DEBRE BERHAN UNIVERSITY

SCHOOL OF POST GRADUATE STUDIES

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF STATISTICS

DETERMENANTS OF INFANT DEATHS IN ETHIOPIA: APPLICATION OF COUNT REGRESSION MODELS

BY:

ZEYED FIKADE

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS

FEBRUARY 2023

DEBRE BIRHAN,

ETHIOPIA

DETERMENANTS OF INFANT DEATHS IN ETHIOPIA: APPLICATION OF COUNT REGRESSION MODELS

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS

BY:

ZEYED FIKADE

ADVISOR: GEBREMEDIHN DESTA (ASSI. PROF)

CO- ADVISOR: WERKU GIRMA (MSc)

FEBRUARY 2023

DEBRE BIRHAN,

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: Zeyed Fikade

____________ ____________

Signature Date

This thesis have been submitted for examination with approval as university advisors

Gebremedihn Desta (Assi. Prof)	____________	____________
Advisor	Signature	Date
Werku Girma (MSc)	____________	____________
Co-advisor	Signature	Date

APPROVAL SHEET

This is to certify that, the thesis prepared by Zeyed Fikade, entitled "**Determinants of infant deaths in Ethiopia: application of count regression models**" and submitted in partial fulfillment of the requirements for the Degree of Masters of Science in Statistics (Biostatistics) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Approved by the Board of Examiners:

____________________	____________________	____________
Name of chairperson	Signature	Date
___________________________	____________________	____________
Name of Advisor	Signature	Date
_____________________	____________________	____________
Name of Co-Advisor	Signature	Date
_____________________	____________________	____________
Name of External examiner	Signature	Date
_____________________	____________________	____________
Name of Internal examiner	Signature	Date
_____________________	____________________	____________
Associate Dean, College Res/CS& Postgraduate	Signature	Date
_____________________	____________________	____________
Dean collage of Graduate Studies	Signature	Date

Final approval and acceptance of the thesis is contingent up on the submission of the thesis of the final copy of the thesis to the college of graduate studies (CGS) through the department of graduate committee (DGC) of the candidate's

Stamp of CGS Date __________

ACKNOWLEDGMENTS

First, and foremost, I would like to thank the almighty God and Kidist dingle Mariam for giving me the opportunity to pursue my graduate study at Department of Statistics, Debre Berhan University.

Second my special thanks also go to Debre Birhan University for sponsoring me.

Finally, I would like to acknowledge my family, friends and all of my instructors in college Science of the department of statistics for all support in different ways to do this thesis for a good manner.

TABLE CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBRVIATIONS

ANC	Antenatal Care
EMDHS	Ethiopia Mini Demographic and Health Survey
EDHS	Ethiopia demographic health survey
EU	European Union
EPHI	Ethiopian Public Health Institute
GTP	Growth and Transformation Plan
IMR	Infant Mortality Rate
MoH	Ministry of Health
HP	Hurdle Poisson
HNB	Hurdle Negative Binomial
NBRM	Negative Binomial Regression model
SDG	Sustainable Development Goal
ZIPM	Zero Inflated Poisson Model
ZIP	Zero Inflated Poisson

ABSTRACT

Infant mortality is defined as the death of a child before the age of one, and it is quantified by the infant mortality rate (IMR). About 4.1 million children worldwide lost their lives in their first year of life in 2018. One of the sustainable development goals is to lower infant mortality to 12 deaths for every 1000 live births by 2030. Studying the factors that contribute to infant mortality is crucial for lowering the rate. Based on the 2019 Ethiopian demographic health census, this study sought to determine the determinant factor of infant mortality per mother. In this study the outcome variable is the number of infant deaths per mother. The survey (2019 EDHS) collected information from total of 8,855 women with the age of 15-49. Out of this we considered 5,679 women in this study which are gave live birth in their life time. Out of the considered women 1,191(20.98%) was experienced one and more infant deaths. Model compression was done using AIC, BIC, likelihood ratio test and Vuong test. HNB regression model was found to be the best model to fit the data. The result showed that the variable: - wealth index, preceding birth interval, Region, Place of delivery, Birth order and antenatal care visit had significant factors on infant mortality. The poorer wealth index (IRR=0.736; 95% CI: 0.548, 0.990), preceding birth interval ≥ 24 *month (IRR=0.744; 95%CI: 0.580, 0.955), infant who born in health facility (IRR=0.777; 95%CI: 0.614, 0.984), birth order (IRR=0.847; 95%CI: 0.792, 0.906), number of antenatal care 5-8 visit (IRR=0.685; 95% CI: 0.476, 0.986) were associated with reduced incidence of infant mortality controlling for other variables in the model. Whereas, being a resident of the Somali and Dire Dawa region (IRR=2.320; 95%CI: 1.022, 5.269) and (IRR=2.402; 95%CI: 1.024, 5.633) respectively were associated with an increased incidence of infant mortality. The implication of this study is that government and other stakeholders should be increasing access of health facility nearest to the community in all regions, encouraging utilization of antenatal care visit and encouraging mothers born in health facility to achieve sustainable development goals.*

Key words: *Infant mortality, Ethiopia, count model and Hurdle negative binomial Model*

1. INTRODUCTION

1.1. Background of the Study

Infant mortality, which is determined by the infant mortality rate (IMR), is the death of a child before the age of one (Weldearegawi et al., 2015). It describes any death that occurs following birth but before the infant turns one (Mulugeta et al., 2022a). The infant mortality rate (IMR) is the number of infant deaths per 1,000 live births. This is the probability that a child under the age of one will pass away for every 1000 live births. Because it is correlated with a number of variables, including maternal health, access to and quality of healthcare, socioeconomic situations, and public health policies, it serves as an indicator of the state of the nation's health. Compared to any other age group in the family, the spatial socioeconomic level of the family has a stronger impact on newborn survival. The leading causes of infant death around the world are infections, lower respiratory infections, diarrheal diseases, complications of preterm birth, neonatal encephalopathy (issues with brain function caused by lack of oxygen during birth), and infections. When compared to older newborns, infants that are just a few days old have different most common causes of death (Lozano et al., 2012).

Infant mortality rates around the world have declined, from an estimated rate of 65 deaths per 1000 live births in 1990 to 29 deaths per 1000 live births in 2018 (Patel et al., 2021). From 8.7 million in 1990 to 4.0 million in 2018, the number of infant deaths annually has decreased (Yu et al., 2022). With a 2.35% decrease from 2021, the global infant mortality rate currently stands at 26.693 deaths per 1000 live births. 2019 had over 14100 infant mortality in the European Union (EU). According to Ritchie and Roser, (2019), there were 6.38 deaths per 1,000 people in Latin American and Caribbean countries, which equates to 3.4 neonatal deaths for every 1000 live births (Herrera-Serna et al., 2019). In lower-middle and lower-income Asia-Pacific nations in 2018, there were 27.2 infant fatalities for per 1,000 live births, which is half the rate seen in 2000. 10 fatalities per 1000 live births were reported in upper-middle class Asia-Pacific nations, down from 18.2 in 2000. Geographically speaking, infant mortality was lower in eastern Asian nations and higher in south and south-east Asia. In 2018, there were no more than three deaths per 1,000 live births in Hong Kong, China, Japan, Singapore, and the Republic of Korea, compared to around six infant deaths per 100 live births in Pakistan (Organization, 2020).

In Africa, The risk of a child dying before his first birthday is (52 per 1000 live births), over six times higher than that in the European (8 per 1000 live births) that is very high death rate as compared to European(Lawn et al., 2014). In 2020, there were 41.6 fatalities for every 1,000 live births in Africa among infants under the age of one. In comparison to 2000, when about 81 newborn infants out of every thousand died before turning one year old, infant mortality across the continent has dramatically dropped (Mboya et al., 2020). The main risk factors for a high rate of infant deaths, particularly in Africa, are: a lack of resources and infrastructure, access to education, a shortage of medical personnel, poverty, and discrimination. Furthermore, the region has a higher prevalence of diseases that infants are especially susceptible to, including malaria, hypoxia, pneumonia, diarrhea, and other birth-related problems (Vakili et al., 2015).

Ethiopia has one of the highest rates of child mortality and morbidity in sub-Saharan Africa, with more than 704 children dying per day from diseases that may be avoided (Kuse et al., 2022). If lasting action is not taken, it is estimated that by 2030 more than 3,084,000 children would have died. In particular, there are 20 deaths per 1000 children who live to be 12 months old (Fenta and Fenta, 2020). Infant mortality was 48 deaths per 1,000 live births in the five years prior to the survey (EPHI, 2019). IMR has decreased less quickly than both child mortality rates for children under five and those who have passed their first birthday (Mulugeta et al., 2022a). In general, determining factors like maternal education, breastfeeding, and others increase infant mortality. The main causes of infant mortality in Ethiopia were examined in this study.

In recent years, Ethiopia health and family planning (EHFP) has successfully implemented in a wide area of fertility and mortality reduction interventions. Besides, the growth and transformation plan (GTP) has been developed and under implementation to improve access and quality of health services (BUZUNEH, 2019). 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 infant mortality, we have considered possible determinants of infant mortality by using count regression model. Therefore, this study explores the demographic factors, biological, environmental, socioeconomic characteristics of infant mortality in Ethiopia based on 2019 EDHS dataset.

1.2. Statement of Problem

Globally, 4.1 million children lost their lives in the first year of life in 2018. Before they reach their first birthday, more than a million youngsters in Africa alone pass away. These numbers correspond to roughly 2,808 deaths each day, or two deaths every minute (UNICEF, 2020). The Sustainable Development Goals (SDG) child mortality objective seeks to bring an end to preventable deaths of newborns and children under-five years of age by 2030, with all countries aiming to reduce newborn mortality to at least as low as 12 deaths per 1000 live births (Organization, 2019). Since the Millennium Development Goals (MDGs) were adopted, the countries of Sub-Saharan Africa have achieved incredible success and increases in infant survival, but the infant mortality rate in Sub-Saharan Africa is still the highest in the global region.

According to studies done from 56 different countries, infant mortality was one of the global public health concern and accounts for 144 deaths per 1000 live births. Overall, the collective global infant mortality rate has significantly decreased in recent decades, dropping from approximately 140 per 1,000 live births in 1950-55 to 52.8 in 2000 and on to 27.4 in 2020 but still infant mortality is world problem (Gampat, 2019). From 1990 to 2017, infant mortality in Sub-Saharan Africa decreased from 182 to 58 deaths per 1000 live births (Masquelier et al., 2021). However, half of all newborn mortality worldwide in 2017 occurred in sub-Saharan Africa. Ethiopia has one of the world's highest rates of infant mortality (Zimmerman et al., 2019). According to (EPHI, 2019) report, the infant mortality rates for the 5 years in the survey were 47 deaths per 1,000 live births. In other words, one in every twenty one children dies before they turn one. This infant mortality is substantially higher than the SDG targets of 12 deaths per 1000 live births. Moreover, infant mortality in Ethiopia is an important concern, where it is essential for monitoring the current health programs and formulating polices for improving the current situation.

A number of studies that have been carried out to investigate the determinants of infant mortality in different countries in the world especially, developing countries including Ethiopia (Baraki et al., 2020), (Kiross et al., 2021a), (Tesema et al., 2021b), (Muluye and Wencheko, 2012), (Terye, 2020)

With a limited variables and statistical methods such as Logistic regression, multilevel analysis and Survival cox-regression survival analysis. Also some of the previous Studies, tried to predict the number of infant deaths in Ethiopia (Mulugeta et al., 2022a) using multilevel log linear count model based on 2016 EDHS dataset. Now in Ethiopia, the Central Statistical Agency (CSA) has conducted new 2019 Ethiopia demographic health survey (EDHS). So far, there is no studies tire to predict the number of infant mortality in Ethiopia by using count regression models on 2019 EDHS dataset.

Moreover, the previous studies have investigated the determinant factors associated with maternal health, infant health and other socioeconomic status of mothers using 2016 EDHS data set. However, thus studies were not considering the factors like postnatal care and infant vaccination. But due to inaccessibility of health facility, weakness of postnatal care and lack of infant vaccination, infant death was high. Therefore, this study tries to investigate the major socio-economic, demographic, health and environmental proximate factor including postnatal care and infant vaccination that might influence infant mortality in Ethiopia. Generally this study will fill the gap and present new and available knowledge for different stakeholders.

1.3. Objective of the Study

1.3.1. General Objective

The general objective of this study was to identify the most important determinants factor that affects the number of infant mortality based on 2019 EDHS by using count regression model.

1.3.2 Specific Objective

- ❖ To examine the socio-economic, demographic, environmental and biological related factors associated with the infant mortality in Ethiopia.
- ❖ To determine the major factors of infant mortality and to assess the magnitude of infant mortality.
- ❖ To identify appropriate count regression models in order to analyze the number of infant mortality data in Ethiopia.

1.4. Significance of the Study

The findings from this study might be 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 might be the following significant

- ❖ In order to provide a better knowledge of the factors impacting infant mortality and a clear impact for the reduction of infant mortality in Ethiopia.
- ❖ The study will serve as a guide to stakeholders in making informed and intelligent policy decisions with regard to infant deaths and the management of the risk factors to avoid the death of infants in Ethiopia.
- ❖ The result can provide an important input for any possible intervention in this area for the future.
- ❖ This study can be baseline for the other interested finders (researchers) for further investigation on the infant mortality in Ethiopia.

1.5. Limitation of Study

In this study there were some challenges that we faced. The 2019 EDHS dataset lacked some crucial variables that might have an impact on infant mortality. The study used data from national surveys, which have inherent flaws like a lack of information on infants for deceased mothers, despite efforts to address them because only surviving women aged 15 to 49 were interviewed and some significant variables are not included because of a high number of missing values.

2. LITRATURE REVIEW

2.1. Concept of Infant Mortality

Infant mortality is the term used to describe infant deaths between the ages of one day and one year (Berrut et al., 2016). There are many causes of infant mortality, ranging from infections to accidents. Globally, the main causes of infant mortality are Neonatal encephalopathy, or problems with brain function after birth. Usually, birth injuries or a shortage of oxygen for the newborn cause neonatal encephalopathy. Illnesses, particularly those that affect the blood Preterm birth complications, diarrheal illnesses, infections of the lower respiratory tract (such as the flu and pneumonia) (Lozano et al., 2012).

A cording to global survey there were 2.5 million and 1.6 million deaths occurred in the first month of life and between ages 1–11 months respectively around the world (Baraki et al., 2020). According to survey conducted by the (UNICEF, WHO, WORLD BANK, UN DESA) the IMR varies from region to region: Africa is ranked highest at 52 per 1000 live births, Europe 8 per 1000 live births, South Asia 32 per 1000 live births, Latin America, Carrabin 14 per 1000 live births and sub-Sahara Africa 50 per 1000 live births.

2.2. International Literature on Infant Mortality

Empirically, many studies have shown that infant mortality is influenced by a number of socio-economic, demographic, nutritional, environmental and maternal health care seeking factors.

For instance, a study conducted to determine the relative significance of demographic and socioeconomic factors with regard to their role in lowering infant mortality in Egypt found that demographic factors have a greater impact on infant mortality than socioeconomic factors do. The study used Logit analyses of data from a nationally representative sample of Egyptian households and separately for urban and rural households (Aly, 1990).

A retrospective cohort study done based on secondary data of births and deaths of infants of mothers living in Brazil in 2011, the study used 207 infants. The study aims to identify, through linkage, factors associated with infant mortality using a Hierarchical logistic regression model. Where the result showed that, maternal age, maternal education, maternal occupation, type of birth and type of delivery are significant impact on infant mortality. The result for maternal age, Mothers aged 20–34 years, who had 2.82 times a higher chance of death compared with mothers aged 35 years or more. The results for maternal education, mothers with low education and high education group were 1.28 times stronger association with the infant mortality as compared with the group with intermediate education. Working mothers have a 2.03 times greater link with infant death than mothers who don't, which is similar to the association found for the variable "maternal education." But sex of infants did not have significant association with infant death (Santos et al., 2016).

The study conducted on infant mortality in Sweden by using Cox regression model, this finding included 13,741 infants, and the study revealed that there was significant association between Single marital status, high parity, LBW, low gestational age and male sex with infant mortality. Infants born to single mothers had 39% higher mortality than infants born to married mothers and Females had 22% lower mortality than males. Infants of the youngest mothers and mothers above age 35 years had the highest mortality. Also, the study shows that more than 50% higher post-neonatal mortality in the offspring of single mothers or 1.56 times higher compared with married mothers (Sovio et al., 2012).

Study conducted on infant mortality in Bangladesh used data from (BDHS-2014) by using both Bivariate and multivariate statistical analysis to investigate the effect of different factors on infant mortality. The result shows there was a significant relationship between birth size and infant mortality as 1.5 times higher risk of mortality in those newborns born with small and very small birth size. Additionally, the findings showed that female infants were less likely than male infants to have 10% mortality. Hence, infant mortality was statistically significantly with institutional delivery, antenatal care, birth size of newborn, child sex and wealth index of the household in this a positive relationship was seen between infant mortality (Vijay and Patel, 2020).

2.3. Literature on Infant Mortality in Ethiopia

Ethiopia is among the countries with the highest infant mortality with the rate of 47 deaths per 1000 live births. By comparing information from the EDHS conducted in 2005, 2011, 2016, and 2019, mortality patterns in Ethiopia can be analyzed. From 123 deaths per 1,000 live births in 2005 to 59 deaths per 1,000 live births in 2019, there was a 52% drop in under-5 mortality. Over the same period, infant mortality declined from 77 to 47 deaths per 1,000 live births, a 39% reduction. Neonatal mortality declined from 39 deaths per 1,000 live births in 2005 to 29 deaths per 1,000 live births in 2016 before increasing to 33 deaths per 1,000 births in 2019 (an overall reduction of 15% over the past 14 years) (EPHI, 2019).

2.3.1. Socio-economic Factors

Socioeconomic factors are the independent variables that affect the degree of illness and death through proximate causes. They can be grouped in to individual level, household level and community variable, socio-economic factors may affect, directly and indirectly, environmental, behavioral, nutritional and demographic risk factors with the exception of age and sex (Honwana and Melesse, 2017).

In worldwide surveys and research, the education of the mother is regularly employed as a stand-in indication of socioeconomic level. However, mother's education is also thought to be associated with hygiene, care seeking, and treatment of illness behavior's pertaining to infant mortality (Asefa et al., 2000).

The findings of a research on infant mortality in rural Ethiopia showed that, out of the 3684 infants tracked, 174 died before turning one. The result of study shows that Infants of mothers who attained a secondary school and above had 56 % lower risk of death (HR = 0.44, 95 % CI: 0.24, 0.81) compared to those whose mothers did not attend formal education. These shows mother's levels of education is highly associated with infant mortality and play an important role for the decreases of infant mortality, since better educated mothers exposed to information about better socioeconomic factors and about children health situations to decrease mortality and increase the ability to use health care resources and facilities but birth order and place of delivery was not significant (Weldearegawi et al., 2015).

In the multivariate cox analysis of infant mortality based on 2016 EDHS dataset, the covariates maternal level of education, preceding birth intervals, and sex of the child showed significant association with infant mortality but wealth index were not significant factor (Abate et al., 2020). Similarly, According to (Muluye and Wencheko, 2012) study was performed on the determinant of infant mortality using survival regression model, shows that mother's level of education is significant predictors of infant mortality. In this case, the infants born to mothers of these educational levels were 3.832 and 3 times higher than infants whose mothers had secondary and above education respectively.

Another study was conducted on EDHS− 2016 dataset to estimate the risk factors associated with infant mortality in Ethiopia by using multilevel ZINB model. the results of study revealed that mother's age, mother's age at first birth, birth order, wealth index and father education level are associated with infant mortality in Ethiopia (Mulugeta et al., 2022b). A study, (Muluye and Wencheko, 2012) indicated that infants belonging to the 5 and more birth order category were about 57% more likely to die relative to the single birth (HR=1.568, CI: 1.237-1.989). In this study wealth index of household were not significant factor. In a study done by Terye, (2020) in Ethiopia, it was indicated that Mother's education, Religion and Mother's wealth index are no significant association with infant mortality. Similarly, in a study done by Abate et al., (2020) and Baraki et al., (2020) in Ethiopia, it was indicated birth order was no significant association with infant mortality.

2.3.2. Environmental Factors

According to study (Kiross et al., 2021b) the Community-level characteristics (such as the ways of life in the regions in Ethiopia) were significantly associated with infant mortality. Infants living in pastoralist regions (Somali and Afar) were 1.4 times more likely to die in their first year compared with infants living in agrarian regions (Amhara; Tigray; Oromia; and Southern Nations, Nationalities and Peoples' Region [SNNPR]). A study, (Baraki et al., 2020) indicated that the rural residents were 1.76 times more likely to die in their first year compared with infants living in urban, from Somali, Harari and Diredawa region were 2.07, 2.14, 1.91 times more likely to die in their first year compared with infants living in other regions respectively.

A study conducted on EDHS− 2016 dataset to estimate the risk factors associated with infant mortality in Ethiopia by using multilevel ZINB model. The results of study show that place of residence is significantly associated with infant mortality in Ethiopia (Mulugeta et al., 2022b).

According to (Muluye and Wencheko, 2012) study was performed to statistical analysis on the determinant of infant mortality using survival regression model. The result of study shows Sources of water are statistically significant association with infant mortality: the risk of dying of infants born in households with access to unprotected water was higher by 47% relative to those born in households with access to water from pipes. The estimated risk of death for infants born in households with access to protected sources of water (wells, springs) was 37% higher compared to those born in households with access to pipe water.

2.3.3. Demographic Factors

The effect of demographic factors on health complex and is conditional by wide range of characteristics and behaviors. For example, maternal age, marital status, type of birth, birth interval. These factors have effect on infant mortality (Tesema et al., 2021a).

The study conducted on determinants of infant mortality by using Bivariable and multivariable multilevel logistic regression models. Results A total of 10,641 live-births from the EDHS data were included in the analysis. Being male infant and Multiple birth, were found to be statistically significantly associated with infant mortality (Baraki et al., 2020). Additionally, Girmay Tsegay Kiross stated that there is a statistically significant relation between infant mortality and individual-level factors like infant sex.

The result of study conducted on determinant factors of infant mortality using Multiple-Cox regression models. The result of a study shows that Infants born to mothers aged 15–19 years old had higher risk of death (HR = 2.68, 95 % CI: 1.74, 4.87) than those born between the ages of 25 and 29. But marital status, type of birth and sex of infant was not significant effect on infant mortality (Weldearegawi et al., 2015).

A study conducted on the determinants of infant mortality by using a mixed-effect logistic regression model in Ethiopia. According to a study's findings, the odds of infant mortality among infant born to mothers aged 20–29 years, 30–39 years, and 40–49 years were decreased by 37%, 48%, and 49% compared to those born to mothers aged < 20 years, respectively. Infants born to mothers having the preceding birth interval of < 24 months were 1.79 (95 CI: 1.46, 2.19, p= 0.05) times higher odds of death than the preceding birth interval of 24 months or above. The odds of infant death among twins were 4.25 (95% CI: 3.01, 6.01, p = 0.0001) times higher than those in single births. Infants who were male had a 1.50 higher chance of dying than infants who were female (Tesema et al., 2021a).

A study conducted on infant mortality in Ethiopia by using survival analysis was employed in order to show weather mother's age and sex of infant are significant predictors of infant mortality or not. In this case, Infants born to mothers of the age group of 15-19 died at a rate 32.3% higher than those born to the age group 35 years and above and infants born to mothers of the age bracket 20-34 died at a rate which was about 25% lower than those born to mothers in the age group 35 years and above. and also in these study sex is significant predictors of infant death (Muluye and Wencheko, 2012).

2.3.4. Health Care Seeking Factors

A study conducted on EDHS− 2016 dataset to investigate the role of maternal health on infant mortality in Ethiopia by using both multivariate logistic regression and hierarchical models. The result of a study shows that Antenatal care, family planning and postnatal care are significant association with infant mortality. In this study, mothers who have not received Antenatal service are 1.25 times higher infant mortality as compared to mothers who have received Antenatal service. similarly mothers who have not Postnatal check-up visits are 1.51 times higher infant mortality and mothers who have not use family planning are 0.48 times increased infant mortality as compared with mothers who use family planning (Terye, 2020). Similarly, (Tesema et al., 2021a) indicated that Infant mortality was 3.14 times more likely to occur in births from women who had no ANC visits during pregnancy compared to mothers who had > four ANC visits.

The result of study conducted on determinant factors of infant mortality using survival regression model. The mother's Breastfeeding is strongly linked to child survival. This study showing that Infants, who were not breastfed died at a rate which was about 4.362 times higher than infants who were breastfed and the risk of death for infants who were not breastfed could be 3.747 times as low and 5 times as large compared to those who were breastfed (Muluye and Wencheko, 2012). Similarly, (Mulugeta et al., 2022b) indicated that breast feeding and contraceptive use are associated with infant mortality in Ethiopia. According to study Kiross et al., (2021b) Infants of mothers who received ANC during the last pregnancy were 50% less likely to die in their first year of life compared with infants whose mothers did not receive ANC. In a study done by Abate et al., (2020) in Ethiopia, it was indicated that place of delivery and antenatal care visit were no significant association with infant mortality. Similarly, In a study done by Kiross et al., (2021b), it was indicated that place of delivery and postnatal care visit were no significant association with infant mortality also Mulugeta et al., (2022b); Baraki et al., (2020) indicates that place of delivery was not significant.

3. DATA AND METHDOLOGY

3.1. Source of Data

The source of data for this study is the 2019 EDHS which are obtained from CSA. It is the fifth significant survey with the aim of providing estimates for the relevant demographic and health variables. The 2019 Ethiopia Mini Demographic and Health Survey (EMDHS) is the country of Ethiopia's second iteration of the Mini Demographic and Health Survey. The survey was conducted from March 21, 2019, to June 28, 2019, based on a nationally representative sample that provided estimates at the national and regional levels and for urban and rural areas.

The survey used a two-stage stratified sampling technique. Each region was stratified into urban and rural areas, yielding 21 sampling strata. In each stratum, samples from the enumeration areas (EA) were chosen separately in two stages. A total of 305 EAs (212 in rural areas and 93 in urban areas) were chosen in the first stage, with probability proportional to EA size and independent selection in each sampling stratum. A household listing operation was carried out for all selected EAs. The generated list of households was used as a sampling frame for the second stage's selection of households. In the second step of the selection process, a specific number of 30 households in each group were chosen with an equal likelihood of systematic selection. The survey interviewed 8,855 women of reproductive age (age 15-49) from a nationally representative sample of 8,663 households. Finally, response on 5,679 women of age 15-49 who gave live birth were analyzed based on count regression model to determine the predictors of infant mortality.

3.2. Variables in the Study

In regression models there are two types of variables called outcome (dependent) and explanatory (independent) variables.

3.2.1. Response Variables

The response variable of the study, Y_i, is a count variable, the number of infant deaths per women of reproductive age (15-49) in Ethiopia, $Y_i = 0, 1, 2, \ldots$ where i refers to the i^{th} individual women.

3.2.2. Explanatory Variables

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

Table 3.1: Independent variables with their label and category

No	Variables	Category	Descriptions of variables
Demographic factors			
1	Sex	0= Male, 1= Female	Sex of infant
2	Marital status	0= Single, 1= Married 2= Widowed, 3= Divorced	Mothers Marital Status
3	Mother age	0= 15-19, 1= 20-24, 2= 25-29, 3= 30-34, 4= 35-39, 5= 40-44, 6= 45-49	Mother's age at the first birth
4	Birth order	-	Birth order
5	Preceding birth interval	0=1-24, 1= >24	Birth interval
Socio economic factors			
1	MEL	0=No education,1=Primary 2= Secondary, 3= Higher	Mother' Education Level
2	Religion	0= Orthodox, 1= Muslim, 4= Traditions 2= Protestant, 3 = Catholic, 5= Others	Religion
3	Wealth index	0= Poorest,1= Poorer, 2= Middle, 3= Richer, 4=Richest	House hold income
Health care seeking factors			
1	Place of delivery	0= Home, 1=Health centre	Place of delivery
2	ANC	0=No antenatal care visit, 1=1-4, 2=5-8, 3= >8	Antenatal service received by the mother
3	Breastfeeding	0=Yes, 1 No	Breastfeeding Status
4	Vaccination	1= Yes, 2= No	infant Vaccination
5	Postnatal care	0= No, 1= Yes	Postnatal check-up visits
Environmental factors			
1	Residence	1= Urban, 2= Rural	Place of residence
2	Region	1= Tigray, 2= Affar, 3= Amhara, 4= Oromiya, 5= Somali, 6=Benshangul-Gumuz, 7= SNNP. 8= Gambela, 9= Harari, 10= Addis Abeba 11= Dire Dawa	Region
3	SDW	0= Protected, 1= Unprotected	Source of drinking water
4	Toilet facility	0= Has toilet facility, 1= No toilet facility	Toilet facility

3.3. Methods of Data Analysis

3.3.1. Introduction on Count Data Model

Counts are non-negative integer numbers, such as 0, 1, 2, and 3, are permitted for the observations. These integers are produced through counting rather than ranking. The Poisson distribution serves as the basis for the creation of count data models. The variable of interest in this investigation is a count variable. When the response or dependent variable (number of infant deaths per woman in this study) is a count (which can take on non-negative integer values), 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 data on counts, the appropriate models are divided in to three parts count model for equal dispersion; Poisson Regression model, count model common case of over dispersion includes; the negative binomial regression model (over-dispersion), Zero-Inflated Count Models (excess zeroes); Zero inflated Poisson model, Zero-inflated Negative Binomial model, Zero- truncated count model and hurdle model. In the oldest time count variables are treated as continuous and linear regression model is applied. However, the linear regression model might not fit count data with a distribution with a positive skew effectively (Moghimbeigi et al., 2008).

3.4. Poisson Regression Model

In order to investigate the relationship between the count outcome variable and covariates, the Poisson regression model is frequently utilized. It is the baseline model for count data analysis. However, because of its limiting presumptions, it frequently fails in practical applications. There are two strong assumptions for Poisson model to be checked: one is the events that occur independently over time or exposure period, the other is the conditional mean and variance is equal. A Poisson regression model allows modeling the relationship between a Poisson distributed response variable and more than one explanatory variables (Hinde, 1982). It is suitable for modeling the number of events that occur in a given time period or area. In practice, counts have greater variance than the mean described as over-dispersion. This suggests that Poisson regression is inadequate. There are two common causes that can lead to over-dispersion: additional variation to the mean or heterogeneity, frequently used is the negative binomial model. And other cause counts with excess zeros or zero-inflated counts, since the excess zeros will give smaller mean than the true value, it can be modeled using zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB). Poisson Regression Model provides a standard framework in order to analyze count data. Let $Y_i$represent counts of events occurring in a given time or exposure periods or area with rateμ_i,Y_i are Poisson random variables with probability mass function (pmf) given below:

$$p(Y_i = y_i, \mu) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \qquad (3.1)$$

where y_i=0,1,2,3,and μ_i=1,2,3…where, Y denotes the number of infants per women is the rate parameter which is non-negative and it is given as:$E(y_i) = \mu_i = exp(X^T\beta)$ Where X^T=$\left(1, X_{1i}{}^T, X_{2i}{}^T \ldots.\right)$ And β=p+1 dimensional column vector of unknown parameters to be estimated and p is number of predictors. The estimation is undertaken by using maximum likelihood method. The first two moments of the Poisson random variable Y are E[Y] = μ and V[Y] = μ.If both are equal that means conditional mean is equal to conditional variance this shows the well-known equ-dispersion (equal mean and variance) property of the Poisson distribution.

Offset Variables

In the general Poisson regression model, we think of μ_i we expected number of infant death from the i^{th} mother and the total number of births ever from the i^{th} mother is N_i. This means parameter will depend on the population size and the total number of births ever from the individual mothers. Thus the distribution of y_i can be expressed as:

y_i~Poisson ($\mu_i N_i$), where $\mu_i = \exp(x_i{}^T\beta)$ and

N_i is the total number of ever born child from the individual mother

The logarithm of the children ever born is introduced when using the regression model as an offset variable. By including ln(children ever born) as offset in the question, it is differentiated from other coefficients when using the regression model by being carried through as a constant and forced to have a coefficient of one (Werner and Guven, 2007). Thus, the GLM with an offset is given by

$$\text{Log}\mu_i = \text{Log}N_i + x_i{}^T\beta$$

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 permits estimate of the 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 total number of births will be equal to the observed deaths if the coefficients of the independent variables, denoted by β, are all equal to zero. Since $\log N_i$ is 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. The model includes the offset variable. As a covariate with a parameter limit of 1.The answer variable can't also be the offset variable.

3.4.1. Estimation of Parameter of Poisson Regression Model

The Poisson regression model is a nonlinear regression model. It is derived from Poisson distribution by allowing the rate parameter μ dependent upon covariates. The most popular formulation is the log-linear regression model as given below:

Log $(\mu_i = X_i{}^T\beta_i)$ Where $X^T = (X_1, X_2, X_3, \dots . X_P)$ is vector of the explanatory variable and vector of unknown regression coefficient. The regression parameters are estimated utilizing the most likely estimate. Using the Poisson mode's likelihood function based on a sample of n independent observations is given by:

$$\ell\,(y, \beta) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i{}^{y_i}}{y_i!} \qquad (3.2)$$

The log-likelihood function of Poisson is $\ell = \text{Log}\left(\text{L}\,(\beta)\right) = \sum_{i=1}^{n}[y_i ln(\lambda_i) - (\lambda_i) - ln(y_i!)]$

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. As a result, we get at the first derivatives of ℓ with regard to the parameters β, which are as follows:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} (y_i - \lambda_i) x_{ij} \tag{3.3}$$

When subsequent events happen independently and at the same rate, the Poisson regression model is suitable for modeling count data. However, in reality, the characteristics of data frequently go against these presumptions. Usually, the variance of count data exceeds its mean, resulting in over-dispersion. Possibly because of unobserved heterogeneity, as the rate parameter is not only determined by a deterministic function of but also by a random (unobserved) component. The over-dispersion may also result from excess zeros, which happens when observed zeros are much more than that predicted by the assumed distribution(Rose et al., 2006). Moreover, over-dispersion will result in deflated standard errors of parameter estimates and therefore inflated t-statistics. Thus, it is always necessary to conduct a test of over dispersion after the development of Poisson regression.

If E(y_i) < Var(y_i) then we speak about over-dispersion, and when E(y_i) > Var(y_i) we say we have under-dispersion. Next, we employed two tests of over dispersion where the Null Hypothesis (HO) is: The response variable's mean and variance are equal against the Alternative Hypothesis ($H1$) :variance exceeds the mean. To check for over dispersion, two fundamental criteria are frequently applied:

Deviance,$D\left(y_i, \hat{\lambda}_i\right)$ is given by

$$D(\left(y_i, \hat{\lambda}_i\right) = 2 * \sum_{i=1}^{n} \left\{ y_i \ln\left(\frac{y_i}{\hat{\lambda}_i}\right) - \left(y_i - \hat{\lambda}_i\right) \right\} \tag{3.4}$$

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

Pearson chi-square test,x^2 also provided by

$$x^2 = \sum_{i=1}^{n} \left(\frac{\left(y_i - \hat{\lambda}_i\right)^2}{\hat{\lambda}_i} \right) \tag{3.5}$$

Over-dispersion may be a result of higher occurrence of zero counts and subject heterogeneity. Deviance and Pearson Chi-square statistic divided by levels of freedom are both roughly equal to one if the model correctly describes the data. Values greater than one indicate the variance is an over dispersion, while values smaller than one indicate an under-dispersion. There is a chance to account for over-dispersion with respect to the Poisson model by introducing a scale (dispersion) parameter into the correlation of the variance with the mean (Mitra and Washington, 2007). Another way of checking the presence of over-dispersion is a statistical test of the hypothesis:

$$H_0:\alpha=0 \; VS \; H_1:\alpha<0$$

If P-value of LRT $\alpha<$ (level of significance), then there is over-dispersion and It is preferred to use the Negative Binomial model. It is more appropriate for over-dispersed data because it relaxes the constraints of equal mean and variance.

3.5. Negative Binomial Regression Model

When over-dispersed count data are present, the NB Regression Model is applied (i.e. when the variance exceeds the mean). 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 problem of over-dispersion is addressed by the Negative Binomial Regression Model. Although it equally models zero and nonzero counts, this could lead to a poor fit for data containing a large number of zeros. Therefore, it is always necessary to check the proportion of zero counts before developing model of negative binomial regression.

This study used the likelihood ratio test to determine the more appropriate model between the Poisson Regression and Negative Binomial Regression. (Gardner et al., 1995) used Negative Binomial Regression to Model over dispersed Poisson data. When the Negative Binomial is used to model over-dispersed Poisson count data, the distribution can be thought of as an extension to the Poisson Model. Model for Negative Binomial Regression uses a log link function between the dependent variable (number of infant deaths per women) and independent variables. The only difference between the Poisson and the NB lies in their variances, regression coefficients tend to be similar across the two models, but standard errors can be very different. The NB regression model is:

A random variable $y_i, i = 1, 2, ...,$ is called a distribution with a negative binomial count with parameter λ and α the probability density function (α>0)

$$P(Y_i = y_i, \lambda_i, \alpha) = \frac{\Gamma\left(y_i + \left(\frac{1}{\alpha}\right)\right)}{y_i\,!\Gamma\left(\frac{1}{\alpha}\right)}(1 + \alpha\lambda_i)^{-\frac{1}{\alpha}}\left(1 + \frac{1}{\alpha\lambda_i}\right)^{-y_i} \qquad (3.6)$$

Then, $E(y_i) = \lambda_i = \exp(x_i{}^T\beta)$ and$var(y_i) = \mu_i(1 + \alpha\mu_i)$, Where, shows the level of over-dispersion and r(.) is the gamma function. If α =0 NB Regression Model will reduce to Poisson Regression Model. Often data will show over-dispersion (Variance > mean) or under-dispersion (Variance < mean). With over-dispersed data we may well use negative binomial regression model. This Model adds unobserved heterogeneity by specifying

$E(y_i) = \lambda_i = \exp(x_i{}^T\beta)$ Where $x_i{}^T$ is 1xp row vector of covariate (including an intercepts), p is the count of covariate in the model and px1 column vector of unknown regression parameters. The probability function of the NB model based on a sample of n independent observations is given by:

$$L(Y_i = y_i, \lambda_i, \alpha) = \prod_{i=1}^{n}\frac{\Gamma\left(y_i + \left(\frac{1}{\alpha}\right)\right)}{y_i\,!\Gamma\left(\frac{1}{\alpha}\right)}(1 + \alpha\lambda_i)^{-1/\alpha}(1 + 1/\alpha\lambda_i)^{-y_i} \qquad (3.7)$$

Similar to the Poisson model, the Newton-Raphson iteration process is employed to determine the regression coefficients and the dispersion parameter.

3.6. Zero-inflated Regression Models

In some instances, there are too many zeros in the count data, which is thought to be the result of over dispersion. In such a case, the NB model cannot be used to handle the over-dispersion which the result of 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 zero (Gurmu and Trivedi, 1996). Zero inflated count models provide a parsimonious yet powerful way to model this type of situation. These models presuppose that the data are the result of two distinct data creation processes, one of which only produces zeros and the other of which is either a Poisson or negative binomial data generation process Count Data that have an incidence of zeros greater a zero inflated distribution can be used to describe deviations from the underlying probability distribution. When modeling zero-inflated count data, zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression are widely utilized.

3.6.1. Zero-inflated Poisson Regression Model

Suppose the mean of the underlying Poisson distribution is μ and the probability of an observation being drawn from the constant distribution that always generates zeros is ω_i. The parameter ω_i is often called the zero inflation probability (Min and Agresti, 2002). In ZIP regression, the responses $Y_i = (Y_1, Y_2, \dots, Y_n)$ are independent distributed. One assumption of this model is that, with probability p, there can be no observation other than zero, and that, with probability (1 - p), a Poisson (λ) an observed random variable in Y. In order to explain the occurrence of extra zeros in the variable Y_i , The ZIP regression model is given by (Michael, 2014),

$$P\ (yi) = \begin{cases} \omega_i(1+\omega_i)e^{-\lambda \mathrm{i}}, & y_i = 0 \\ (1-\omega_i)\frac{e^{\lambda i}\lambda i^{yi}}{y_i!}, & yi = 1,2 \dots \end{cases} \quad 0 \le \omega i \le 1 \dots, \qquad (3.8)$$

Where, Y_i ~ ZIP (λ_i,ω_i)

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

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

The excess zeros are a type of over-dispersion, and they can be explained by fitting a zero inflated Poisson model, but there are also other sources of over-dispersion should be considered. Failure to account for sources of over-dispersion that cannot be attributed to the excess zeros results in a model misspecification, which leads to biased standard errors. In the case of the first subpopulation, the underlying Poisson distribution in ZIP Models is predicated on having a variance equal to the distribution's mean. The data show over-dispersion if this supposition is false (or under-dispersion) (Carruthers et al., 2008). The Pearson chi-square statistic, which is defined as, is a helpful diagnostic tool that can help identify over dispersion.

$$x^2 = \sum_{i=0}^{n} \frac{(y_i - \mu_i)^2}{V(\mu_i)} \qquad (3.9)$$

Comparing the computed Pearson chi-square statistic to an appropriate chi-squared distribution with n-p Df constitutes a test of over-dispersion. If over dispersion is detected, the ZINB Model often provides an adequate alternative.

3.6.1.1. Parameter Estimation of Zero-inflated Poisson Regression Model

The parameter λ_i and ω_i can be obtained by using the link functions,

Log $(\lambda_i)= \ x_i{}^T\beta$ and log $(\frac{\omega_i}{1-\omega_i}) = z_i{}^T\gamma,$ i= $1, 2, \dots$

Where, $x_i{}^T$ and $z_i{}^T$ are covariate matrices, β and γ are the (p+1)$\times$ 1 and (q+1)$\times$ 1 accordingly, unknown parameter vectors.

The ZIP model's likelihood function is provided by

$$\ell(\lambda, y) = \sum_{i=1}^{n} \begin{Bmatrix} \ln(\left[\omega_i + (1-\omega_i)e^{-\lambda_i}\right] I_{(y_i=0)} \\ +[\ln(1-\omega_i) - \lambda_i + y_i \ln(\lambda_i) - ln y_i!]_{(y_i>0)} \end{Bmatrix} \quad (3.10)$$

Where I (.) is a function that indicates whether an event is true or false, returning 1 in the former case and 0 in the latter. The Newton-Raphson approach is applicable to get the parameter estimates of ZIP regression models. Regarding and, the first derivatives are

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

$$\frac{\partial \ell}{\partial \gamma r} = \frac{\partial \ell\, \partial \omega_i}{\partial \omega \partial \gamma r_i} = \left\{ (y_i = 0) \left[\frac{-(1-\omega_i)\lambda_i e^{-\lambda_i}}{\omega_i + (1-\omega_i)e^{-\lambda_i}}\right] - (y_i > 0) \left[\frac{1}{(1-\omega_i)}\right] \right\} z_{ir}, \text{ r = 0, 1, 2.., q;}$$

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. This model provides away 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 (Moghimbeigi et al., 2008). The ZINB distribution is a mixture distribution, similar to ZIP distribution, where the probability ‘p’ for excess zeros and with probability (1- p) the remaining counts followed negative binomial distribution.

Note that Poisson distributions are mixed to create the negative binomial distribution which allows the Poisson, mean λ to be distributed as Gamma, and in this way over dispersion is modeled (Mwalili et al., 2008).

The ZINB regression the model presented by

$$P(y_i|\omega_i, \alpha, \lambda) = \begin{cases} \omega_i + (1-\omega_i)(1+\alpha\lambda_i)^{-\frac{1}{\alpha}}, & y_i = 0 \\ (1-\omega_i)\frac{\Gamma\left(yi+\left(\frac{1}{\alpha}\right)\right)}{y_i!\,\Gamma\left(\frac{1}{\alpha}\right)}(1+\alpha\lambda i)^{-\frac{1}{\alpha}}\left(1+\frac{1}{\alpha\lambda_i}\right)^{-y_i}, & y_i > 0 \end{cases} \quad (3.11)$$

Where, λ_i is the mean of the underlying negative binomial distribution, $\alpha > 0$ is the over dispersion parameter and is presumed not to be affected by factors and $0 \leq \omega_i \leq 1$. Also the parameters λ_i and ω_i depend on vectors of covariates xi and zi, respectively. The formulations for λi and ωi are the same as those used Poisson regression model with zero inflation. In this instance, the Y_i's the mean and variance are given by

$$E\,(Y_i) = (1 - \omega_i\,\lambda_i) \text{ And } Var\,(Y_i) = (1 - \omega_i\,\lambda_i\,(1 + \omega_i\,\lambda_i + \alpha\lambda_i \quad (3.12)$$

Considering that α = 0 and ω_i = 0, respectively, ZINB approaches ZIP and NB. ZINB reduces to Poisson when α and ω_i both equals 0. The parameter λ_i is modeled as a function of a linear predictor, that is,

$$\lambda_i = exp\,(x_i{}^T\beta)$$

Where, β is the $(p+1)\times 1$ vector of unknown parameters associated with the known covariate vector $x_i{}^T = (1, x_{i1}, \ldots, x_{ip}), p$ is the number of covariates not including the intercept. The parameter ωi, which is often referred as the zero-inflation is the likelihood of zero counts from the binary process. For common choice and simplicity, ω_i is characterized in terms of a logistic regression model by writing as

$logit\,(\omega_i) = \log\left(\frac{\omega_i}{1-\omega_i}\right) = Z_i{}^T\gamma$ Where, γ is the $(q+1)\times 1$ vector of zero-inflated coefficients to be estimated, associated with the known zero-inflation covariate vector $Z_i{}^T = (1, Z_{i1}, \ldots, z_{iq})$, where’$q$‘is the quantity of covariates without including the intercept. ZINB is also used to analyze exploratory data. When all the Variables are incorporated into 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 parameter value for which the probability of the observed data takes on its highest value is known as the parameter's maximum likelihood estimation. Even when the over dispersion parameter is known and standard GLM fitting techniques are not used, the ZINB distribution is not an exponential family distribution of the standard GLM type. The Newton-Raphson approach is applicable to get the parameter estimates of ZINB regression models.

The log likelihood function $\ell = \ell(\alpha, \lambda_i, \omega_i;\ y)$, for the ZINB model is given below

$$\ell = \ell(\alpha, \lambda_i, \omega_i;\ y) = \sum_{I=1}^{n} \left\{ I_{(y_i=0)} \log(\omega_i + (1-\omega_i) + (1+\alpha\lambda_i)^{-1/\lambda} + I_{(y_i>0)} \log(1-\omega_i) \frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{y_i!\Gamma\left(\frac{1}{\alpha}\right)} \left((1+\alpha\lambda_i) - \frac{1}{\alpha}\left(1+\frac{1}{\alpha\lambda_i}\right)^{-y_i} \right] \right\} \quad (3.13)$$

Since $\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{y_i!\Gamma\left(\frac{1}{\alpha}\right)} = \prod_{k=1}^{yi}(y_i + \frac{1}{\alpha} - k) = \alpha^{-yi}\prod_{k=1}^{yi}(\alpha y_i - \alpha k + k + 1)$

Furthermore, it can be expressed as

$$\ell = \sum_{I=1}^{n} \left\{ I_{(y_i=0)} \log(\omega_i + (1-\omega_i) + (1+\alpha\lambda_i)^{-1/\lambda} + I_{(yi>0)} \log(1-\omega_i) - \log(yi!) + \sum_{k=1}^{n}(\alpha y_i - \alpha k + k + 1) - \left(y_i + \frac{1}{\alpha}\right)\log(1+\alpha\mu_i) + \log y_i! + y_i \log\mu_i \right\} \quad (3.14)$$

Newton-Raphson iteration procedure can be applied to estimating the parameter of ZINB regression models.

3.7. Hurdle Regression Models

In a hurdle model, statistical models where a random variable is modeled using two parts, the first which is the probability of attaining value 0, also the second section models the probability of the non-zero values. The usage of hurdle models is frequently driven by a surplus of zeros in the data that cannot be adequately accounted for by more traditional statistical models. It was introduced by

(Mills, 2013), where a probit model was used to model the zeros and a normal model to model the non-zero values of x. The hurdle model got its name because the probit component of the model was supposed to simulate the existence of "hurdles" that must be surmounted in order for the values of x to reach non-zero values. Later, to construct hurdle models for count data, non-zero counts were modeled using Poisson, geometric, and negative binomial models.

The Hurdle model may be defined as a two part model where the first part is binary outcome model, and the second component is a truncated count model. As per (Cameron et al., 2004) such as a partition allows for the conclusion that successful observation results from passing through zero barriers or thresholds. 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 portions the population in to subpopulations in a meaningful way. So, a data set is split in to zero and non-zero (positive) values to fit two different models with associated covariates in regression.

There are many other probability distributions that can be taken into account for zero counts, and real-world data frequently use Poisson distribution, binomial distribution, and negative binomial distribution. Geometric distribution, Poisson distribution, and negative binomial distribution are other probability distributions that are frequently used for positive count data. An approach to model the excessive number of zero values and allow for over dispersion is using zero-inflated models and hurdle models.

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 and Trivedi, 2013). Suppose that $g_1(0)$ is the chance when the response variable's value is 0 and (k), where k = 1, 2... When the response variable is a positive integer, is a probability function.

Consequently, the hurdle-at-zero model's probability function is given by:

$$P(Y_i = \mathrm{k}) = \begin{cases} g_1(0) \text{for k} = 0 \\ (1 - g_1(0))g_2(\mathrm{k}) for\ k = 0,1, \end{cases} \quad (3.15)$$

(Mullahy, 1986) mentioned the hurdle-at-zero concept, and he believes that both of the components of the hurdle model are based on probability functions for non-negative integers such as f_1 and f_2. Using the general model presented above,

Let $g_1(0) = f_1(0)$ and g_2(k) = $\frac{f_2(k)}{(1- f_2(0))}$. In the situation g_2, normalization is necessary because f_2 is defined over the non-negative integers (k=0, 1, 2 …) While the backing of g_2 should exceed positive integers (k = 1, 2…). As a result, the probability function f_2 must be truncated. This is only a theoretical idea, thus truncation on f_2 does not always imply that the population is reduced in this area. All that is required is that we employ a distribution with positive support; the second hurdle model component may also utilize a displaced distribution or another distribution with positive support. The probability distribution of the hurdle at zero models is given by under the assumptions in (Mullahy, 1986) and is given by

$$f\ (y = 0) = f_1(0)$$

$$f\ (y = k) = \frac{f_1(0)}{(1-f_2(0))} f_2(k) = \theta f_2(k), \quad k = 1, 2 \ldots \tag{3.16}$$

Where f_2 to be known as percent-process. The numerator of θ shows the likelihood of succeeding, while the denominator provides normalization to take into account the (strictly technical) truncation of f_2 . It follows that if $f_1 = f_2$ or, equivalently, $\theta = 1$ then the hurdle model degrades to the parent model. The hurdle model's predicted value is given by

$$E(Y) = \theta \sum_{k=1}^{\infty} k f_2(k) \tag{3.17}$$

This expected value differs from the parent model's expected value due to the factor.θ. Additionally, the hurdle model's variance value is provided by

$$Var\ (Y) = \theta \sum_{k=1}^{\infty} k^2 f_2(k) - [\theta \sum_{k=1}^{\infty} k\ f_2(k)]^2 \tag{3.18}$$

If θ is larger than 1, the probability of overcoming the hurdle is greater than the percent model's total probability of favorable outcomes. The expected value of the model is therefore correlated with an increase in the expected value of the hurdle model. In contrast, if θ is smaller than1.

3.7.1. Hurdle Poisson Regression Model

The hurdle approach is adaptable and can address issues with both under- and over-dispersion. Gurmu, (1998) introduces a generalized hurdle model for the investigation of over dispersed or under dispersed count data. Greene (2005) compared the hurdle and zero-inflated models as two separate models.

In the case of count data, a Poisson model is commonly used. Given that the response variable contains there are many zeros frequently prevents the mean from being equal to the dependent variable's variance value. As a result, this type of data no longer fits the Poisson model. To solve the over dispersion issue, we advise employing a hurdle Poisson regression model.

We start with the binomial process, which determines whether the dependent variable takes on the value zero or a positive value and Probability mass function is given by

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

The probability mass formula for the zero-truncated Poisson process

$$P\,(Y = y/Y = 0) = \begin{cases} \frac{\lambda^y}{(e^{\lambda}-1)y!} & Y = 1, 2, 3 \\ 0 & otherwise \end{cases} \tag{3.19}$$

Thus, the unconditional likelihood mass function for Y is

$$P\,(Y = y) = \begin{cases} \pi, & y = 0 \\ (1-\pi)\frac{\lambda^y}{(e^{\lambda}-1)y!}, & y = 1, 2, 3, \ldots \end{cases} \tag{3.20}$$

Additionally, under the assumption that each observation has an independent distribution, the log likelihood for the i^{th} observation is

$$\text{Ln L}\,(\pi_i, \lambda_i, y_i) = = \begin{cases} \ln(\pi), & y = 0 \\ ln\left\{(1-\pi)\frac{\lambda^y}{(e^{\lambda}-1)y!}\right\}, & y = 1, 2, 3, \ldots \end{cases} \tag{3.21}$$

3.7.2. Hurdle Negative Binomial Regression Model

In many instances, the mean does not equal the variance of the dependent variable because the response variable contains a large number of zeros. To solve the over dispersion issue in this situation, we advise employing a hurdle negative binomial regression model. Let $Y_i(i = 1, 2, \ldots n)$ be anon negative integer-valued suppose a random variable Y_i = 0 is observed much more frequently than the standard model would predict. A hurdle negative binomial regression model is one that we take into consideration $Y_i(i = 1, 2, \ldots n)$ has the distribution

$$P(Y_i = y_i) = \begin{cases} \omega_0, \quad y_i = 0 \\ (1 - \omega_0) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})} \frac{(1+\alpha\mu_i)\alpha^{-1} - y_i\alpha^{y_i} \mu_i{}^{y_i}}{1-(1+\alpha\mu_i)-\alpha^{-1}} \quad , y_i > 0 \end{cases} \qquad (3.22)$$

3.8. Methods of Variable Selection

A variable selection method is a way of selecting a particular set of independent variables for use in a regression model. The goal is to choose the most effective subset of predictors. This selection possibly an effort to find a best model, or it possibly an effort to limit the number of independent variables when there are too many potential independent variables. The main objective of a variable selection procedure is to identify the correct predictor variables, which have an important influence regarding the response variable and could provide robust model prediction. There are a number of commonly used methods. These are forward selection, backward selection and stepwise selection.

3.8.1. Forward (step-up) Selection

This method is often used to provide an initial screening of the candidate variables when a large group of variables exists. It adds the most significant variable. Stop adding variables when none of the remaining variables are significant. Note that once a variable centers the model, it cannot be deleted.

3.8.2. Backward (Step-down) Selection

The backward selection model starts with all candidate variables in the model. At each step, the variable that is the last significant is removed. This process continues until no non-significant variables remain. The user decides at what level of relevance variables can be dropped from the model. This method is less popular because it begins with a model in which all candidate variables have been included.

3.8.3. Stepwise Selection

The forward selection method and the backward selection method are combined in stepwise regression. Stepwise regression, which was once highly popular, modifies forward selection so that all candidate variables in the model are tested to see if their significance has decreased below the designated tolerance threshold after each step in which a variable was included. A non-significant variable is eliminated from the model if it is discovered. Two significance levels are necessary for stepwise regression: one for adding variables and one for eliminating variables. In order to prevent the method from entering an infinite loop, the cutoff probability for adding variables should be lower than the cutoff probability for removing variables.

In this study we used Stepwise variable Selection which incorporates both forward selection and backward elimination to identify the predictors in the model. This was doing on Poisson regression model as it is the bench mark for other count regression models. Stepwise selection method address where added or removed with respect to p-value in the process.

3.9. 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 (LR), Akakie information criteria (AIC) and Bayesian information criteria (BIC). The most popular method for determining which model fits the data the best when comparing two or more models is the AIC. The following is the formula:

$AIC = -2\ell + 2k$, Where ℓ is the model's log-likelihood, which will be used to compare it to other models, and "k" is how many parameters there are in the model, including the intercept. The Bayesian information criterion (BIC) takes the volume of the data into account.

BIC is given by,

$$BIC = -2\ell + k\,log(n) \qquad (3.23)$$

Where ℓ 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.

3.9.1. Tests for the Comparison of the Models

3.9.1.1. Tests for Comparison of Nested Models

1. Likelihood Ratio Test (LRT)

The likelihood-ratio test is useful for evaluating 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 successfully fit 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 according to the negative binomial regression).The likelihood-ratio test is given by:

$$G^2 = -2(\ell_{null} - \ell k) \sim \chi^2 p - 1 \qquad (3.24)$$

Where: ℓ_{null} is the log-likelihood of the null model and ℓ_k is the log-likelihood of the full model comprising k predictors, is number of parameters and χ^2 (p −1) is a chi-square distribution that includes p −1 degree of freedom. When the test statistic is exceeds the critical value, rejecting the null hypothesis. The statistic of likelihood ratio test for is given by the following equation:

$$\text{LRT}\alpha = -2(LL_1 - \text{LL}_2) \qquad (3.25)$$

This statistic has a Chi-squared distribution with only one degree of freedom and LL is log-likelihood. If the statistic higher than the critical value then, the model 2 is better than the model.

3.9.1.2. Test for Comparison of Non-nested Models

1. Vuong Test

The Vuong testis anon-nested test that is founded on a comparison of the estimated 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. For model comparison, this test is applied. For testing the relevance of using zero-inflated models versus Poisson and NB regression models, the Vuong statistic is used. Let's define:

$$M_i = log\left(\frac{p1(Y_i/X_i)}{p2(Y_i/X_i)}\right) \tag{3.26}$$

Where, $P1\ (Y_i\ /X_i)$ and $p2(Y_i/X_i)$ are probability mass functions of zero-inflated and Poisson or NB models, respectively. In general, $(Yi\ /X)$ is the predicted chances of observed count for case$i$ from model N, then the Vuong test statistic is simply the average log-likelihood ratio suitably normalized.

There is a test statistic of

$$V = \sqrt{n}\frac{\frac{\sum_{i=1}^{n} mi}{n}}{\sqrt{\sum_{i=1}^{n}(mi-\bar{m})^2}} = \frac{\sqrt{n}}{sm}(\bar{m}) \tag{3.27}$$

The hypotheses of the Vuong test are:

$$Ho: [mi] = 0 \quad vs \quad H1: [mi] \neq 0$$

The two models are equivalent, and this is the test's null hypothesis. Asymptotically, Vuong demonstrated that the distribution is of the standard kind (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

3.9.2. Information Criteria (AIC and BIC)

The goodness of fit standard the best model is chosen using AIC and BIC. The Poisson model and the NB model were compared using the likelihood ratio test. Numerous Monte-Carlo simulations demonstrate the necessity of combining the AIC and BIC selection criteria. It is better to use a model's lowest AIC or BIC value. A common likelihood information criterion is frequently used to choose the best model. Consider the Akakie information criterion. Alternatively, the Bayesian information criterion, denoted by the initials AIC and BIC, respectively.

AIC = -2log likelihood + 2k

BIC = -2log likelihood + k log (n), Where k = number of parameters and n = number of observations.

3.9.3. Test for Individual Predictors

Let β denote an arbitrary parameter. Consider a significance test of $H0$: β=0. The simplest test statistic uses the large-sample normality of the ML estimatorβ, let SE (β) denote the typical error of β, evaluated by substituting the ML estimate for the unknown parameter in the expression for the true standard error. When $H0$ is true, the test statistics

$$Z = \frac{\hat{\beta} - \beta o}{SE(\hat{\beta})} \tag{3.28}$$

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

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

Z^2 is Assuming a chi-square distribution with one degree of freedom when H_0 is true. For tiny samples, use Wald statistics. Most people agree that likelihood-ratio tests are superior.

3.10. Statistical Software and Packages

In this study, R version 4.2.2, statistical software is used for statistical analysis and graphics.

4. STATISTICAL DATA ANALYSIS

4.1. Result of Descriptive Statistics

We perform a descriptive analysis on the data in order to get a broad sense of the distribution of the number of infant deaths before moving on to design a suitable count model. So, we start by describing the response variables.

4.1.1. Number of Infant Mortality per Mother in Ethiopia

The data to be analyzed for this study were obtained from Ethiopian Demographic and Health Survey (EDHS) 2019. This study includes total of 8,855 women in the age of 15-49. From the total 5,679 women who gave live birth in their life time. The result showed that, descriptive statistics of number and percentage of infant deaths per mothers. Based on information from 5,679 mother's 4,488(79.03%) of the mothers have not experienced any infant death in their life time and 802(14.12%), 278(4.90%), 81(1.43%) and 30(0.53%) of mothers lost 1, 2, 3 and ≥ 4 of their infants respectively. which indicates excess zero and les percentage of non-zero counts.

Table 4.1: Frequency distribution of number of infant deaths in Ethiopia

number of infants death per mother	Freq.	Percent
0	4,488	79.03
1	802	14.12
2	278	4.90
3	81	1.43
≥4	30	0.53
Total	5,679	100.00
Mean	0.306	
Variance	0.485	

Further screening of the number of infant death calculated showed that the variance (0.485) is greater than the mean (0.306) indicating there is an over-dispersion and hence the standard Poisson regression model was not appropriate to fit infant deaths data. This is an indication that the data could be fitted better by count data models which takes into account excess zeroes (79.03%).

As Figure 4.1 showed, there are massive counts of zero outcomes, the histogram are highly picked at the beginning (zero values). This leads to have a positively (right) skewed distribution. This is an indication that the data could be fitted better by count data models which takes in to account excess zeros models.

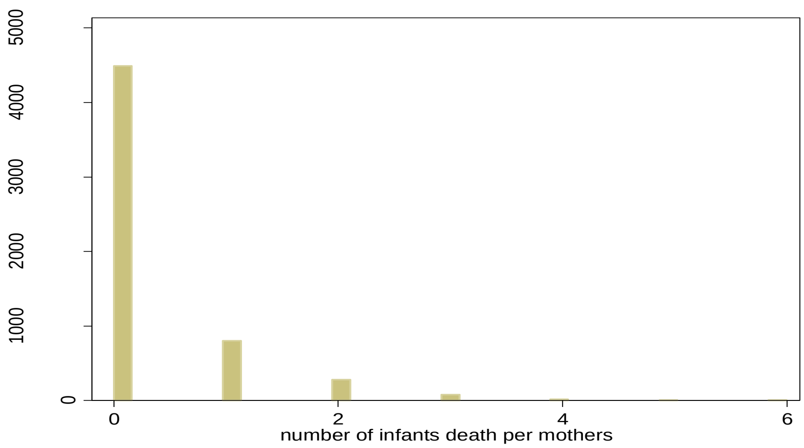

Figure 4.1: histogram of number of infant deaths per mother

4.1.2. Summary Statistics for Explanatory Variables

Socioeconomic, demographic, health and environmental related factors to the number of infant death per mother are summarized in Table 4.2. The highest mean number of infant death per mother was occurred in Benishangul (0.435) and Dire Dawa (0.432) whereas, the lowest mean number of infant death was occurred in Tigray (0.130) and Addis Abeba 0.132). Regardless of place of residence highest mean number of infant death per mother was occurred in Rural (0.307) as compared to urban areas (0.301).

Table 4.2 revealed that the mean number of infant death for Poorest, Poorer, Middle and Richer income level were 0.388, 0.314, 0.296 and 0.255, respectively, whereas, the mean number of infant death per mother for Richest was 0.142. Therefore, mothers living in better and standard economic situation experienced to have less number of infant deaths as compared with mothers who have low income level. From the result it can be observed that the mean number of infant death for infants who born at home (0.320) is higher than born at health facility (0.291).

The result also indicated that mothers who had no postnatal checkup have highest mean number of infant death (0.325) as compared to mothers who had postnatal checkup (0.285). Similarly, mothers who had no antenatal care visit have highest mean number of infant death per mother (0.359) as compared to mothers who had one or more antenatal care visits.

From the result can also observe that highest mean number of infant death per mother was occurred with families who did not toilet facility (0.335) as compared to families have toilet facility (0.281). According to preceding birth interval, the highest mean number of infant death is occurred infants born less than or equal to 24 months (0.328) and the lowest mean number of infant death per mother occurred infant born greater than or equal to 24 months (0.296). Also the descriptive results show that the mean number of infants Birth orders is approximately 4.

Table 4.2: Summery statistics of predictor variables related to infant death in Ethiopia

Variable	Category	number of infant death per mother		
		Observations	Mean	Std. Dev.
Region	Tigray	438	0.130	0.399
	Afar	665	0.208	0.572
	Amhara	522	0.297	0.660
	Oromia	712	0.396	0.763
	Somali	623	0.424	0.854
	Benishangul	538	0.435	0.794
	SNNPR	654	0.287	0.650
	Gambela	446	0.294	0.738
	Harari	435	0.207	0.524
	Addis Abeba	266	0.132	0.418
	Dire Dawa	380	0.432	0 .874
Place of residence	Urban	1,255	0.301	0.699
	Rural	4,424	0.307	0.696
Mother education level	No education	3,129	0.318	0.719
	Primary	1,796	0.291	0.656
	Secondary	468	0.312	0 .761
	Higher	286	0.255	0.581

Wealth index	Poorest Poorer Middle Richer Richest	1,717 1,795 835 628 704	0.388 0.314 0.296 0.255 0.142	0.787 0.685 0.706 0.641 0.455
breastfeeding	No Yes	2,247 3,432	0.313 0.302	0.689 0.701
Sex of infant	Male Female	2,934 2,745	0.305 0.306	0.698 0.695
Place of delivery	Home Health facility	2,804 2,875	0.320 0.291	0.735 0.657
Postnatal care	No Yes	3,047 2,632	0.325 0.285	0.715 0.674
No of antenatal care visit	No ANC 1-4 ANC 5-8 ANC >8 ANC	2,152 1,907 1,149 471	0.359 0.275 0.287 0.234	0.732 0.683 0.645 0.697
Source of drinking water	Protected Unprotected	2,822 2,857	0.304 0.308	0.701 0.692
Toilet facility	Has toile facility No toilet facility	3,007 2,672	0.281 0.335	0.677 0.717
Preceding birth interval	1-24 >24	1,847 3,832	0.328 0.296	0.752 0.668
Marital status	Single Married Widowed Divorced	32 5,329 57 261	0.406 0.307 0.193 0.295	0.875 0.696 0.515 0.724
Infant vaccination	No Yes	3,016 2,663	0.307 0.305	0.697 0.696
age of mothers	15-19 20-24 25-29 30-34 35-39 40-44 45-49	383 1,184 1,814 1,168 725 305 100	0.413 0.316 0.280 0.316 0.286 0.282 0.360	0.791 0.717 0.642 0.735 0.656 0.711 0.772

Religion	Orthodox	1,585	0.257	0.629
	Catholic	32	0.406	0.875
	Protestant	1,029	0.312	0.697
	Muslim	2,951	0.332	0.733
	Tradition	62	0.209	0.410
	Others	20	0.100	0.308

4.2. Count Regression Model Result for Infant Mortality

4.2.1. Variable Selection Method

In order to select the best subset of variables to include in multivariable analysis, Stepwise variable selection was used. The result recognized that predictors, with respect to the P-value in the process: such as wealth index, Toilet facility, preceding birth interval, Region, Place of delivery, Place of residence, Birth order, Postnatal care and antenatal care visit had statistically significant variable. Hence these significant variables are considered in the multivariable count regression models.

4.2.2. Goodness of fit and Test of Over dispersion

The result in table 4.3, 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. The result show that the ratio of the deviance and Pearson Chi-square statistic to their corresponding degrees of freedom are greater than one, indicating an existence of over dispersion in the data and the negative binomial model more appropriate than the Poisson model.

Table 4.3: The result of over-dispersion test

Statistics	Value	Df	Value/Df	P-value
Deviance test statistic	5910.947	5655	1.05	0.0000
Pearson Chi-square statistic	11988.77	5655	2.12	0.0000

4.2.3. Model Selection Criteria

4.2.3.1. Information Criteria's

In order to select the best model which fits the data well, six different models were considered, namely; Poisson, Negative binomial, Zero-inflated Poisson, Zero-inflated negative binomial, Hurdle Poisson and Hurdle negative binomial models. In this study, different model selection criteria were considered like the log-likelihood, AIC and BIC in order to identify the most fitted model.

The result in table 4.4 showed that, HNB model is the most appropriate than the other count models to fit number of infant death per mother. Due to the fact that HNB has a lower AIC (7893.771) and BIC (8219.353) value and a higher Log-likelihood (-3897.886) value.

Table 4.4: model selection criteria for the count regression models

Model	Def.	Selection criteria		
		AIC	BIC	Log-likelihood
Poisson	24	8633.293	8792.762	-4292.647
NB	25	8140.157	8306.271	-4045.079
ZIP	48	8145.218	8464.156	-4024.609
ZINB	49	8006.922	8332.504	-3954.461
HP	48	7951.485	8270.423	-3927.743
HNB	49	**7893.771**	**8219.353**	**-3897.886**

4.2.3.2. Likelihood Ratio Test (LRT)

The result in table 4.5 showed that, with p_{value} =0.0001 it is smaller than α-value it implies that NB is better than the Poisson model, ZINB is better than zero inflated Poisson model and HNB model is better than hurdle Poisson regression model. Therefore, HNB model is the most appropriate model than the other count models to fit number of infant death per mother. The AIC, BIC and log likelihood also supported HNB model from the others count model.

Table 4.5: Likelihood ratio test for nested models

Model	LRT test statistic(p-value)	Preferable model
NB versus Poisson	0.0001	NB
ZINB versus ZIP	0.0001	ZINB
HNB versus HP	0.0001	HNB

4.2.3.3. Vuong Test

To compare the performance of each model we use Vuong test as the non-nested. If models are non-nested like: ZIP vs. Poisson, ZINB vs. NB, HP vs. ZIP and HNB vs. ZINB regression models were identified using the Vuong test statistic. The first comparison is made between the ZIP model and Poisson model, with a Vuong test statistic of 8.899 is greater than 1.96 and p-value < 0.05, implying that the ZIP model preferred to the Poisson model for predicting the number of infant death per mother. The Vuong test statistic for the ZINB versus NB (4.069 is greater than 1.96 and p-value < 0.05) it implying that the ZINB model is preferred and the Vuong statistics for HP versus ZIP (5.491 > 1.96, P-value = 0.0001) from this result HP model is preferred to ZIP.

After a series test of model comparison as shown in table 4.6 the calculate value of Vuong statistic for comparing HNB versus ZINB model is 2.544 (2.544> 1.96, p-value = 0.0054) this indicated that HNB model is preferred to ZINB regression model. Thus we might select HNB model. The AIC, BIC and log likelihood were also supported to fit infant death.

Table 4.6: Vuong test for non-nested models

Model comparison	Vuong test statistic	p-value	Preferable model
ZIP vs. Poisson	8.899	0.0001	ZIP
ZINB vs. NB	4.069	0.0001	ZINB
HP vs. ZIP	5.491	0.0001	HP
HNB vs. ZINB	2.544	0.0054	HNB

4.2.3.4. Predicted and Observed Value

The result showed that the Poisson model was under-estimated zero counts while, the NB and zero inflated Poisson model over-estimated zero counts and the ZINB and hurdles captured all zero values. Based on predicted outcomes, the differences in model fit between the six models are remarkable. Still the standard Poisson model and the NB model do not fit the data reasonably well. From table 4.7 result we conclude that HNB model is a better choice than the other count models, since the predicted zero count for HNB model is closed to the observed zero count. Therefore, it is possible to conclude that the HNB model is more appropriate than the other count model to fit the number of infant death per mother.

Table 4.7: Zero count capturing in count model

	Observed	Poisson	NB	ZIP	ZINB	HP	HNB
Number of zeros	4488	4250.3	4511.486	4488.608	4520.735	4488	4488

From table 4.8 result we conclude that HNB model is a better choice than the other count models, since the predicted probability for HNB model is closed to the observed probability. Therefore, it is possible to conclude that the HNB model is more appropriate than the other count model to fit the number of infant death per mother.

Table 4.8: Observed and predicted probability from count model for infant death

Number of infant death per mother	Observed probability	Predicted probability					
		Poisson	NB	ZIP	ZINB	HP	HNB
0	0.7903	0.7484	0.7944	0.7904	0.7960	0.7903	0.7903
1	0.1412	0.2059	0.1358	0.1372	0.1328	0.1397	0.1412
2	0.0489	0.0383	0.0413	0.0494	0.0416	0.0496	0.0400
3	0.0143	0.0062	0.0155	0.0160	0.0159	0.0149	0.0139
4	0.0033	0.0010	0.0066	0.0049	0.0068	0.0041	0.0055
5	0.0009	0.0002	0.0030	0.0015	0.0032	0.0012	0.0024
6	0.0011	0.0003	0.0015	0.0004	0.0016	0.0003	0.0011

4.3. Parameter Estimation of HNB Model for Infant Mortality in Ethiopia

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 unite increase in a unit's of the significant predictors, the log number of infant 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. To interpret the count data we used the incidence rate ratios ($IRR= exp^{coef}$).

According to table 4.9 the result showed that truncated HNB model variable like: - wealth index, preceding birth interval, Region, Place of delivery, Birth order and antenatal care visit were significant factor on number of infant death per mother but Toilet facility, Place of residence and Postnatal care were not significant factor on number of infant death per mother. Moreover, the zero inflation part analysis showed that variables such as wealth index, Region, Toilet facility, Postnatal care, Place of residence and antenatal care visit are significantly associated with the experience of infant death.

4.3.1. Result of HNB Regression Model for Count Part of Model

The result in table 4.9 showed that, Region has significant factors on number of infant deaths in the non-zero group. The expected number of infant deaths for mothers from Somali and Dire Dawa region were increased by a factor of 2.320, 2.402 as compared to the expected number of infant deaths for mother in Tigray, controlling for the other variables in the model.

From table 4.9, wealth index of the household has statistically significant influence to reducing the number of infant death per mother. The expected numbers of infant deaths for mother in the poorer household were decreased by a factor of 0.736 as compared to the expected number of infant deaths for mother in the poorest households, while holding all other variables in the model constant.

The finding of this study also revealed that the estimated coefficient of birth order is negative and had significant effect with infant death per mother. That means the expected number infant death decrease by a factor of 0.847 for every one unit increase infant birth order, holding other variable constant in the model.

The result also show that, the number of antenatal care visits of mothers 5-8 ANC is statistically significant with the infant death per mother that expected number of infant death decrease by 0.685 when compared to mothers who are no antenatal care visit keeping other variables constant in the model.

In this study delivery place has significant factor on the number of infant mortality. The expected numbers of infant deaths for infant who are born in health facility was decreased by a factor of 0.777 as compared infants who born in home keeping other variables constant in the model.

Finally, preceding birth interval has significant factor on the number of infant mortality. The expected number of infant deaths for those infant born with preceding birth interval more than 24 months had decreased by 0.744, as compared to infant born with preceding birth interval less than 24 months keeping other variables constant in the model.

Table 4.9: parameter estimation of HNB for count part of infant death in Ethiopia

Count part(non-zero part)							
	Estimate	Std. Error	z value	p-value	IRR	95% Conf. Interval	
Intercept	-2.007	0.579	-3.464	0.001	0.134	0.043	0.418
Region(Tigray)							
Afar	0.284	0.446	0.638	0.524	1.329	0.555	3.182
Amhara	0.477	0.440	1.085	0.278	1.611	0.680	3.817
Oromia	0.552	0.415	1.330	0.183	1.736	0.770	3.916
Somali	0.842	0.418	2.011	0.044	2.320	1.022	5.269
Benishangul	0.784	0.426	1.842	0.065	2.191	0.951	5.048
SNNPR	0.308	0.432	0.714	0.475	1.361	0.584	3.176
Gambela	0.413	0.446	0.926	0.354	1.511	0.630	3.623
Harari	-0.009	0.485	-0.019	0.985	0.991	0.383	2.561
Addis Abeba	0.069	0.625	0.111	0.912	1.072	0.315	3.650

Dire Dawa	0.876	0.435	2.015	0.043	2.402	1.024 5.633
Wealth index (Poorest)						
Poorer	-0.306	0.151	-2.029	0.042	0.736	0.548 0.990
Middle	-0.108	0.186	-0.581	0.561	0.897	0.623 1.293
Richer	0.024	0.225	0.108	0.914	1.025	0.659 1.594
Richest	0.144	0.304	0.474	0.636	1.155	0.636 2.098
Toilet facility (Has toile facility)						
No toilet facility	0.035	0.121	0.293	0.770	1.036	0.817 1.314
Birth order	-0.166	0.034	-4.837	0.001	0.847	0.792 0.906
Preceding birth interval (1-24)						
>24	-0.295	0.127	-2.322	0.020	0.744	0.580 0.955
No of antenatal care visit (No ANC)						
1-4 ANC	0.143	0.157	0.911	0.362	1.154	0.848 1.570
5-8 ANC	-0.378	0.186	-2.038	0.042	0.685	0.476 0.986
>8 ANC	0.241	0.268	0.899	0.369	1.272	0.753 2.149
Place of delivery (Home)						
Health facility	-0.252	0.120	-2.094	0.036	0.777	0.614 0.984
postnatal care (No)						
Yes	0.087	0.122	0.709	0.479	1.091	0.858 1.386
Place of residence (Urban)						
Rural	0.010	0.169	0.057	0.955	1.010	0.725 1.406

4.3.2. Result of HNB Regression Model for Covariate of Zero Counts

The covariates zero count of HNB model indicates that Region, wealth index, Toilet facility, No of antenatal care visit, Postnatal care and Place of residence have a significant effect on the probability of being in the always zero group.

From table 4.10 the odds of no occurrence of infant deaths (being in the always zero group) for infant born in Afar, Amhara, Oromia, Somali Benishangul, SNNPR, Gambela, Harari and Dire Dawa increased by a factor of 1.424, 2.212, 2.976, 2.782, 3.490, 2.373, 2.288, 1.922 and 3.559 respectively, as compared to reference category (Tigray) considering other variables constant in the model, Similarly, wealth index has a significant effect on probability of being an excess zero. The odds of being in the zero groups are decreased by a factor of 0.695, 0.586 and 0.311 for infants born from Middle, Richer and Richest mothers as compared infant born from poorest mothers controlling for other variables in the model.

The odds of being in zero group infant death increased by a factor of 1.253 mothers who have no toilet facility as compared to mothers who have toilet facility considering other variables constant in the model. Similarly, the odds of being in zero group infant death number of antenatal care visit: 1-4 ANC, 5-8 ANC and >8 ANC decreased by the factor of 0.744, 0.786 and 0.502 respectively as compared to no antenatal care visit considering other variable constant in the model.

The odds of being in zero group infant death for mothers who have postnatal check-up decreased by the factor of 0.874 as compared to mothers who have no postnatal check-up considering other variable constant in the model.

The odds of being in zero group infant death for infants born in rural areas decreased by a factor of 0.661 as compared to infants born in urban considering other variable constant in the model.

Table 4.10: parameter estimation of HNB for Zero-inflation part of infant death in Ethiopia

	Estimate	Std. Error	z value	p-value	IRR	95% Conf. Interval
Zero inflation part						

Intercept	-1.445	0.209	-6.910	0.000	0.236	0.156 0.355
Region(Tigray)						
Afar	0.354	0.195	1.812	0.070	1.424	0.972 2.088
Amhara	0.794	0.191	4.159	0.000	2.212	1.522 3.217
Oromia	1.090	0.180	6.065	0.000	2.976	2.092 4.233
Somali	1.023	0.183	5.593	0.000	2.782	1.944 3.981
Benishangul	1.250	0.185	6.757	0.000	3.490	2.429 5.015
SNNPR	0.864	0.187	4.615	0.000	2.373	1.644 3.424
Gambela	0.828	0.200	4.139	0.000	2.288	1.546 3.386
Harari	0.653	0.212	3.082	0.002	1.922	1.269 2.911
Addis Abeba	0.116	0.271	0.427	0.669	1.123	0.660 1.911
Dire Dawa	1.270	0.202	6.300	0.000	3.559	2.398 5.283
Wealth index (Poorest)						
Poorer	-0.157	0.081	-1.943	0.052	0.854	0.729 1.001
Middle	-0.363	0.105	-3.456	0.001	0.695	0.566 0.854
Richer	-0.535	0.122	-4.393	0.000	0.586	0.461 0.744
Richest	-1.168	0.149	-7.822	0.000	0.311	0.232 0.417
Toilet facility (Has toile facility)						
No toilet facility	0.225	0.067	3.369	0.001	1.253	1.099 1.429
Birth order	0.003	0.015	0.219	0.827	1.003	0.974 1.033
Preceding birth interval (1-24)						
>24	0.039	0.071	0.546	0.585	1.040	0.904 1.196
No of antenatal care visit (No ANC)						
1-4 ANC	-0.296	0.087	-3.380	0.001	0.744	0.627 0.883
5-8 ANC	-0.241	0.096	-2.509	0.012	0.786	0.651 0.949
>8 ANC	-0.689	0.141	-4.891	0.000	0.502	0.381 0.662

Place of delivery (Home) Health facility	-0.015	0.069	-0.229	0.819	0.985	0.864 1.123
postnatal care (No) Yes	-0.134	0.067	-1.989	0.047	0.874	0.766 0.998
Place of residence (Urban) Rural	-0.414	0.093	-4.436	0.000	0.661	0.550 0.794

5. DISCUSSION, CONCLUSION AND RECOMENDATION

5.1. Discussion of the Results

The purpose of this study was to identify, socioeconomic, demographic, and environmental related determinants on infant mortality in Ethiopia based on EDHS data. Consequently, count regression models were used to identify the most important significant variables that affect infant death in Ethiopia.

The descriptive result show that among the total number of women includes in the study (79.03%) of mothers did not face any infant death through survey period of time and (20.98%) of mothers experienced infant deaths due to different factors. The most appropriate count regression model was selected from six possible count models. Among six count models the Hurdle Negative Binomial regression model was selected as the most appropriate model for infant mortality in Ethiopia. This discussion part aims some explanation of the results of Hurdle Negative Binomial regression model of proximate and socioeconomic, demographic, environmental, and health care related determinants impact on infant mortality in related to theoretical background and previous researches.

The findings are also consistent with modernization theory, which contends that industrialization (and the ensuing economic growth) improves human welfare and lowers infant mortality. Another way in which poverty and underdevelopment affect people's well-being is via raising infant mortality. One of the most significant indicators of human well-being is health. The infant mortality factors considered in the current study include antenatal care visits, place of delivery, postnatal care, toilet facility, wealth index, birth order, preceding birth interval, region and place of residence. The empirical analysis of this study reveals that six out of nine of these infant death factors, namely, antenatal care visits, place of delivery, wealth index, birth order, preceding birth interval and region were found to be the determinants of infant mortality in Ethiopia.

According to the results, infant mortality was influenced by geographic location. Infants born in the regional states of Somali and Dire Dawa were more likely to die than those born in Tigray. It is supported by other findings (Kiross et al., 2021b) and (Baraki et al., 2020). The potential explanation for this may be the regional disparities in socioeconomic status, health-care coverage, and other amenities. But this finding is Contradict with (Abate et al., 2020).

The finding of this study showed that number of antenatal care visit is an important health care seeking predictor of infant mortality that is, the number of antenatal care visits of mothers 5-8 ANC is statistically significant with the infant death per mother that expected number of infant death decrease 0.685 times as compared to mothers who are no antenatal care visit. This result is consistent with (Terye, 2020), (Tesema et al., 2021a), (Kiross et al., 2021b) and (Vijay and Patel, 2020). This may be due to the fact that antenatal care appointments offer health benefits like folic acid, iron, and tetanus shots, which may lower the risk of infant mortality. Additionally, ANC gives women and newborns the chance to receive a variety of therapies, such as anti-D, childhood vaccinations, and nutritional supplementation (Agu et al., 2015). But this finding is Contradict with (Abate et al., 2020).

According to the results, the estimated coefficient of birth order is negative and had significant effect with infant death per mother. That means number of infant death decrease by a factor of 0.847 times when increase in the birth order. This result is consistent with (Mulugeta et al., 2022b) and (Muluye and Wencheko, 2012) but contradict with (Weldearegawi et al., 2015), (Abate et al., 2020) and (Baraki et al., 2020).

Infants that are born in health facility is statistically significant with the infant death per mother that expected number of infant death decrease by a factor of 0.777 times when comparing to infants that are born in home as shown positive count table 4.9. This result consistent with (Santos et al., 2016, Vijay and Patel, 2020). This outcome could be explained by the fact that the place of delivery is required to promote the health of women and fetuses by lowering birth complications. But this finding is Contradict with (Weldearegawi et al., 2015), (Abate et al., 2020), (Baraki et al., 2020) and (Mulugeta et al., 2022b).

According to the results, wealth index of household has significant influence in reducing the risk of infant mortality. Income of the household is one of the commonly identified social determinant healths. In Previous studies, countries with higher national income were associated with lower infant mortality rates. Similarly, in this study a significant reduction in risk of infant mortality was observed among births to mothers residing in poorer household as compared to poorest household. This finding was consistent with (Mulugeta et al., 2022b) suggested that infants born to low-income families were more likely to die. But this finding is Contradict with (Abate et al., 2020), (Muluye and Wencheko, 2012) and (Baraki et al., 2020).

In this finding, the number of infant death per mother did not depend on postnatal care (IRR=1.091, 95% CI= (0.858, 1.386), p-value=0.479) suggested that mothers who have Postnatal check-up visits was not statistically associated with increase infant mortality. The finding of this study is contradict with (Terye, 2020).

Although there is a significant difference between urban and rural category, there is no statistical difference between urban and rural residence (IRR=1.010, 95% CI= (0.725, 1.406), p-value=0.955) suggested that infant born rural area was not statistically associated with increasing infant mortality when compared to infant born urban. This is may be due to access of health facility and socio-economic characteristics of the study participants. The finding of this study is contradict with (Baraki et al., 2020) and (Mulugeta et al., 2022b).

Finally, in this study we revealed that preceding birth interval had significant factor on infant mortality. The expected number infant deaths for those infant born with preceding birth interval more than 24 months had decreased by 0.744 times, as compared to infant born with preceding birth interval less than 24 months. This result is consistent with (Tesema et al., 2021a) and (Abate et al., 2020) suggested that mothers increase preceding birth interval was statistically associated with the decreasing of infant mortality.

5.2. Conclusion

The purpose of this study was to identify socioeconomic, demographic, health and environmental related determinant factors of infant mortality per mother in Ethiopia based on 2019 EDHS dataset. This study considered 5,679 women who gave live birth in their life time. The descriptive result showed that 4,488 (79.03%) of mothers had no experienced infant death and 1,191(20.98%) of mothers experienced infant deaths due to different factors. Only 0.53% of them lost ≥ 4 of their infants.

In this study, the most fitted model was selected from different six count regression models: Poisson, Negative binomial (NB), Zero-inflated Poisson (ZIP), Zero-inflated negative binomial (ZINB), Hurdle Poisson (HP) and Hurdle negative binomial (HNB) regression model using different compression techniques. The compression was conducted by using log-likelihood, likelihood ratio test (LRT), information criteria AIC, BIC for nested model and Vuong test for non-nested model.

The result also revealed that HNB model was found to be the most appropriate model to predict the number of infant death per mother in Ethiopia. Hurdle negative binomial regression model is batter 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 also used to identify predictor variables that had significant effects on infant deaths per mother under truncated HNB models variable like:- region, birth order, place of delivery, wealth index, antenatal care visit and preceding birth interval had significant effect on number of infant death per mother. Moreover under inflation (logistic) part model variable like: - region, wealth index, toilet facility, antenatal care visit, postnatal checkup and residence had significant effect on number of infant death per mother. While, variable like source of drinking water, age of mothers, sex, religion, infant vaccination, mother's education level, breastfeeding and marital status are not statistically significant.

5.3. Recommendation

Based on the finding of this study, we forward the following possible recommendations

- Effort are needed to extend educational programs aimed at educating mothers on the benefits of antenatal care, place of delivery and preceding birth interval in order to reduce infant mortality.
- Wealth index of mothers play an important role to reducing infant mortality in Ethiopia. Therefore, government create job opportunities for poorest women's in order to address the problem through improving income and also enhancing the quality of care and attention they can provide to their children.
- The concerned body should work closely with both the private sector and civil society to teach households to have sufficient knowledge and awareness on infant mortality and mechanisms of reduction.

- The government, concerned institutions and other involved stakeholders should be make compressive prevention strategies; commitment and leadership are needed to ensure that child health receives the attention and resources needed to accelerate progress to archive sustainable development in 2030, to reducing infant mortality in to 12 per 1000 child in Ethiopia.
- Further researchers can extend this study by using multilevel count regression models.

Reference

ABATE, M. G., ANGAW, D. A. & SHAWENO, T. 2020. Proximate determinants of infant mortality in Ethiopia, 2016 Ethiopian demographic and health surveys: results from a survival analysis. Archives of Public Health, 78**,** 1-10.

ADHIKARI, R. & SAWANGDEE, Y. 2011. Influence of women's autonomy on infant mortality in Nepal. Reproductive health, 8**,** 1-8.

AHMED, I., ALI, S. M., AMENGA-ETEGO, S., ARIFF, S., BAHL, R., BAQUI, A. H., BEGUM, N., BHANDARI, N., BHATIA, K. & BHUTTA, Z. A. 2018. Population-based rates, timing, and causes of maternal deaths, stillbirths, and neonatal deaths in south Asia and sub-Saharan Africa: a multi-country prospective cohort study. The Lancet Global Health, 6**,** e1297-e1308.

ALY, H. Y. 1990. Demographic and socioeconomic factors affecting infant mortality in Egypt. J Biosoc Sci, 22**,** 447-51.

AGU, U., AGU, M. C. & NNAJI, G. A. 2015. Utilisation of antenatal services at comprehensive health center umunya, anambra state: a retrospective study. Habert Open Access Journal.

BAIRD, S., FRIEDMAN, J. & SCHADY, N. 2011. Aggregate income shocks and infant mortality in the developing world. Review of Economics and statistics, 93**,** 847-856.

BARAKI, A. G., AKALU, T. Y., WOLDE, H. F., LAKEW, A. M. & GONETE, K. A. 2020. Factors affecting infant mortality in the general population: evidence from the 2016 Ethiopian demographic and health survey (EDHS); a multilevel analysis. BMC pregnancy and childbirth, 20**,** 1-8.

BERRUT, S., POUILLARD, V., RICHMOND, P. & ROEHNER, B. M. 2016. Deciphering infant mortality. Physica A: Statistical Mechanics and its Applications, 463**,** 400-426.

BUZUNEH, T. 2019. DETERMINANTS OF UNDER-FIVE DEATHS IN ETHIOPIA, WITH AN APPLICATION OF COUNT REGRESSION MODELS.

CALDWELL, J. C. 1979. Education as a factor in mortality decline an examination of Nigerian data. Population studies**,** 395-413.

CAMERON, A. C. & TRIVEDI, P. K. 2013. Regression analysis of count data, Cambridge university press.

DEFO, B. K. 1996. Areal and socioeconomic differentials in infant and child mortality in Cameroon. Social science & medicine, 42**,** 399-420.

EPHI, I. 2019. Ethiopia mini demographic and health survey 2019: key indicators. Rockville, Maryland, USA: EPHI and ICF.

GAMPAT, R. 2019. Guyana: Leading Causes of Death During the Last 120 years. Available at SSRN 3399003.

GURMU, S. & TRIVEDI, P. K. 1996. Excess zeros in count models for recreational trips. Journal of Business & Economic Statistics, 14**,** 469-477.

HERTEL-FERNANDEZ, A. W., GIUSTI, A. E. & SOTELO, J. M. 2007. The Chilean infant mortality decline: improvement for whom? Socioeconomic and geographic inequalities in infant mortality, 1990-2005. Bull World Health Organ, 85**,** 798-804.

KEELEY, B., LITTLE, C. & ZUEHLKE, E. 2019. The State of the World's Children 2019: Children, Food and Nutrition--Growing Well in a Changing World. UNICEF.

KIROSS, G. T., CHOJENTA, C., BARKER, D. & LOXTON, D. 2021a. Individual-, household- and community-level determinants of infant mortality in Ethiopia. PLoS One, 16**,** e0248501.

KIROSS, G. T., CHOJENTA, C., BARKER, D. & LOXTON, D. 2021b. Individual-, household-and community-level determinants of infant mortality in Ethiopia. PloS one, 16**,** e0248501.

KLEINSCHMIDT, I., SCHWABE, C., BENAVENTE, L., TORREZ, M., RIDL, F. C., SEGURA, J. L., EHMER, P. & NCHAMA, G. N. 2009. Marked increase in child survival after four years of intensive malaria control. The American journal of tropical medicine and hygiene, 80**,** 882.

KUSE, K. A., CHIKAKO, T. U., HAGAN, J. E., SEIDU, A.-A. & AHINKORAH, B. O. 2022. Assessing Individual-and Community-Level Variability in Predictors of Neonatal, Infant, and Under-Five Child Mortality in Ethiopia Using a Multilevel Modeling Approach. Children, 9**,** 1071.

LAWN, J. E., BLENCOWE, H., OZA, S., YOU, D., LEE, A. C., WAISWA, P., LALLI, M., BHUTTA, Z., BARROS, A. J. & CHRISTIAN, P. 2014. Every Newborn: progress, priorities, and potential beyond survival. The lancet, 384**,** 189-205.

LORENZ, J. M., ANANTH, C. V., POLIN, R. A. & D'ALTON, M. E. 2016. Infant mortality in the United States. Journal of Perinatology, 36**,** 797-801.

LOZANO, R., NAGHAVI, M., FOREMAN, K., LIM, S., SHIBUYA, K., ABOYANS, V., ABRAHAM, J., ADAIR, T., AGGARWAL, R. & AHN, S. Y. 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet, 380**,** 2095-2128.

MASQUELIER, B., HUG, L., SHARROW, D., YOU, D., MATHERS, C., GERLAND, P., ALKEMA, L. & FOR CHILD, U. I.-A. G. 2021. Global, regional, and national mortality trends in youth aged 15–24 years between 1990 and 2019: a systematic analysis. The Lancet Global Health, 9**,** e409-e417.

MATHEWS, T., MACDORMAN, M. F. & THOMA, M. E. 2015. Infant mortality statistics from the 2013 period linked birth/infant death data set.

MIN, Y. & AGRESTI, A. 2002. Modeling nonnegative data with clumping at zero: a survey.

MONDAL, K. K. & MANI, C. 2012. Investigation of the antibacterial properties of nanocopper against Xanthomonas axonopodis pv. punicae, the incitant of pomegranate bacterial blight. Annals of microbiology, 62**,** 889-893.

MULUGETA, S. S., MULUNEH, M. W., BELAY, A. T., MOYEHODIE, Y. A., AGEGN, S. B., MASRESHA, B. M. & WASSIHUN, S. G. 2022a. Multilevel log linear model to estimate the risk factors associated with infant mortality in Ethiopia: further analysis of 2016 EDHS. BMC Pregnancy and Childbirth, 22**,** 1-11.

MULUGETA, S. S., MULUNEH, M. W., BELAY, A. T., MOYEHODIE, Y. A., AGEGN, S. B., MASRESHA, B. M. & WASSIHUN, S. G. 2022b. Multilevel log linear model to estimate the risk factors associated with infant mortality in Ethiopia: further analysis of 2016 EDHS. BMC Pregnancy and Childbirth, 22**,** 597.

MULLAHY, J. 1986. Specification and testing of some modified count data models. Journal of econometrics, 33**,** 341-365.

MULUYE, S. & WENCHEKO, E. 2012. Determinants of infant mortality in Ethiopia: A study based on the 2005 EDHS data. Ethiopian Journal of Health Development, 26**,** 72-77.

ORGANIZATION, W. H. 2018. Standards for improving the quality of care for children and young adolescents in health facilities.

ORGANIZATION, W. H. 2020. Infant mortality.

ORGANIZATION, W. H. 2019. World health statistics overview 2019: monitoring health for the SDGs, sustainable development goals. World Health Organization.

PATEL, K. K., RAI, R. & RAI, A. K. 2021. Determinants of infant mortality in Pakistan: evidence from Pakistan Demographic and Health Survey 2017–18. Journal of Public Health, 29, 693-701.

SANTOS, S. L., SANTOS, L. B., CAMPELO, V. & SILVA, A. R. 2016. Factors Associated with Infant Mortality in a Northeastern Brazilian Capital. Rev Bras Ginecol Obstet, 38, 482-491.

SOVIO, U., DIBDEN, A. & KOUPIL, I. 2012. Social determinants of infant mortality in a historical Swedish cohort. Paediatr Perinat Epidemiol, 26, 408-20.

TESEMA, G. A., SERETEW, W. S., WORKU, M. G. & ANGAW, D. A. 2021a. Trends of infant mortality and its determinants in Ethiopia: mixed-effect binary logistic regression and multivariate decomposition analysis. BMC Pregnancy and Childbirth, 21, 362.

TESEMA, G. A., SERETEW, W. S., WORKU, M. G. & ANGAW, D. A. 2021b. Trends of infant mortality and its determinants in Ethiopia: mixed-effect binary logistic regression and multivariate decomposition analysis. BMC Pregnancy Childbirth, 21, 362.

VAKILI, R., EMAMI MOGHADAM, Z., KHADEMI, G., VAKILI, S. & SAEIDI, M. 2015. Child mortality at different world regions: A comparison review. International Journal of Pediatrics, 3, 809-816.

VIJAY, J. & PATEL, K. K. 2020. Risk factors of infant mortality in Bangladesh. Clinical Epidemiology and Global Health, 8, 211-214.

WELDEAREGAWI, B., MELAKU, Y. A., ABERA, S. F., ASHEBIR, Y., HAILE, F., MULUGETA, A., ESHETU, F. & SPIGT, M. 2015. Infant mortality and causes of infant deaths in rural Ethiopia: a population-based cohort of 3684 births. BMC public health, 15, 1-7.

YU, X., WANG, Y., KANG, L., MIAO, L., SONG, X., RAN, X., ZHU, J., LIANG, J., LI, Q. & DAI, L. 2022. Geographical disparities in infant mortality in the rural areas of China: a descriptive study, 2010–2018. BMC pediatrics, 22, 1-9.

ZIMMERMAN, L. A., SHIFERAW, S., SEME, A., YI, Y., GROVE, J., MERSHON, C.-H. & AHMED, S. 2019. Evaluating consistency of recall of maternal and newborn care complications and intervention coverage using PMA panel data in SNNPR, Ethiopia. PloS one, 14, e0216612.

APPENDEX

Table 1: Poisson regression model estimates

	Estimate	Std. Error	z value	p-value
Intercept	-1.751	0.166	-10.532	0.000
Region(Tigray)				
Afar	0.289	0.160	1.800	0.072
Amhara	0.768	0.156	4.917	0.000
Oromia	0.951	0.147	6.488	0.000
Somali	0.953	0.148	6.457	0.000
Benishangul	1.173	0.149	7.862	0.000
SNNPR	0.760	0.153	4.952	0.000
Gambela	0.799	0.162	4.943	0.000
Harari	0.527	0.174	3.033	0.002
Addis Abeba	0.183	0.225	0.816	0.415
Dire Dawa	1.249	0.158	7.926	0.000
Wealth index (Poorest)				
Poorer	-0.167	0.058	-2.903	0.004
Middle	-0.247	0.075	-3.291	0.001
Richer	-0.364	0.089	-4.079	0.000
Richest	-0.919	0.116	-7.934	0.000
Toilet facility (Has toile facility)				
No toilet facility	0.157	0.048	3.259	0.001
Birth order	-0.166	0.034	-4.837	0.001
Preceding birth interval (1-24)				
>24	-0.055	0.051	-1.090	0.276
No of antenatal care visit (No ANC)				
1-4 ANC	-0.161	0.063	-2.556	0.011
5-8 ANC	-0.232	0.070	-3.311	0.001
>8 ANC	-0.475	0.103	-4.628	0.000
Place of delivery (Home)				
Health facility	-0.099	0.048	-2.052	0.040
postnatal care (No)				
Yes	-0.091	0.049	-1.867	0.062
Place of residence (Urban)				
Rural	-0.346	0.065	-5.357	0.000

Table 2: Negative binomial regression model estimates

	Estimate	Std. Error	z value	p-value
Intercept	-1.666	0.202	-8.261	0.000
Region(Tigray)				
Afar	0.302	0.187	1.608	0.108
Amhara	0.849	0.184	4.616	0.000
Oromia	1.011	0.174	5.816	0.000
Somali	1.039	0.175	5.927	0.000
Benishangul	1.331	0.178	7.485	0.000
SNNPR	0.788	0.181	4.355	0.000
Gambela	0.819	0.193	4.243	0.000
Harari	0.557	0.205	2.721	0.007
Addis Abeba	0.233	0.258	0.901	0.367
Dire Dawa	1.296	0.193	6.733	0.000
Wealth index (Poorest)				
Poorer	-0.214	0.077	-2.781	0.005
Middle	-0.314	0.099	-3.183	0.001
Richer	-0.405	0.114	-3.555	0.001
Richest	-0.916	0.137	-6.666	0.000
Toilet facility (Has toile facility)				
No toilet facility	0.191	0.063	3.029	0.002
Birth order	-0.222	0.015	-15.078	0.000
Preceding birth interval (1-24)				
>24	-0.060	0.067	-0.900	0.368
No of antenatal care visit (No ANC)				
1-4 ANC	-0.229	0.082	-2.795	0.005
5-8 ANC	-0.318	0.092	-3.453	0.001
>8 ANC	-0.540	0.128	-4.203	0.000
Place of delivery (Home)				
Health facility	-0.085	0.063	-1.353	0.176
postnatal care (No)				
Yes	-0.073	0.064	-1.151	0.249
Place of residence (Urban)				
Rural	-0.366	0.088	-4.157	0.000

Table 3: Zero inflated poison regression model estimates

Count part (non- zero part)				
	Estimate	Std. Error	z value	p-value
Intercept	-1.203	0.291	-4.139	0.000
Region(Tigray)				
Afar	0.118	0.284	0.416	0.678
Amhara	0.004	0.278	0.015	0.988
Oromia	0.251	0.255	0.985	0.324
Somali	0.548	0.276	1.984	0.047
Benishangul	0.212	0.279	0.758	0.449
SNNPR	0.168	0.289	0.580	0.562
Gambela	0.466	0.287	1.621	0.105
Harari	-0.305	0.318	-0.960	0.337
Addis Abeba	-0.351	0.405	-0.867	0.386
Dire Dawa	0.654	0.273	2.392	0.017
Wealth index (Poorest)				
Poorer	-0.459	0.108	-4.250	0.000
Middle	-0.211	0.119	-1.758	0.079
Richer	-0.110	0.158	-0.698	0.485
Richest	0.205	0.199	1.028	0.304
Toilet facility (Has toile facility)				
No toilet facility	0.059	0.0735	0.810	0.418
Birth order	-0.056	0.031	-1.827	0.068
Preceding birth interval (1-24)				
>24	-0.216	0.0759	-2.842	0.004
No of antenatal care visit (No ANC)				
1-4 ANC	0.129	0.101	1.283	0.199
5-8 ANC	-0.342	0.125	-2.727	0.006
>8 ANC	0.191	0.186	1.028	0.304
Place of delivery (Home)				

Health facility	-0.226	0.073	-3.093	0.002
postnatal care (No) Yes	 0.052	 0.074	 0.710	 0.477
Place of residence (Urban) Rural	 -0.074	 0.098	 -0.753	 0.451

Zero inflation part				
Intercept	-0.552	0.491	-1.123	0.261
Region(Tigray) Afar Amhara Oromia Somali Benishangul SNNPR Gambela Harari Addis Abeba Dire Dawa	 -0.319 -1.473 -1.334 -0.809 -2.043 -1.084 -0.681 -1.615 -0.972 -1.101	 0.416 0.468 0.385 0.410 0.513 0.446 0.424 0.573 0.720 0.419	 -0.767 -3.146 -3.465 -1.973 -3.985 -2.431 -1.606 -2.819 -1.349 -2.622	 0.443 0.002 0.001 0.048 0.000 0.015 0.108 0.005 0.177 0.009
Wealth index (Poorest) Poorer Middle Richer Richest	 -0.556 0.171 0.554 1.898	 0.217 0.213 0.269 0.295	 -2.556 0.806 2.063 6.427	 0.011 0.420 0.039 0.000
Toilet facility (Has toile facility) No toilet facility	 -0.216	 0.136	 -1.586	 0.113
Birth order	0.299	0.063	4.722	0.000
Preceding birth interval (1-24) >24	 -0.326	 0.139	 -2.351	 0.019
No of antenatal care visit (No ANC) 1-4 ANC 5-8 ANC >8 ANC	 0.608 -0.206 1.181	 0.181 0.271 0.273	 3.364 -0.761 4.332	 0.001 0.447 0.000
Place of delivery (Home) Health facility	 -0.249	 0.135	 -1.854	 0.064
postnatal care (No) Yes	 0.268	 0.136	 1.980	 0.048
Place of residence (Urban) Rural	 0.619	 0.194	 3.197	 0.001

Table 4: Zero-inflated negative binomial regression model estimates

Count part (non- zero part)				
	Estimate	Std. Error	z value	p-value
Intercept	-1.773	0.219	-8.092	0.000
Region(Tigray)				
Afar	0.359	0.222	1.621	0.105
Amhara	0.433	0.214	2.025	0.043
Oromia	0.707	0.203	3.474	0.001
Somali	0.709	0.211	3.366	0.001
Benishangul	0.757	0.210	3.598	0.000
SNNPR	0.395	0.215	1.841	0.066
Gambela	0.722	0.229	3.138	0.002
Harari	-0.008	0.229	-0.035	0.972
Addis Abeba	-0.070	0.283	-0.248	0.804
Dire Dawa	0.761	0.217	3.509	0.001
Wealth index (Poorest)				
Poorer	-0.347	0.087	-3.983	0.000
Middle	-0.359	0.109	-3.273	0.001
Richer	-0.388	0.125	-3.109	0.002
Richest	-0.823	0.155	-5.308	0.000
Toilet facility (Has toile facility)				
No toilet facility	0.167	0.069	2.424	0.015
Birth order	-0.052	0.022	-2.341	0.019
Preceding birth interval (1-24)				
>24	-0.057	0.072	-0.796	0.426
No of antenatal care visit (No ANC)				
1-4 ANC	-0.161	0.089	-1.800	0.072
5-8 ANC	-0.318	0.102	-3.115	0.002
>8 ANC	-0.148	0.177	-0.836	0.403
Place of delivery (Home)				
Health facility	-0.073	0.068	-1.068	0.286
postnatal care (No)				
Yes	-0.009	0.069	-0.133	0.894
Place of residence (Urban)				
Rural	-0.424	0.096	-4.398	0.000

Zero inflation part				
Intercept	-6.688	1.062	-6.299	0.000
Region(Tigray)				
Afar	0.428	0.788	0.542	0.587
Amhara	-2.642	0.815	-3.243	0.001
Oromia	-2.530	0.759	-3.331	0.001
Somali	-2.351	0.778	-3.021	0.003
Benishangul	-4.911	1.029	-4.772	0.000
SNNPR	-3.722	1.402	-2.654	0.008
Gambela	-0.979	0.868	-1.129	0.259
Harari	-11.268	6.336	-1.779	0.075
Addis Abeba	-1.535	1.629	-0.942	0.346
Dire Dawa	-9.491	2.825	-3.359	0.001
Wealth index (Poorest)				
Poorer	-1.083	0.467	-2.319	0.020
Middle	-0.225	0.483	-0.465	0.642
Richer	0.347	0.539	0.643	0.520
Richest	1.358	0.810	1.675	0.094
Toilet facility (Has toile facility)				
No toilet facility	-0.213	0.331	-0.642	0.521
Birth order	1.387	0.164	8.443	0.000
Preceding birth interval (1-24)				
>24	-0.013	0.325	-0.040	0.968
No of antenatal care visit (No ANC)				
1-4 ANC	0.456	0.376	1.213	0.225
5-8 ANC	-0.534	0.623	-0.857	0.391
>8 ANC	2.267	0.703	3.227	0.001
Place of delivery (Home)				
Health facility	0.143	0.314	0.456	0.649
postnatal care (No)				
Yes	0.655	0.315	2.080	0.038
Place of residence (Urban)				
Rural	-0.089	0.429	-0.210	0.834

Table 5: Hurdle poison regression model estimates

Count part (non- zero part)				
	Estimate	Std. Error	z value	p-value
Intercept	-1.062	0.339	-3.128	0.002
Region(Tigray)				
Afar	0.258	0.343	0.752	0.452
Amhara	0.435	0.341	1.278	0.201
Oromia	0.406	0.322	1.260	0.208
Somali	0.779	0.325	2.400	0.016
Benishangul	0.696	0.330	2.107	0.035
SNNPR	0.389	0.338	1.152	0.249
Gambela	0.517	0.345	1.498	0.134
Harari	0.129	0.379	0.341	0.733
Addis Abeba	0.142	0.488	0.292	0.770
Dire Dawa	0.858	0.334	2.570	0.010
Wealth index (Poorest)				
Poorer	-0.309	0.101	-3.054	0.002
Middle	-0.071	0.124	-0.573	0.566
Richer	-0.030	0.154	-0.198	0.843
Richest	0.083	0.211	0.392	0.695
Toilet facility (Has toile facility)				
No toilet facility	-0.004	0.081	-0.050	0.959
Birth order	-0.167	0.025	-6.791	0.000
Preceding birth interval (1-24)				
>24	-0.249	0.084	-2.970	0.003
No of antenatal care visit (No ANC)				
1-4 ANC	0.159	0.106	1.502	0.133
5-8 ANC	-0.246	0.129	-1.908	0.056
>8 ANC	0.256	0.176	1.454	0.146
Place of delivery (Home)				
Health facility	-0.274	0.081	-3.362	0.001
postnatal care (No)				
Yes	0.052	0.082	0.637	0.524
Place of residence (Urban)				
Rural	0.006	0.111	0.051	0.959

Zero inflation part				
Intercept	-1.445	0.209	-6.910	0.000
Region(Tigray)				
Afar	0.354	0.195	1.812	0.069
Amhara	0.794	0.191	4.159	0.000
Oromia	1.090	0.179	6.065	0.000
Somali	1.023	0.183	5.593	0.000
Benishangul	1.249	0.185	6.757	0.000
SNNPR	0.864	0.187	4.615	0.000
Gambela	0.828	0.199	4.139	0.000
Harari	0.653	0.212	3.082	0.002
Addis Abeba	0.116	0.271	0.427	0.669
Dire Dawa	1.269	0.202	6.300	0.000
Wealth index (Poorest)				
Poorer	-0.157	0.081	-1.943	0.052
Middle	-0.363	0.105	-3.456	0.001
Richer	-0.535	0.122	-4.393	0.000
Richest	-1.168	0.149	-7.822	0.000
Toilet facility (Has toile facility)				
No toilet facility	0.225	0.067	3.369	0.001
Birth order	0.003	0.015	0.219	0.827
Preceding birth interval (1-24)				
>24	0.039	0.071	0.546	0.585
No of antenatal care visit (No ANC)				
1-4 ANC	-0.296	0.087	-3.380	0.001
5-8 ANC	-0.241	0.096	-2.509	0.012
>8 ANC	-0.689	0.141	-4.891	0.000
Place of delivery (Home)				
Health facility	-0.015	0.067	-0.229	0.819
postnatal care (No)				
Yes	-0.134	0.067	-1.989	0.047
Place of residence (Urban)				
Rural	-0.414	0.093	-4.436	0.000

R package and command for count regression models

```
library(AER)
library(MASS)
library(pscl)
library(readr)
library(haven)
library(tidyverse)
library(foreign)
library(lme4)
library(ggplot2)
Fzeydstata <- read_dta("C:/Users/Bethel/Desktop/zeyd final stata data.dta"")
Fzeydstata
View(Fzeydstata)
names(Fzeydstata)
head(Fzeydstata)
plot(table(Fzeydstata$death))
sum(Fzeydstata$death)
ftable(Fzeydstata$death)
summary(death)
poisson <- glm(death ~ factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
, offset= log(TCEB),data=Fzeydstata,family=poisson(link="log"))
summary(poisson)
negbin <- glm.nb(death ~ factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
+offset(log(TCEB)),data=Fzeydstata)
summary(negbin)
zip <- zeroinfl(death ~ factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
|factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
,offset= log(TCEB),link="logit",dist="poisson",data=Fzeydstata)
summary(zip)
zinb <- zeroinfl(death ~ factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
|factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
, offset= log(TCEB),dist="negbin",data=Fzeydstata)
summary(zinb)
hurdle.poisson <- hurdle(death  ~ factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
|factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
, offset= log(TCEB),link="logit",dist="poisson",data=Fzeydstata)
```

```
summary(hurdle.poisson)
summary(negbin)
hurdle.nb <- hurdle(death ~ factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
|factor(Region)+factor(Wealth)+factor(T_facility)+
BORD+factor(PB_interval)+factor(ANCvisit)+factor(Pdelivery)+factor(Postnatal)+factor(Residence)
, offset= log(TCEB),link="logit",dist="negbin",data=Fzeydstata)
summary(hurdle.nb)
AIC(poisson,negbin,hurdle.poisson,hurdle.nb,zip,zinb)
BIC(poisson,negbin,hurdle.poisson,hurdle.nb,zip,zinb)
logLik(poisson)
logLik(negbin)
logLik(hurdle.poisson)
logLik(hurdle.nb)
logLik(zip)
logLik(zinb)
vuong(zip,poisson)
vuong(zinb,negbin)
vuong(hurdle.poisson, zip)
vuong(hurdle.nb,zinb)
lrtest(negbin,poisson)
lrtest(zinb,zip)
lrtest(hurdle.nb,hurdle.poisson)
phat.poisson = predprob(poisson)
phat.poisson.mn = apply(phat.poisson,2,mean)
phat.negbin = predprob(negbin)
phat.negbin.mn = apply(phat.negbin,2,mean)
phat.zip = predprob(zip)
phat.zip.mn = apply(phat.zip,2,mean)
phat.zinb = predprob(zinb)
phat.zinb.mn = apply(phat.zinb,2,mean)
phat.hurdle.poisson = predprob(hurdle.poisson)
phat.hurdle.poisson.mn = apply(phat.hurdle.poisson,2,mean)
phat.hurdle.nb = predprob(hurdle.nb)
phat.hurdle.nb.mn = apply(phat.hurdle.nb,2,mean)
m1=sum(Fzeydstata$death < 1)
p1=sum(dpois(0, fitted(poisson)))
p2=sum(dnbinom(0, mu = fitted(negbin), size = negbin$theta))
p3=sum(predict(zip, type = "prob")[,1])
p4=sum(predict(hurdle.poisson, type = "prob")[,1])
p5=sum(predict(hurdle.nb, type = "prob")[,1])
p6=sum(predict(zinb, type = "prob")[,1])
exp(cbind(IRR = coef(hurdle.nb),confint(hurdle.nb)))
```