STATISTICAL ANALYSIS OF NUMBER OF ANTENATAL CARE VISITS AMONG PREGNANT WOMEN IN RURAL ETHIOPIA

BY:
DAWIT SEKATA

Department of Statistics
School of Graduate Studies
College of Natural Science
Jimma University, Ethiopia

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biostatistics

December, 2014

STATISTICAL ANALYSIS OF NUMBER OF ANTENATAL CARE VISITS AMONG PREGNANT WOMEN IN RURAL ETHIOPIA

BY:

DAWIT SEKATA

Advisor: Wondwosen Kassahun (PhD)

Co-advisor: Geremew Muleta (MSc)

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

December, 2014

Jimma, Ethiopia

Acknowledgment

Primarily and foremost I would like to express my appreciation and gratitude to my Lord and Savior, Jesus Christ, for giving me life breath and the strength to complete this thesis. I owe you my life Lord!

Next, I want to thank my main advisor Dr Wondwosen Kassahun, who provided valuable guidance and inspiration, encouragement, sound advice, and good teaching throughout my studies. His inspiring and enthusiastic being and accessible help has been a great support in writing up of this paper. I would also like to thank my co-advisor Geremew Muleta (MSc.) who gives me all the necessary advices and comments regarding my thesis development.

Discussion with my fellow students has been important to me during the months of studies. I want to thank my friend Alemu Bekele for suggesting to meet on a daily basis and discuss our progress and problems concerning the paper.

Finally my sincere thanks are extended to all my beloved and cooperative families and friends for their continuous encouragement. Particularly, I must make mention of my wife Obse Tesfaye and my son, Siyosan Dawit. I thank my wife for her constant support, patience and encouragement to me as my life partner. I also express my appreciation to her for all the time she availed to me to complete the paper, by taking care of our son. Thank you Siyosan! Your loving care and splendid sense of funniness is of great importance to me.

Dawit Sekata
sekatad@gmail.com

Dedication

This thesis is dedicated to the Almighty God, my family, my advisor, Dr. Wondwosen Kassahun, Wollega University, the basement of mine to be here and anybody doing anything to reduce maternal mortality in Ethiopia and/or in Africa.

Table of Contents **Page**

List of Tables

List of Figures

Acronyms

AIC	Akaike’s Information Criterion
AICC	Akaike Information Criterion Corrected
AIDS	Acquired Immune Deficiency Syndrome
ANC	Antenatal Care
CSA	Central Statistical Agency
DF	Degrees of Freedom
EDD	Expected Date of Delivery
EDHS	Ethiopian Demographic and Health Survey
EHNRI	Ethiopian Health and Nutrition Research Institute
GLM	Generalized Linear Model
GOF	Goodness of Fit
HIV	Human Immune Virus
HNB	Hurdle Negative Binomial
HP	Hurdle Poisson
HSDP	Health Sector Development Program
IPTp	Intermittent Preventive Treatment for Malaria during Pregnancy
MDG	Millennium Development Goal
ML	Maximum Likelihood
MoH	Ministry of Health
NB	Negative Binomials
NGO	Non Governmental Organization
NLMIXED	Non Linear Mixed
OR	Odds Ratio
PMF	Probability Mass Function
PMTCT	Prevention of Mother-To-Child Transmission
PIH	Pregnancy Induced Hypertension
ROC	Receiver Operating Characteristic
SE	Standard Error
SNNPR	South Nation Nationalities and Peoples Republic
UNFPA	United Nations Population Fund
UNICEF	United Nations Children’s Fund
USAID	United States Agency for International Development
VIF	Variance Inflation Factor
WHO	World Health Organization
ZINB	Zero Inflated Negative Binomials
ZIP	Zero Inflated Poisson

Abstract

Background: *The ANC service is used to ensure a normal pregnancy with delivery of a healthy baby from a healthy mother. Even if WHO recommends a minimum of four ANC visits, existing evidence from developing countries including Ethiopia indicates that few women utilize it due to different determinants such as lack of education, awareness, nearby health post, residence, etc.*

Objective: *The main objective of the study was statistically to analyze the determinants of the barriers in number of antenatal care service visits among pregnant women in rural Ethiopia.*

Methods: *A cross sectional data from EDHS-2011 was used and 1127 pregnant women who had 9 months of pregnancy was included to the study. Several count models were fitted to select the model which best fits the data, these are: Poisson, NB, ZIP, ZINB, HP, and HNB regression models. Each of these models was compared by likelihood ratio test (LR), Voung test and the information criteria's. The data were Analyzed using SAS version 9.2.*

Results: *In this study there were excess zeros, 51.5%; the variance of the data, 7.196, was much higher than its mean, 1.85. Women from better progressed regions were about one (OR=0.8048) times more likely to have positive ANC visits than women from other regions, educated women were one times (OR=1.44) more likely to have positive ANC visits than non educated ones, women who had seen danger signs of pregnancy was one times (OR=1.42) the rate of positive ANC visits than women who hadn't seen it, women with heavy workload had less likely to visit positive ANC attendance (OR =0.36) than women who had no workload problems, there is a greater likelihood of a positive number of ANC visits for rich women than poor women (OR=1.15), and women who had nearby health post have a greater likelihood of a positive ANC visits than women with lack of nearby health post (OR=0.48) holding all other predictors constant.*

Conclusion: *Though the government effort is to improve access to modern ANC visits during pregnancy, it was low in rural Ethiopia than the national value. Lack of awareness, absence of education, heavy workload, poverty, and shortage of health post were significantly associated with not attending ANC visits. Hence, institutions that act on maternal and children's health care should do well to apply the minimum of four ANC visits scheduled by WHO mainly on rural areas so that all perpetrators of maternal care shall be brought to book to deter others from repeating such absences and thus move the country closer to MDG targets for maternal health by 2015. Hurdle Poisson regression model was found to be better fitted with data which is characterized by excess zeros and high variability in the non-zero outcomes.*

Key words: *ANC; Poisson regression model; Negative binomial (NB) regression model; Zero-inflated Poisson (ZIP) regression model; Zero-inflated negative binomial (ZINB)regression model; Hurdle Poisson regression model; Hurdle NB regression model.*

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

The World Health Organization [55] estimates that about 536,000 women of reproductive age die each year because of pregnancy related complications. Nearly all of these deaths (99%) occur in the developing world [56]. Ethiopia is one of the countries with an unacceptably highest maternal mortality and the infant mortality rate in the world [12]. In Ethiopia, 85% of the population lives in rural areas, availability of health services, especially maternal health care services, is extremely difficult. Overall, access for maternity care is on average 26% for rural and 76% for urban areas [55].

One of the Millennium Development Goals (MDGs) targets is to reduce by three quarters, between 1990 and 2015, the maternal mortality ratio in all countries. Maternal mortality is the most important indicator of maternal health and well-being in any country. As a result, it has been central to government health sector policies aimed at improving the overall health of the Ethiopian population especially that of the women. The World Health Organization [57] has defined maternal mortality as “the death of a woman while pregnant or within 42 days of a termination of a pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental and incidental causes.”

Periodic and regular supervision including examination and advice of a woman during pregnancy is called antenatal care. In other word, antenatal care is a preventive obstetric health care program aimed at optimizing maternal fetal outcome through regular monitoring of pregnancy, [55]. In general, the main objective of ANC is to ensure a normal pregnancy with delivery of a healthy baby from a healthy mother.

World Health Organization [53] advocates an improved model for antenatal care use for women without complicated pregnancy in developing countries. This model recommends at least four antenatal care visits which would include compulsory blood pressure measurement, urine and blood tests and non-compulsory weight and height check at each visit [55].

The Four-Visit ANC Model Outlined in WHO Clinical Guidelines: First visit (8-12 weeks) confirm pregnancy and EDD, classify women for basic ANC (four visits) or more specialized care. Screen, treat and give preventive measures. Develop a birth and emergency plan. Advice and counsel; Second visit **(**24-26 weeks) assess maternal and fetal well-being, exclude PIH and anemia, give preventive measures, review and modify birth and emergency plan; Third visit (32 weeks) assess maternal and fetal well-being, exclude PIH, anemia, multiple pregnancies, give preventive measures, review and modify birth and emergency plan; Fourth visit (36-38 weeks), assess maternal and fetal well-being exclude PIH, anemia, multiple pregnancy, and mal-presentation, give preventive measures, review and modify birth and emergency plan and advice and counsel.

While some studies have looked at different risk factors for antenatal care (ANC) and delivery service utilization in the country, information coming from community-based studies related to the Health Extension Programme (HEP) in rural areas is limited. This study aimed to determine the prevalence of maternal health care utilization and explore its determinants among 9th month pregnant women in rural Ethiopia.

1.2 Statement of the Problem

The data from WHO confirms that in developing countries as a whole, educated women are more likely to receive antenatal care and the likelihood of their using antenatal care is associated with their level of education. Well oriented women’s about pregnancy complications are also more likely to report four or more visits. In most countries, the greatest proportionate difference occurs between women following socioeconomic, demographic, health and environmental related factors [14].

In Ethiopia, the maternal mortality was estimated to be 673 deaths per 100,000 live births and infant mortality rate was 77 per 1,000 live births, which is among the highest in the world [9]. As emphasized in the 2005 Ethiopian Demographic and Health Survey (EDHS), the use of antenatal care services is very low and ranges between 26% in rural area to 76% in the urban parts of the country [9]. Therefore, even if there are efforts to improve access to modern antenatal care visits during pregnancy, it remained very limited by international standards.

Though ANC service utilization is very essential for improvement of maternal and child health, the use of the service is still very limited in rural areas of Ethiopia [9]. There could be several factors that limit the utilization of ANC in the region in general, in the zone in particular which requires further study. Therefore, it is important to explore and describe the status of ANC service visits utilization in rural areas of Ethiopia and describe the influencing determinants.

In addition, this study will provide valuable information how to model count data when assumption of the standard Poisson regression is violated (when there is greater variability in the response counts than one would expect if the response distribution truly were Poisson). In such occasions, it is of interest to examine the applicability of the Zero Inflated models (ZIP, ZINB) and Hurdle models (Hurdle Poisson as well as Hurdle NB) in addition to Negative Binomial and Poisson regression models and compare their performances in terms of their goodness-of-fit statistics, AIC, BIC, likelihood ratio test and theoretical soundness.

1.3 Objectives of the Study

1.3.1 General Objectives

The general objective of this study is statistically to analyze the determinants of the barriers in number of antenatal care service visits among pregnant women in rural Ethiopia.

1.3.2 Specific Objectives

1. To estimate the mean number of ANC visits of pregnant women in rural Ethiopia
2. To examine the key socio-economic and demographic factors influencing the utilization of antenatal care services in rural Ethiopia.
3. To fit an appropriate statistical model for the number of ANC visits of pregnant women in rural Ethiopia using the appropriate GOF measurements.

1.4 Significance of the Study

The findings of this study may identify the determinants of the barriers in number of ANC service visits among pregnant women in rural Ethiopia. This has a great importance in providing timely booking, awareness of risks and early seeking to care and birth preparedness. Information on which factor determines the time of ANC booking in the area could be helpful for policy makers, program implementers, monitoring and evaluation activities. Since the study will attempt to reveal the major factors for barriers in ANC in rural Ethiopia, it will help to guide the end user governmental and non-governmental organizations to develop maternal care programs and set appropriate plans to tackle the existing health and antenatal care problems. The other significance of this study will be to provide the appropriate model that aid researchers in determining the appropriate model to use given zero-inflated data.

CHAPTER TWO

LITERATURE REVIEW

2.1 Coverage and Trends of ANC

Currently, 71 percent of women worldwide receive any ANC; in industrialized countries, more than 95 percent of pregnant women have access to ANC. In sub-Saharan Africa, 69 percent of pregnant women have at least one ANC visit, more than in South Asia, at 54 percent [42]. Coverage for ANC is usually expressed as the proportion of women who have had at least one ANC visit during her pregnancy. However, according to the report of MoH of Ethiopia in 2007, about 52% Ethiopian women received one or more ANC visits, less than 17% received professionally assisted delivery care and 19% received postnatal care [40]. Trends indicate slower progress in sub-Saharan Africa than in other regions, with an increase in coverage of only four percent during the past decade. In Africa, 80 percent of women in the richest quintile have access to three or more ANC visits, while only 48 percent of the poorest women have the same level of access [56]. A similar disparity exists between urban and rural women. Within the continuum of care, however, there is a smaller gap between the rich and the poor in ANC than in skilled attendance during childbirth, which is available to only 25 percent of the poorest women in sub-Saharan Africa, while reaching 81 percent of the richest. Coverage of four or more ANC visits as well as the number of visits disaggregated by trimester is important to assess, because the effectiveness of certain ANC interventions such as tetanus vaccination, IPTp for malaria, and prevention of mother-to-child transmission (PMTCT) of HIV depend on repeated visits and the trimester in which they occur. In Africa, the proportion of pregnant women who attended the recommended four or more visits increased by six percent over 10 years [55].

Similarly, the proportion of women who received ANC in the first six months of pregnancy increased by 10 percent over 10 years, faster than the increase of overall ANC coverage. Measuring coverage alone does not provide information on quality of care, and poor quality in ANC clinics, correlated with poor service utilization, is common in Africa. This is often related to an insufficient number of skilled providers (particularly in rural and remote areas), lack of standards of care and protocols, few supplies and drugs, and poor attitudes of health providers. An assessment conducted in Tanzania found twice as many poorly qualified health workers in rural facilities than in urban facilities [42].

In Ethiopia, according to EDHS, 2005, only 6 percent of women make their first ANC visit before the fourth month of pregnancy [55]. The median duration of pregnancy for the first ANC visit was 5.6 months. The median duration of pregnancy for the first ANC visit was 4.2 months for urban women compared with 6.0 for rural women. In urban area where the health services are physically accessible and ANC at the public services are provided free of charge, only 32.4% of women seek the service before 16 weeks of gestation [9].

The report identified that, 72% of mothers with at least secondary school education received ANC compared to 45% and 21% of mothers' with primary and no education respectively. The EDHS, 2005 and community and family survey conducted in SNNPR to assess maternity care utilization, also reflected the above situation [18].

2.2 The Effects of Inadequate ANC During Pregnancy

Good care during pregnancy is important for the health of the mother and the development of the unborn baby. Pregnancy is a crucial time to promote healthy behaviours and parenting skills. Good ANC links the woman and her family with the formal health system, increases the chance of using a skilled attendant at birth and contributes to good health through the life cycle [5]. Inadequate care during this time breaks a critical link in the continuum of care, and effects both women and babies.

It has been estimated that 25 percent of maternal deaths occur during pregnancy, with variability between countries depending on the prevalence of unsafe abortion, violence, and disease in the area [55]. Between a third and a half of maternal deaths are due to causes such as hypertension (pre-eclampsia and eclampsia) and antepartum haemorrhage, which are directly related to inadequate care during pregnancy. In a study conducted in six west African countries, a third of all pregnant women experienced illness during pregnancy, of whom three percent required hospitalisation. Certain pre-existing conditions become more severe during pregnancy. Malaria, HIV/AIDS, anaemia and malnutrition are associated with increased maternal and newborn complications as well as death where the prevalence of these conditions is high. New evidence suggests that women who have been subject to female genital mutilation are significantly more likely to have complications during childbirth, so these women need to be identified during ANC [14].

In sub-Saharan Africa, an estimated 900,000 babies die as stillbirths during the last twelve weeks of pregnancy. It is estimated that babies who die before the onset of labour, or antepartum stillbirths, account for two-thirds of all stillbirths in countries where the mortality rate is greater than 22 per 1,000 births – nearly all African countries. Antepartum stillbirths have a number of causes, including maternal infections notably syphilis and pregnancy complications, but systematic global estimates for causes of antepartum stillbirths are not available. Newborns are affected by problems during pregnancy including preterm birth and restricted fetal growth, as well as other factors affecting the baby's development such as congenital infections and fetal alcohol syndrome.

2.3 Determinants of ANC Uptake

Disparities in ANC uptake between urban and rural areas, across regions, and by women socio-economic status and women's fertility behaviors have been documented. Women with shorter preceding birth interval were less likely to uptake ANC. Lower ANC use was also recorded among women whose pregnancy was unintended [15]. A study reported that wealth status, age, ownership of health insurance (especially for rural women), educational attainment, birth order, religion and administrative region of residence were significant predictors of the intensity of antenatal care services utilization. In particular, the utilization rate increases in wealth status. Utilization of these services was very low among rural women as compared to those living in urban areas [9].

In Ethiopia, educational status of the mother, household wealth, place of residence, birth order of the child and educational and occupational status of the husband were found to be strong indicators of utilization of antenatal care service visits in the total sample of women [15]. Antenatal care use was found to be a strong indicator of use of assistance during delivery. The report made an advice that to increase women's utilization of health care services and improve maternal health in Ethiopia some crucial steps should be taken on educating women and strengthening antenatal care services. Furthermore, great attention should be given to the most vulnerable group of women in the country this includes those who are living in rural areas with no education and in the low economic status group [15]. In Metekel zone, Northwest Ethiopia, 49.8% of pregnant women had received at least one antenatal care visit during the pregnancy of their last delivery [18]. According to the study report, lack of awareness, low educational status and socio-economic characteristics, place of residence, educational status, husband's educational status, possessing radio, monthly income and knowledge about antenatal care were found to have a statistically significant reasons mentioned for not attending antenatal care utilization in the zone [18].

The proportion of women who received antenatal care for their recent births in Samre Saharti District, Tigray, Ethiopia was 54% [58]. According to the study, education, parity, family education, history of obstructed labor and ANC visit were significant predictors for the selection of delivery place. About (55.7%) of the married women used ANC service compared single 32.3%; about 78.5% of women with primary education and 86% with secondary education received ANC while it was 52% among those who were illiterate [58]. Mothers with primary education were three times higher to receive ANC than those who were illiterate, and mothers with secondary education were six times more likely to receive ANC than those who were illiterate[58]. Similarly, in Maichew Town, Southern Tigray Ethiopia, 80% of pregnant women had at least one antenatal visit during their pregnancy period [22]. The study reported that among the antenatal user's 6.3% had only one or two antenatal contacts and 15.8% had three antenatal visits. Majority of the attendees (77.9%) reported to have four or more antenatal visits at the time of the interview [22]. The main reasons for nonattendance in this area were found to be absence of illness, being too busy, long waiting time, husbands disapproval, poor quality of services, and others [22]. On the contrary, a study conducted in Southwestern Ethiopia in 2009 [3] showed that 28.5% of pregnant women in Yem Special Woreda received ANC at least once but the majority 71.5% reported that they did not attend ANC up to their last pregnancy. The study reported that no illness experienced during pregnancy, lack of awareness about ANC, far distance from health facility, being too busy and husband disapproval as the major reasons for not attending ANC visits [3]. In 2010, 86.3% in Hadiya Zone of Southern Ethiopia had received at least one antenatal visit during their last pregnancy [59]. Maternal age, husband attitude, family size, maternal education, and perceived morbidity were major predictors of antenatal care service utilization [59].

2.4 Count Data Models

Count data often arise as a counting process in which the counts are nonnegative, discrete, and constrained by a lower bound, which is typically zero. The lower bound constraint presents the greatest obstacle for analyzing count data when assuming a normal distribution. It is common for this type of data to have a skewed distribution with variance that increases as the count levels increase. Therefore, standard models, such as ordinary least squares regression, are not appropriate. Cameron clarified that the use of standard OLS regression leads to significant deficiencies unless the mean of the counts is high [19]. Several models have been proposed for analyzing data characterized by a preponderance of zeros. Substantively, the choice between these models should be based solely on the data generating process. However, datasets can vary as a function of both the proportions of zeros and the distribution for the non-zeros.

Sometimes overdispersion of a data may not be significant if the percentage of zeros is too high (might be 80% or more) and in such case ZIP and ZINB have nearly identical estimate of the parameters [39]. But the paper suggests that ZIP does not fit the data well, if there is over-dispersion with moderate percentage of zeros. Hurdle model has a higher flexibility to fit a model with mixture of distribution for zeros and positive counts. And it performs in a competitive way with ZIP and ZINB [39].

The best-fitting zero-inflated model sometimes depends on the proportion of zeros and the distribution for the non-zeros [19]. For the positively skewed distribution, Cameron suggests that the negative binomial Hurdle model should be chosen regardless of the proportion of zeros. This was also true for the negatively skewed distribution. However, for the normal distribution, the more complicated negative binomial Hurdle model may not be necessary. This provides a guideline for choice between the Hurdle and negative binomial Hurdle models for the distributions.

According to some study, the negative binomial and ZIP model appears to be superior when the event -stage distribution is positive and when there is moderate to moderately-high zero-inflation but not extreme zero –inflation [19, 39].

CHAPTER THREE
METHODOLOGY

3.1 Sources of Data

The data used for this study was taken from the 2011 Ethiopian Demographic and Health Survey which is a nationally representative survey of women in the 15-49 years age groups which was taken from the Central Statistical Agency (CSA), Ethiopia. Women who had 9 months pregnancy during the survey interview were included in the analysis.

The 2011 Ethiopian Demographic and Health Survey (EDHS) is the third compressive survey designed to provide estimates for the health and demographic variables of interest for the following domains: Ethiopia as a whole; urban and rural areas of Ethiopia (each as a separate domain); and 11 geographic areas (9 regions and 2 city administrations).

This study aimed to analyze responses from 1127 (27.03%) rural women (only those who had at least 9 months of pregnancy period during survey) out of 37431 (82.19%) rural women of age 15-49 interviewed in 2011 DHS. Since the main target of the study is rural Ethiopia, Addis Ababa city administrations, majority of Dire Dawa city administrations and Harari region were not included to this study. The rest pregnat women who had 9 months of pregnancy in Dire Dawa city administrations and Harari region were included under the better pregressed regions assuming they are sorrounded by Oromiya region which one among petter progressed regions.

3.2 Variables Included in the Model

The response variable of this study is a count, which is the number of antenatal care visits of pregnant women from early pregnancy to their 9 months of pregnancy period in rural Ethiopia. Thus, number of ANC visits takes descrete values starting from zero to number of visit counts in last. This paper attempts to include the potential (barriers) in the count number of antenatal care service visits, adopted from literature reviews and their theoretical justification from the source data. The explanatory variables at individual mothers to be analyzed are grouped as socioeconomic, demographic and health and environmental related factors.

The Socio economic variables under consideration are economic status of mother, workload of mother, if the mother residing with her husband or not, mother education, and region are the demographic variable considered; and availability and accessibility of health post and awareness about the use of ANC and pregnancy complications are considered as health and environmental variables. Whether the pregnancy is wanted and preceding birth intervals are considered in women's fertility behaviors. Detailed descriptions of these are presented in Table 1.

3.2.1 Descriptions of the Variables

Table 1: Variable Description for the Analyzed ANC Visits Dataset

Dependent Variable	Description
ANC	The number of Antenatal service visits
Independent Variable	
MEDUC	Mother educational: (0) if she has no education, (1) otherwise.
REGION	Region: (0) Women from better progressed regions, (1) Otherwise
RESID	Pregnant mother residing with husband/partner: (0) if No, (1) Yes
WLOAD	Workload inside and/or outside home: (0) if no problem, (1) else
WEALTH	Wealth index: (0) if poor, (1) if middle, (2) if rich
HPOST	Availability & accessibility of health post: (0) if no problem, (1) else
AWARN	Awareness about ANC & pregnancy complication: (0) if no, (1) yes
SIGN	Had seen sign of pregnancy complications: (0) if no, (1) yes
PWANTD	Pregnancy wanted when became pregnant: (0) if no, (1) yes

3.2.2 Count Data

An event count refers to the number of times an event occurs within a fixed interval such as the number of failures of electronic components per unit of time, the number of traffic accidents per day, or the number of patents applied for and received, the number of individuals arriving at a serving station and etc. In such type of situations, the response variable of interest is often measured as a nonnegative integer or count.

The Poisson regression is commonly used method to model count data formed under two principal assumptions: one is that events occur independently over given time or exposure period and the other is that the conditional mean and variance are equal. However, in practice, the equality of the mean and variance rarely occurs; the variance may be either greater or less than the mean. If the variance is greater than the mean, it means that counts are more variable than specified by the Poisson events and are described as overdispersion. If the variance is less than the mean, it means that counts are less variable than specified by the Poisson events and are described as underdispersion. However, in practice, underdispersion is less common [38].

One general cause of overdispersion is excess number of observed zero counts, since the excess zeros will give smaller conditional mean than the true value. The count data with excess zeroes is known as zero-inflated Poisson counts. Of course it is possible to have fewer zero counts than expected, but this is again less common in practice [47].

In the literature of statistical modeling for counts there are number of models proposed to handle zero-inflated counts, for example, Hurdle model [19], Two-part model [23], Zero-modified distributions [13], and Zero-inflated Poisson (ZIP) models and Zero-inflated Negative Binomial (ZINB) models [29]. This thesis focus on Poisson, Negative Binomial, ZIP, ZINB, Hurdle Poisson models, Hurdle Negative Binomial models and accessed different tests for comparing their performances. The choice between the models should be guided by the researcher's beliefs about the source of the zeros. Beyond this substantive concern, the choice should be based on the model providing the closest fit between the observed and predicted values. Unfortunately, the literature presents anomalous findings in terms of model superiority [11].

3.3 Statistical Models

Even though there are several statistical models, some models may not be appropriate to deal with some specific types of data. Their use is solely depending on the types and nature of the data. In this study, the variable of interest is a count data, which is most often characterized as non-normal distribution. Thus, to deal with the data and methodological issues associated with number of ANC visits, a wide variety of statistical methods which can be used to model count data was discussed in the next subsections.

3.3.1 Generalized Linear Regression Models

The GLM is defined in terms of a set of independent random variables $Y_1, Y_2, \dots, Y_N$, satisfies two properties:

(1) The distribution of each Y_i belongs to the exponential family in same canonical form and depends on a single parameter μ_i, though μ_i do not have to be the same for all i.

(2) The distribution of all the Y's are of the same form.

Usually, the parameters μ_i does not serve as parameters of our interest since there will be too many unknown parameters to be estimated. For model specification we focus more on a smaller set of parameters $\beta_1, \beta_2, \dots, \beta_p$, where p<<N [1].

For a GLM there is a transformation of μ_i such that $E(Y_i) = \mu_i$ and $g(\mu_i) = x_i^T \beta$. Function

g is a monotone, differentiable function called the link function, which provides the relationship between the linear predictor and the mean of the distribution function. So the parameter

θ_i are replaced by the parameter β , which makes the estimation process easier [38].

For most analyses of continuous data, the linear models are set under assumption that the random variables Y_i are independent and $Y_i \sim N(\mu_i,\ \delta^2)$ then $E(Y_i) = \mu_i = x_i^T \beta$. Compared to linear models, GLMs are more applicable to solve problems under more general situations as follows:

(1) Dependent variable can have a distribution other than the Normal distribution. It can have any distribution belong to exponential family in canonical form.

(2) Relationship between dependent and predictor variables need not be of the simple linear form as above.

There are several advantages to introduce GLMs.

(1) We don’t have to transform dependent variable Y to normality.

(2) Many “nice” properties of the Normal distribution are shared by the exponential family of distributions.

(3) There can be some non-linear function relating $E(Y_i) = \mu_i$ to $x_i^T \beta$, that is, $E(Y_i) = x_i^T \beta$.

Such models have now been further generalized to situations where functions may be estimated numerically.

There are a range of techniques which had been developed for analyzing data with count or frequency response variables. For this study, some extension of generalized linear models such as poisson regression, Negative Binomial regressions and other modes of them like Zero-Inflated poisson regression, Zero-Inflated Negative Binomial regression, and Hurdle Poisson model and Hurdle Negative Binomial models was applied [48].

3.3.1.1 Poisson Regression Model

Because antenatal care visits-frequency data are non-negative integers, the application of standard ordinary least-squares regression (which assumes a continuous dependent variable) is not appropriate. Given that the dependent variable is a non-negative integer, most of the recent thinking in the field has used the Poisson regression model as a starting point. In a standard Poisson regression model, the probability of pregnant women i having y_i antenatal care service visits until nine (9) months of pregnancy period (where y_i is a non-negative integer) is given by:

$p(y_i) = \frac{Exp(-\mu_i)\mu_i{}^{y_i}}{y_i!}$, y_i= 0, 1, 2, … and ($\mu_i > 0$) [44] …………………………….. (**Eq 1**)

Where $p(y_i)$ is the probability of 9 month pregnant women entity i having y_i antenatal care service visits in nine (9) months of pregnancy period and μ_i is the Poisson parameter for pregnant women i, which is equal to 9 month pregnant women entity i's expected number of antenatal care service visits in nine (9) months, $E(y_i)$. Poisson regression models are estimated by specifying the Poisson parameter μ_i (the expected number of antenatal care service visits) as a function of explanatory variables, the most common functional form being $\mu_i = Exp(\beta X_i)$, where X_i is a vector of explanatory variables and β is a vector of estimable parameters.

The log-likelihood function is: $l(\mu_i) = l(\mu_i; y) = \sum_{i=1}^{n}\{y_i \ln(\mu_i) - \mu_i - \ln(y_i!)\}$. …….. (**Eq 2**)

Let X be a $n\times(p+1)$ matrix of explanatory variables. The relationship between y_i and i^{th} row vector of X, x_i linked by $l(\mu_i)$ is: $ln(\mu_i) = \eta_i = x_i^T\beta = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ [44]

There are two principal assumptions in the Poisson model we need to regard: one is that events occur independently over time or exposure period, the other is that the conditional mean and variance are equal [4]. The latter assumption is quite important. If it fails, the fitted model should be reconsidered.

Although the Poisson model has served as a starting point for count or frequency analysis for several decades, researchers have often found that count data exhibit characteristics that make the application of the simple Poisson regression (as well as some extensions of the Poisson model) problematic. Specifically, Poisson models cannot handle over- and under-dispersion and they can be adversely affected by low sample means and can produce biased results in small samples.

There are two basic criteria commonly used to check the presence of over-dispersion: the deviance, $D(y;\widehat{\mu_i})$ or the Pearson (χ^2) statistic be greater than its degrees of freedom [20]. For the Poisson regression, $D(y;\widehat{\mu_i})$ and χ^2 are respectively defined in expression $D(y;\widehat{\mu_i}) = 2 \times \sum_{i=1}^{n}\left\{y_i ln\left(\frac{y_i}{\widehat{\mu_i}}\right) - (\mu_i - \widehat{\mu_i})\right\}$; $\chi^2 = \sum_{i=1}^{n}\frac{(\mu_i-\widehat{\mu_i})}{\widehat{\mu_i}}$ [53]. …………………………….. (**Eq 3**)

However, these two rules of thumb can yield misleading inference from a direct likelihood point of view. Therefore, selecting between Poisson regression and an over-dispersed Poisson model should be performed using some appropriate modeling procedure.

3.3.1.2 Negative Binomial Regression Model

The negative binomial (or Poisson-gamma) model is an extension of the Poisson model to overcome possible over-dispersion in the data. The negative binomial/Poisson-gamma model assumes that the Poisson parameter follows a gamma probability distribution. The model results in a closed-form equation and the mathematics to manipulate the relationship between the mean and the variance structures is relatively simple.

The negative binomial model is derived by rewriting the Poisson parameter for each observation i as $\mu_i = Exp(\beta X_i + \varepsilon_i)$ where $Exp(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance k. The addition of this term allows the variance to differ from the mean as:

$Var[y_i] = E[y_i][1 + kE[y_i]] = E[y_i] + kE[y_i]$. The probability mass function for the negative binomial distribution is: $p(Y_i = y_i) = \binom{y_i + r - 1}{y_i} p^r (1-p)^{y_i}, r = 0,1,2, \ldots$ [44] …….. (**Eq 4**)

The parameter p is the probability of success in each trial and it is calculated as:

$p = \frac{r}{\mu_i + r}$ where, $\mu_i = E(Y) =$ mean of the observations; and $r =$ inverse of the dispersion parameter k $(i.e. r = \frac{1}{k})$. When the parameter r is extended to a real, positive number, its PMF can be rewritten using the gamma function:

$p(Y_i = y_i) = \frac{\Gamma(y_i + r)}{\Gamma(r)\Gamma(y_i + 1)} p^r (1-p)^{y_i}, y_i \epsilon \{0\} U Z^+$ …………………………….. (**Eq 5**)

Where $\Gamma(.)$ is the gamma function. The mean and variance of the negative binomial are $E[y_i] = \mu = r\frac{1-p}{p}$ and $Var[y_i] = \frac{1-p}{p^2}$ [8]. It is common to parameterize r and p in the terms of k and μ. Define $k = \frac{1}{r}$, $\mu = \frac{1-p}{kp}$, solving yields $p = \frac{1}{(1+k\mu)}$. After the re-parameterization, the above model becomes $p(Y_i = y_i) = \frac{\Gamma\left(y_i + \frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i + 1)} (\frac{1}{(1+k\mu)})^{\frac{1}{k}} \left(\frac{k\mu}{1+k\mu}\right)^{y_i}$…………………….. (**Eq 6**)

The mean of this parameterization is $E[y_i] = \mu$ and $Var[y_i] = \mu + k\mu^2$. This is known as the “NB-2” model because it has a quadratic variance function. In this model $k \geq 0$ and if $k = 0$, then it reduces to a Poisson.

The negative binomial model can be estimated using maximum likelihood. The NB2 likelihood function is: $l(\mu_i|k, y_i) = \sum_{i=1}^{n}[y_i \ln(\frac{k\mu_i}{k\mu_i+1}) - \frac{1}{k}\ln(k\mu_i + 1) + ln\,\Gamma\left(y_i + \frac{1}{k}\right) - ln\,\Gamma(y_i + 1) - ln\Gamma\left(\frac{1}{k}\right)$ ………………………………………**(Eq 7)**

The NB2 model is less robust to distributional misspecification than the Poisson model where one could use a pseudo-maximum likelihood estimator.

In the NB regression model, μ_i is linked to the covariates: $\mu_i = Exp(x_i\beta)$.

In the context of the NB GLM, the mean response for the number of antenatal care service visits is assumed to have a log-linear relationship with the covariates and is structured as:

$ln(\mu_i) = \beta_0 + \sum_{i=1}^{p}\beta_i\, x_i$ …………………………….. **(Eq 8)**

Where, $x_i =$ selected determinants of the barriers in number of ANC; $\beta's =$ regression coefficients to be estimated; and, p = total number of covariates in the model [44].

The Poisson regression model is a limiting model of the negative binomial regression model as k approaches zero, which means that the selection between these two models is dependent upon the value of k. The parameter k is often referred to as the overdispersion parameter.

The Poisson-gamma/negative binomial model is the probably the most frequently used model in crash-frequency modeling. However, the model does have its limitations, most notably its inability to handle under-dispersed data, and dispersion-parameter estimation problems when the data are characterized by the low sample mean values and small sample sizes [3, 32, 35]. Although the negative binomial model can solve an overdispersion problem, it may not be enough flexible to handle when there are excess zeros. In such cases, one can use the zero-inflated models (zero-inflated Poisson or zero inflated negative binomials) as well as hurdle models (Hurdle Poisson or Hurdle negative binomial model) to solve the problem.

3.3.2 Zero-Inflated Models

There are situations where a major source of overdispersion is a preponderance of zero counts, and the resulting overdispersion cannot be modeled accurately with negative binomial model. In such scenarios, one can use zero-inflated Poisson or zero-inflated negative binomial model to fit the data. The first concept of a zero–inflated distribution originated from the work of [47] who examined the characteristics of mixed Poisson distributions [36].

According to Lord, Zero-inflated techniques permit the researcher to answer two questions that pertain to low base rate-dependent variables: (a) what predicts whether or not the event occurs, and (b) if the event occurs, what predicts frequency of occurrence? In other words, two regression equations are created: one predicting whether the count occurs and a second one predicting the occurrence of the count [32]. Moreover, zero-inflated models have statistical advantage to standard Poisson and negative binomial models in that they model the preponderance of zeros as well as the distribution of positive counts simultaneously[41]. In next sections, zero-inflated Poisson and zero-inflated negative binomial models will be discussed briefly.

3.3.2.1 Zero-Inflated Poisson (ZIP) Regression Models

Zero-inflated models have been developed to handle data characterized by a significant amount of zeros or more zeros than the one would expect in a traditional Poisson or negative binomial model. Zero-inflated models operate on the principle that the excess zero density that cannot be accommodated by a traditional count structure is accounted for by a splitting regime that models a women who are not visited for antenatal care versus a women who have visited for antenatal care during their pregnancy period. The probability of an antenatal care visitation entity being in zero or non-zero states can be determined by a binary logit or probit model [29, 54].

The essential idea is that the data come from two regimes. In one regime (R_1) the outcome is always a zero count, while in the other regime (R_2) the counts follow a standard Poisson process. Suppose that: $p[y_i \epsilon R_1] = \omega_i$; $p[y_i \epsilon R_2] = (1 - \omega_i); i = 1, 2, \ldots\ldots, n.$ ω_i = Inflation Probability

Then, this two-state process gives a simple two-component mixture distribution with PMF [26].

$$p(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\mu_i};\ when\ y_i = 0 \\ (1-\omega_i)\frac{e^{-\mu_i}\mu_i{}^{y_i}}{y_i!};\ \ when\ y_i > 0 \end{cases} \mu_i > 0; and\ 0 \le \omega_i \le 1 \dots\dots\dots\dots\dots\dots.. (\textbf{Eq 9})$$

As before, covariates enter the model through the conditional mean, μ_i, of the Poisson distribution: $\mu_i = Exp(x_i^T\beta)$, where x_i^T is a $(1 \times p)$ vector of the i^{th} observation on the covariates, and β is a $(p \times 1)$ vector of coefficients.

Clearly, $E(y_i) = (1-\omega_i)\mu_i = \mu_i$ and $Var(y_i) = \mu_i + \left(\frac{\omega_i}{1-\omega_i}\right)\mu_i{}^2 = (1-\omega_i)(\mu_i + \omega_i\mu_i{}^2)$ indicating that the marginal distribution of y_i exhibits over-dispersion of the data $(if\ \omega_i > 0)$. It is clear that this reduces to the standard Poisson model when $\omega_i = 0$. This over-dispersion does not arise from heterogeneity, as is case when the Poisson model is generalized to the Negative Binomial model. Instead, it arises from the splitting of the data into the two regimes. In practice, the presence of over-dispersion may come from one or both of these sources [17, 30].

Following Lambert, 1992, it is common, and convenient, to model ω_i using a Logit model, so: $\omega_i = \frac{\exp(Z_i^T\gamma)}{1+\exp(Z_i^T\gamma)}$, where Z_i is a $(1 \times p)$ vector of the i^{th} observation on some covariates, and γ is a $(p \times 1)$ vector of additional parameters [29]. Of course, the elements of Z_i may include elements of x_i, and a Probit (or other) specification may be substituted for the Logit specification. The covariates can be incorporated by using a log link for μ_i and a logit link for ω_i, $ln(\mu_i) = x_i^T\beta$ and $ln\left(\frac{\omega_i}{1-\omega_i}\right) = Z_i^T\gamma$; Where x_i and Z_i are the vectors of explanatory variables, and γ and β are the vectors of regression parameters. Maximum likelihood estimates can be obtained by maximizing the log likelihood which may be written as

$$logL(\beta,\gamma) = \sum_{y_i=0}\log\left[\exp(Z_i^T\gamma) + \exp\left(-\exp(x_i^T\beta)\right)\right] + \sum_{y_i\neq 0}[y_i\, x_i^T\beta - \exp(x_i^T\beta) - \log(y_i!)] - \sum_{i=1}^{n}\log\left[1 + \exp(Z_i^T\gamma)\right]\ [4]\ .\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.\ (\textbf{Eq 10})$$

To code up the above log-likelihood function for use in R packages, we need to take account of the different ranges of summation. The third term in the log-likelihood requires no modification as the range of summation is for all n. To deal with the ranges of summation in the first two terms, we can construct a dummy variable, D_i, which takes the value unity if $y_i = 0$, and zero otherwise as follows. $D_i = \begin{cases} 1\ if\ y_i = 0 \\ 0\ Otherwise \end{cases}$.

The i^{th} observation on the log-likelihood would then be coded as:

$$logL_i(\beta,\gamma) = D_i \log[\exp(Z_i^T\gamma) + \exp(-\exp(x_i^T\beta))] + (1-D_i)[y_i x_i^T\beta - \exp(x_i^T\beta) - \log(y_i!)] - \log[1+\exp(Z_i^T\gamma)] \text{ .... } \textbf{(Eq 11)}$$

Since its inception, the zero-inflated model (both for the Poisson and negative binomial models) has been popular among transportation safety analysts [7, 27, 30, 48, 49]. Despite its broad applicability to a variety of situations where the observed data are characterized by large zero densities, others have criticized the application of this model in highway safety. For instance, Lord et al. argued that, because the zero or safe state has a long-term mean equal to zero, this model cannot properly reflect the crash-data generating process [33, 34].

3.3.2.2 Zero-Inflated Negative Binomial (ZINB) Regression Models

Similar to ZIP regression above, Zero-Inflated Negative Binomial (ZINB) regression model assumes there are two distinct data generation processes. The result of a Bernoulli trial is used to determine which of the two processes is used. For mother i, with probability ω_i the only possible response of the first process is zero counts, and with probability of $(1-\omega_i)$ the response of the second process is governed by a negative binomial with mean μ_i. The zero counts are generated from both the first and second processes, where a probability is estimated for whether zero counts are from the first or the second process. The overall probability of zero counts is the combined probability of zeros from the two processes.

A ZINB model for the response y_i (the number of ANC visits during pregnancy) can be written as:

$$P(Y = y_i) = \begin{cases} \omega_i + (1-\omega_i)(1+v\mu_i)^{\frac{1}{k}}; & when\ y_i = 0 \\ (1-\omega_i)\frac{\Gamma(y_i+\frac{1}{k})}{\Gamma(y_i+1)}\frac{(k\mu_i)^{y_i}}{(1+k\mu_i)^{y_i+\frac{1}{k}}}; & when\ y_i > 0 \end{cases} \text{ .... } \textbf{(Eq 12)}$$

In this case, the mean and variance of the y_i are: $E[y_i] = (1-\omega_i)\mu_i$ and $Var[y_i] = (1-\omega_i)\mu_i(1+\mu_i(\omega_i+k))$. Where μ_i is the mean of the underlying negative binomial distribution, and v is the over-dispersion parameter [29]. The ZINB distribution reduces to the ZIP distribution as $k \to 0$. The parameter μ_i is modeled as a function of a linear predictor, that is, $\mu_i = Exp(x_i^T\beta)$. β is the $(p+1)\times 1$ vector of unknown parameters associated with the known covariate vector $x_i^T = (1, x_{i1}, \dots, x_{ip})$, where p is the number of covariates not including the intercept. The parameter ω_i, which is often referred as the zero-inflation factor, is the probability of zero counts from the binary process. For common choice and simplicity, ω_i is characterized in terms of a logistic regression model by writing as $logit(\omega_i) = Z_j^T \gamma. \gamma$ is the $(q+1) \times 1$ vector of zero-inflated coefficients to be estimated, associated with the known zero-inflation covariate vector $Z_j^T = (1, Z_{j1}, \dots, Z_{jq})$, where q is the number of the covariates Z's not including the intercept. In the terminology of generalized linear models (GLMs) log (μ_i) and logit (ω_i) are the natural links for the negative binomial mean and Bernoulli probability of success [29].

$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ and

$\text{logit}\ (\omega_i) = \gamma_0 + \gamma_1 z_{i1} + \cdots + \gamma_p z_{iq}$ …………………………….. **(Eq 13)**

where X_i and Z_i are respectively vectors of covariates for the negative binomial and the logistic components, and β and γ are the corresponding vectors of regression coefficients.

3.3.3 Hurdle Regression Models

A hurdle model is “a modified count model in which the two processes generating the zeros and the positives are not constrained to be the same” [4].

Originally developed by Mullahy (1986), Hurdle regression is also known as two-part model [41]. Mullahy states, “The idea underlying the hurdle formulations is that a binomial probability model governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is non- zero (positive), the “hurdle is crossed”, and the conditional distribution of the positives is governed by a truncated-at-zero count data model.” The attraction of Hurdle regression is that it reflects a two-stage decision-making process in most human behaviors and therefore has an appealing interpretation. For instance, it is pregnant mother’s decision whether to contact the doctor’s office and to make the initial visit. However, after the pregnant mother’s first visit, doctor plays a more important role in determining if the pregnant mother needs to make follow-up visits. Therefore, in a regression setting, the first decision might be reflected by a Logit or Probit regression, while the second one can be analyzed by a truncated Poisson or Negative binomial regression. Moreover, different explanatory variables are allowed to have different impacts at each decision process.

3.3.3.1 Hurdle Poisson (HP) Regression Model

The most popular formulation of a Hurdle regression is called Logit-Poisson model, which is the combination of a Logit regression modeling zero vs. nonzero outcomes and a truncated Poisson regression modeling positive counts conditional on nonzero outcomes. Its probability density function is given as:

$$p(y_i/x_i) = \begin{cases} \omega_i & for \quad y_i = 0 \\ \frac{(1-\omega_i)Exp(\mu_i){\mu_i}^{y_i}}{(1-Exp(-\mu_i))y_i!} & for \ \ y_i > 0 \end{cases} \qquad Where: \omega_i = p(y_i = 0), \ \mu_i = Exp(x_i\beta),$$

$log\left(\frac{\omega_i}{1-\omega_i}\right) = {Z_i}^T\gamma \ \ and \ log(\mu_i) = \ {X_i}^T\beta$ [41] . …………………………….. **(Eq 14)**

The log-likelihood function of a Logit-Poisson regression therefore can be expressed as the sum of log-likelihood functions of two components as below:

$LL = \sum_{i=1}^{n}\left[I_{y_i=\mathbf{0}} \log(\boldsymbol{\omega_i}) + I_{y_i>0} \log(\mathbf{1} - \ \boldsymbol{\omega_i}) - \boldsymbol{\mu_i} + \boldsymbol{y_i} \log(\boldsymbol{\mu_i}) - \log(\mathbf{1} - \boldsymbol{Exp}(-\boldsymbol{\mu_i}) -\right.$ $\left.\log(\boldsymbol{y_i}!))\right]$. …………………………….. **(Eq 15)**

Unlike Poisson and Negative binomial regressions, Hurdle regression can only be modeled through log-likelihood function.

3.3.3.2 Hurdle Negative Binomial (HNB) Regressions Model

We consider a hurdle negative binomial regression model in which the response variable $y_i(i = 1, \dots, n)$ has the distribution

$$p(Y_i = y_i) = \begin{cases} \omega_i, & when \ \ y_i = 0 \\ (1-\omega_i)\frac{\Gamma(y_i+k^{-1})}{\Gamma(y_i+1)\Gamma(k^{-1})}\frac{(1+k\mu_i)^{-k^{-1}-y_i}k^{y_i}{\mu_i}^{y_i}}{1-(1+k\mu_i)^{-k^{-1}}}, y_i > 0 \end{cases} \quad \text{…………………. } \mathbf{(Eq\ 16)}$$

Where $(k \geq 0)$ is a dispersion parameter that is assumed not to depend on covariates [41]. In addition, we suppose $0 < \mu_i < 1$ and $\omega_i = \omega_i(z_j)$ satisfy

$logit(\theta_i) = \log\left(\frac{\boldsymbol{\omega_i}}{1-\boldsymbol{\omega_i}}\right) = \sum_{\mathrm{i}=1}^{\mathrm{q}} \boldsymbol{Z_i\gamma}$,

$log(\mu_i) = \sum_{\mathrm{i}=1}^{\mathrm{p}} \boldsymbol{X_i\beta}$ …………………………….. **(Eq 17)**

Where Z_i and X_i are the i^{th} row of covariate matrix Z and X as well as β and γ are the independent variables in the regression model. We now obtain the log-likelihood function for the hurdle negative binomial regression model, we have:

$LL = \sum_{i=1}^{n}\{(1-d_i)[I_{y_i=0}\log\omega_i + I_{y_i>0}\{log(1-\omega_i) + log - log(1-(1+k\mu_i)^{-k^{-1}})\}] + d_i log \sum_{j=y_i}^{\infty} pr(Y_j = j)\}$ …………………………….. **(Eq 18)**

In many applications, extra zeros (relative to the Poisson model) generated by the above models are insufficient to account for the full amount of zeros in the data. All single index models have to compromise between the large proportion of zeros, which tends to lower the mean, and a right-skewed distribution of counts with large non-zero values, which tends to increase it. Moreover, one often has a substantive interest in treating the zero-generating process separately from the process for strictly positive outcomes, which requires different sets of parameters.

3.4 Goodness of Fit

3.4.1 Likelihood and Deviance Residual

The likelihood function can be used to assess the goodness of fit of a model, and several further measures of model performance are based on it. It is to note that this assumes mutual independence of observations. In case the observations are not mutually independent, the likelihood will be overestimated. This will have the effect of exaggerating differences in log-likelihood and so will tend to favor elaborate models unduly.

Deviance provides an alternative to likelihood. The deviance is used as a measure of discrepancy of a generalized linear model; each unit i of observation contributes an amount D_i as an increment to total deviance. For the Poisson model with observed number y_i and corresponding estimated number u_i, residual deviance is given by:

$D_i = sign(y_i - u_i)\sqrt{d_i^2}$ [21]. …………………………….. **(Eq 19)**

Where d_i^2 is the squared deviance residual which can be obtained according to the distribution as follows:

Poisson regression: $d_i^2 = \begin{Bmatrix} 2u_i & if\ y_i = 0 \\ 2\left\{y_i \ln\left(\frac{y_i}{u_i}\right) - (y_i - u_i)\right\} & otherwise \end{Bmatrix}$ [21]. ………….. **(Eq 20)**

NB regression: $d_i^2 = \begin{Bmatrix} \frac{2ln\,(1+\alpha u_i)}{\alpha} & if\ y_i = 0 \\ 2y_i\, ln\left(\frac{y_i}{u_i}\right) - \frac{2}{\alpha}(1+\alpha y_i)ln\left(\frac{1+\alpha y_i}{1+\alpha u_i}\right) & otherwise \end{Bmatrix}$ [21] …….. **(Eq 21)**

Where α is the over-dispersion parameter. The standardized residuals were obtained by multiplying the deviance residual D_i by the factor $(1-h_i)^{-\frac{1}{2}}$ where h_i is the leverage, which indicates the influence of observation i.

The total residual deviance D of the model is given by summation over all units:

$D = \sum_{i=1}^{n} D_i$. For Poisson, a properly fitted model the expected value of residual deviance should be approximately equal to the residual degrees of freedom [37].

3.4.2 Likelihood Ratio Test

The maximum likelihood estimation method is used to assess the adequacy of any two or more than two nested models by using the likelihood ratio test. it compares the maximum likelihood under the alternative hypothesis with the null hypothesis. For instance, the null hypothesis can be the overdispersion parameter is equal to zero (i.e. the Poisson distribution can be fitted well the data) and the alternative hypothesis can be the data would be better fitted by the Negative binomial regression (i.e. the overdispersion parameter is different from zero). The likelihood ratio test is defined as:

$$R_\omega = -2 \times [l(\hat{\mu}) - l(\hat{\mu}, \hat{\omega})] \text{ ……………………………..} \textbf{ (Eq 22)}$$

$l(\hat{\mu})\ and\ l(\hat{\mu}, \hat{\omega})$ are the maximized log-likelihood of models under the alternative and null hypothesis respectively. From the earlier computations, this likelihood ratio test can be written as

$$R_\omega = 2 \times \{n_0\, ln\left(\frac{n_0}{n}\right) + (n - n_0)\left(\ln\left(\frac{\bar{y}}{\hat{\mu}}\right) - \hat{\mu}\right) + n\bar{y}\,(ln\,\hat{\mu} + 1 - ln\bar{y}], \text{ ………………..} \textbf{ (Eq 23)}$$

Where $\bar{y}$ is the mean of the observations under H_0 and $\hat{\mu}$ is the estimated positive mean counts under H_1. This test statistic R_ω approximately follows chi-square distribution on 1 degree of freedom (d.f) under the null hypothesis.

This has a chi-square distribution. As a result this test of statistics will be compare with the tabulated chi-square with a degree of freedom, the difference between the degree of freedom of the model under null hypothesis and the alternative hypothesis respectively. This method is not appropriate for models which are not nested one on the other, in such situation; we will use another method such as the Akaike information criteria (AIC) and Bayesian information criteria (BIC) [25].

In this study a likelihood ratio was used to compare the Poisson with the negative binomial and zero-inflated Poisson with zero-inflated negative binomial as well as Hurdle Poisson with Hurdle Negative Binomial since Poisson is nested on negative binomial and zero-inflated Poisson is nested in zero-inflated negative binomial; However this will not be used to compare Poisson or negative binomial with the zero inflated Poisson and negative binomial as long as these models are not nested one on the other.

3.4.3 Variance Inflation Factor

The variance inflation factor (VIF) is used to quantify multicollinearity among the explanatory variables. Stata estimated the values of VIF which can be used to adjust the standard errors of the parameter estimates, due to the presence of collinearity. A maximum acceptable value of 10 as proposed by Kutner (2004) is adopted in this study [28]. The following formula is used in Stata to estimate the value of VIF.

$VIF = \frac{1}{\left(1-R_j^2\right)}$ [28]. ……………………………….. **(Eq 24)**

Where $j = 1, 2, \ldots.., p$ and R_j^2 is the multiple correlation coefficient of x_j on the other explanatory variables.

3.4.4 Information Criteria

If there are several models to be compared in order to select the best model which fits the data instead of using the likelihood ratio test, it can be easily select by using the Akaike information criteria (AIC) and Bayesian information criteria (BIC).

3.4.4.1 Akaike Information Criteria (AIC)

AIC is the most common means of identifying the model which fits well by comparing two or more than two models. It is trying to balance the goodness of fit against the complexity of the model It is similar as of the coefficient of multiple determination (R^2); however, it penalized by the number of parameter included in the model (i.e. the complexity of the model). Unlike the R^2, the good model is the one which has the minimum AIC value. It is given by the following formula:

$AIC = -2l + 2k.$ ……………………………….. **(Eq 25)**

Where l are the log likelihood of a model that will compare with the other models and k is the number of parameter in the model including the intercept [21].

3.4.4.2 Bayesian Information Criteria (BIC)

Unlike the Akaike information criteria the Bayesian information matrix (BIC) takes in to account the size of the data under considered. It is given by:

$BIC = -2l + klog(n)$. [21] …………………………….. **(Eq 26)**

Where l are 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.

For this study the AIC is preferred over the BIC as it is more stringent and has a stricter entry requirement than BIC for additional parameters when large datasets are used. This helps to resolve over-fitting of models where many additional parameters are added to increase the likelihood, so AIC helps to promote a parsimonious model [51].

3.4.5 Chi-square Test

The chi-square statistic χ^2 is used to test if a sample of data came from a population with a specific distribution. The χ^2 is commonly defined by:

$\chi_\omega^2 = \sum_{k=1}^{c} \frac{(O_k - E_k)^2}{E_k}$. …………………………….. **(Eq 27)**

Where c denotes the number of classes (categories) decided for a given data set, O_k and E_k are observed frequencies and expected frequencies under the null hypothesis of the k^{th} class, respectively. When the null hypothesis is valid, χ_ω^2 follows an asymptotic chi-square distribution on $c - 1$ d.f.

3.4.6 Voung Test

The Vuong test is a non-nested test that is based on a comparison of the predicted probabilities of two models that do not nest [53]. 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 Voung test. This test is used for model comparison. Let's define: $m_i = \left(\frac{P_1(Y_i|X_i)}{P_2(Y_i|X_i)}\right)$. Where $P_N(Y_i|X_i)$ is the predicted probability of observed count for case i from model N, then Vuong test statistic test the hypothesis of $E(m_i = 0)$ given as:

$V = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})^2}}$. …………………………….. **(Eq 28)**

The test statistic provides evidence of the superiority of model 1 over model 2. If $V > 1.96$, the first model is preferred. But if $V < 1.96$, the second model is preferred.

3.5 Software

Almost all statistical computation was carried out using SAS version 9.2. For all regression modeling we used Proc NLMIXED, specifying the likelihood equations, and maximizing them directly using numerical methods. Maximization began from various starting points and the final gradient vectors and hessian matrices were investigated to ensure proper convergence of estimated model parameters. In addition, all hypotheses were tested at 0.05 level of significance. R statistical software version 3.0.3 was used for graphical purpose.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 RESULTS

4.1.1 Descriptive Statistics and Exploratory Analysis

Table 2 shows descriptive statistics of the number and percentage of ANC visits that the pregnant mothers in the sample have encountered in their nine months of pregnancy period. It can be seen that 580 (51.5%) of the pregnant mothers have not visited antenatal care service during their periods of pregnancy months, whereas 125 (11.1%) of them visited only once, 91(8.1%) of them visited twice, 77(6.8%) visited three times, 82 (7.3%) visited four times and etc. Figure 8 (at appendix B) presents the distribution of the number of ANC visits per nine months of pregnancy period. Since there is large number of zero outcomes, the histograms are highly picked at the very beginning (about the zero values). However large observations (i.e. large number of ANC visits) are less frequently observed. This leads to have a positively (or right) skewed distribution. This could be fitted better by count data models which takes into account excess zeros like zero-inflated models.

Table 2: Number of mothers that experienced ANC visits

Number of ANC visits	Percent	Cumulative Percent
0	51.5	51.5
1	11.1	62.6
2	8.1	70.6
3	6.8	77.5
4	7.3	84.7
5	4.2	88.9
6	3.5	92.4
7	2.5	94.9
8	2.0	96.8
9	1.1	97.9
10	0.4	98.3
11	0.6	98.9
12	0.8	99.7
13	0.3	100

Table 3 presents summary statistics of the variables that are assumed to affect the number of ANC visits and its distributions for each levels of the variables. The variables included were Pregnant mother's education status, Region, whether the pregnant mother is currently residing with her husband/partner, Workload inside/outside home, Wealth index, Availability and accessibility of health post, Awareness about the use of ANC and pregnancy complications, whether the pregnant women ever seen signs of pregnancy complications and If the pregnancy is wanted when became pregnant.

Table 3: Descriptive Statistics of ANC Services Utilization among Pregnant Women in Rural Ethiopia

Variable	Category	Min	Max	N (%)	Median	Mean (St. Dev)
MEDUC	No education	0	12	782 (69.4)	0.00	0.98 (1.713)
	Can read and write	0	13	345 (30.6)	4.00	3.84 (3.348)
REGION	Better progressed regions	0	13	712 (63.2)	1.00	2.43 (3.008)
	Regions wait for special aids	0	11	415 (36.8)	0.00	0.86 (1.570)
RESID	No	0	13	476 (42.2)	0.00	1.22 (2.443)
	Yes	0	12	651 (57.8)	1.00	2.32 (2.755)
WLOAD	No problem	0	13	468 (41.5)	3.00	3.62 (3.189)
	Problem	0	8	659 (58.5)	0.00	0.60 (1.138)
WEALTH	Poor	0	12	588 (52.2)	0.00	0.61 (1.295)
	Middle	0	13	426 (37.8)	2.00	2.90 (3.050)
	Rich	0	12	113 (10.0)	4.00	4.40 (3.061)
HPOST	No problem	0	13	526 (46.7)	3.00	3.28 (3.175)
	Problem	0	6	601 (53.3)	0.00	0.60 (1.152)
AWARN	No	0	9	597 (53.0)	0.00	0.51 (1.088)
	Yes	0	13	530 (47.0)	3.00	3.37 (3.102)
SIGN	No	0	4	169 (15.0)	0.00	0.62 (1.134)
	Yes	0	13	958 (85.0)	1.00	2.07 (2.815)
PWANTD	No	0	13	174 (15.4)	4.00	4.19 (3.516)
	Yes	0	12	953 (84.6)	0.00	1.43 (2.255)

Accordingly, less than one-third, 345 (30.6%) of the respondents (mothers) can read and write while more than two-third, 782 (69.4%) of them have no education. Figure 9a & 9b above again confirms that the distribution of the number of ANC visits per Region and Education status of Mother in each group differs considerably. Since there are a number of ANC visit outcomes for mothers from better progressed regions and educated mothers, the plots looks like obese after 5 ANC in both groups. However large observations (i.e. large number of ANC visits) are less frequently observed. The number of participated pregnant women from regions that need special aids (Afar, Somali, Benishangul-Gumuz, and Gambella) found to be lower 415 (36.8%) than the number of mothers from better progressed regions (Tigray, Amhara , Oromiya, and SNNPR), 712 (63.2%). It was also observed that husband or partner of 476 (42.2%) pregnant mothers were not living with them, 651 (57.8%) of them were residing with their husband or partner during the time of their pregnancy periods. About 588 (52.2%) of sampled pregnant mothers were poor, 426 (37.8%) had middle income, and 113 (10.0%) were rich [Figure 5]. The number of pregnant mothers who had a problem of workload inside and/or outside home was 659 (58.5%) and those pregnant mothers who had no problem of workload inside and/or outside home was found to be 468 (41.5%). The frequency of pregnant mothers who became pregnant unexpectedly was 174 (15.4%) and majority of them, 953 (84.6%) became pregnant eagerly.

This table (Table 3) again reflects that nationally pregnant mothers use ANC visits approximately twice (1.85 visits) per their duration of pregnancy periods with standard deviation of 2.683, which is more than the mean indicating overdispersion. The number of ANC service visits during pregnancy for educated pregnant mothers is 3.84 ≈ 4visits and 0.98≈ 1 visits for that of non-educated mothers. The average number of ANC service visits for pregnant mothers from better progressed regions is two (2.43) times while the average number of ANC visits for pregnant mothers who are from regions who need special aids such as Afar, Somali, Benishangul-Gumuz and Gambella is found to be only once (0.86). The Table reveals that the happening of the signs of pregnancy complications during their pregnancy periods such as such as vaginal bleeding, vaginal gush of fluid, severe head ache, blurred vision, fever, abdominal pain had made variations among pregnant women. Hence, pregnant women who had ever seen the signs of pregnancy complications during their pregnancy periods used ANC service visits for more than twice (2.07) at average, despite the fact that the pregnant mother who had not seen the signs were visited only less than once, (0.62).

The descriptive statistics of Table 3 further illustrates that pregnant mother who had no radio or television at home and who was not visited by family planning worker last 12 months as well as not told about pregnancy complications(average ANC visits of 0.51), mother who had a problem of the availability of nearby health post and/or a problem of access to means of transportation (average ANC visits of 0.60), mothers who had a problem of workload inside and /or outside home (average ANC visits of 0.60), and poor pregnant mothers (average ANC visits of 0.61) were found to be the least ANC service users respectively. Therefore, the average number of ANC utilization ranges from pregnant mothers who had lack of awareness about the use of ANC and pregnancy complications (0.51 visits at average) to rich pregnant mothers (4.40 ≈4 visits) correspondingly.

4.1.2 Modeling the Number of ANC Service Visits

4.1.2.1 Model Identification and Selection Summary Information

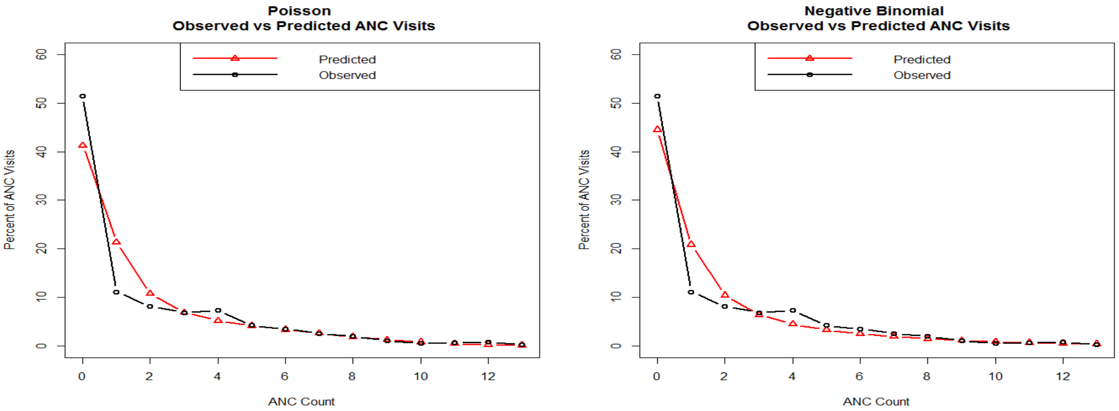


Figure 1: Observed Vs Predicted Values of Poisson and Negative Binomial Regression Models

Then, by penalizing a model with additional parameters, ten (10) models were refitted again under NB regression and compared with their AIC and BIC. After fitting the model, covariates with the largest p-value of Wald test is removed and refitted the model with the rest of the covariates sequentially. Then, the status of the pregnant mother, either she is residing with her husband or not, (RESID) and whether the pregnancy is wanted when become pregnant (PWANTD) are the covariates excluded from the model; with Wald test p-value for the given covariates are large (P-value > 0.05). Hence, as it was found in Table 8 (at appendix), new 10 models were fitted and the negative Binomial model with the smallest AIC (AIC=3254.0411) containing three types of interactions were selected. The last model of NB regression model is as follows:

$$\begin{aligned} log(\mu_i) = \beta_0 + \beta_1 * MEDUC_{(no\ educ)} + \beta_2 * REGION_{(better\ prog)} + \beta_3 * WLOAD_{(no\ prob)} + \beta_4 \\ * WEALTH_{(poor)} + \beta_5 * WEALTH_{(Middler)} + \beta_6 * HPOST_{(no\ prob)} + \beta_7 \\ * AWARN_{(no)} + \beta_8 * SIGN_{(no)} + \beta_9 * WLOAD_{(no\ prob)} * AWARN_{(no)} + \beta_{10} \\ * HPOST_{(no\ prob)} * RESID_{(not)} + \beta_{11} * MEDUC_{(no\ educ)} * WEALTH_{(Middler)} \end{aligned}$$

Graphs of the observed and predicted proportions of recurrent ANC visit counts for the two models fitted with the offset for the Ln of the follow-up time are provided in [Figure 1]. Though the fit of the NB model is slightly improved compared with the Poisson, It does not provide an acceptable fit to the data overall since it over estimates the proportion of mothers who had 4 ANC visits.

Finally, the best model of the refitted NB above was compared with the rest five models again based the values of their corresponding 2 log likelihood and the various information criteria AIC, BIC, and AICC.

4.1.2.2 Overdispersion and Poisson Regression

In Poisson regression analyses, Table 4, deviance and Pearson Chi-square goodness of fit statistics indicating over dispersion was obtained as 1688.1931 and 1758.7784, respectively. Since the Pearson chi-square statistic divided by the degrees-of-freedom is higher than one and the observed value of 1.1268 is significantly different from one, with P-value 0.0019, then the mentioned goodness of statistics represents that there was an overdispersion in data set. Even if the Deviance and Pearson chi-square goodness of fit statistics of 1210.3476 and 1257.4983 respectively in Negative Binomial regression is dropped considerably but still an indication of significant overdispersion exists; because we would like this value divided by the degrees of freedom to be close to 1.

Table 4: Test for Overdispersion

Criteria	Models	DF	Value	Value/DF	p-value
Deviance	Poisson	1116	1688.1931	1.5127	<.0001
	NegBin	1116	1210.3476	1.0845	0.0252
Scaled Deviance	Poisson	1116	1688.1931	1.5127	<.0001
	NegBin	1116	1210.3476	1.0845	0.0252
Pearson Chi-Square	Poisson	1116	1758.7784	1.5760	<.0001
	NegBin	1116	1257.4983	1.1268	0.0019
Scaled Pearson X2	Poisson	1116	1758.7784	1.5760	<.0001
	NegBin	1116	1257.4983	1.1268	0.0019

4.1.2.3 Model Fitting and Selection

The results of applying the model selection paradigm for the series of models fitted to the subset of Antenatal data are provided in Table 10 (on Appendix A) and Table 5, and detailed parameter estimates and standard errors for each model are provided in Table 9, (on Appendix A). In this study we have considered different possible count data models. Likelihood ratio test (LR), Akaike information criterion (AIC), Bayesian information criterion (BIC) and Vuong test were used to compare the candidate models to identify the most parsimonious model.

The overdispersion parameter (k^{-1}) is significantly different from zero in NB and in both hurdle models (HP and HNB) regression models. Hence there is an overdispersion problem in the data. As a result of this the standard error of the standard Poisson regression model is smaller than the standard error of the other models.

As it can be seen from Table 9, all covariates included in the standard Poisson model such as: mother education, region, work load, wealth, heath post, awareness, signs of pregnancy complications and the interaction between mother education and wealth, heath post and residing, and work load and awareness are significantly associated with the number of ANC visits even at 1% significance level; however in the case of the NB model only some of them are significantly associated number of ANC visits at 1% significance level.

Table 5: Model Selection Criteria for PR, NB, ZIP, ZINB, PH and NBH Regression Models

Criteria	P	NB	ZIP	ZINB	HP	HNB
-2 Log Likelihood	3310.9762	3230.0	3049.6	3063.9	3048.2	3059.0
AIC (smaller is better)	3347.1444	3254.0	3093.6	3109.9	3092.2	3105.0
AICC (smaller is better)	3347.3812	3254.3	3094.5	3110.9	3093.1	3106.0
BIC (smaller is better)	3402.4449	3314.4	3204.2	3225.5	3202.8	3220.6

ZIP and ZINB regression models as well as HP and HNB were better fitted than Poisson and NB respectively based on their corresponding AIC as well as BIC. -2log likelihood, AIC and BIC selection criteria for the models of PR, NB, ZIP, ZINB, PH and NBH are given in Table 5. It was found out that the model with the smallest AIC and BIC was HP regression followed by ZIP regression model since their LR, χ^2 =3048.2 and $\chi^2 = 3049.6$ both were highly significant (p-value<0.0001) supported by the information criteria's.

The plots of predicted probability from each model against the observed probability of the outcome (Figure 2) show that the Poisson and the NB model under-estimated zero counts and the zero inflated and the hurdle models captured almost all zero values. Based on predicted probabilities, the differences in model fit between the six models were remarkable. Still the standard Poisson model and the NB model do not fit the data reasonably well; the standard Poisson predicted about 42% zeros and NB model predicted about 45% zeros compared to 51.5% observed zeros.

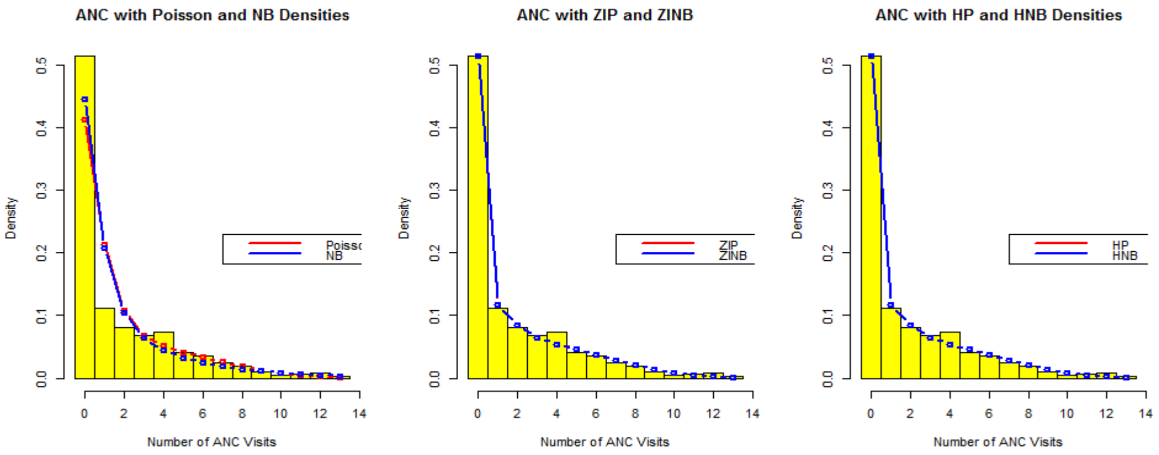


Figure 2: Comparison of the Densities of Each Model Fits

The overdispersion parameter(k^{-1}) in the HP regression model is significantly different from zero since there is a high variability in the non-zero outcomes. In such situation, it would be better to use the model which takes into account the excess zeros and high variability due to non-zero outcomes. Therefore, since it has the smallest AIC (3092.2) as well as BIC (3202.8) values as presented in Table 5, HP regression model was chosen as the most parsimonious model which fits the data better than the other possible candidate models.

Table 6: Model Selection: Vuong test, AIC, Log-Likelihood, and Inflation Probabilities

	Poi	NB	ZIP	ZINB	HP	HNB
Poi	AIC= 3359.7 LL=-0.2270 InfPr= 0.1476					
NB	V=3.3374 P=0.0008 Prefers NB	AIC= 3261.6 LL=-0.1876 InfPr= 0.1574				
ZIP	V=7.4161 P=1.206E^{-13} Prefers ZIP	V=8.0639 P=6.661E^{-16} Prefers ZIP	AIC= 3093.1 LL= -0.1284 InfPr= 0.5727			
ZINB	V=7.0799 P=1.44E^{-12} Prefers ZINB	V=7.9618 P=1.776E^{-15} Prefers ZINB	V=1.9815 P=0.0307 Prefers ZIP	AIC= 3107.7 LL= -0.1328 InfPr=0.6179		
HP	V=7.4068 P=1.295E^{-13} Prefers HP	V=8.1066 P=4.441E^{-16} Prefers HP	V=1. 9704 P=0.04781 Prefers HP	V=2.0304 P=0.0423 Prefers HP	AIC= 3091.9 LL= -0.1171 InfPr= 0.8757	
HNB	V=7.2661 P=3.6993E^{-13} Prefers HNB	V=7.9646 P=1.554E^{-15} Prefers HNB	V=1.9934 P=0.0464 Prefers ZIP	V=-0.0423 P=1.0337 Prefers ZINB	V=-2.1681 P=1.9698 Prefers HP	AIC=3107.9 LL= -0.1345 InfPr= 0.8756

Note: V=Vuong Test, P= P-value, LL=Log-Likelihood, InfPr= Estimated Proportion of Zeros

With respect to model performance, the log-likelihood (LL) was used as a measure of each model's performance. Table 6 clearly shows an improvement in model fitting from Poisson (LL = -0.2270) and negative binomial (LL = -0.1876) to zero-inflated Poisson (LL = -0.1284), zero-inflated negative binomial (LL= -0.1328), hurdle Poisson (LL= -0.1171) and hurdle negative Binomial (LL= -0.1345) models. The Vuong test statistic [53] found on Table 6 result reflected that all the candidate models, NB, ZIP, ZINB, HP, and HNB performed better than the standard Poisson model. zero-inflated Poisson performed better than NB (V=8.0639, P=6.661E^{-16}), better than Zero Inflated NB (V=1.9815, P=0.0307), better than Hurdle NB (V=1.9934, P=0.0464), which also holds for zero-inflated negative binomial vs. Hurdle negative binomial (V=-0.0423, P=1.0337). However, the hurdle Poisson model performed better than the ZIP model (V=1. 9704, P=0.04781).

The estimated value of k^{-1} (the overdispersion parameter) is 0.1476, 0.1574, 0.5727, 0.6179, 0.8757, and 0.8756 for NB, ZIP, ZINB, PH, and NBH respectively. This suggests that Zero Inflated models are better in handling zero counts, but the AIC and LL values for zero-inflated Poisson and hurdle Poisson models were smaller compared to the others. Therefore, ZIP and HP models predicted each count outcome very close to the observed counts, suggesting better fit than standard Poisson, negative binomial, ZINB, HNB and models.

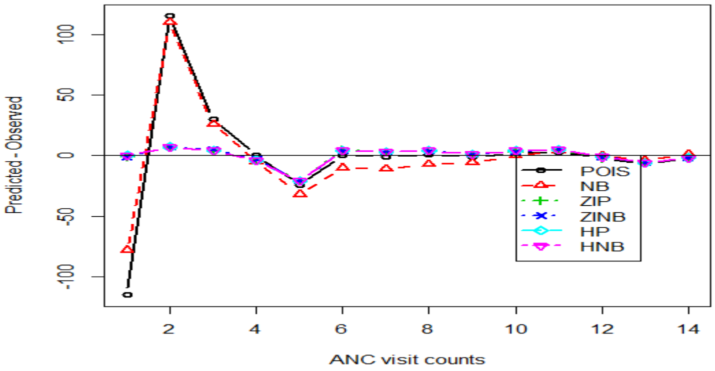


Figure 3: Observed vs Predicted Plots of Poisson, NB, ZIP, ZINB, HP and HNB Model Fits

To better illustrate this fact, as well as to provide a more intuitive presentation of variables' influence on the expected count in the different models, Figure 3 presents the change in the value of μ as the value of the variable indicating a declaration of unconstitutionality goes from zero to one. While this variable is not, in fact, continuous, the graph does serve the useful purpose of outlining the general shape the variable's effects on the observed count. Each line was calculated holding all other variables at their mean values. We therefore turn to the hurdle and ZIP specifications, both to obtain more accurate results, and to examine the properties of the models described above.

4.1.3 Zero-Inflated Poisson and Hurdle Poisson Estimation Results

For purposes of comparison both zero-inflated Poisson and hurdle Poisson regressions are estimated by reducing the interaction between mother education status and wealth index. This is because in both transitions the interaction term is not significant as presented in Table 9. The AIC values of the full model above and the reduced one is 3093.6 and 3092.2 (which is found in Table 6) and 3093.1 and 3091.9 for ZIP and HP respectively. Then it turned out that the model with pregnant mother education status, region, work load inside and/or outside home, wealth index, awareness about pregnancy complications and ANC utilizations, availability and access ability of health post, signs of pregnancy complications, work load and awareness interaction as well as health post and residing interaction as covariates was the most parsimonious model. Based on the above mentioned criteria for model selection and evaluation, especially, vuong test, AIC and log likelihood, we opted for the Hurdle Poisson model for fitting the ANC data. The cumulative evidence suggests that the HP model provides an adequate fit to the data and that it is at least as good as, or superior to, the ZIP model for these data. With no evidence of overdispersion, it is reasonable to assume that the standard errors of the HP model's parameter estimates are unbiased and that the model's estimates are suitable for statistical inference.

Therefore, the final hurdle Poisson regression model proposed for number of ANC service utilization of pregnant mothers was given as:

$$\begin{aligned} logit(\mu_i) = {} & 0.6407 + 0.3596 * MEDUC_{(no\ educ)} - 0.2312 * REGION_{(prog\ reg)} - 1.0168 \\ & * WLOAD_{(no\ prob)} - 0.7201 * WEALTH_{(poor)} - 0.1407 * WEALTH_{(middle)} \\ & - 0.7579 * HPOST_{(no\ prob)} + 0.2935 * AWARN_{(no)} + 0.3844 * SIGN_{(no)} \\ & + 0.6469 * WLOAD_{(no\ prob)} * AWARN_{(no)} + 0.4936 * HPOST_{(no\ prob)} \\ & * RESID_{(not)} \end{aligned}$$

$$\begin{aligned} log(\mu_i) = {} & 0.6993 - 0.7763 * MEDUC_{(no\ educ)} + 0.6352 * REGION_{(prog\ reg)} + 0.6260 \\ & * WLOAD_{(no\ prob)} - 0.7633 * WEALTH_{(poor)} - 0.1945 * WEALTH_{(middle)} \\ & + 0.6267 * HPOST_{(no\ prob)} - 0.9758 * AWARN_{(no)} - 0.2290 * SIGN_{(no)} \\ & + 0.2057 * WLOAD_{(no\ prob)} * AWARN_{(no)} - 0.7375 * HPOST_{(no\ prob)} \\ & * RESID_{(not)} \end{aligned}$$

Table 7: Results for Zero Inflated Poisson and Hurdle Poisson Model Estimates

Variables	Zero Inflated Poisson(ZIP)		Hurdle Poisson (HP)	
	Poisson	Inflated part	Poisson	Inflated part
	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)	Estimate (s.e)
Intercept	0.4792 (0.3748)	0.6511 (0.1553)***	0.6993 (0.3066)*	0.6407 (0.1583)***
$\mathrm{MEDUC}_{\mathrm{educ}}$	-0.6590 (0.2243)**	0.3661 (0.05578)***	-0.7763 (0.1713)***	0.3596 (0.05593)***
$\mathrm{REGION}_{\mathrm{other}}$	0.6808 (0.2032)***	-0.2171 (0.06965)**	0.6352 (0.1499)***	-0.2312 (0.07010)**
$\mathrm{WLOAD}_{\mathrm{prob}}$	-0.2154 (0.3786)	-1.0136 (0.1700)***	0.6260 (0.2506)*	-1.0168 (0.1668)***
$\mathrm{WEALTH}_{\mathrm{rich}}$	-0.4553 (0.1657)**	0.1435 (0.04046)***	-0.4593 (0.1308)***	0.1407 (0.04028)***
$\mathrm{HPOST}_{\mathrm{prob}}$	0.02906 (0.3731)	-0.7426 (0.1549)***	0.6267 (0.1978)**	-0.7579 (0.1535)***
$\mathrm{AWARN}_{\mathrm{yes}}$	-0.9398 (0.3046)**	0.3047 (0.1084) **	-0.9758 (0.2662)***	0.2935 (0.1085)**
$\mathrm{SIGN}_{\mathrm{yes}}$	-0.6941 (0.2796)*	0.3506 (0.1208) **	-0.7375 (0.2116)***	0.3844 (0.1234)**
$\mathrm{WLOAD}_{\mathrm{prob}}$*$\mathrm{AWARN}_{\mathrm{yes}}$	0.9664 (0.4489)*	0.6410 (0.1866) ***	0.2057 (0.3297)	0.6469 (0.1844)***
$\mathrm{HPOST}_{\mathrm{prob}}$*$\mathrm{RESID}_{\mathrm{yes}}$	0.2933 (0.3731)	0.4864 (0.1645) **	-0.2290 (0.1937)	0.4936 (0.1648)**

* refers to p<0.05. ** refers to p<0.01. *** refers to p<0.001.

In the binary (logistic) portion of the ZIP model in Table 7 provides that all variables emerged as statistically significant predictors of number of ANC visits: MEDUC, REGION, WLOAD, WEALTH, HPOST, AWARN, and SIGN and WLOAD×AWARN interaction effect as well as HPOST×RESID since their p-values are less than 5%. It must be kept in mind that the interpretation of the binary portion of the model is different from the interpretation of the count portion. The sign of the parameters in the positive part of the ZIP model is different from the Poisson model. The percentage changes in the factors are largely changed; and are more realistic than that of the Poisson model. Although we are still trying to estimate the relationship between each of the ANC variables and a binary outcome, here the two levels of the binary variable consist of either structural (or true) zeroes or sampling zeroes that follow the Poisson distribution. The percentages changes of the factors pregnant mother who have no education, mother from better progressed region, mother with a problem of work load inside and/or outside home, poor pregnant mother, mother with a problem of access to health post, mother who have lack of awareness about pregnancy complications as well as ANC utilization and pregnant mother who had not seen sign of pregnancy complications are 44.21%, 80.48%, 36.29%, 15.43%, 47.59%, 35.62%, 41.99%, 89.83%, and 62.26% respectively.

Consequently, the negative relationship between WLOAD since workload problem and the "no ANC visits" portion of our outcome indicates an inverse relationship between no workload problem of the women and "true" zeroes. That is, as workload problem inside outside home decreases, there is a greater likelihood of a positive number of ANC visits in the future. (OR =0.36). Similarly, the REGION and HPOST have a negative signs. A positive change in these factors induces then an increase in the number of ANC visits. The percentage change of the factor REGION is 80.48% (OR=0.8048) this means that the number of ANC service visited by the regions that need special aids is about one more likely to have zero visits than the better progressed region. The fitted model again suggests that the rate of non-zero ANC visits in educated mother was exp(0.3661)=1.44 times the rate of non zero ANC visits in non educated holding all other predictors constant. The rate of non-zero ANC visits for women who had seen signs of pregancy complications was exp(0.3506)=1.42 times the rate of non zero ANC visits in women who had no signs pregancy complications holding all other predictors constant. The presence of a statistically significant interaction term indicates that rich women with having awareness about ANC utilization have a 89.84% higher odds of having non zero ANC visits compared to poor women with lack awareness about ANC utilization in this study population.

Accordingly the percentages change of the factor HPOST is around 47.59%. That means, the number of non-zero ANC service visited by pregnant mother that have a problems of accessibility and/or availability of health post is 47.59% less non-zero ANC visits than that of women who have no problems related to health post. Whereas the percentages change of the factor MEDUC is around 44.21%. Hence, pregnant mothers who can read and write are 1.44 times more user of ANC service than mothers who have no education. For AWARN, the percentage change is around 35.62%. Thus, mothers who had lack of awareness about ANC utilization and pregnancy complications are showing less participation in ANC service to that of pregnant mothers who had awareness of ANC use. The presence of a statistically significant interaction term also indicates that pregnant mother in rural Ethiopia who had no health post related problems with not residing with her husband/ partner have 62.26% less likely ANC visits than pregnant mothers that have health post related problems and residing with her husband/

partner. If we consider the significance level of 5%, we conclude easily that there is no striking difference between zero-inflated Poisson regression fits and the hurdle Poisson (HP) model fits and they are better than the standard Poisson regression, Negative Binomial, ZINB, and NBH. But, the ZIP model is suitable only for handling zero inflation. However, the hurdle model is also suitable for modeling zero deflation. This tells us that even when a test shows significant evidence of zero inflation, the ZIP model may still not be suitable to fit the data. Since the hurdle Poisson (HP) model had the best fit than all the rest models, we interpreted the results from this model (Table 7).

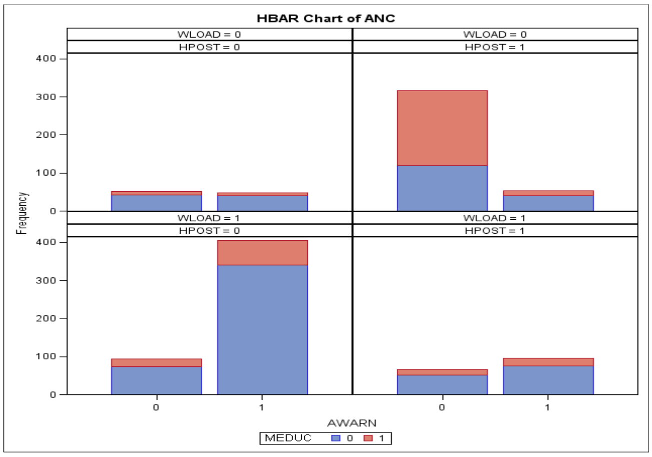


Figure 4: Bar Chart of WLOAD vs HPOST nested in AWARN

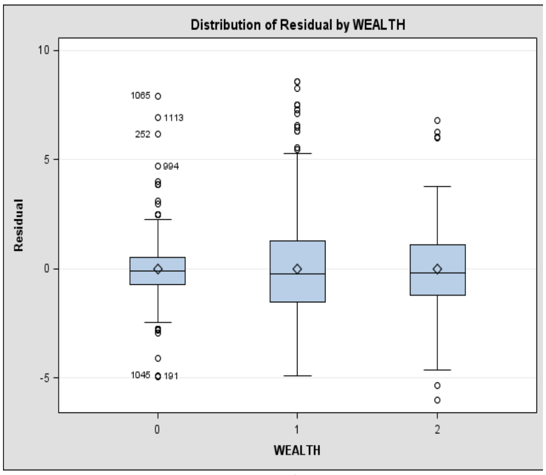


Figure 5: Both Plots of Wealth Index

Figure 4 shows that health post and workload is nested in awareness so that there are large numbers of women with adequate awareness about ANC usage with heavy workload and lack nearby health post. In addition, many mothers who had no awareness about ANC usage and pregnancy complications have no workload problem inside and /or outside home but they have shortage of nearby heath post. In the same way, Figure 5 indicates that large numbers of women who utilize ANC have middle income and there were smaller numbers of women who make use of ANC visits having poverty. It certainly looks as if median ANC visit numbers are higher in rich women than in middle income, but the range of counts is very large in middle income earner women, so the significance of the difference is certain.

4.1.3.1 Hurdle Poisson (HP) Model Parameter Estimate

The hurdle Poisson-logit model suggested that non educated pregnant mothers have a higher probability of not visiting ANC service and a higher expected number of zero visits than educated pregnant mothers. The non-zero part of HP-logit model fitting confirms this conclusion, as $\hat{\beta} = 0.3596$ has a standard error of 0.05593 (found in Table 7). The estimated odds that the number of ANC visits become zero with non educated (mothers who cannot read and write) are $\exp(0.3596\ ) = 1.43$ times the estimated odds for educated pregnant mothers. This estimate has the almost same order of magnitude as the estimate from the binary part of the ZIP model. The impact of covariates on the odds of visiting the ANC service for a less visit versus a more visits is quite different. For example, being residing with her husband/partner is not associated with the likelihood of ANC service utilization in the analysis characterized by slight number of ANC visits. However, better progressed regions such as Tigray, Amhara, Oromiya and SNNPR are statistically significantly associated with increased odds of at least one ANC visits in the analysis than not good enough regions. That means, better progressed regions are 0.79 times positive ANC visits than those regions that need special aids. This result indicates the importance of stratifying our analyses according to the severity of the growth level of rural Ethiopian regions, as the factors influencing the pregnancy complications and ANC service utilization.

The impact of access to a severe workload inside and/or outside home on ANC service utilization is an interesting finding in our analysis. Not having access to a severe workload inside and/or outside home did not influence the odds of ANC service utilization in those who demonstrated a number of ANC service visits over the study interval. For less number of ANC service visits, we estimate that having access to a primary care provider significantly reduces the likelihood (OR = 0.47) of a visit. Having signs of pregnancy complications such as vaginal bleeding, vaginal gush of fluid, severe head ache, blurred vision, fever, abdominal pain and others significantly increases the likelihood (OR = 1.47) ANC visits. That means, pregnant mothers who have cases of pregnancy complications have 1.47 times more non zero ANC visits than those who did not seen the signs of pregnancy complications. The interaction between access to a severe workload inside and/or outside home and lack of awareness about ANC utilization and pregnancy complications are statistically significant. Hence, a pregnant mother from rural Ethiopia who has no problem of workload inside and/or outside home as well as have good awareness about ANC utilization is 1.91 times more positive ANC visits than a mother with severe workload inside and/or outside home as well as lack of awareness about ANC utilizations.

4.2 Model Diagnostics

From Figure 14 (found at Appendix B), it seems that the variance stay constant as the fitted values vary, while there exist 3 outliers as labeled on Figure 14. The visual inspection plot of equal Cook's distance are shown in Figure 14 and Figure 6d to identify if any problem existed in the model. There are points having cook's distance larger as labeled.

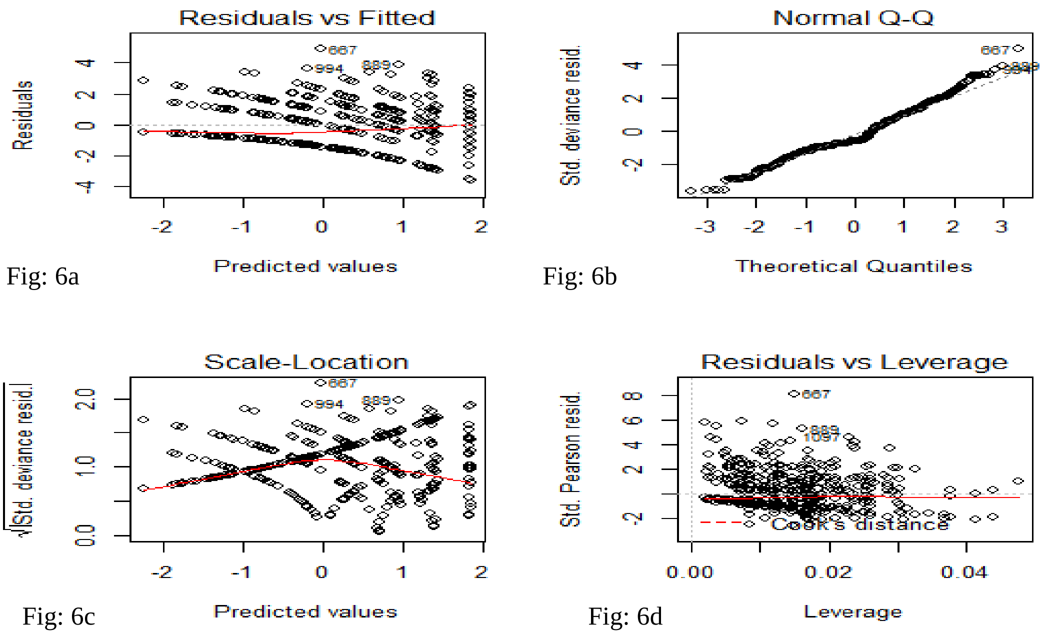


Fig: 6a

Fig: 6b

Fig: 6c

Fig: 6d

Figure 6: Residuals versus Fitted for Poisson Regression

We plot the standardized deviance residuals (SDR) against the fitted rates. An informal procedure that is used to check for systematic departures from the Poisson regression is based on four regression diagnostic plots. Figure 6 contains the four regression diagnostic plots. A plot of the standardized deviance residuals (SDRs) against the fitted rate is shown in Figure 6a. The R function was used to calculate the solid line, and the two dashed lines correspond to the 0.005 and 0.995 quantities of the standard normal distribution, i.e. if the SDRs are approximately N(0,1) about 99% of these residuals should be between the dashed lines. The seven SDRs outside the 99% limits are identified with their observation numbers (also found at Appendix B, Figure 14). For model checking purposes, a normal Q-Q plot is used to identify extreme values which would appear in the upper right and/or lower left portion of the plot (Figure 6b). The solid line in Figure 6b corresponds to the standard normal distribution. The SDR-Leverage plot in Figure 6c identifies four points (especially observation number 667, 994, 889 and 1097.) with both $h_i > 2$ (i.e. to the right of the solid vertical line. A plot of the absolute value of the SDRs against the fitted values (see Figure 6d) gives an informal check of on the adequacy of the assumed variance function. The null pattern will not show a trend , and smoothing (shown by the solid line) is used to identify a possible pattern , in this case a positive trend. Though 3 outlier cases need more investigation, the results from diagnostics of the Poisson model in Figure 1 indicate that Poisson did not fit well. Conclusively speaking, the apparent trend in Figure 6d indicate over -dispersion, and the other three diagnostic plots do not indicate that “outliers” are a problem.

The parameter estimates of the final model before and after excluding the outlying observations were close to each other. Thus in this study the zero-inflated Poisson (ZIP) regression model was robust to the outlying observation.

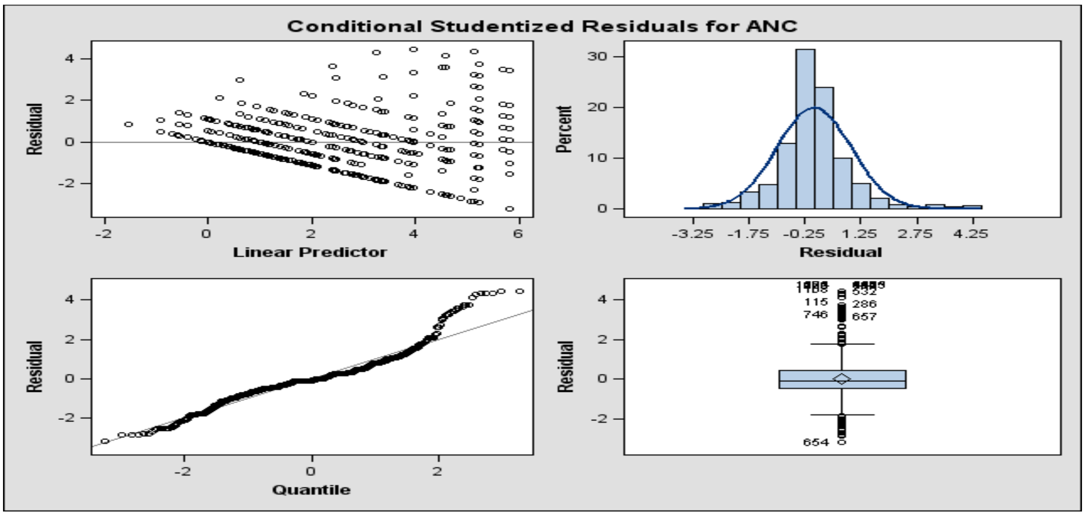


Figure 7: Conditional Studentized Residuals for ANC Visits

The first graph of Figure 7 shows the deviance residuals plotted against fitted values. It is observed that the plot of deviance residuals against fitted values appears to show some trend of falling variation with increase in estimated value. In the second graph of Figure 7, the normal density plot and in the third a normal quantile plot of standardized deviance residuals is shown. The quantile plot appears to follow a reference line except in the upper right portion. This verifies the assumptions of normality of the residuals for most of the range of values. Some deviations are observed especially at the high end which suggests the data distribution has a long tail at that end.

4.3 Discussion

In health service studies, Antenatal care service visits could be a relevant metric to quantify efficiency of maternal care utilization. This thesis, which focused on an efficient statistical modeling for number of Antenatal care service visits, propose a GLM, zero-inflated and hurdle modeling approach to estimate parameters of demographic, socio-economic, health and environmental related factors. World Health Organization recommends a minimum of four ANC visits initiated during their pregnancy periods. In this study the ANC service utilization rate in rural Ethiopia was found to be 48.5%. Although this shows a low level of ANC service utilization, educated pregnant mothers, and mothers from better progressed regions, mothers who have no severe workloads, mothers that could get a nearby health post and have awareness about ANC use attends more than 3 times. Moreover, a significant proportion (77.5%) of the attendees had less than four visits which is less than the recommended.

The finding of this study significantly differs with that of EDHS 2005 which showed 21.6% attendance of ANC in the rural areas of Ethiopia. This could be attributed to the fact that DHS covered more remote areas where distance from health institution could be a major predictor of ANC utilization. It is also important to note the time gap between the EDHS and the current study. A study conducted in Northern Ethiopia (2004) showed that the magnitude of ANC attendance was 45%. With regard to the determinants of ANC service utilization; this study revealed that ANC service utilization is significantly influenced by mother education, region, workload, economic status, access to health post, awareness about pregnancy complications, and manifestation of pregnancy complications. Non educated mothers were less likely to utilize ANC service than educated women. This finding is consistent with the findings of previous studies conducted in Addis Ababa [52]. Moreover, in this study the use of antenatal care was found to be related to economic status; Mothers with middle and rich economic level were more likely to attend ANC than poor women. This is also in line with other studies conducted in Southern Ethiopia [59].

This finding differs with the study conducted in Metekel zone which confirms that being awared about ANC utilization were more than two times (OR=2.32) more likely to use ANC visits [18]. In our case, awared women were one times (OR=1.47) more likely to utilize ANC visit than non awared ones. This study again found that education status of secondary school and above had three times (OR=3.68) more ANC visits than non educated ones, but here 1.44 times more likely to be visited. Gurmessa’s report determined that having monthly family income of 500 Ethiopian Birr and above (OR=1.53) were positively associated with antenatal care service utilization [18]. This agrees with our finding that rich and middle incomer women were more likely to attend ANC visits; the odds ratio in our case is 1.15 (slightly different with this report).

A report from Samre Saharti district in Tigray region of Ethiopia found differently from our finding that being resing with her husband or partner have significant association with ANC utilization [58]. According to this finding, this is not the factor. The reason for the difference could be that our study was conducted in a rural area, while the regional report included urban areas. Unwanted pregnancy were not the determinant to receive ANC visits in this study. This was dissimilar to other studies conducted in the Saharti Samre district, Tigray, Ethiopia [58]. This could be due to fear of stigma because a pregnancy without marriage is not accepted by the community in the study area. Therefore it appears rational to see that most of single and widowed mothers might be faced unwanted pregnancies. In addition mothers from low socio-economic status (poor mothers) are unlikely to afford the cost of transport and could have limited access to ANC utilization, and low health seeking behavior [58]. Other studies have shown comparable results with this [2, 18, 23, 58, 59]. As part of enabling factors, distance from health post were found to be predictor of antenatal care service utilization where women who live within nearby distance from the health facility were about 0.48 times more likely to visit ANC than women who live at distance far from health post(OR=0.48). this was line with the study Yem special woreda, southwestern Ethiopia [3].

Hurdle Poisson model assumes that all zero data are from one “structural” source. The positive (i.e., non-zero) data have “sampling” origin, following either truncated Poisson or truncated negative binomial distribution. For example, consider a study of ANC visit users in which a secondary outcome is a number of ANC visits during last nine months. In this case, it is safe to assume that only non ANC users will visit zero ANC visits during the last nine months and ANC users will score some positive (non-zero) number of ANC service visits during last nine months. Hence the zero observations can come from only one “structural” source, the non ANC users. If a pregnant mother is considered as ANC user, they do not have the ‘ability’ to score zero ANC visits during the last nine months and will always score a positive number of ANC visits in a hurdle model with either truncated Poisson or truncated negative binomial distributions.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

In conclusion, the antenatal care service utilization rate in rural Ethiopia is lower than the national figures available to date. In addition, it is worth nothing that majority of the mothers who attend ANC did not receive adequate number of visits recommended by the World Health Organization. Furthermore, maternal education, workload inside and/or outside home, availability and accessibility of health post, regions, and awareness about pregnancy complications were major predictors of ANC service utilization. Therefore, efforts to bring about changes in these major predictors at individual and community level through behavioral change communication are recommended.

In this study, it was found that ZIP and hurdle Poisson regression models were better fitted the data than NB, ZINB, HNB and Poisson. This may be due to the high variability of the number of ANC visits. Hurdle Poisson regression model was better fitted the data which is characterized by excess zeros and high variability in the non-zero outcome than any other models and therefore it was selected as the best parsimonious model.

5.2 Recommendation

Looking at the state of pregnant mothers and the number of ANC visits in rural Ethiopia, it is recommended that;

1. Environmental factors and Social activities such as workload of pregnant mothers inside and/or outside home, lack of finance and problems of availability as well as accessibility of health post should be reduced to help maximize the number of ANC visits during pregnancy.
2. Since the pregnant mothers who have awareness about ANC utilizations and pregnancy complications were attained more ANC visits than the mothers with shortage of awareness, education on ANC usage should be intensified especially among women’s in fertile age group in rural areas. Hence, concerning bodies including mass Medias and health extension workers should give special attention in raising awareness to be able to avoid preventable complications, especially in rural areas of Ethiopia.

3. Institutions that act on maternal and children's health should do well to apply the minimum of four (4) ANC visits scheduled by WHO for developing countries especially on rural areas so that all perpetrators of maternal care shall be brought to book to deter others from repeating such absences and thus move the country closer to MDG targets for maternal health by 2015.
4. Women who have no problem of the accessibility and availability of health post were more likely to receive ANC visits than women with a problem of that. Hence, there is a need to increase the availability and accessibility of health post in order to ease antenatal care services to the needy, particularly to those rural women.
5. Even if the EDHS dataset used for this study was not the latest, three years later on the date data was collected, The EDHS data base of the country should be expanded to include more variable so that researcher could really determine the actual factors contributing the casualties' in absence of ANC utilizations during their pregnancy periods.
6. Finally, from this study we can recommend that as this study is a small study, the result may not be generalizable, that is its external validity may not be valid. So that it would be better to examine in a large data set.

References

[1] Agresti, A., & Finlay, B. (1997). Statistical methods for the social sciences. (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

[2] Alemayehu Tariku. (June, 2008). *Why Pregnant Women Delay to Attend Prenatal Care?* Cross Sectional Study on Timing of First Antenatal Care Booking at Public Health Institutions in Addis Ababa, Ethiopia; Journal

[3] Bahilu Tewodros, et al. (March 2009). *Factors Affecting Antenatal Care Utilization in Yem Special Woreda, Southwestern Ethiopia*; Journal

[4] Cameron, A. C. and P. K. Trivedi (1998). *Regression Analysis of Count Data*, Cambridge University Press, Cambridge.

[5] Cameron, A. C., & Trivedi, P. K. (2010). *Micro econometrics using stata* (revised Ed.). Stata Press Books. Consul, P. C., & F

[6] Carla A, Tessa W, Blanc A, Van P, et al. ANC in developing countries, promises, achievements and missed opportunities; *an analysis of trends, levels and differentials, 1990-2001*. WHO Geneva, 2003.

[7] Carson, J., and Mannering, F. (2001). "*The effect of ice warning signs on ice-accident frequencies and severities*." Accid. Anal. Prev., 33(1), 99-109.

[8] Casella G. and R.L. Berger (2002). *Statistical Inference*, (2nd ed). Thomson Learning: Pacific Grove, CA.

[9] Central Statistical Authority (CSA) and ORC Macro. *Ethiopia Demographic and Health Survey 2005*, Addis Ababa, Ethiopia, and Calverton, Maryland, USA: CSA and ORC Macro, 2006.

[10] Chatterjee S and Hadi AS, (2006). *Regression analysis by example*, Fourth edition, Wiley Interscience.

[11] Cohen, A. (1963). "*Estimation in mixtures of discrete distributions*." Proc., Int. Symp. On Classical and Contagious Discrete Distributions, Pergamon Press, New York,351-372.

[12] CSA, ORC Macro, author. *Ethiopia Demographic and Health Survey*, Addis Ababa, Ethiopia and Calverton, Maryland, USA. Sep, 2006.

[13] Dietz, E. and Bohning, D. (2000). on estimation of the Poisson parameter in zero modified Poisson models. *Computational Statistics and Data Analysis*. 34(4): 441-459.

[14] Edward N., Bernardin S. and Eric A. (2012). *Determinants of utilization of antenatal care services in developing countries Recent evidence from Ghana*. Department of *Economics*, University of Ghana, Accra, Ghana.

[15] Eyerusalem Dagne. (2010). *Role of socio-demographic factors on utilization of maternal health care services in Ethiopia*. UMEA Universitet.

[16] Germu, S. and Trivedi, P. K. (1996). *Excess zeros in count models for Recreational Trips*. Journal of Business and Economic Statistics. 14: 469-477.

[17] Greene, William H. (1994). "*Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models*." New York University Department of Economics Working Paper EC-94-10.

[18] Gurmesa Tura. (July 2009). *Antenatal Care Service Utilization and Associated Factors in Metekel Zone*, Northwest EthiopiaJournal.

[19] Gurmu, S. and P.K. Trivedi (1996). “*Excess Zeros in Count Models for Recreational Trips*”, Journal of Business and Economic Statistics, 14, 469-477.

[20] Hardin, J & Hilbe, J(2001).*Generalized linear models & extensions*. Stata Press, Texas, USA.

[21] Hardin, J and Hilbe, J (2003). *Generalized estimation equations*. Chapman and Hall/CRC

[22] Hayelom Kassyou. (2008). *Factors Affecting Antenatal Care Attendance in Maichew Town, Southern Tigray*. Addis Ababa University, school of graduate studies.

[23] Heilbron, D. (1994). *Zero-altered and other regression models for count data with added zeros*. Biometrical Journal, 36, 531–547.

[24] Hinde, J. P. and Demetrio, C. G. B. (1998). Overdispersion: models and estimation. *Computational Statistics and Data Analysis*. 27: 151-170.

[25] Ismail, N. and Jemain, A. A. (2007). *Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models*. Casualty Actuarial Society Forum,103-158.

[26] Jansakul N and Hinde JP (2002). *Score tests for zero-inflated Poisson models*. Computational Statistics and Data Analysis40, 75–96.

[27] Kumara, S., and Chin, H. (2003). "*Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros*." Traffic Inj. Prev., 4(1), 53-57.

[28] Kutner, M and Neter J (2004). *Applied linear regression models*. McGraw-Hill Irwin.

[29] Lambert, D. (1992), *Zero-inflated Poisson regression, with an application to defects in manufacturing*, Technometrics, 34, 1-14.

[30] Lee, J., and Mannering, F. (2002). "*Impact of roadside features on the frequency and severity of run-off-roadway accidents*: an empirical analysis." Accid. Anal. Prev., 34(2), 149-161.

[31] Long, S. (1997). *Regression models for categorical and limited dependent variables*. Newbury Park, CA: SAGE.

[32] Lord, D., (2006). *Modeling motor vehicle crashes using Poisson-gamma models*: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention 38 (4), 751–766.

[33] Lord, D., Bonneson, J.A., (2005). *Calibration of predictive models for estimating the safety of ramp design configurations*. Transportation Research Record 1908, 88–95.

[34] Lord, D., Bonneson, J.A., (2007). *Development of accident modification factors for rural frontage road segments in Texas*. Transportation Research Record 2023, 20-27.

[35] Lord, D., and Mahlawat, M. (2009). "*Examining Application of Aggregated & Disaggregated Poisson-Gamma Models Subjected to Low Sample Mean Bias*." Transp. Res. Rec., 2136, 1-10

[36] Lord, D., Washington, S. P., and Ivan, J. N. (2005). "*Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes:* balancing statistical fit and theory." Accid. Anal. Prev., 37(1), 35-46.

[37] McConway, KJ, Jones, MC and Taylor PC (1999). *Statistical modeling using GenStat*. Arnold, United Kingdom.

[38] McCullagh, P., & Nelder, J. A. (1989).*Generalized linear models*. London: Chapman & Hall.

[39] Md Abdullah al Mamun. (May 2014). *Zero-inflated regression models for count data*: an application to under-5 deaths. Ball State University, Muncie, Indiana.

[40] Ministry of Health (2006).*AIDS in Ethiopia*, Sixth Report, Addis- Ababa, Ethiopia

[41] Mullahy, J. (1986). *Specification and testing of some modified count data models*. Journal of Econometrics, 33, 341–365.

[42] Nishat Fatema.(2010). *Importance of Antenatal care*. MO, CMH, Chittagong, Bangladish.

[43] Ornella L. et al. (2011). Antenatal Care. *Opportunities for Africa's Newborns*; Journal.

[44] Paternoster R, Brame R, Bachman R, Sherman L (1997). *Do fair procedures matter?* The effect of procedural justice on spouse assault. Law Soc Rev 31:163–204

[45] Regassa N. (2011). *Antenatal and postnatal care service utilization in southern Ethiopia*: a population-based study. African Health Sciences. 11(3): 390 – 397

[46] Rider, P. R. (1961). "*Estimating the parameters of mixed Poisson, binomial and Weibull distributions by the method of moments*." Bulletin De l'Institut International De Statistiques.

[47] Ridout, M.S., Demétrio, C.G.B. and Hinde, J.P. (1998). *Models for counts data with many zeroes*. Proceedings of the XIXth International Biometric Conference, Cape Town, Invited Papers, pp. 179-192. Paper retrieved March 13, 2006 from http://www.kent.ac.uk/ims/personal/msr/zip1.html

[48] Shankar V, Milton J and Mannering F (1997). *Modeling accident frequencies as zero-altered probability processes*: an empirical inquiry. Accident Analysis and Prevention29, 829–37.

[49] Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., Nebergal, M.B., (2003). *Modeling crashes involving pedestrians and motorized traffic*. Safety Science 41(7), 627-640.

[50] STATA Press (1985). Reference manual. *Longitudinal/ Panel data*, Release 9, USA.

[51] STATA Press (2001). Reference manual. *Generalized Linear Models and Extensions*. USA.

[52] Trends in Maternal Health in Ethiopia (December 2012), *Challenges in achieving the MDG for maternal mortality*. Addis Ababa; Journal.

[53] Vuong, Q. H. (1989). *Likelihood ratio tests for model selection and non-nested hypotheses*. Econometrica: Journal of the Econometric Society, 57, 307–333.

[54] Washington, S.P., Karlaftis, M., Mannering, F.L., (2003). *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall, Boca Raton, FL.

[55] WHO, UNICEF, and UNFPA, author. *Maternal Mortality in 2000*: Estimates Developed by WHO, UNICEF, and UNFPA. Geneva: 2003.

[56] WHO & UNICEF. Antenatal care in developing countries: promises, achievements and missed opportunities. *An analysis of trends, levels, and differentials 1990-2001*. Geneva: WHO & UNICEF, 2003.

[57] World Health Organization, (2004). *World Report on Antenatal Care Service Visits*: Summary. World Health Organization, Geneva.

[58] Yalem Tsegay Assfaw. (2010). *Determinants of Antenatal Care, Institutional Delivery and Skilled Birth Attendant Utilization in Samre Saharti District*, Tigray, Ethiopia. Umeå University, Sweden

[59] Zeine Abosse, et al. (July 2010). *Factors Influencing Antenatal Care Service Utilization in Hadiya Zone*, Southern Ethiopia; Journal.

Appendix A: SAS Output Tables

Table 8: Comparison of the Final Ten Models Involved in the Selection Criteria

Model	Deviance	Full LogLik	AIC	BIC
Model 1	1058.0453	-2007.3227	4018.6454	4028.7000
Model 2	1209.2201	-1626.5486	3275.0971	3330.3976
Model 3	1215.8687	-1629.4282	3276.8564	3322.1022
Model 4	1211.1484	-1621.7896	3263.5792	3313.8524
Model 5	1217.1069	-1623.4388	3266.8776	3317.1508
Model 6	1209.5207	-1617.6377	3257.2753	3312.5758
Model 7	1208.1321	-1617.2171	3256.4342	3311.7347
Model 8	1212.0614	-1615.1787	3254.3575	3314.6853
Model 9	1207.1569	-1615.3895	3254.7790	3315.1067
Model 10	1210.3476	-1615.0205	3254.0411	3314.3688

Table 9: Parameter Estimations and S.E for the Models of PR, NB, ZIP, ZINB, HP & HNB

Parameters	Basic Count Models		Zero Inflated Models		Hurdle Models	
	Poisson	NB	ZIP	ZINB	PH	NBH
	Estimation (St. Error)	Estimation (St. Error)	Estimation (St. Error)	Estimation (St. Error)	Estimation (St. Error)	Estimation (St. Error)
Intercept	-0.2638 (0.1388)	-0.2621 (0.1654)	0.5872 (0.1608)***	0.2036 (0.1937)	0.5756 (0.1640)***	0.1781 (0.0687)
MEDUC	0.7542 (0.0867) ***	0.7782 (0.1121) ***	0.4987 (0.0999)***	0.5507 (0.1027)***	0.4937 (0.1009)***	0.5820 (0.1022)***
REGION	-0.4352 (0.0614) ***	-0.4287 (0.0769) ***	-0.2249 (0.0698)***	-0.2482 (0.0744)***	-0.2381 (0.0703)***	-0.1997 (0.0708)**
WLOAD	-0.9568 (0.1212) ***	-0.9600 (0.1391)***	-1.0231 (0.1726)***	-0.7014 (0.1744)***	-1.0142 (0.1669)	-0.8168 (0.1644)***
WEALTH	0.3844 (0.0565) ***	0.4271 (0.0732)**	0.2292 (0.0660)***	0.2612 (0.0679)***	0.2253 (0.0663)***	0.2850 (0.0666)***
HPOST	-0.7527 (0.1061)***	-0.7551 (0.1197)**	-0.7251 (0.1555)***	-0.6621 (0.1529)***	-0.7430 (0.1538)***	-0.6813 (0.1543)***
AWARN	0.5495 (0.0980)***	0.5437 (0.1212)***	0.2976 (0.1083)***	0.4899 (0.1276)***	0.2853 (0.1086)**	0.4866 (0.1034) ***
SIGN	0.6106 (0.1022)***	0.5723 (0.1209)***	0.3432 (0.1204)***	0.5318 (0.1480)***	0.3785 (0.1234)**	0.6775 (0.1055) ***
WLOAD*AWARN	0.3877 (0.1398)**	0.3814 (0.1658)*	0.6358 (0.1891)***	0.2617 (0.1931)	0.6320 (0.1847)***	0.4212 (0.1814)*
HPOST*RESID	0.3708 (0.1131) **	0.4093 (0.1266)**	0.4804 (0.1647)***	0.4105 (0.1633)*	0.4870 (0.1649)**	0.4660 (0.1664)**
MEDUC*WEALTH	-0.2693 (0.0703)***	-0.3015 (0.0974)**	-0.1305 (0.0796)	-0.1972 (0.0820)*	-0.1289 (0.0804)	-0.1947 (0.0807)*
Dispersion (k^{-1})		0.2862 (0.0471)***	0.5670 (0.1085)	0.0071 (0.0154)	0.8792 (0.0264)***	0.0077 (0.0152)*

* refers to p<0.05. ** refers to p<0.01. *** refers to p<0.001.

Table 10: Inflation and Dispersion Probabilities

Model	Inflation Probability (π)					Dispersion probability (κ^{-1})				
	Beta	S.E	DF	t value	Pr >|t|	Beta	S.E	DF	t value	Pr >|t|
NB	0.1476	0.0232	1127	6.38	<.0001	0.2977	0.0481	1127	6.19	<.0001
ZIP	0.5727	0.1047	1127	5.47	<.0001					
ZINB	0.6179	0.0960	1127	6.44	<.0001	0.0071	0.0153	1127	0.47	0.6397
PH	0.8757	0.0267	1127	32.78	<.0001					
NBH	0.8756	0.0267	1127	32.77	<.0001	0.0093	0.0154	1127	0.60	0.5472

Table 11: Estimates of Hurdle Negative Binomial Models with Logit Link Function

Variables	Poisson Hurdle(PH)		NegBin Hurdle (NBH)	
	Poisson part	Inflated part	Negative Binomial	Inflated part
	Beta(S.E)	Beta(S.E)	Beta(S.E)	Beta(S.E)
Intercept	0.6993(0.3066)*	0.6407(0.1583) ***	-0.6992(0.3066)*	0.2541(0.2636)
MEDUC	-0.7763(0.1713)***	0.3596(0.0559) ***	0.7761(0.1713)***	0.3794 (0.0578)
REGION	0.6352(0.1499) ***	-0.2312(0.0701) ***	-0.6355(0.1499)***	-0.1847 (0.0708)
WLOAD	0.6260(0.2506)*	-1.0168 (0.1668) ***	-0.6260(0.2506)*	-0.7841 (0.1651)
WEALTH	-0.4593(0.1308) ***	0.1407 (0.0403) ***	0.4593(0.1308)***	0.1590 (0.0417)
HPOST	0.6267(0.1978) ***	-0.7579 (0.1535) ***	-0.6264(0.1978)**	-0.6976 (0.1544)
AWARN	-0.9758(0.2662) ***	0.2935 (0.1085)**	0.9759(0.2662)***	0.5355 (0.1038)
SIGN	-0.7375(0.2116) ***	0.3844 (0.1234)**	0.7375(0.2116)***	0.7474 (0.1032)
WLOAD*AWARN	0.2057(0.3297)	0.6469 (0.1844)***	-0.2059(0.3297)	0.4071 (0.1827)
HPOST*RESID	-0.2290(0.1937)	0.4936 (0.1648)**	0.2287(0.1937)	0.4747 (0.1665)

Appendix B: Plots

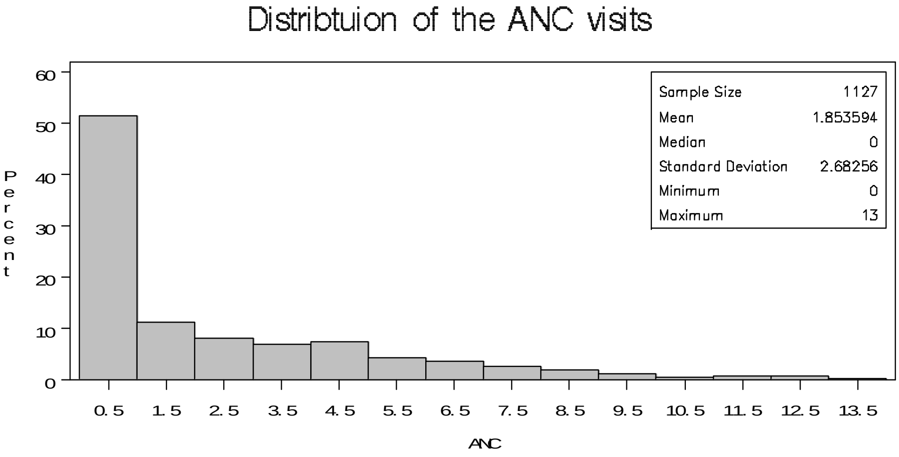


Figure 8: Histogram of Number of ANC Visits per Pregnant Women

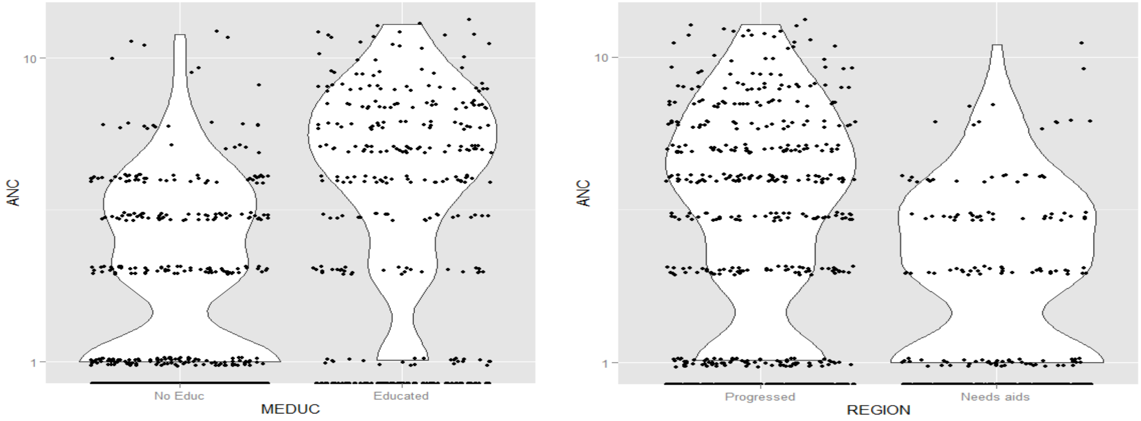
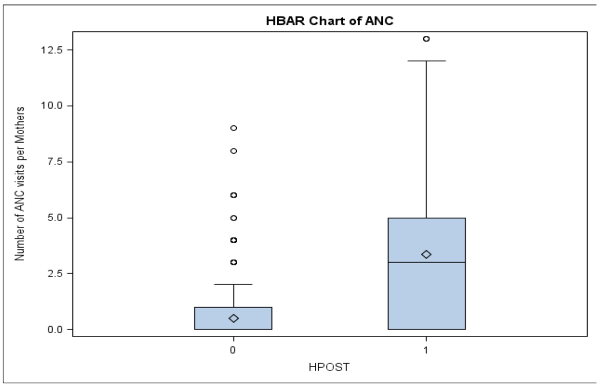


Figure 9: Profile Plot of ANC Visits in REGION



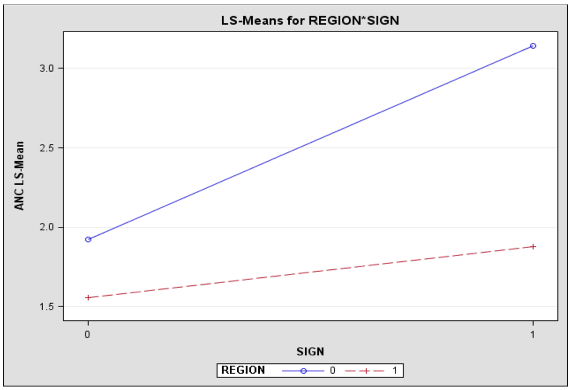


Figure 10: Box Plot of Health Post

Figure 11: The Mean Plot of ANC Utilization

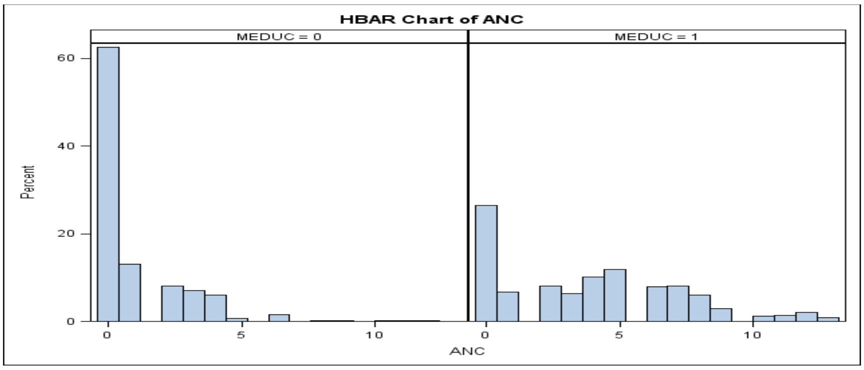


Figure 12: Histogram of Mother Education Status and their ANC Utilization Experience

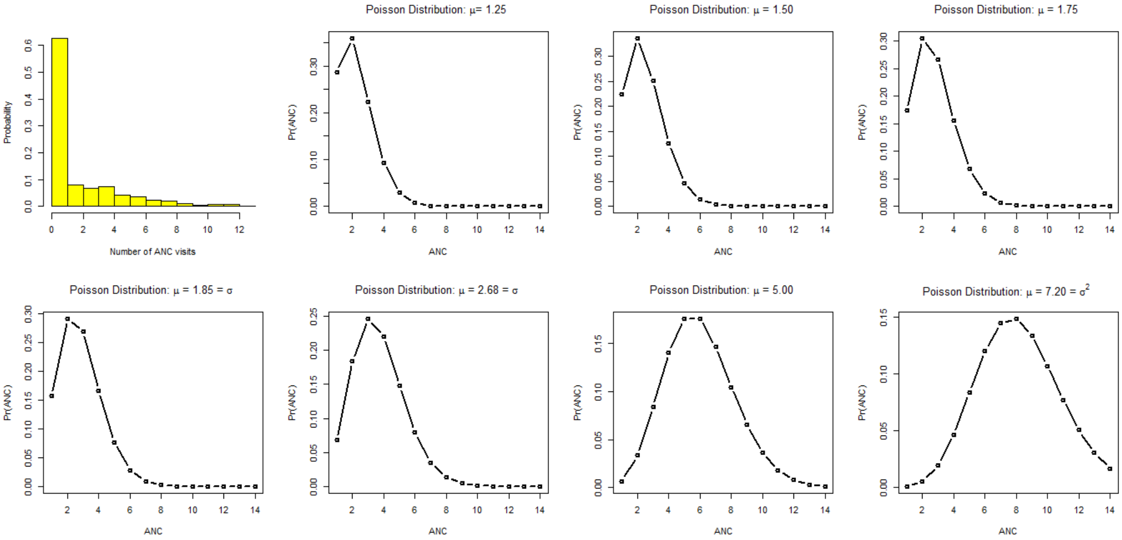


Figure 13: An Illustration of How the Shape of a Poisson distribution Changes as its Mean Changes

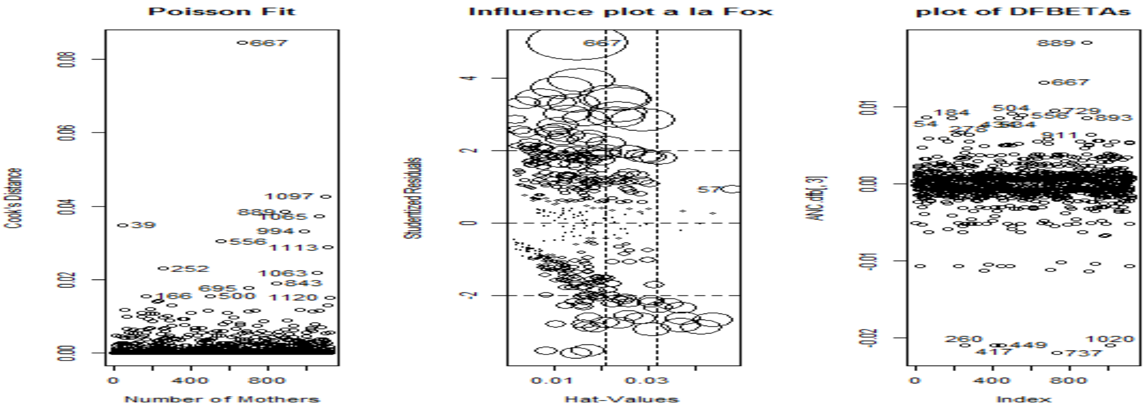


Figure 14: Cook's Distance for Poisson Fit

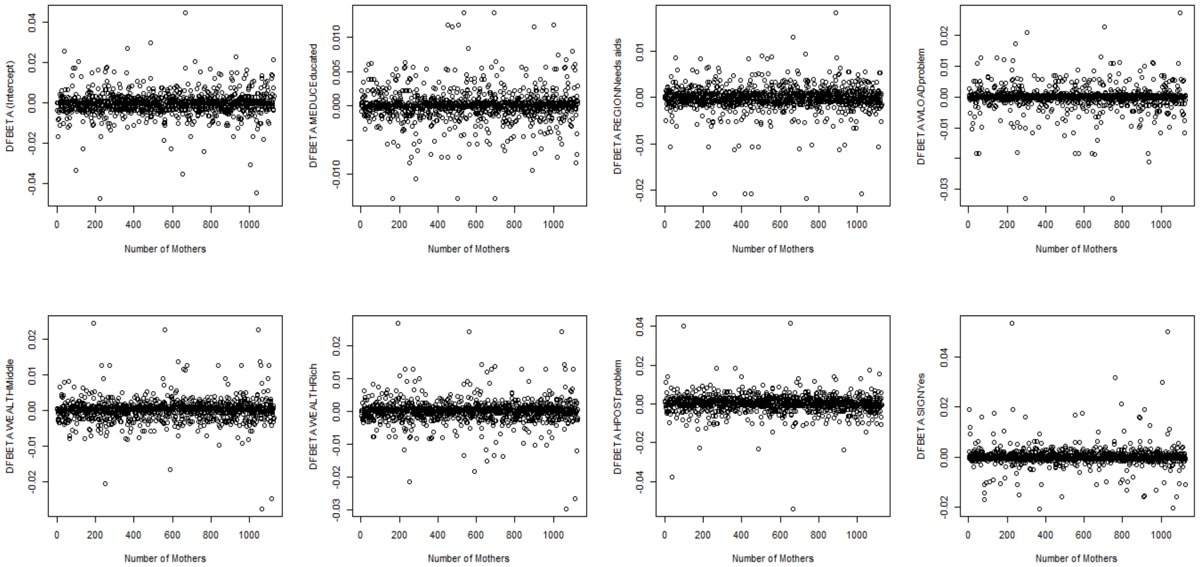


Figure 15: DFBETAs of Different Explanatory Variables

Appendix C: SAS Code

```
*Standard Poisson;
Proc nlmixed data = Antenatal;
parms a0 =-0.2638 a1 = 0.7542 a2 = -0.4352 a3=-0.9568 a4 =0.3844
a5 = -0.7527 a6 =0.5495 a7 = 0.6106 a8 = 0.3877 a9 = 0.3708;
lambda = exp(a0 + a1 * MEDUC + a2 * REGION + a3*WLOAD + a4 * WEALTH +
a5 * HPOST + a6 * AWARN + a7 * SIGN + a8 * WLOAD*AWARN + a9*RESID*HPOST);
ll = -lambda + ANC * log(lambda) - log(fact(ANC));
model ANC ~ general(ll);
predict lambda out = poi_out (rename = (pred = Yhat));
title1 "Poisson Regression"; run;

*Negative Binomial;
Proc nlmixed data = ANTENATAL;
parms b0 =-0.2638 b1 = 0.7542 b2 = -0.4352 b3=-0.9568 b4 =0.3844
b5 = -0.7527 b6 =0.5495 b7 = 0.6106 b8 = 0.3877 b9 = 0.3708;
etanb= b0 + b1 * MEDUC + b2 * REGION + b3*WLOAD + b4 * WEALTH + b5 * HPOST +
b6 * AWARN + b7 * SIGN + b8 * WLOAD*AWARN + b9*RESID*HPOST;
lambda = exp(etanb);
ll=lgamma(ANC+1/k)-lgamma(ANC+1)-lgamma(1/k)+ANC*log(k*lambda)-
(ANC+1/k)*log(1+k*lambda);
ESTIMATE "inflation probability" lambda;
model ANC ~ general(ll);
ods output Modelfit=fit;
title1 "Negative Binomial Regression"; run;

*ZIP;
Proc nlmixed data=ANTENATAL;
parms a0 =-0.2638 a1 = 0.7542 a2 = -0.4352 a3=-0.9568 a4 =0.3844 a5 = -0.7527
a6 =0.5495 a7 = 0.6106 a8 = 0.3877 a9 = 0.3708 b0 = -0.1577 b1 = 0.4882
b2 = -0.4276 b3=-0.9733 b4 =0.2199 b5 = -0.7822 b6 =0.5632 b7 =0.6307
b8 =0.4249 b9 =0.3942;
etazip = a0 + a1 * MEDUC + a2 * REGION + a3 * WLOAD + a4 * WEALTH + a5 *
HPOST  + a6 * AWARN + a7 * SIGN + a8 *WLOAD*AWARN  + a9 * RESID*HPOST;
infprob = 1/(1+exp(-etazip));
lambda = exp(b0 + b1 * MEDUC + b2 * REGION + b3 * WLOAD + b4 *WEALTH  +
b5 *HPOST  + b6 * AWARN  + b7 * SIGN + b8 * WLOAD*AWARN  + b9 *RESID*HPOST);
if ANC=0 then ll = log(infprob + (1-infprob)*exp(-lambda));
else ll = log((1-infprob)) + ANC *log(lambda)-lgamma(ANC+1)-lambda;
model ANC ~ general(ll);
predict _ll out=LL_3;
ods output Modelfit=fit;
ESTIMATE "inflation probability" infprob;
ESTIMATE "lambda" lambda;
title1 "ZIP Regression model";run;

*ZINB;
Proc nlmixed data = ANTENATAL tech = dbldog;
parms a0=-0.2638 a1=0.7542 a2=-0.4352 a3=-0.9568 a4=0.3844 a5=-0.7527
a6=0.5495 a7=0.6106 a8=0.3877 a9=0.3708 b0=-0.1577 b1=0.4882 b2=-0.4276
b3=-0.9733 b4=0.2199 b5=-0.7822 b6=0.5632 b7=0.6307 b8=0.4249 b9=0.3942;
etazinb = a0 + a1 * MEDUC + a2 * REGION + a3 * WLOAD + a4 * WEALTH + a5 *
HPOST  + a6 * AWARN + a7 * SIGN + a8 *WLOAD*AWARN  + a9 * RESID*HPOST;
lambda = exp(etazinb)/(1+exp(etazinb));
etap = b0 + b1 * MEDUC + b2 * REGION + b3 * WLOAD + b4 *WEALTH  +
b5 *HPOST + b6 * AWARN + b7 * SIGN + b8 * WLOAD*AWARN + b9 *RESID*HPOST;
mu = exp(etap);
```

```
if ANC = 0 then ll = log(lambda+(1-
lambda)*(((1/k)**(1/k))/((mu+(1/k))**(1/k))));
else ll = log(1-lambda) + lgamma(ANC +(1/k)) + ANC*log(k*mu)
-lgamma(ANC+1)-lgamma(1/k)-(ANC+(1/k))*log(1+k*mu);
model ANC ~ general(ll);
ods output Modelfit=fit;
title1 "ZINB model regression Model"; run;

*HP;
proc nlmixed data = ANTENATAL tech = dbldog;
parms a0 =-0.2638 a1 = 0.7542 a2 = -0.4352 a3=-0.9568 a4 =0.3844 a5 = -0.7527
a6 =0.5495 a7 = 0.6106 a8 = 0.3877 a9 = 0.3708 b0 = -0.1577 b1 = 0.4882 b2 =
-0.4276 b3=-0.9733 b4 =0.2199
b5 = -0.7822 b6 =0.5632 b7 =0.6307 b8 =0.4249 b9 =0.3942;
etahp = a0 + a1 * MEDUC + a2 * REGION + a3 * WLOAD + a4 * WEALTH + a5 * HPOST
+ a6 * AWARN + a7 * SIGN + a8 *WLOAD*AWARN  + a9 * RESID*HPOST;
exp_eta0 = exp(etahp);
lambda = exp_eta0 / (1 + exp_eta0);
etap = b0 + b1 * MEDUC + b2 * REGION + b3 * WLOAD + b4 *WEALTH  +
b5 *HPOST  + b6 * AWARN  + b7 * SIGN + b8 * WLOAD*AWARN  + b9 *RESID*HPOST;
exp_etap = exp(etap);
if ANC = 0 then ll = log(lambda);
else ll = log(1 - lambda) - exp_etap + ANC * etap - lgamma(ANC + 1)
- log(1 - exp(-exp_etap));
model ANC ~ general(ll);
predict _ll out=LL_5;
predict exp_etap out = hdl_out1 (keep = pred ANC rename = (pred = Yhat));
predict lambda out = hdl_out2 (keep = pred rename = (pred = lambda));
title "Hurdle Poisson";
run;

 *HNB;
Proc nlmixed Data= ANTENATAL TECH=NRRIDG;
parms a0 =-0.2638 a1 = 0.7542 a2 = -0.4352 a3=-0.9568 a4 =0.3844 a5 = -0.7527
a6 =0.5495 a7 = 0.6106 a8 = 0.3877 a9 = 0.3708 b0 = -0.1577 b1 = 0.4882
b2 = -0.4276 b3=-0.9733 b4 =0.2199 b5 = -0.7822 b6 =0.5632 b7 =0.6307 b8
=0.4249 b9 =0.3942;
eta1 = a0 + a1 * MEDUC + a2 * REGION + a3 * WLOAD + a4 * WEALTH + a5 * HPOST
+ a6 * AWARN + a7 * SIGN + a8 *WLOAD*AWARN + a9 * RESID*HPOST;
expeta1 = exp (eta1);
eta2 = b1 * MEDUC + b2 * REGION + b3 * WLOAD + b4 *WEALTH +
b5 *HPOST + b6 * AWARN + b7 * SIGN + b8 * WLOAD*AWARN + b9 *RESID*HPOST;
expeta2 = exp (eta2); m=expeta2;
p=k*m/ (1+k*m); P1=k*m; p2=1+P1;
P_negbin_0=(1/p2)**(1/k);
P_binom_0 =1-(expeta1/(1+expeta1));
Pred2 =m*(1-p_binom_0)/(1-P_negbin_0);
P_nb=lgamma(ANC+1/k)-lgamma(ANC+1)-lgamma(1/k)+ 1/k*log(1-p) + ANC*log(p);
P_nb0=(1/k)*log(1-p);
P_ztnb=eta1-log(1+expeta1)+P_nb-log(1-exp(P_nb0));
if ANC=0 then ll=-log(1+expeta1);
else ll= P_ztnb;
model ANC~general(ll);
ESTIMATE "inflation probability" expeta1;
ESTIMATE "Exp(mu)" P_binom_0;
title "Hurdle Negative Binomial";
run;
```

```
/* Vuong Test: ZIP VS. Hurdle Poisson **/
title1 'Vuong test for ZIP VS. Hurdle Poisson';
title2 'H0 = no improvement of ZIP over Hurdle Poisson';
data ll_diff;
merge ll_3 (rename= (pred=ll_zip))
ll_5 (rename= (pred=ll_hp));
run;
data ll_diff;
set ll_diff;
lr_i = ll_hp - ll_zip;
keep ll_zip ll_hp lr_i;
run;
proc means data=ll_diff vardef=n;
var lr_i;
output out=vuong_stats mean=LR var=V_lr_i n=n;
run;
data vuong_stats;
set vuong_stats;
Vuong = (LR /sqrt(V_lr_i/n));
p = 2*(1-probnorm(vuong));
put vuong= p=;
run;

*Test for one sided overdispersion;
data fit;
set fit(where=(criterion="Scaled Pearson X2"));
format pvalue pvalue6.4;
pvalue=1-probchi(value,df);
run;
proc print data=fit noobs;
var criterion value df pvalue;
run;
```