corresponding standard error of the Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB) regression models.

Table 4. 5. Estimated Coefficients and Std.err of the Poisson, NB, ZIP and ZINB models

Parameter	Poisson		NB		ZIP		ZINB	
	coef	St.err	Coef	Sd.err	coef	Sd.err	Coef	St.err
<b>Mother age</b>								
20-24	-.0021	.0401	-.00212	.04077	.00133	.0425	-.02089	.04105
25-29	.0406	.0514	.0406	.0534	-.0418	.05561	-.11121	.05399
30-34	.0196	.0279	.01955	.0289	-.0196	.03012	-.028025	.02919
35-39	.0627	.0224	.06267	.0264	.0643	.0275	.054639	.02658
≥ 40	.0089	.0241	.0089	.0250	.0107	.02604	.007680	.02524
<b>Sex of child</b>								
Female	.0155	.0142	.0155	.0179	0.0140	.018728	.00346	.02696
<b>Place delivery</b>								
Health center	.0094	.0112	.0094	.0181	.01094	.0188	.0104	.01826
<b>Women educat</b>								
Primary	.2814	.0187	.2813	.0218	.1967	.0238	-.1874	.02194
Secondary	-.3773	.0322	-.377	.0365	-.1943	.0434	-.5481	.03489
Higher	-.4174	.0423	-.418	.0538	.0779	.0814	-.5843	.05282
<b>Employment S</b>								
No	.0056	.0192	-.00589	.01958	.00526	.0204	.00143	.01975
<b>Family size</b>								
4-6	.0015	.0221	.00096	.02366	-.00153	.0246	.010893	.02388
≥ 7	-.0013	.0158	-.00029	.0215	.00223	.0224	.0048314	.02191
<b>Breastfeeding</b>								

No	.0183	.0112	.0182	.0158	.01987	.0189	.0203056	.01835
<b>Wealth index</b>								
Medium	-.688	.0196	-.6877	.0219	.00618	.0229	-.00631	.02221
Rich	-1.11	.0236	-1.107	.0277	-.0105	.0288	-.01928	.02790
<b>Marital status</b>								
Married	-.1258	.0215	-.1258	.0256	-.1204	.0265	-.08564	.02578
Widowed	-.0721	.0227	-.0721	.0256	-.0691	.0267	-.05097	.02577
Divorced	-.1140	.0235	-.1141	.0248	-.1079	.0258	-.06220	.0250
<b>Residence</b>								
Rural	.3632	.0234	.3639	.02844	.36788	.0296	.4049	.02831
<b>Religion</b>								
Muslim	-.0007	.0213	.00017	.02169	.00113	.0225496	.00268	.02185
Protestant	.07085	.0232	.0712	.0253	.07384	.026292	.074009	.02546
Other	-.0223	.1356	-.0226	.1377	-.0489	.14094	-.04276	.1359
<b>Type of birth</b>								
Multiple	-.0093	.0137	-.0092	.0187	-.0117	.0194541	-.0202	.0188
<b>Anemia status</b>								
Mild	-.0083	.0293	-.0058	.06928	-.0066	.0288021	.015827	.02767
Sever	.07885	.0277	.0788	.06928	.06928	.0212846	.0508	.01277
<b>Source water</b>								
Unprotected	.0482	.0189	.04802	.0199	.05092	.020691	.0584	.02009
<b>Region</b>								
Affar	.20399	.0412	.2043	.0416	.2136	.04323	-.23412	.04207
Amhara	-.2209	.0413	-.2208	.0428	-.219	.04422	-.45189	.0433
Oromia	.17759	.0395	.1778	.0395	.1790	.04101	-.0559	.04038
Somali	.34622	.0372	.3467	.0392	.3699	.04102	.14058	.03886
Benishan-Gum	.13154	.0362	.1317	.0392	.1375	.04068	.10045	.04053
SNNPR	.09726	.0402	.0975	.0406	.0993	.04209	.12307	.04164
Gambela	.11798	.0434	.1179	.0428	.1098	.04460	.14569	.04386
Harari	.08059	.0481	.0809	.0484	.0851	.05020	-.12904	.04933
Addis Ababa	-.1204	.0621	-.1201	.0645	-.127	.06696	-.37359	.06517
Dire Dawa	.0529	.0516	.0528	.0506	.0565	.05238	-.19852	.0516
<b>Toilet facility</b>								
Has no toilet	.0994	.0202	.0993	.0209	.1026	.0217	.11347	.02109
<b>Birth interval</b>								
2-3 years	-.0167	.0201	-.0167	.0202	-.0131	.0209	.0231239	.02032
≥ 4 years	-.0113	.0231	-.012	.0261	-.0096	.0272	.0222448	.02534
<b>Vaccination</b>								
No	.00954	.01501	.00953	.0180	.01106	.0188	.00923	.01825
<b>Father educat</b>								
Primary	-.0195	.00338	.00338	.0224	.00401	.0233	-.6867	.0226
Secondary	-.0019	.02241	-.00186	.0307	-.0007	.0319	-1.096	.0310
Higher	-.0328	.03066	-.0195	.03282	-.0197	.0341	-1.4157	.03304
<b>Birth order</b>								

2-3	.00013	.02376	.00013	.0238	.00303	.0247	-.00535	.02394
4-6	-.0023	.02612	-.0023	.0261	-.0008	.0272	-.00018	.02633
≥ 7	-.0184	.03493	-.0184	.03494	-.0112	.0364	-.02195	.03525

Table 4.6 presented parameter estimates with their corresponding standard error of the excess number of zero model, for zero-inflated generalized Poisson, Hurdle Poisson (HP) and Hurdle negative binomial (HNB) regression models.

**Table 4. 6.** Estimated Coefficients and Std.Err of the ZIGP, HP and HNB count models

Parameters (variables)	ZIGP		Hurdle Poi (HP)		Hurdle (HNB)	
	Coef	St.err	Coef	St.err	Coef	Sd.err
<b>Mother age</b>						
20-24	-.0398	.0254	-.0492	.0389	-.05860	.04464
25-29	-.0952	.0425	-.0872	.0513	-.06875	.05172
30-34	.0357	.0196	-.0264	.0264	-.02745	.02768
35-39	.0695	.0236	.0738	.0251	.07264	.02532
≥ 40	.8523	.0221	.0271	.0237	.90554	.02389
<b>Sex of child</b>						
Female	.0126	.0156	.00739	.0171	.00507	.01723
<b>Place of delivery</b>						
Home	.0145	.01612	.01358	.01721	.01177	.01733
<b>Women educat level</b>						
Primary	-.146	.0123	-.1319	.02077	-.13218	.02088
Secondary	-.302	.0245	-.2033	.0329	-.20169	.03312
Higher	-.232	.0651	-.1614	.04940	-.16802	.04960
<b>Family size</b>						
4-6	.0196	.0321	.011429	.02254	.010724	.02267
≥ 7	.0256	.0296	.009348	.02065	.50569	.02079
<b>Employment mother</b>						
Yes	0.0256	0.0256	.01889	.01873	-.02057	.01885
<b>Breastfeeding</b>						
No	.0231	.0852	.01320	.01732	.01807	.01743
<b>Wealth index</b>						
Medium	-.0098	.0231	.001802	.02090	-.00579	.02104
Rich	-.0452	.0561	-.03418	.02620	-.02980	.02636
<b>Place residence</b>						
Rural	0.265	.0123	.16323	.02622	.15257	.02654
<b>Marital status</b>						
Married	-.0452	.0362	-.03317	.02433	-.15856	.02955
Widowed	-.059	.0189	-.00536	.02429	-.09554	.02956
Divorced	.156	.0269	.00417	.02352	-.14606	.02886
<b>Religion</b>						
Muslim	0.0496	.0345	.00379	.02056	.07129	.02074

Protestant	0.0654	.0269	.03679	.02413	.03896	.02429
Other	-0.0258	.1356	-.1867	.1253	-.20488	.12864
<b>Type of birth</b>						
Multiple	.0245	0.0396	.01580	.0177	1.0710	.017883
<b>Anemia status</b>						
mild	.0125	.0654	-.0086	.07528	.0833	.028252
several	.0875	.0584	.0988	.06928	.06387	.02650
<b>Source water</b>						
Unprotected	.0345	.0256	.04011	.01886	.169	.01896
<b>Region</b>						
Affar	.1402	.0452	.12039	.03994	.06684	.05043
Amhara	-.1521	.0425	-.17052	.03946	-.16710	.04557
Oromia	.1356	.0352	.1611	.03719	.17583	.03737
Somali	.4723	.0563	.49480	.03788	.48759	.03826
Benishan-Gumuz	.5213	.0356	.15844	.03673	.554828	.03697
SNNPR	.0654	.0254	.06932	.03794	.074106	.03805
Gambela	.1154	.0258	.10691	.04011	.105470	.04031
Harari	.1342	.045	.14712	.04522	.15094	.04540
Addis Ababa	-.0521	.0463	.03343	.05849	-.04044	.05907
Dire Dawa	.1423	.0489	.15312	.04700	.04214	.04729
<b>Toilet facility</b>						
no toilet facility	0.0714	.0245	.06363	.01981	.06248	.01997
<b>Birth interval</b>						
2-3 years	.0345	.0212	-.02773	.0191	-.03033	.01926
≥ 4 years	-.0458	.0145	-.01560	.0239	-.01778	.024087
<b>Vaccination</b>						
No	.0189	.0542	.0095369	.0180	.04721	.017317
<b>Father education</b>						
Primary	.0156	.0321	.00683	.02134	.00657	.02145
Secondary	-.0415	.0235	-.00865	.02908	-.01519	.02925
Higher	-.0354	.0452	-.02362	.0311	-.02609	.03140
<b>Birth order</b>						
2-3	.0324	.0341	.02181	.02259	.02139	.02275
4-6	.0154	.0236	.01334	.0248	.0102	.02495
≥ 7	.0198	.0421	.02475	.03337	.02138	.03355

### 4.3. Comparison of Models

Poisson and negative binomial regression models are designed to analyze count data. However, Poisson and negative binomial regression models differ in terms of their assumptions. Poisson model assume, that the mean and variance of the outcome variable are equal while negative binomial regression model does not assume an equal number of mean and variance. The main

objectives of this study was identify the most appropriate count regression models that can provide better fit to the under-five mortality dataset. Several criteria can be used to compare and select among considered count models. In this study, seven different count regression models, namely; Poisson, negative binomial, zero-inflated Poisson, zero-inflated negative binomial, zero-inflated generalized, Poisson hurdle Poisson and hurdle negative binomial models were considered. Different model selection criteria were considered like the likelihood ratio test (LR), Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to compare and identify the most appropriate count regression model that can provide better fit to the dataset. For non-nested models: ZIP versus Poisson, ZINB versus NB, ZIGP versus Poisson, hurdle Poisson versus Poisson and hurdle NB versus NB regression models were identified using the Vuong test statistic.

**Table 4. 7.** Model selection criteria for PR, NB, GP, ZIP, ZINB, ZIGP, HP and HNB models for the number of under-five death dataset.

<b>Selection criteria</b>	<b>Models</b>							
	<b>Poisson</b>	<b>NB</b>	<b>GP</b>	<b>ZIP</b>	<b>ZINB</b>	<b>ZIGP</b>	<b>HP</b>	<b>HNB</b>
<b>Likelihood</b>	-18943.0	-18737.1	-18635	-18707.5	-18775.6	-18712.3	-18698.1	<b>-18677.3</b>
<b>AIC</b>	37928.15	37821.15	37635.3	37531.09	37445.26	36971.26	36721.24	<b>36542.36</b>
<b>BIC</b>	38090.27	38067.15	37989.5	37978.86	37886.13	37735.24	37602.12	<b>37531.65</b>

#### **4.3.1. Vuong Test for non-nested models**

The Vuong test is a non-nested test that is based on a comparison of the predicted probabilities of two models that do not nest (Vuong, 1989). That means vuong test statistics are needed to provide the appropriateness of zero-inflated and hurdle models against the standard count models. We can determine this by running the corresponding Vuong test of the two models, in our case a zero-inflated and hurdle models and its non-zero inflated analog. The values of the Vuong test were provided in the Table 4.8

**Table 4. 8.** Vuong non- nested tests results

<b>Model Comparison</b>	<b>Vuong Test Statistic</b>	<b>P- value</b>	<b>Preferable Modes</b>	
ZIP vs Poisson	11.83	0.0001	<b>ZIP</b>	<b>Preferable</b>
ZINB vs NB	15.25	0.0001	<b>ZINB</b>	<b>Preferable</b>
ZIGP vs GP	13.15	0.0000	<b>ZIGP</b>	<b>Preferable</b>
HP vs ZIGP	5.32	0.0001	<b>HP</b>	<b>preferable</b>
HNB vs ZINB	8.63	0.0000	<b>HNB</b>	<b>Preferable</b>

Table 4.7 and table 4.8 showed the criteria in order to select the best model among the seven count regression models. First, the calculated value of the Vuong test was 11.83 and p values = 0.0001 was greater than the hypothetical value (1.96) for ZIP versus Poisson model. This value revealed that ZIP model was preferred to Poisson model in order to estimate the number of under-five mortality. In the second case, comparison of ZINB versus NB models, the calculated value of the Vuong test was 15.25 revealed that the ZINB model was preferred to NB regression model. In the third case, comparison of ZIGP versus GP models, the calculated value of the Vuong test was 13.15 revealed that the ZIGP model was preferred to GP regression model. In the fourth case, comparison of HP versus ZIGP models, the calculated value of the Vuong test was 5.32 revealed that the HP model was preferred to ZIGP regression model. After a series of tests and model comparisons (as shown in Table 4.8), the calculated value of the Vuong test statistic for comparing HNB versus ZINB models was 8.63, indicating that the HNB model was preferred to ZINB regression model.

Finally, to compare the Poisson, NB, GP, ZIP, ZINB, ZIGP, HP and HNB models, AIC, BIC and Log likelihood were used as shown in Table 4.7. The model with the smallest AIC, smallest BIC and largest Log likelihood is preferred. Since Hurdle Negative Binomial model has the smallest AIC, smallest BIC and maximum Log likelihood as compare other count regression models. Hurdle Negative Binomial model is the most appropriate and preferred model among the seven count models. Thus, the Hurdle Negative Binomial regression model with the lowest value of AIC, lowest value of BIC and the highest value of Log likelihood among other count models. Therefore, hurdle negative binomial regression model is the most appropriate and preferred model to analysis the number of under-five mortality in Ethiopia.

### 4.3.2. Predicted value and Probability

The result showed that the Poisson and the NB model under-estimated zero counts, the zero inflated models over-estimated zero counts and the hurdle models captured all zero values. Based on predicted outcomes, the differences in model fit between the seven models are remarkable. Still the standard Poisson model and the NB model do not fit the data reasonably well, but the hurdle regression model can fit the dataset shown in Table 4.9.

**Table 4. 9.** Zero count capturing in count models for under-five mortality

	<b>observed</b>	<b>Poisson</b>	<b>NB</b>	<b>GP</b>	<b>ZIP</b>	<b>ZINB</b>	<b>ZIGP</b>	<b>HP</b>	<b>HNB</b>
<b>No_of zero</b>	8584	8392.48	8489.68	863.26	8745.62	8749.35	8593	8584	8584

**Table 4. 10.** Observed and predicted probability from Poisson, NB, ZIP, ZINB, ZIGP, HP and HNB model for under-five deaths

<b>Numb death</b>	<b>Observation</b>	<b>Observed Probability</b>	<b>Predicted probability</b>						
			<b>Poisson</b>	<b>NB</b>	<b>ZIP</b>	<b>ZINB</b>	<b>ZIGP</b>	<b>HP</b>	<b>HNB</b>
0	8584	0.51560	0.5230	0.5094	.5258	.5016	.5102	0.5156	0.5156
1	4416	0.26500	0.2936	0.241	.2451	.2723	.2432	0.2641	0.2651
2	2810	0.16900	0.1594	0.1637	.1589	.1484	.1689	0.1699	0.1691
3	745	0.0450	0.0407	0.0325	.0497	.0492	.0485	0.0497	0.0450
4	81	0.0050	0.0009	0.0009	.00722	.0053	.0049	0.0052	0.0051
5	10	0.0010	.0.0025	0.0002	.00287	.0032	.00021	0.0024	0.0012
6	4	0.00001	.00141	0.0013	.00001	.0001	.00002	0.0002	0.00001

Table 4.10 showed the observed and predicted probability of number of under-five mortality. In the table 4.10, under-five deaths data set results indicates that predicted probabilities for Hurdle Negative Binomial regression model were the closest to the observed probabilities, the HNB model is the most appropriate and preferred model than the other count regression models.

### The Observed and predicted probability of under-five death

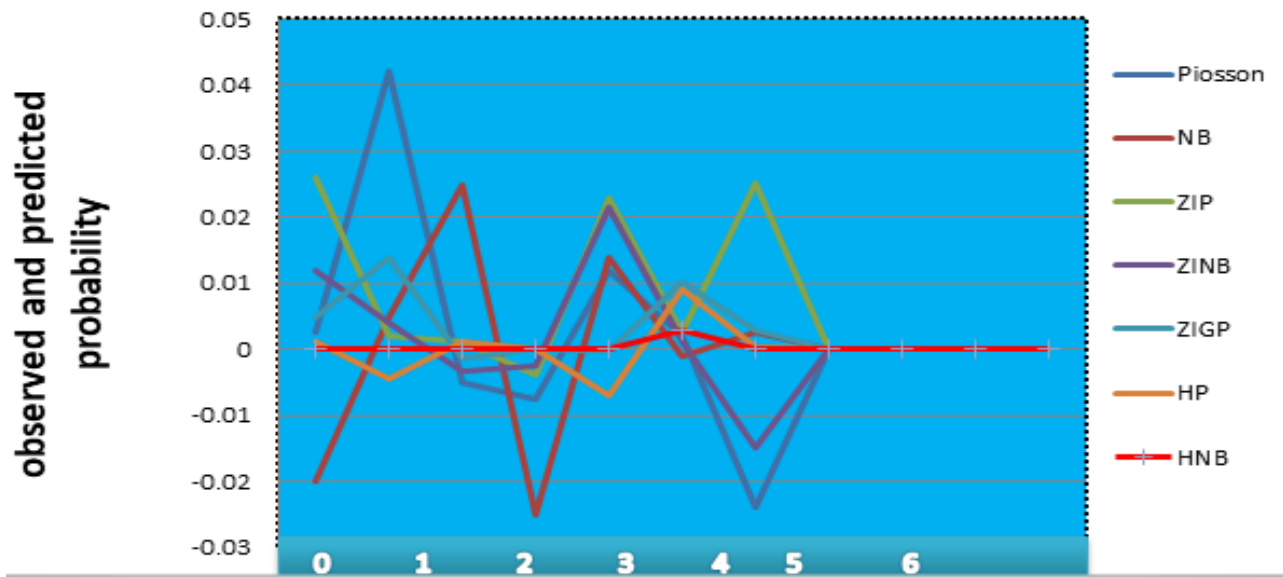


Figure 4. 2. The Observed and predicted probability of Poisson, NB, ZIP, ZINB, ZIGP, HP and HNB models.

#### 4.4. Parameter Estimation of HNB regression Model for under-five mortality

Table 4. 11. Estimates and standard errors for HNB model for under five death in Ethiopia

	Estimate	S. E	Z value	P> z	IRR	[95% Conf. Inte]	
<b>Intercept</b>	-1.28001	.05968	.44	0.000	0.27821	-1.396	-1.163
<b>Women age (15-19)Ref</b>							
20-24	-.05860	.04464	-8.51	0.0001	0.943	-.146	-.0289
25-29	-.06875	.0517	-13.86	0.0321	0.934	-0.170	-0.0326
30-34	-.02744	.02769	0.99	0.022	0.973	-0.136	-.0268
35-39	.07263	.0253	2.87	0.004	1.075	-.122	-0.0230
≥40	.90255	.0239	-17.51	0.005	2.466	0.856	0.949
<b>Place deliver (healthc)Ref</b>							
Home	.0118	0.173	-2.499	0.013	1.125	0.327	0.351

<b>Women educat (No) Ref</b>							
Primary	-.13219	.0209	-6.33	0.000	1.141	-.1732	-.0912
Seconder	-.20169	.0331	-6.09	0.000	0.817	-.2666	-.13678
Higher	-.16802	.0496	-3.39	0.001	0.845	-.2652	-.0708
<b>Family size (<math>\leq 3</math>) Ref</b>							
4-6	.020724	.02267	2.1563	0.031	1.021	0.0237	.0652
$\geq 7$	.50569	.02079	2.90	0.004	1.66	0.465	0.5464
<b>Breastfeeding (yes) Ref</b>							
No	.1807	.0174	-1.04	0.025	1.198	0.1466	.2148
<b>Wealth index</b>							
Medium	-.05794	.0211	1.50	0.028	0.944	-.0993	-0.0165
Rich	-.098	.0263	2.92	0.0013	0.97	-0.149	-.0465
<b>Marital status(single) Ref</b>							
married	-.15856	.02955	-5.03	0.000	0.853	-.2165	-.1007
Widowed	-.09554	.0296	-3.23	0.001	0.909	-.1535	-.0376
Divorced	-.1460	.0289	-5.41	0.000	0.864	-.2127	-.0994
<b>Types birth (single) Ref</b>							
Multiple	1.071	.0179	20.2	0.0001	2.92	1.035	1.106
<b>Anemia status (no) Ref</b>							
Mild	.0833	.0283	0.53	0.041	1.09	.0482	.1388
Several	.06387	.0265	-2.41	0.016	1.07	0.012	.1158
<b>Source water (prote) Ref</b>							
Unprotected	.169	.0189	2.37	0.018	1.184	0.132	.206
<b>Region</b>							
Affar	.0668	.0505	1.33	0.185	1.07	-.0320	.1657
Amhara	-.1671	.0456	-3.67	0.000	0.85	-.2564	-.0778
Oromia	.1758	.0374	4.44	0.000	1.19	.09258	.23908
Somali	.4876	.0383	12.74	0.000	1.63	.4126	.56259
Benishan	.5548	.0369	4.19	0.000	1.74	.0825	.2278
SNNPR	.0741	.0380	1.95	0.051	1.08	-.0005	.14868
Gambela	.1055	.04031	2.62	0.009	1.11	.02646	.1845
Harari	.1509	.0454	3.32	0.001	1.16	.06195	.2399
Addis Ababa	-.0405	.0592	0.68	0.049	0.67	-.0753	.1562
Dire Daw	.0422	.04728	3.01	0.003	1.04	.0495	.23481
<b>Toilet (yes) Ref</b>							
Has no toil	.06248	.01997	3.13	0.002	1.064	.02335	.10162
<b>Birth interv (&lt;2year) Ref</b>							
2-3 year	-.0304	.0193	-2.91	0.003	0.97	.062	.0751
$\geq 4$ yea	-.01778	.02409	-1.74	0.043	0.98	-.065	-.0294
<b>Vaccination (yes) Ref</b>							
No	.0472	.01732	2.73	0.006	1.05	0.0963	.1907
<b>Estimates for Zero-Inflation part model</b>							
<b>Intercept</b>	2.5750	.34366	-7.20	0.000	13.13	1.9014	3.2486

<b>Women educatio (no) Ref</b>							
Primary	.28124	.02468	-11.39	0.000	1.32	0.2328	0.3297
Secondar	.30153	.0325	-18.49	0.000	1.35	0.2378	0.3652
Higher	.4255	.04614	-13.93	0.000	1.53	0.3052	0.5458
<b>Family size</b>							
4-6	-.00817	.02720	-0.30	0.764	1.01	-.0615	.0452
≥ 7	-.2219	.02499	-0.79	0.029	.801	-0.271	-0.1729
<b>Breastfeeding (Yes) Ref</b>							
No	-.01024	.0209	0.49	0.016	0.989	-.0512	-.0308
<b>Wealth index (poor) Ref</b>							
Medium	.0359	.02556	1.60	0.048	1.036	0.0142	.0860
Rich	.0166	.0314	0.53	0.037	1.017	0.0449	.0781
<b>Resident (urban) Ref</b>							
Rural	-.3209	.0295	10.90	0.000	0.725	-0.379	-.2631
<b>Marital status (single)</b>							
Married	.14856	.02755	-5.03	0.000	1.16	0.0946	0.2026
Widowed	.1554	.0285	-3.33	0.001	1.17	0.097	0.2138
Divorced	-.1660	.0259	-5.21	0.000	0.847	-.2168	-0.1152
<b>Type of birth (single) Ref</b>							
Multiple	-.0438	.0215	-0.87	0.036	0.957	-.0859	-.0017
<b>Anemia</b>							
Mild	.1006	.0339	2.97	0.003	1.106	.03415	.16709
Severl	-.0769	.0312	2.47	0.014	0.927	-0.138	-0.0157
<b>Source water (prote) Ref</b>							
Unprotected	-.0338	.0215	-0.87	0.036	0.967	-.0759	-.0074
<b>Toilet (yes) Ref</b>							
Has no toil	-.2013	.0248	4.08	0.000	0.817	-.2499	-.1527

#### 4.5. Interpretations of HNB Model result for under-five deaths

The result in table 4.11 showed that, estimated Hurdle Negative Binomial regression model fit results of incident counts, the coefficients can be interpreted as follows: for a one unit change in the predictor variable, the log of the response variable is expected to change by the value of the regression coefficient (coef). In HNB model, for every one unit increase in a unit's of the significant predictors, the log number of under-five death is expected to increase or decrease by approximately the corresponding coefficient in the column of coefficient (coef). In this model the variables whose p-value < 0.05, were considered statistically significant. The count data we use the incidence rate ratios ( $IRR = \exp^{(coef)}$ ) to interpret the coefficient directly interpreted (Hilb, 2008). This is important to explain the change in percentage ( $IRR - 1$ ) of significant predictors.

**As shown in table 4.11, results of “for non-zero part or truncated part”, HNB model for under-five death interpreted as follows.**

Women ages at first birth, the estimated coefficients of age groups of women are statistically significant for the number of under-five death. The results in Table 4.11 show that age category of women has a significant impact on the number of under-five death per women. The expected number of under-five deaths those women aged 20-24 years had decreased by 5.7% as compared to the expected number of under-five deaths in the age group 15-19 while holding all other variables in the model constant. Similarly, the expected number of under-five deaths those women aged 25-29 and 30-34 years had decreased by 6.6 % , 2.7% respectively as compared to the expected number of under-five deaths in the age group 15-19 while holding all other variables in the model constant . But the expected number of under-five deaths those women aged 35-39 and above 40 years had increased by 7.5% and 146.6% respectively as compared to the expected number of under-five deaths in the age group 15-19 while holding all other variables in the model constant. The finding also showed that estimated coefficients of Place of deliver are statistically significant for the number of under-five death. The expected number of under-five death for children who are born at home increased by 12.5% as compared to children who are born at health sector holding all other variables in the model constant.

The finding of this study also revealed that mother’s level of education had a significant factor to reducing of the number of under-five mortality. The expected number of under-five mortality for women with primary education was decreased by 14.1% as compared to those with no education (reference group) controlling other variables in the model. In addition, the expected number of the under-five mortality for women with secondary and higher level of education were decreased by 18.3% and 15.5% as compared to those women with no education, respectively, holding all other variables in the model constant. Similarly, the finding of this study, wealth index of the household has a significant influence to reducing the number of under-five mortality. The expected numbers of under-five deaths for women in the medium and rich households were decreased by 5.63% and 3% as compared to the expected number of under-five deaths for women in the poor households, respectively, while holding all other variables in the model constant.

The result also suggests, household family size was a significant determinant of under-five mortality. The risk of under-five death increases as family size increased. The expected number of under-five mortality for family size group 4-6 was increased by 2.1% as compared to those with family size less than or equal to three household member (reference group) controlling other variables in the model. The expected number of under-five mortality for family size greater than or equal to seven was increased by 66% as compared to those with family size less than or equal to three household member holding other variables in the model constant.

In this study, currently breastfeeding has a significant effect to reducing the number of under-five mortality. The expected number of under-five mortality for mothers who were not breastfeeding was increased by 20% as compared to those mothers who were breastfeeding their child holding other variables in the model constant. In addition, mother marital status had a significant factor on the number of under-five deaths. The expected number of under-five deaths decreased by 14.7% for married mothers as compared to that for single mothers while holding all other variables in the model constant. Similarly, the expected number of under-five mortality for widowed and divorced mother had reduced by 9.1% and 13.6% respectively as compared to that for single mothers while holding all other variables in the model constant.

The finding of this study also revealed that types of birth child had statistically significant impact on the number of under-five mortality. The expected numbers of under-five deaths for multiple births was increased by factor of 1.071 as compared to the expected number of under-five mortality for the single birth, while holding all other variables in the model constant. Table 4.11 showed that anemia status was an important factor for predicting the number of under-five mortality. The expected numbers of under-five deaths for mild and severe anemia status of child were increased by 9% and 7% , respectively, as compared to that for non-anemia status of child, while holding all other variables in the model constant.

As shown in Table 4.11 toilet facility had a significant effect on the number of under-five deaths. The expected number of under-five deaths increased by 6.4% for a woman with no access to toilet facility as compared to a woman with access to toilet facility holding all other variables in the model constant. According to the findings of this study, source of water drinking has a significant influence on the number of under-five mortality. The expected number of under-five

deaths for women those use unprotected source of water was increased by 18.4% as to women use protected water source holding all other variable in the model constant. The result also shows that vaccination is the most important determinants of under-five mortality. The expected number of under-five death for non-vaccinated children was increased by 5% as compared to vaccinated children holding all other variable in the model constant.

Regarding the regional differences in under-five mortality, the expected number of under-five deaths for women from Oromia, Benshangul-Gumuz, Somali, SNNP and Gambela were increased by 19%, 74%, 63%, 8% and 11%, respectively, as compared to the expected number of under-five deaths for women in Tigray, controlling for the other variables in the model. Conversely, the expected number of under-five deaths for women from Addis Ababa was decreased by 33% as compare to women in Tigray controlling other variables in the model. The finding of this study also revealed that the birth interval was statistically a significant factor on the number of under-five mortality. The expected number of under-five deaths for those children born with preceding birth interval between two and three years and above four years had decreased by 3% and 2%, respectively, as compare to children born with preceding birth interval less than or equals to two years controlling other variables in the model.

However, the variable in this study like father education level, mother work status, sex of child, birth order and religion were no statically significant impact on under-five mortality. This insignificant effect of father education, mother work status, birth order and religion might be due to the correlation with other variables in this study and this needs further investigations.

#### **4.5.1. Interpretation of HNB model for covariates of zero parts**

The bottom half of table 4.11, labeled “Estimates for Zero-Inflation part model”, contains coefficients for the factor change in the odds of being in the always-0 group compared to the not always-0 group. As shown in table 4.11, women education level has a significant impact on the probability of being in the always zero group. The odds of no occurrence of under-five deaths (being in the always zero group) increase by a factor of 1.32 for women primary education level as compared to women no education (reference group) controlling other variables in the model. Similarly place of residence has a significant impact on the probability of being in the always

zero group. The odds of no occurrence of under-five deaths decrease by a factor of 0.725 for rural women as compared to those women in urban controlling other variables in the model.

The finding of this study also revealed that marital status significant factor on the probability of being in the always zero group. The odds of no occurrence of under-five death (being in the always zero group) increase by a factor of 1.16 for married women as compared to no married women controlling other variables in the model. According to the findings of this study, the types of birth was significant impact on the probability of being in the always zero group. The odds of no occurrence of under-five deaths decrease by a factor of 0.957 for multiple births as compare to single birth. The result also suggests, toilet facility was a significant determinant on the probability of being in the always zero-0 group. The odds of being in the always zero group decrease by a factor of 0.817 for a women with no access to toilet facility as compared to a women with access to toilet facility holding all other variables in the model constant.

Similarly source of drinking water has a significant impact on the probability of being in the always zero group. The odds of no occurrence of under-five death (being in the always zero group) decrease by a factor of 0.967for women those use unprotected source of water as to women use protected water source controlling other variables in the model.

## **5. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS**

### **5.1 Discussion of the Results**

The purpose of this study was to identify, socioeconomic, biological, demographic, and environmental related determinants on under-five mortality in Ethiopia based on the 2016 EDHS data. The most appropriate count regression model was selected from seven possible count models. Among those seven count models the Hurdle Negative Binomial regression model was selected as the most appropriate model for under-five mortality in Ethiopia. In this discussion part aims some explanation of the results of Hurdle Negative Binomial regression model of proximate and socioeconomic, biological, demographic, and environmental related determinates impact on under-five child mortality in related to theoretical background and previous researches. The results obtained from HNB are discussed as follows.

The findings of the study show that being born to a mother with education schooling was associated with decreased risk of under-five death compared to being born to mothers with no education. Thus, educational level of mothers is an important and significant factor of under-five mortality risks in Ethiopia. The mother's level of education is strongly linked to child survival. Higher levels of educational attainment are generally associated with lower mortality rates, since education exposes mothers to information about better nutrition, use of contraceptives to space births, and knowledge about childhood illnesses and treatment. In addition to this indicated that education improves the ability of mothers to implement simple health knowledge and facilitates their capacity to manipulate their environment including health care facilities, interact more effectively with health professionals, comply with treatment recommendations and keep their environment clean. This finding was consistent with (Alemu, 2015, Abdullah, 2014, Stalling, 2004, Lemani, 2013, Root, 2001, Goro, 2007, Sampson, 2014, Abimbola et al 2012)

Similarly, the findings of this study, wealth index of the household has a significant influence in reducing the risk of under-five mortality. Income of the households is one of the commonly identified social determinants of health. Previous study has indicated that income and under-five mortality had negatively statistical association. Even countries with higher national income were associated with lower under-five mortality rates. Similarly, in this study a significant reduction in risk of under-five mortality was observed among births to mothers residing in richest households.

Even though, basic health services are free in Ethiopia, other opportunistic costs and in case of severe health problems; catastrophic costs cannot be affordable to less income households for advanced and better health services. This finding was consistent with (Lawn, 2017, Alemu, 2015, Mamun, 2014, Mondal et al. 2009, Sampson, 2014, Getiye, 2011, Lemani, 2013)

According to the findings of this study, type of birth was also a significant predictor of under-five mortality. As expected, multiple births positively associated with under-five mortality. Children of multiple births were found 1.071 times more likely to die before age of five relative to the reference group of single birth. Multiple births are also a serious issue in Ethiopia and getting worse and also it exposed to higher economic burden and affects the quality of nutrition and health care of children. Several studies also identified birth type to be related with under-five child death as multiple births were associated with a higher risk of child mortality. This finding was consistent with (Berhie & Yirtaw, 2017, Gebretsadik & Gabreyohannes, 2016).

The findings this study suggested that, birth intervals was also a significant predictor of under-five mortality. As expected, the effects of birth interval have a negative impact on under-five mortality. Long birth intervals are largely reduced the risk of under-five mortality relative to the short birth interval and as the increase the previous birth interval the risk of child mortality decreased. Short birth interval increases the risk of under-five child mortality due to physiological and nutrition depletion of the mothers which relate to premature child birth and the mothers exposed to pregnancy complication. Shorter birth interval births were more risky in this study also. Beyond to biological risks associated with short birth intervals, there may be also economic, emotional and other burdens associated to the mothers and influencing the survival of the under-five children as well. This finding was in line with what was found by (Kummar & Gemechis, 2010, Rutstein, 2005, Mutunga, 2004) that the risk of under-five deaths is higher for children born less than two years of the previous period and lower for children born more than 4 years of the previous births. Similar result also obtained in Ghana (Goro, 2007 & Zimbabwe, Kombo & Ginneken, 2009).

According to the results, mother's current breastfeeding status was a significant determinant for reducing under-five mortality showing that children born to non-breastfeeding mother experience higher risk of mortality than children born to breastfeeding mothers. That children breastfed for

more than six months highly reduced the risk of under-five mortality. These are in line with the majority rural parts of Ethiopian mother's economical statuses that very poor and have no access to provide alternative nutrition choice for children. In many research the breastfeeding occurs not only as a response to hunger, but also as a mechanism of decreasing a child's stress and discomfort. Therefore would be expected to play an important role in the child's psychosocial development. Breastfeeding also providing the child with comfort, love, security and communication. This result is consistent with the findings by (Bello & Joseph, 2014, Gebretsadik & Gabreyohannes, 2016, Tesfaye, 2011, Uddin et al. 2009, Alemu, 2015).

Similarly, this study found that, Vaccination status of child was found to be a significant predictor of under-five mortality. Those non vaccinated children were 0.0472 times higher risk of mortality than that of vaccinated children. Vaccines can prevent infectious diseases that once killed or harmed many infants, children, and adults. Without vaccines, your child is at risk for getting seriously ill and suffering pain, disability, and even death from diseases like measles and whooping cough. This finding was inconsistent with (Kwabena, 2011, Alemu, 2015).

The empirical result shows that, source of drinking waters was found to be the most important significant predictor of under-five mortality. The women those use unprotected source of drink water were 18.4% times higher risk of mortality than that of women use protected drink water source. Improved sources of drinking water are less likely to be contaminated and likely to prevent the spread of water-related diseases, such as infections and cholera. This finding is in line with (Aemu, 2015, Fayehun, 2010, Berhie & Yirtaw, 2017, Mesike & Mojekwu, 2012). In addition, the under-five mortality risk was lower for women with access to toilet facility than that for women without toilet facility. This study have shown that children living in households with some kind of toilet facility are less likely to be sick than children in households which do not have toilet facilities. Access to modern sanitation facilities, flush toilets, reduces diarrhea prevalence and ultimately reduces the under-five mortality. This finding was consistent with (Gebretsadik & Gabreyohannes, 2016, Bottenheim, Genser et al; 2008, Berhie & Yirtaw, 2017, Alemu, 2015).The finding of the study revealed that marital status is the important determinant for the death of under-five children. These results show that, single women increase the risk of under-five mortality relative to married, divorced and widowed. In line with this, the main finding of this study that children born to married, divorced and widowed women play an

important significant role for the reduction of under-five mortality in comparison to children born non-married or single women. This might be due to socioeconomic factors, traditions and the lifestyle effects of the non-married women. Similar result was observed in another study done by (Tesfaye, 2011, Gideon, 2012, Seyoum, 2012, Sampson, 2014).

According to the results, place of residence was found to statistically significant impact on under-five mortality, such that children living in the rural areas had an increased risk to death compared to those children living in the urban areas. Children living in urban areas should be associated with a higher standard of living, better education, better sanitation, better health facilities, water are available and where modern treatment is more frequent will have lower incidences of under-five mortality. This finding was in agreement with studies by which shows a significant association between place of residence, (Stallings, 2004, Alemu, 2015, Berhie & Yirtaw, 2017).

A result also shows that children born from mothers whose age at first birth is between (15-19) have a significantly higher risk of mortality compared to those born from mothers whose age at first birth was between (20-24, 25-29 and 29-34). The estimated result also show that mothers age at first birth and intercourse increases reduced the risk of under-five mortality and mothers born their first child and mothers start sexual intercourse at younger age face high under-five mortality risk due to social and reproduction immaturity. This result in line with many studies, in Ethiopia (Kumar & Gemechis, 2010, Alemu, 2015, Seyoum, 2012, Berhie & Yirtaw, 2017).

Finding of this study also suggest that, family size was found to statistically significant impact on under-five mortality. Family size was positively correlated with under-five mortality. Family size has been found to influence infections. When many people live together, the chance of contact with pathogens increases, and hygiene may deteriorate. This has been confirmed by different studies (Woldemicael, 2001; Tarihu & Eshetu, 2013, Berhie & Yirtaw, 2017).

The study also revealed that, place of delivery was an important variable that affects the number of under-five mortality. Children born in health sector were at lower risk than those born at home. This might be due to the proper health care and attention they received during and after delivery. This has been confirmed by different studies (Berhie & Yirtaw, 2017, Gebretsadik & Gabreyohannes, 2016).

## 5.2 Conclusions

In this study determinants on under-five mortality in Ethiopia were identified by using count regression models. Data from the 2016 Ethiopia Demographic and Health Survey (EDHS) were used for analysis. The 2016 EDH result showed that there is a remarkable decline in under-five mortality rate in Ethiopia from 166 per 1000 in 1990 to 67 per 1000 in 2016. This study includes 16650 women under-fifty years of age among those 8584 (51.56 %) never experienced under-five deaths and about 8066 (48.44%), women had experienced under-five deaths due to different factors. The descriptive results suggested that there is high variability in the non-zero values. Similarly, the data contain about 51.56% of the number of under-five mortality without death event happened indicate the occurrence of excess number of zeros in the dataset.

In this study, the most fitted model was selected from different seven count regression models: Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-inflated generalized Poisson (ZIGP), hurdle Poisson (HP) and hurdle negative binomial (HNB) using different comparison techniques. The comparison was conducted by using likelihood-ratio test (LRT), Akaike information criteria (AIC), Bayesian information criteria (BIC) for nested model and Vuong test for non-nested model. The result revealed that HNB model was found to be the most appropriate model to predict the number of under-five mortality in Ethiopia. Hurdle negative binomial regression model is better fitted the data which is characterized by excess zeros and high variability in the non-zero outcome than any other count regression models.

This study captured predictor variables that had significant factors influencing the number of under-five deaths have been identified. For selected ordinary HNB model, the truncated negative binomial part, predictor variables like women education level, women age at first birth, place deliver, family size, breastfeeding status, wealth index, marital status, types birth, anemia status, source of water drinking, region, toilet, preceding birth interval and vaccination status were statistically significant factors influencing the number of under-five mortality. While for logistic part, women education level, family size, breastfeeding status, source of water drinking, toilet, wealth index, marital status and type of birth were found to be statistically significant determinants of the number of under-five deaths in Ethiopia.

### **5.3 Recommendations**

**Based on the result of the study we recommend the following issues:**

- ❖ The Ethiopian governments should be increasing access to improved water and sanitation strengthen policies and support technical capacities in programmers for sanitation and water quality, particularly for poor rural and urban families to reducing under-five mortality.
- ❖ The Federal Ministry of Health should work properly to increase the awareness of women for shorter duration of exclusive breastfeeding practices, vaccination child, and efforts should be made to improve spacing birth interval, mother age at first birth, place of delivery, low healthcare-use for children and strategies lacking area-wise focus on child mortality in order to reducing under-five mortality.
- ❖ Educational level of mothers plays an important role to reducing under-five mortality in Ethiopia. Therefore, concerned effort should be made to improve mothers' access to education, particularly for rural women in order to address the problem through improving their income and also enhancing the quality of care and attention they can provide to their children.
- ❖ The government, concerned institutions and other involved stakeholders should be make comprehensive prevention strategies; commitment and leadership are needed to ensure that child health receives the attention and resources needed to accelerate progress to achieve sustainable development goals 3 in 2030, to reducing under-five mortality in to 25 per 1000 child in Ethiopia.

## REFERENCE

- AGRESTI, A. 2010. *Analysis of ordinal categorical data*, John Wiley & Sons.
- AGRESTI, A. 2007. *An introduction to Categorical data analysis*, 2nd edition, John Wiley and sons; Hoboken, New Jersey.
- ALEMU, Y. 2015. *Count Regression Models for Under-Five Deaths in Ethiopia*. Addis Abeba university.
- AKLILU Z. W. 2011. *Use of Zero-Inflated Models for analysis of Immunological Data*. Inter university Institute for Biostatistics and Statistical Bioinformatics.
- AMARE, D. & FASIL, T. 2007. *Determinants of under-five mortality in Gilgel Gibe Field Research Center, Southwest Ethiopia*.
- AMIRI, A. & GERDTHAM, U.-G. 2013. *Impact of maternal and child health on economic growth: New evidence based granger causality and DEA analysis*. Newborn and Child Health, Study Commissioned by the Partnership for Maternal, Lund University, Sweden.
- AMPONSAH, M. 2017. *Trends and Differentials in Infant Death in Ghana, 1988-2014*. University Of Ghana.
- ANTAI, D. 2011. *Regional inequalities in under-5 mortality in Nigeria: a population-based analysis of individual-and community-level determinants*. *Population Health Metrics*, 9, 6.
- BEISE, J., GROJEC, A., HANSEN, C., MATTHES, J., ROSENGAERTNER, S. & YOU, D. 2017. *A child is a child: protecting children on the move from violence abuse and exploitation*.
- BELLO, R. & JOSEPH, A. 2014. *Determinants of child mortality in Oyo State, Nigeria*. *African Research Review*, 8, 252-272.
- BERHIE, K. A. & YIRTAW, T. G. 2017. *Statistical Analysis on the Determinants of Under Five Mortality in Ethiopia*. *American Journal of Theoretical and Applied Statistics*, 6, 10-21.
- BLENCOWE, H., LAWN, J., VANDELAER, J., ROPER, M. & COUSENS, S. 2010. *Tetanus toxoid immunization to reduce mortality from neonatal tetanus*. *International journal of epidemiology*, 39, i102-i109.
- BUTTENHEIM, A. M. 2008. *The sanitation environment in urban slums: implications for child health*. *Population and environment*, 30, 26-47.
- CAMERON, A. C. & TRIVEDI, P. K. 2013. *Regression analysis of count data*, Cambridge university press.
- CLARKE, K. A. 2007. *A simple distribution-free test for nonnested model selection*. *Political Analysis*, 15, 347-363
- DALRYMPLE, M. L., HUDSON, I. L. & FORD, R. P. K. 2003. *Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS*. *Computational Statistics & Data Analysis*, 41, 491-504.
- DESTA, M. 2011. *Infant and Child Mortality in Ethiopia: The role of socioeconomic, demographic and biological factors in the previous five years period of 2000 and 2005*. Lund University.

- ESPO, M. 2002. *Infant mortality and its underlying determinants in rural Malawi*, Tampere University Press.
- ERMIAS, D. 2013. *Determinants of Child Survival Chances in Rural Ethiopia*. Adama Science and Technology University, Adama, Ethiopia. *Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session CPS105)*.
- EZRA, M. & GURUM, E. 2002. *Breastfeeding, birth intervals and child survival: analysis of the 1997 community and family survey data in southern Ethiopia*. *Ethiopian Journal of Health Development*, 16, 41-51.
- FAYEHUN, O. A. 2010. *Household environmental health hazards and child survival in sub-Saharan Africa*.
- GARDNER, W., MULVEY, E. P. & SHAW, E. C. 1995. *Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models*. *Psychological bulletin*, 118, 392.
- GEBRETSADIK, S. & GABREYOHANNES, E. 2016. *Determinants of under-five mortality in high mortality regions of Ethiopia: an analysis of the 2011 Ethiopia demographic and health survey data*. *International Journal of Population Research*, 2016.
- GETACHEW, Y. & BEKELE, S. 2016. *Survival analysis of under-five mortality of children and its associated risk factors in Ethiopia*. *J Biosens Bioelectron*, 7, 2.
- GETIYE, T. 2011. *identification of risk factors and regional differentials in under-five mortality in ethiopia using multilevel count model*. *citeseer*.
- GORO, M. *The stalling child mortality: the case of three northern regions*. *The 5th conference of union for Africa population, Tanzania, 2007*.
- GOULDING, R. H., CAUGHMAN, J., RAPP, J., BIEWER, T. M., BIGELOW, T. S., CAMPBELL, I. H., CANESES, J., DONOVAN, D., KAFLE, N. & MARTIN, E. H. 2017. *Progress in the development of a high power helicon plasma source for the materials plasma exposure experiment*. *Fusion Science and Technology*, 72, 588-594.
- GRIGGS, D., STAFFORD-SMITH, M., GAFFNEY, O., ROCKSTRÖM, J., ÖHMAN, M. C., SHYAMSUNDAR, P., STEFFEN, W., GLASER, G., KANIE, N. & NOBLE, I. 2013. *Policy: Sustainable development goals for people and planet*. *Nature*, 495, 305.
- GURMU, S. 1998. *Generalized hurdle count data regression models*. *Economics Letters*, 58, 263-268.
- GURMU, S. & TRIVEDI, P. K. 1996. *Excess zeros in count models for recreational trips*. *Journal of Business & Economic Statistics*, 14, 469-477.
- HARRIS, T., HILBE, J. M. & HARDIN, J. W. 2014. *Modeling count data with generalized distributions*. *The Stata Journal*, 14, 562-579.
- HILBE, J. M. 2011. *Negative binomial regression*, Cambridge University Press.
- HILL, K., BICEGO, G. & MAHY, M. 2001. *Childhood mortality in Kenya: An examination of trends and determinants in the late 1980s to mid 1990s*.
- HUG, L., SHARROW, D. & YOU, D. 2017. *Levels & trends in child mortality: report 2017. Estimates developed by the UN Inter-agency Group for Child Mortality Estimation*.

- IRAM, U. & BUTT, M. S. 2008. *Socioeconomic determinants of child mortality in Pakistan: Evidence from sequential probit model. International Journal of Social Economics*, 35, 63-76.
- JOHANSSON, A. 2014. *A comparison of regression models for count data in third party automobile insurance.*
- JOHNSON, N. L., KEMP, A. W. & KOTZ, S. 2005. *Univariate discrete distributions*, John Wiley & Sons.
- KITTUR, I. 2014. *factors influencing infant mortality in urban kenya.*
- KASAHUN, T. & TESHOME, K. 2014. *Correlation of under-five child mortality in Ethiopia.*
- KUMAR, S., KUMAR, N. & VIVEKADHISH, S. 2016. *Millennium development goals (MDGS) to sustainable development goals (SDGS): Addressing unfinished agenda and strengthening sustainable development and partnership. Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, 41
- KYEI, K. A. 2012. *Determinants of childhood mortality in South Africa: using categorical data modeling. Journal of Human Ecology*, 37, 47-56.
- LAMBERT, D. 1992. *Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics*, 34, 1-14.
- LEMANI, C. 2013. *Modelling covariates of infant and child mortality in Malawi. University of Cape Town.*
- LONG, J. S. & FREESE, J. 2006. *Regression models for categorical dependent variables using Stata, Stata press.*
- LORD, D., WASHINGTON, S. P. & IVAN, J. N. 2005. *Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis & Prevention*, 37, 35-46.
- LORD, D. 2006. *Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention*, 38, 751-766.
- MAMUN, M. A. A. 2014. *Zero-inflated regression models for count data: an application to under-5 deaths.*
- MANUNEBO, M. N., HAGGERTY, P. A., GAIE, M. K., ASHWORTH, A. & KIRKWOOD, B. R. 1994. *Influence of demographic, socioeconomic and. J trop med hyg*, 97, 31-8.
- MCGUIRE, J. W. 2006. *Basic health care provision and under-5 mortality: a cross-national study of developing countries. World Development*, 34, 405-425.
- MEKONNEN, D. 2011. *Infant and Child Mortality in Ethiopia: The role of Socioeconomic, Demographic and Biological factors in the previous five years period of 2000 and 2005. Lund, Sweden: Lund University*, 68.
- MESIKE, C.G.AND MOJEKWU J. N. 2012. *Environmental Determinants of Child Mortality in Nigeria. Journal of Sustainable Development*, 5(1): 65-75.

- MOHAMMED, A. R., MOHAMMED, N. S. & BYOUMI, M. H. 2013. *Supportive strategies regarding accidents prevention for mothers of children under five years old. JBAH, 3, 1-11.*
- MONDAL, M. N. I., HOSSAIN, M. K. & ALI, M. K. 2009. *Factors influencing infant and child mortality: A case study of Rajshahi District, Bangladesh. Journal of Human Ecology, 26, 31-39.*
- MOSLEY, W. H. & CHEN, L. C. 1984. *An analytical framework for the study of child survival in developing countries. Population and development review, 10, 25-45.*
- MULLAHY, J. 1986. *Specification and testing of some modified count data models. Journal of econometrics, 33, 341-365.*
- MUSYOKA, P. K., OMOLO, J., KORIR, J. & NZAI, C. 2018. *An econometric analysis of effect of poverty on health status in Kenya.*
- NYINAWAJAMBO, M. R. 2018. *Survival Analysis of Time to Event Data: An application to child mortality in Sub-Saharan Africa Region using Their Demographic and Health Surveys.*
- ORGANIZATION, W. H. 2016. *World health statistics 2016: monitoring health for the SDGs sustainable development goals, World Health Organization.*
- ROOT, G. P. 2001. *Sanitation, community environments, and childhood diarrhoea in rural Zimbabwe. Journal of health, population and nutrition, 73-82.*
- RUTSTEIN, S. O. 2005. *Effects of preceding birth intervals on neonatal, infant and under-five years mortality and nutritional status in developing countries: evidence from the demographic and health surveys. International Journal of Gynecology & Obstetrics, 89, S7-S24.*
- SEYOUM, S. & WENCHEKO, E. 2013. *Determinants of mortality among one to four years old children in Ethiopia: A study based on the 2011 EDHS data. Ethiopian Journal of Health Development, 27, 8-15.*
- STALLINGS, R. Y. 2004. *Child morbidity and treatment patterns.*
- SUZUKI, E., KASHIWASE, H., 2017. *New child mortality estimates show that 15,000 children died every day in 2016. Work. World Free Poverty.*
- TEKLEMARIAM, S., GETANEH, T. & BEKELE, F. 2000. *Environmental determinants of diarrheal morbidity in under-five children, Keffa-Sheka zone, south west Ethiopia. Ethiopian Medical Journal, 38, 27-34.*
- TESFAYE, A. 2011. *“Modeling children of under-five mortality in Ethiopia using Bayesian approach”.*
- UNITED NATIONS, 2011. *The Millennium Development Goals Report 2011, New York*
- UNICEF, 2010. *Levels and Trends in Child Mortality: Report 2010: Estimates Developed by the UN Inter-Agency Group for Child Mortality Estimation, United Nations Children's Fund.*
- UNICEF, 2016. *Female Genital Mutilation/cutting: a Global Concern. UNICEF's Data Work on FGM/C, Unicef.*
- USAID, 2018. *Maternal, Neonatal and Child Health In Ethiopia.*
- VAN DER KLAUW, B. & WANG, L. 2004. *Child mortality in rural India.*

- VUONG, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307-333.
- WHO, U. & MATHERS, C. 2016. *Global strategy for women's, children's and adolescents' health (2016-2030)*. Organization.
- WHO, 2016. *Under five mortality rates*. Global Health Observatory (GHO) Data.
- WHO, 2017. *Children: reducing mortality*.
- WOLDEMICAEL, G. 2001. Diarrhoeal morbidity among young children in Eritrea: environmental and socioeconomic determinants. *Journal of health, population and nutrition*, 83-90.
- YOU, D., HUG, L., EJDEMYR, S. & BEISE, J. 2015. *Levels and trends in child mortality. Estimates developed by the UN Inter-agency Group for Child Mortality Estimation (IGME)*. Report 2015.
- ZUUR, A. F., IENO, E. N., WALKER, N. J., SAVELIEV, A. A. & SMITH, G. M. 2009. *Zero-truncated and zero-inflated models for count data. Mixed effects models and extensions in ecology with R*. Springer.

## APPENDICES

**Table A 1.** Estimates and standard errors for NB model

	Coef.	Std. Err.	z	P> z	[95% Conf. Inte]	
<b>Intercept</b>	-.7482753	.0643295	-11.63	0.000	-.8743588	-.6221919
<b>sex child</b>						
Female	.0155039	.017991	0.86	0.389	-.0197579	.0507656
<b>Family size</b>						
1-3	.0009646	.0236616	0.04	0.967	-.0454112	.0473405
>= 7	-.000297	.0215424	-0.01	0.0189	.0112519	.0419252
<b>Place delivery</b>						
Health c	.009397	.0181019	0.52	0.604	-.026082	.044876
<b>Vaccination</b>						
Yes	.0095369	.0180856	0.53	0.598	-.0259103	.0449841
<b>Women education level</b>						
No educa	.2813877	.0218136	12.90	0.000	.2386338	.3241416
Seconder	-.3773071	.0364862	-10.34	0.000	-.4488187	-.3057954
Higher	-.4174572	.05383	-7.76	0.000	-.522962	-.3119524
<b>Anemia status</b>						
Mild	.0788411	.0292549	2.69	0.007	.0215026	.1361796
Severall	.015827	.027676	0.57	0.567	-.0384169	.07007
<b>Birth order</b>						
2-3	.0001294	.0237595	0.01	0.996	-.0464382	.0466971
4-6	-.0022262	.026124	-0.09	0.932	-.0534284	.0489759
≥7	-.0183634	.0349344	-0.53	0.599	-.0868336	.0501068

<b>Birth type</b>						
Multiple	-.0092261	.018694	-0.49	0.0022	.0158657	.0274134
<b>Marital status</b>						
Married	-.1257773	.0255076	-4.93	0.000	-.1757713	-.075783
Widowed	-.0720381	.0255776	-2.82	0.005	-.1221692	-.021907
Divorced	-.1140105	.0247627	-4.60	0.000	-.1625446	-.065476
<b>Women Work status</b>						
Yes	-.0058897	.0195704	-0.30	0.763	-.044247	.0324675
<b>Father Education</b>						
Primary	.0033776	.0224119	0.15	0.880	-.0405489	.0473042
Seconder	-.001864	.0306613	-0.06	0.952	-.0619591	.0582311
Higher	-.01951	.0328178	-0.59	0.012	.0238387	.0448048
<b>Birth interval</b>						
< 2 year	-.0166545	.0201227	-0.83	0.008	.0160943	.0227853
≥ 4 yea	-.0113137	.0261818	-0.43	0.0166	.032629	.0400017
<b>Income household</b>						
Medium	-.6877109	.0219947	-31.27	0.000	-.7308197	-.6446021
Rich	-1.107022	.0276559	-40.03	0.000	-1.161227	-1.052818
<b>Women age</b>						
20-24	-.0021234	.0407678	-0.05	0.958	-.0820268	.0777801
25-29	.040626	.0534478	0.76	0.047	.0641297	.1453817
30-34	.0195493	.028969	0.67	0.030	.0372289	.0763275
35-39	.0626672	.0263964	2.37	0.018	.0109311	.1144032
≥ 40	.0089524	.0250425	0.36	0.721	-.0401301	.0580348
<b>Religion</b>						

Muslim	.0001629	.0216895	0.01	0.994	-.0423479	.0426736
Protestant	.0711526	.0252453	2.82	0.005	.0216727	.1206325
Others	-.0225517	.137652	-0.16	0.870	-.2923447	.2472412
<b>Breastfeed status</b>						
No	.0182041	.0181999	1.00	0.017	.0174671	.0538752
<b>Source of drinking water</b>						
Unprotected	.04801	.0199222	2.41	0.016	.00897	.08706
<b>Toilet of facilitated</b>						
Has no toilet	.0993309	.0208627	4.76	0.000	.0584408	.140221
<b>Resident</b>						
Rural	.3638772	.0284402	12.79	0.000	.3081354	.419619
<b>Region</b>						
Afar	.2040829	.0416031	4.91	0.000	.1225424	.2856234
Amhara	-.2208547	.0428475	-5.15	0.000	-.3048342	-.1368751
Oromia	.1778207	.0395247	4.50	0.000	.1003537	.2552878
Somali	.3467126	.0392272	8.84	0.000	.2698288	.4235964
Benishan	.1316616	.0391794	3.36	0.001	.0548713	.208451
SNNPR	.0974773	.0405592	2.40	0.016	.0179827	.176971
Gambela	.1179888	.0428352	2.75	0.006	.0340334	.201944
Harari	.0808083	.0483917	1.67	0.095	-.0140378	.1756544
Addis Ababa	-.1200492	.064428	-1.86	0.062	-.2463258	.0062274
Dire Dawa	.0528737	.0505607	1.05	0.296	-.0462234	.1519707
Lnalpha	-16.16553	156.3703	-	-	-322.6457	290.3147
Alpha	9.54e-08	.0000149	-	-	7.5e-141	1.2e+126

Table A 2. Estimates and standard errors for ZIP model

	<b>Coef.</b>	<b>Std. Err.</b>	<b>z</b>	<b>P&gt;z</b>	<b>[95% Conf. Interval]</b>	
<b>Intercept</b>	-.3954327	.0867475	-4.56	0.000	-.5654547	-.2254108
<b>Sex</b>						
Female	.0034565	.0269618	0.13	0.898	-.0493876	.0563005
<b>Family size</b>						
4-6	-.0009524	.0358836	-0.03	0.979	-.0712829	.069378
≥ 7	-.0031234	.0319939	-0.10	0.922	-.0658304	.0595835
<b>Place delivery</b>						
Health c	.0131592	.0268089	0.49	0.624	-.0393852	.0657037
<b>Birth order</b>						
2-3	.0225784	.0363211	0.62	0.534	-.0486096	.0937665
4-6	.0187165	.04016	0.47	0.641	-.0599956	.0974285
≥7	.0769717	.0543876	1.42	0.157	-.029626	.1835695
<b>Type of birth child</b>						
Multiple	-.0233368	.0196088	-1.19	0.234	-.0617693	.0150958
<b>Marital</b>						
Married	-.0834696	.0267999	-3.11	0.002	-.1359964	-.0309427
Widowed	-.0431799	.0268546	-1.61	0.108	-.0958139	.0094541
Divorced	-.0586567	.0259468	-2.26	0.024	-.1095116	-.0078019
<b>Women Work</b>						
Yes	.0017578	.0205776	0.09	0.932	-.0385736	.0420891
<b>Income of household</b>						
Medium	.01306	.0305475	0.43	0.669	-.0468121	.072932
Rich	-.0272576	.0411402	-0.66	0.508	-.107891	.0533758

<b>Women age</b>						
20-24	.1082517	.0626426	1.73	0.084	-.0145255	.2310289
25-29	.1070909	.077872	1.38	0.169	-.0455355	.2597172
30-34	.0429378	.0470409	0.91	0.361	-.0492607	.1351362
35-39	.1230861	.0409517	3.01	0.003	.0428222	.2033499
≥ 40	.0416127	.0379716	1.10	0.273	-.0328103	.1160357
<b>Religion</b>						
Muslim	.0041578	.0227309	0.18	0.855	-.0403939	.0487095
Protestant	.0720949	.0265129	2.72	0.007	.0201306	.1240591
Others	-.0729695	.1399312	-0.52	0.602	-.3472297	.2012908
<b>Breastfeeding</b>						
No	.019121	.0190714	1.00	0.316	-.0182583	.0565003
<b>Source of water</b>						
Unprotected	.0593172	.020862	2.84	0.004	.0184286	.1002059
<b>Toilet of</b>						
Has no toilet	.1127896	.021904	5.15	0.000	.0698585	.1557208
<b>Resident</b>						
Rural	.3982381	.0292818	13.60	0.000	.3408468	.4556294
<b>Region</b>						
Afar	-.2369577	.0438077	-5.41	0.000	-.3228193	-.1510961
Amhara	-.460041	.044854	-10.26	0.000	-.5479533	-.3721286
Oromia	-.0631902	.042102	-1.50	0.133	-.1457085	.0193281
Somali	.1654724	.0409568	4.04	0.000	.0851986	.2457462
Benishan	.701180	.0421941	-2.40	0.016	-.1838797	-.0184819
SNNPR	.127703	.0433508	-2.95	0.003	-.212669	-.042737

Gambela	.1429257	.0456719	-3.13	0.002	-.2324409	-.0534105
Harari	-.1322437	.0511996	-2.58	0.010	-.232593	-.0318944
Addis Ababa	-.3758027	.0670731	-5.60	0.000	-.5072635	-.2443419
Dire Dawa	-.1980753	.0533633	-3.71	0.000	-.3026655	-.0934851
<b>Women education</b>						
Primary	-.1800562	.0229155	-7.86	0.000	-.2249698	-.1351426
Seconder	-.5366496	.0361466	-14.85	0.000	-.6074957	-.4658036
Higher	-.5685238	.054519	-10.43	0.000	-.675379	-.4616687
<b>Anemia status</b>						
Mild	.043383	.0133214	3.26	0.001	.0172735	.0694924
several	.0788411	.0292549	2.69	0.007	.0215026	.1361796
<b>Father Education</b>						
Primary	.0085074	.023513	0.36	0.717	-.0375772	.0545919
Seconder	.0002365	.0321718	0.01	0.994	-.062819	.063292
Higher	-.0309339	.0343402	-0.90	0.368	-.0982395	.0363716
<b>Birth interval</b>						
2-3 year	.0174664	.0211197	0.83	0.408	-.0239276	.0588603
>=4 year	.0181483	.0263074	0.69	0.490	-.0334133	.0697099
<b>Vaccination</b>						
Yes	.0142676	.0189648	0.75	0.452	-.0229027	.05143
<b><i>Zero-Inflation part model</i></b>						
intercept	-1.475004	.3436609	-7.20	0.000	-3.148567	-1.801441
<b>Income household</b>						
Medium	.154087	.1731242	0.89	0.373	-.1852302	.493404
Rich	-.061104	.2713705	-0.23	0.822	-.5929804	.4707724

<b>Women age</b>						
20-24	.9175416	.3399065	2.70	0.007	.251337	1.583746
25-29	-.0454648	.6280179	-0.07	0.942	-1.276357	1.185428
30-34	.138255	.3693312	0.37	0.708	-.5856209	.8621308
35-39	.5383697	.2671828	2.01	0.044	.014701	1.062038
≥ 40	.2949429	.2585451	1.14	0.254	-.2117962	.8016821
<b>Birth order</b>						
2-3	.2449442	.2446785	1.00	0.317	-.2346168	.7245053
4-6	.1724977	.2870998	0.60	0.548	-.3902076	.7352029
≥7	.7282107	.3261364	2.23	0.026	.0889951	1.367426
<b>Place delivery</b>						
Home	.0357426	.1656506	0.22	0.829	-.2889266	.3604117
<b>Sex</b>						
Female	-.0786445	.167932	-0.47	0.640	-.407785	.2504961
<b>Family size</b>						
4-6	.0744586	.2224635	0.33	0.738	-.3615617	.510479
≥ 7	.0104296	.2000098	0.05	0.958	-.3815825	.4024417

**Table A 3.** Estimates and standard errors for ZINB model

	<b>Coef.</b>	<b>Std. Err.</b>	<b>z</b>	<b>P&gt;z</b>	<b>[95% Conf.</b>	
<b>Interval]</b>						
Intercepted	-.4784064	.0760575	-6.29	0.000	-.6274763	-.3293366
<b>Sex of child</b>						
female	.0034565	.0269618	0.13	0.898	-.0493876	.0563005
size						
4-6	-.010893	.0238824	-0.46	0.648	-.0577016	.0359156

≥ 7	-.0048314	.0219123	-0.22	0.825	-.0477788	.0381159
<b>Place delivery</b>						
Health c	.0104032	.0182588	0.57	0.569	-.0253834	.0461898
<b>Birth order</b>						
2-3	-.0053543	.0239432	-0.22	0.823	-.052282	.0415734
4-6	-.0001826	.0263259	-0.01	0.994	-.0517804	.0514152
≥7	-.0219498	.0352514	-0.62	0.534	-.0910414	.0471417
<b>Vaccination</b>						
No	.0092738	.0182456	0.61	0.011	.0264869	.0450345
<b>Types birth child</b>						
Multiple	-.0202294	.0188526	-1.07	0.283	-.0571799	.016721
<b>Marital</b>						
Married	-.0856364	.0257801	-3.32	0.001	-.1361646	-.0351083
Widowed	-.050974	.0257755	-1.98	0.048	-.1014931	-.000455
Divorced	-.0622034	.0250049	-2.49	0.013	-.1112122	-.013194
<b>Women Work</b>						
Yes	.0014029	.0197057	0.07	0.943	-.0372196	.0400253
<b>Wealth index</b>						
Medium	-.0063086	.0222089	-0.28	0.776	-.0498373	.03722
Rich	-.0192857	.0279022	-0.69	0.489	-.073973	.0354017
<b>Women age</b>						
20-24	-.020894	.0410501	-0.51	0.611	-.1013508	.0595628
25-29	.1112107	.0539945	2.06	0.039	.0053835	.2170379
30-34	.0280258	.0291877	0.96	0.337	-.029181	.0852326
35-39	.0546398	.02658	2.06	0.040	.002544	.1067356

>= 40	.0076803	.0252391	0.30	0.761	-.0417874	.057148
<b>Religion</b>						
Muslim	.0026855	.0218554	0.12	0.902	-.0401502	.0455212
Protestant	.0740093	.0254628	2.91	0.004	.0241031	.1239155
Others	-.0427609	.1359632	-0.31	0.753	-.3092439	.2237221
<b>Breastfeeding</b>						
No	.0203056	.0183542	3.11	0.039	.0156679	.0862791
<b>Source of water</b>						
Unprotected	.0582677	.020091	2.90	0.004	.01889	.0976453
<b>Toilet</b>						
Has no toil	.1134651	.0210899	5.38	0.000	.0721297	.1548005
<b>Resident</b>						
Rural	.4049278	.0283132	14.30	0.000	.3494349	.4604207
<b>Region</b>						
Afar	-.2341196	.0420782	-5.56	0.000	-.3165913	-.1516479
Amhara	-.4518979	.0433108	-10.43	0.000	-.5367856	-.3670102
Oromia	-.0559384	.0403811	-1.39	0.166	-.1350839	.0232071
Somali	.1405839	.0388609	3.62	0.000	.064418	.2167499
Benishan	.1004589	.0405386	-2.48	0.013	-.1799131	-.0210046
SNNPR	.1230794	.0416496	-2.96	0.003	-.2047111	-.0414477
Gambela	.1456984	.0438626	-3.32	0.001	-.2316676	-.0597293
Harari	-.129037	.0493339	-2.62	0.009	-.2257296	-.0323444
Addis Ababa	-.3735995	.0651754	-5.73	0.000	-.501341	-.245858
Dire Dawa	-.1985221	.0515937	-3.85	0.000	-.2996438	-.097400
<b>Women education</b>						

Primary	-.1874842	.0219468	-8.54	0.000	-.2304991	-.1444692
Secunder	-.5481002	.0348995	-15.71	0.000	-.616502	-.4796984
Higher	-.5843183	.0528189	-11.06	0.000	-.6878414	-.4807953
<b>Anemia status</b>						
Mild	.015827	.027676	0.57	0.567	-.0384169	.0700709
Several	.0508	.01277	3.98	0.000	.0257713	.0758288
<b>Father education level</b>						
Primary	-.6867022	.0226361	-30.34	0.000	-.7310682	-.6423363
Secunder	-1.096205	.0310062	-35.35	0.000	-1.156976	-1.035434
Higher	-1.41571	.0330443	-42.84	0.000	-1.480475	-1.350944
<b>Birth interval</b>						
2-3 year	.0231239	.0203165	1.14	0.255	-.0166958	.0629435
≥4 year	.0222448	.0253378	0.88	0.380	-.0274164	.0719061

**Table A 4.**Estimates and standard errors for HP model

	<b>Coef.</b>	<b>Std. Err.</b>	<b>z</b>	<b>P&gt; z</b>	<b>[95% Conf. Interval]</b>	
<b>Intercept</b>	1.253196	.0542674	23.09	0.000	1.146834	1.359558
<b>Sex</b>						
Female	.0073949	.0171272	0.43	0.666	-.0261738	.0409636
<b>Vaccination</b>						
No	.0095369	.0180856	0.53	0.0298	.0259103	.0449841
<b>Family size</b>						
4-6	.011429	.0225445	15.1	0.041	.0327574	.0556153
>= 7	.0093489	.0206591	0.45	0.023	.0311422	.0498401
<b>Place of delivery</b>						
Home	.0135871	.0172192	0.79	0.043	.0201618	.047336

<b>Birth Order</b>						
2-3	.0218156	.0225976	0.97	0.334	-.0224749	.0661061
4-6	.0133424	.0248	0.54	0.591	-.0352647	.0619496
≥7	.0247509	.0333792	0.74	0.458	-.0406711	.090173
<b>Types of birth</b>						
Multiple	-.0158048	.017777	-0.89	0.037	-.0506471	-.0190375
<b>Marital</b>						
Married	-.0331722	.0243362	-1.36	0.173	-.0808704	.0145259
Widowed	-.0053644	.0242966	-0.22	0.825	-.0529849	.042256
Divorced	.0041747	.0235296	0.18	0.859	-.0419425	.0502919
<b>Women Work</b>						
No	.0187956	.0187386	-1.00	0.316	-.0555226	.0179314
<b>Wealth index</b>						
Medium	.001802	.0209068	0.9	0.0431	.0391747	.0827786
Rich	-.034184	.0262048	1.36	0.0392	.0155444	.0871765
<b>Women age</b>						
20-24	.0492348	.0389603	1.89	0.006	.0271259	.1255955
25-29	.0872021	.0513519	1.70	0.089	-.0134458	.18785
30-34	.0264202	.0275187	0.96	0.337	-.0275154	.0803558
35-39	.0738351	.0251447	2.94	0.003	.0245524	.1231178
≥ 40	.0271045	.0237462	1.14	0.254	-.0194373	.0736462
<b>Religion</b>						
Muslim	.0037996	.0205676	0.18	0.043	.0365121	.0741114
Protestant	.0367972	.0241334	1.52	0.127	-.0105035	.0840979
Others	-.1867873	.125328	-1.49	0.136	-.4324256	.058851

<b>Breastfeeding</b>						
yes	.0132016	.0173274	0.76	0.046	.0207594	.0671627
<b>Source of water</b>						
Unprotected	.0401153	.0188657	2.13	0.033	.0031393	.0770912
<b>Toilet</b>						
has toil	.0636374	.0198118	3.21	0.001	.0248069	.1024678
<b>Resident</b>						
Rural	.1632359	.026221	6.23	0.000	.1118436	.2146282
<b>Region</b>						
Afar	.1204456	.0399431	6.84	0.000	.1947874	.3513613
Amhara	-.1705253	.0394699	-4.32	0.000	-.2478849	-.0931658
Oromia	.1611061	.0371919	4.33	0.000	.0882114	.2340008
Somali	.4948073	.0378873	13.06	0.000	.4205496	.5690651
Benishan	.1584485	.0367391	4.31	0.000	.0864411	.2304558
SNNPR	.0693166	.0379475	1.83	0.068	-.0050592	.1436924
Gambela	.1069185	.0401193	2.67	0.008	.0282862	.1855508
Harari	.1471236	.0452259	3.25	0.001	.0584825	.2357648
Addis Ababa	.033438	.0584974	0.57	0.568	-.0812148	.1480909
Dire Dawa	.1531248	.0470081	3.26	0.001	.0609905	.2452591
<b>Women education</b>						
Primary	-.1319696	.0207784	-6.35	0.000	-.1726945	-.0912446
Seconder	-.2033163	.0329904	-6.16	0.000	-.2679763	-.1386564
Higher	-.1614263	.0494043	-3.27	0.001	-.2582569	-.0645957
<b>Father education</b>						
Primary	.0068348	.0213436	0.32	0.749	-.0349979	.0486675

Secunder	-.0086532	.0290878	-0.30	0.766	-.0656642	.0483578
Higher	-.023608	.0311172	-0.76	0.448	-.0845965	.0373806
<b>Birth interval</b>						
2-3 year	-.027733	.0191616	-1.45	0.148	-.065289	.009823
>=4 year	-.0156013	.0239052	-0.65	0.514	-.0624545	.031252
<i>Zero-Inflation part model</i>						
<b>Income household</b>						
Medium	.174087	.1731242	0.89	0.373	-.1852302	.493404
Rich	.061104	.2713705	-0.23	0.822	-.5929804	.4707724
<b>Women age</b>						
20-24	.9175416	.3399065	2.70	0.007	.251337	1.583746
25-29	.0454648	.6280179	-0.07	0.042	1.076357	1.185428
30-34	.138255	.3693312	0.37	0.038	.5856209	.8621308
35-39	.5383697	.2671828	2.01	0.044	.014701	1.062038
≥ 40	.2949429	.2585451	1.14	0.024	.2117962	.8016821