DEALING WITH EXCESS ZEROS IN A DISCRETE DEPENDENT VARIABLE IN THE REPORTED NUMBER OF ANTENATAL CARE VISITS

BY

AGBAJE, SHAKIRAT AYOOLA

BSc. STATISTICS (UNIVERSITY OF IBADAN)

MATRIC NUMBER: 138783

A DISSERTATION SUBMITTED TO

THE DEPARTMENT OF EPIDEMIOLOGY AND MEDICAL STATISTICS,

FACULTY OF PUBLIC HEALTH, COLLEGE OF MEDICINE,

UNIVERSITY OF IBADAN

IN PARTIAL FULFILMENT OF

THE REQUIREMENT FOR THE AWARD OF

MASTERS OF SCIENCE DEGREE IN MEDICAL STATISTICS

JANUARY, 2016

CERTIFICATION

I certify that this work was carried out by Agbaje, Shakirat Ayoola in the Department of Epidemiology and Medical Statistics, Faculty of Public Health, University of Ibadan under my supervision.

....

Supervisor

Dr. (Mrs.) O.B. Yusuf

BSc. Statistics, MSc. (Ibadan)

PhD (Ibadan)

Department of Epidemiology and Medical Statistics,

Faculty of Public Health,

College of Medicine, University of Ibadan

....

Co-Supervisor

Dr. R.F. Afolabi

BSc. (Ilorin), MSc. (Ilorin) PhD (Ilorin)

Department of Epidemiology and Medical Statistics,

Faculty of Public Health,

College of Medicine, University of Ibadan

DEDICATION

To the Almighty God, in whom are all treasures of wisdom, knowledge and understanding. To him be all the glory and also to my parents Alhaji M.A Agbajelola and Mrs. S.A Agbaje.

ACKNOWLEDGEMENTS

My first appreciation goes to the Almighty God for granting me the grace, wisdom and knowledge in the completion of this project.

I am very indebted to my parents Alhaji M.A Agbajelola and Mrs. S.A Agbajefor their financial support, moral support, prayers and love. I say 'thank you'.

I would like to express my heartfelt gratitude to my supervisorsDr O.B Yusuf and Dr, A.F Afolabi, who helped me with their advice and supervision all through the period of the work. God will reward them abundantly.

I am also grateful to all my lecturers in the Department of Epidemiology and Medical statistics, College of Medicine, University of Ibadan who have in one way or the order imparted knowledge in me, the likes of Prof Ayeni,. Dr. O.B Yusuf, Dr. Fawole, Dr. M.D Dairo, Dr. Fagbamigbe, Dr. Adebowale, Dr. Adeoye, Dr. Adedokun, Dr. Akinyemi, Dr. Akpa, Mr.Gbadebo, Dr.R.F Afolabi and Mr. Afolabi.

Thanks to members of my family, Mr and Mrs Ajose Alabi and their children Favour, Micheal and Success, my younger ones Abayomi and Funmilayo Agbaje for their prayers and support.

My sincere gratitude also goes to Oluwabodunrin, Oladimeji Opeyemi, for his love and care and for providing me with moral, financial and spiritual support. Thanks for being there for me; your support was all I needed to complete this project work.

To my course mates, Mrs. Kehinde Akinpelu, Fowobaje Kayode, Pst. Samuel Olorunju, Oyedapo Abimbola, Michael Ekholuenetale, Mrs. Grace Odunlami-Livingwater, Mrs. Funmi Adelakun, Mr. Koya Inubile, Miss Olawumi Adewumi, Maryam Jaiyeola, Opeyemi Ojo, Mr. Wegbom Anthony and others members of EMSEH who made my stay interesting. I say a big thank you.

Some people have made this project succeed; others watched it being done, while some wonder how it was done.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

LIST OF ACRONYMS

ANC	Antenatal Care
ZIP	Zero Inflated Poisson
ZINB	Zero Inflated Negative Binomial
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
WHO	World Health Organisation
NDHS	National Demographic and Health Survey
MHC	Maternal Health Care
MDG	Millennium Development Goals
PR	Poisson Regression
NBR	Negative Binomial Regression
GPR	Generalized Poisson Regression
LRT	Likelihood Ratio Test
MLE	Maximum Likelihood Estimate
EA	Enumeration Areas

ABSTRACT

Background: Data with excess zeros have often been associated with count regression models. Poisson regression and Negative Binomial regression models have been used as a standard for modelling count outcomes but these methods do not take into account the problems associated with excess zeros and over-dispersion of count data. Failure to account for these extra zeros may result in biased parameter estimates and wrong inferences while over-dispersion causes the standard error of the estimates to be underestimated. Therefore, this study was designed to evaluate the performance of the Zero Inflated Poisson and Zero Inflated Negative Binomial models in determining the factors associated with the number of antenatal care visits in Nigeria.

Methods: Data for this study was obtained from the National Demographic and Health Survey (NDHS 2013). The survey made use of a cross-sectional population based study design. A sample of 31,482 women within the reproductive age of 15-49yrs who gave birth five years prior to the survey and provided information about antenatal care visits were utilised. Number of antenatal care visits was used as the dependent variable while the explanatory variables include age, region, residence, parity, educational level, religion, wealth index, employment status, husband/partners employment status and husband/partners occupation. Data were analysed using descriptive statistics, chi square test, Zero Inflated Poisson (ZIP) regression analysis and Zero Inflated Negative Binomial (ZINB) regression models, Kolmogorov-Smirnov test was used to check for over-dispersion and three test criteria (AIC, BIC and -2LogL)were used to assess model fit using SPSS version 19 and STATA version 12.

Results: Mean age of women was 29.5 ± 7.0yrs and median number of ANC visits was 4 visits. (Range = 30). Findings revealed an urban-rural differential in antenatal care utilisation; with respondents living in urban areas (78.2%) having higher proportion compared to those in rural areas (41.2%). About 53.5% had at least 4 antenatal visits while 46.5% had no ANC visit. The Zero Inflated Negative Binomial regression analysis revealed that age, region, education, wealth index, husband/partners education, respondent's employment, religion, residence and parity were significant determinants of number of antenatal care visit ($p<0.001$). The ZINB model fits the data better than ZIP model with AIC values of 85,707.75 and 94,635.19 respectively.

Conclusion: Antenatal care utilisation was relatively low among women living in rural areas compared to women living in urban areas. The Zero-Inflated Negative Binomial regression model provided the better fit for the data on number of antenatal care visits. This study suggested Zero-Inflated Negative Binomial model for count data with excess zeros and over-dispersion.

Keywords: Over-dispersion, Antenatal care utilisation, Zero inflated Poisson model, Zero inflated Negative Binomial model, Excess Zeros.

Words count: 435

CHAPTER ONE

INTRODUCTION

1.1 ANTENATAL CARE UTILISATION

Antenatal care (ANC) utilisation is an important factor that contributes to the burden of maternal mortality. Proper utilisation of antenatal care promotes safe motherhood and delivery with improved maternal outcome. The major objective of antenatal care is to ensure optimal health outcomes for the mother and her baby. Antenatal care can be defined in various ways. WHO defines antenatal care as having had one or more visits to a trained person during pregnancy. It includes routine follow up provided to all pregnant women at primary care level from screening to intensive life support during pregnancy and up to delivery.

Proper Utilisation of antenatal care from a trained provider is important to monitor the pregnancy as it indirectly influences the lives of mothers and babies by promoting safe motherhood and delivery with improved maternal and neonatal outcome, thus impacting positively on maternal and foetal health. It is often the first contact opportunity for a pregnant woman to connect with health service, this offers an entry point for integrated care thereby promoting healthy home practices, influencing care-seeking behaviours and linking women with pregnancy complications to a referral system.

Antenatal care from a trained provider is important to monitor the pregnancy and reduce morbidity risks for the mother and child during pregnancy and delivery. Antenatal care provided by a skilled health worker enables

(1) Early detection of complications and prompt treatment (e.g., detection and treatment of sexually transmitted infections)

(2) Prevention of diseases through immunisation and micronutrient supplementation,

(3) Birth preparedness and complication readiness, and

(4) Health promotion and disease prevention through health messages and counselling for pregnant women (NDHS 2013).

In recent years the country has embarked on measures to reform the healthcare system, including maternal healthcare (MHC) delivery, in a bid to attain Millennium Development

Goals (MDGs) 4 and 5. Most health reform efforts have been geared towards increasing availability of healthcare services, without appropriate increase in quality of antenatal care delivery. Studies have shown that increased availability of service does not always translate to increased access to healthcare.

The low rate of maternal/infant morbidity and mortality rates reported for developed countries compared with the extremely high figures in developing countries have been attributed to the higher utilisation of proper antenatal care by the developed countries.

The burden of maternal mortality in Nigeria is 576 maternal deaths per 100,000 live births (2013 NDHS). Nigeria makes up 1.7% of total world population yet contributes 10% of global maternal mortality and ranks second globally (to India) in number of maternal deaths.

Currently, 71% of women worldwide utilise ANC services, 95% in industrialised countries, 54% in non-industrialised countries and 64% in Sub-Sahara Africa.

Utilisation of antenatal care facility is measured by the number of visits, timing of the first visits and characteristics of users and non-users. It is assumed to greatly reduce pregnancy related complications, deaths and the general well-being of mother and child.

Optimum utilisation of antenatal care according to WHO is attending Antenatal Care at least four times during pregnancy in an antenatal care facility to attain full life saving potentials for women and their babies (WHO 2012).

Studies in developing countries have shown that the use of health-care services is related to the availability, quality and cost of services as well as social structure, health beliefs and personal characteristics of the user.

1.2 POISSON REGRESSION MODEL

The Poisson regression model is used in a situation where the dependent variable is a count variable. A Poisson distribution is derived as a limiting form of a binomial distribution when the number of trials is large and the probability of success is small. It assumes that the variance is equal to the mean which is the often violated because the assumption of equal mean and variance does not apply to real life data. Real life data are often characterised by over-dispersion of data, a situation where the variance is greater than the mean. When the variance is too large because there are many 0s as well as a few very high values, the negative binomial model is an extension that can handle the extra variance.

1.3 NEGATIVE BINOMIAL REGRESSION MODEL

The Negative Binomial regression model is an extension to the Poisson regression model; it is a more generalized model than the Poisson regression. The Negative Binomial model enables more flexibility in modelling the relationship between the conditional variance and the conditional mean compared to the Poisson model. An important characteristic of count data is the number of zeros in the sample; it can exceed the number of zeros predicted by either Poisson or negative binomial model.

1.4 ZERO-INFLATED COUNT REGRESSION MODELS

Zero-Inflated count regression models provide a very powerful way to model data with excess zeros. It assumes that the data are a mixture of two separate data generating processes, one generates only zeros and the other is either a Poisson or a negative binomial data-generating process. There are two distinct processes driving zeros, one is sampling zeros which occur by chance and can be assumed a result of a dichotomous process and the other one is structural zeros (true zeros) which are inevitable and are part of the counting process.

The ZIP and ZINB are two models that have been applied to account for the excess zeros observed in count regression models.

1.4.1 Assumption of Zero-Inflated Regression Models

These models assumes that the sample is a "mixture" of two sorts of individuals: one group whose counts are generated by the standard Poisson regression model or a standard Negative Binomial regression model, and another group (call them the absolute zero group) who have zero probability of a count greater than 0. Observed values of 0 could come from either group. Although not essential, the model is typically elaborated to include a logistic regression model predicting which group an individual belongs to.

So the usual assumptions for a logistic regression model (for the absolute zero group) and the usual assumptions for a Poisson regression and a negative binomial regression model are needed.

1.5 ZERO-INFLATED POISSON REGRESSION MODEL

Zero Inflated Poisson regression is a model for count data with excess zeros, it was first introduced by Lambert in 1992 and she applied this model to the data collected from a quality control study in which the response typically is the number of defective products in a sampling unit. It assumes that with probability p the only possible observation is 0, and with probability 1-p a Poisson distribution (λ) is observed. ZIP models are easy to fit and the MLE estimates are approximately normal in large samples and confidence intervals can be constructed by inverting likelihood ratio test. They are easy to interpret and lead to a more refined data analysis.

Zero Inflated Poisson model assumes there are two processes at work—the first process determines if an individual is even eligible for a non-zero response and the other process determines the count of event for eligible individuals. The tricky part is either process can result in a 0 count. Since one cannot tell which zeros were eligible for a non-zero count, one cannot tell which zeros were results of which process. The ZIP model fits, simultaneously, two separate regression models. One is a logistic or probit model that models the probability of being eligible for a non-zero count. The other models the size of that count.

Both models use the same predictor variables, but estimate their coefficients separately. So the predictors can have vastly different effects on the two processes.

1.6 ZERO-INFLATED NEGATIVE BINOMIAL REGRESSION MODEL

Zero-inflated negative binomial regression is for modelling count variables with excessive zeros and it is usually for over-dispersed count outcome variables. Furthermore, theory suggests that the excess zeros are generated by a separate process from the count values and that the excess zeros can be modelled independently.

Greene described an extended version of the negative binomial regression model for excess zero count data which may be more appropriate than the ZIP. It has been established that the ZIP parameter estimates can be severely biased if the non-zero counts are over-dispersed in relation to the Poisson distribution.

Moreover, the non-zero observations may be over-dispersed in relation to the Poisson distribution, biasing parameter estimates and underestimating standard errors. In such a circumstance, Zero-Inflated Negative Binomial (ZINB) regression model better accounts for these characteristics compared to a Zero-Inflated Poisson (ZIP).

But sometimes it is just a matter of having too many zeros than a Poisson would predict. In this case, a better solution is often the Zero-Inflated Poisson (ZIP) model. (And when extra variation occurs too, its close relative is the Zero-Inflated Negative Binomial model).

1.7 PROBLEM STATEMENT

There has been considerable research done on the number of antenatal care visits and antenatal care uptake. The type of model that has been used to fit the data includes, but is not limited to, the Poisson regression model, the negative binomial regression model and the logistic regression model. Given a range of possible modelling approaches and a lot of assumptions with each modelling approach, making an intelligent choice for modelling the number of antenatal care visits becomes tedious especially when the data contains extra zeros.

Outcome variables determine the type of models to be used and there are quite a few types of outcome variables that do not meet the assumption of ordinary linear model's being normally distributed residuals. An outcome variable that is not normally distributed can have normally distributed residuals, but this outcome variable needs to be continuous and measured on an at least an interval or ratio scale. Categorical dependent variables clearly do not fit this requirement of being measured on at least an interval or ratio scale, so it is easy to see that an ordinary linear model is not appropriate for it and neither does count variables. Count variable is less obvious because they are measured on a ratio scale so it is easier to think of them as continuous, or close to it. But they are not continuous and this really affects the assumption of ordinary linear models.

Continuous variables measure how much; while count variables measure how many. Continuous variables can be negative and count variables cannot be negative, 0 is the lowest possible value, and they are often skewed because of the excess zeros that they possess.

Poisson Regression and Negative Binomial Regression have been used to model count regression data and it does not take into account the problem associated with count data, especially due to excessive number of zeros (Hinde and Demetrio, 1998).In order to fit Poisson Regression model transformation of the response variable can be done which does not guarantee to fix the over-dispersion problem. (Gideon et al, 2008).Failure to account for extra zeros in count data may result in biased parameter estimates and wrong inferences, while over-dispersion causes the standard error of the estimates to be underestimated.

1.8 JUSTIFICATION OF THE STUDY

Count variables often follow a Poisson or a negative binomial distribution. The Poisson distribution assumes that each count is the result of the same Poisson process a random process that says each counted event is independent and equally likely. If this count variable is used as the outcome of a regression model, we can use Poisson regression or Negative Binomial regression to estimate the predictors of the outcome.

Zero-Inflated Regression models offers a number of advantages including modelling excess zeros and accounting for over-dispersion. The use of Zero inflated regression as a tool for modelling data with excess zeros has been used at increasing level in recent years (Hinde and Demetrio, 1998).

Several studies have been done on factors affecting the utilisation of antenatal care and often times the excess zeros found in the data are not taken into consideration. This study takes into consideration the excess zeros in the data which is an improvement on the study carried out by Yusuf and Ugalahi (2015) using PR, NBR and GPR, the models PR, NBR and GPR were used to analyse the data to find the one that fits better in the presence of over-dispersion but it did not take into account the excess zeros in the data.

Previous studies done on urban-rural differentials in antenatal care service in Nigeria had not been nationally representative as they concentrated on small communities, examining the factors that determine the urban-rural differences in antenatal care utilisation can assist in decision making as regards the type of intervention programs that will encourage women to seek antenatal care and how to bridge the gap of urban-rural differences.

The main reason for zero-inflated count models is that real-life data mostly have excess zeros and may display over-dispersion, Zero-inflated count models provide a way for modelling excess zeros as well as allowing for over-dispersion.

1.9 OBJECTIVES

1.9.1 Main Objective

To evaluate the performance of the ZIP and ZINB models in determining the model that best fit the data on the number of antenatal care visits in Nigeria.

1.9.2 Specific Objectives

1) To examine socio-economic factors affecting antenatal care utilisation across urban and rural areas.
2) To identify the better model that fits the data on the number of antenatal care visits using AIC, BIC and -2LogLikelihood.
3) To identify factors affecting utilisation of antenatal care using the better model discovered in objective 2.

CHAPTER TWO

LITERATURE REVIEW

2.1 BRIEF HISTORICAL REVIEW OF ANTENATAL CARE

Antenatal Care Services refers to the regular medical care received by women during pregnancy. It is a type of preventative care with the goal of providing regular check-ups that allow doctors or midwives to treat and prevent potential health problems throughout the course of the pregnancy while promoting healthy lifestyles that benefit both mother and child (Ahmed and Mosely 2002). Antenatal care is an important determinant of high maternal mortality rate and one of the basic components of maternal care on which the life of mothers and babies depend (Nisar and White 2003). Antenatal care service is a medical attention given to the expectant mother and her developing baby.

Systematic antenatal care was first introduced early in the 20th century in Europe and North America and is now almost universal in the developed world, it has been in existence for over 100 years (Nisar and White 2003). Antenatal care was studied by William Farr in 1914, he concluded that it reduced fetal mortality by 40 percent. It was later introduced to Nigeria in the 1950s and 1960s, but the usage dropped in the 1900s due to falling of Health system (Ekabua et al, 2011).

2.2 COVERAGE AND TREND OF ANTENATAL CARE

Globally the number of women estimated to have had at least one antenatal care visit is 71 percent worldwide, while it is more than 95% in industrialised countries In sub-Saharan Africa, 69 percent of pregnant women have at least one Antenatal care visit which is more than what is estimated in South Asia at 54 percent. Coverage of at least four ANC visits is lower at 44 percent. 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. (Ornella et al, 2006).

CHAPTER TWO

LITERATURE REVIEW

2.1 BRIEF HISTORICAL REVIEW OF ANTENATAL CARE

Antenatal Care Services refers to the regular medical care received by women during pregnancy. It is a type of preventative care with the goal of providing regular check-ups that allow doctors or midwives to treat and prevent potential health problems throughout the course of the pregnancy while promoting healthy lifestyles that benefit both mother and child (Ahmed and Mosely 2002). Antenatal care is an important determinant of high maternal mortality rate and one of the basic components of maternal care on which the life of mothers and babies depend (Nisar and White 2003). Antenatal care service is a medical attention given to the expectant mother and her developing baby.

Systematic antenatal care was first introduced early in the 20th century in Europe and North America and is now almost universal in the developed world, it has been in existence for over 100 years (Nisar and White 2003). Antenatal care was studied by William Farr in 1914, he concluded that it reduced fetal mortality by 40 percent. It was later introduced to Nigeria in the 1950s and 1960s, but the usage dropped in the 1900s due to falling of Health system (Ekabua et al, 2011).

2.2 COVERAGE AND TREND OF ANTENATAL CARE

Globally the number of women estimated to have had at least one antenatal care visit is 71 percent worldwide, while it is more than 95% in industrialised countries In sub-Saharan Africa, 69 percent of pregnant women have at least one Antenatal care visit which is more than what is estimated in South Asia at 54 percent. Coverage of at least four ANC visits is lower at 44 percent. 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. (Ornella et al, 2006).

2.3 FACTORS AFFECTING THE UTILISATION OF ANTENATAL CARE SERVICES

Studies in developing countries have shown that the use of health-care services is related to the availability, quality and cost of services, as well as to the social structure, health beliefs and personal characteristics of the users (Awusi et al, 2009).

Maternal and husband's education: Higher education attainment of both the woman and the husband has a positive impact on the rate of ANC services utilisation. The level of education of the pregnant woman and her husband often shows a significant positive association with the use of antenatal care (Simkhada *et al.* 2008,Awusi *et al.*2009, Nisar and White 2003).

Parity: The number of children ever born also affects the use of antenatal care services. It has a statistically significant negative effect on adequate attendance of antenatal care, women of lower parity tend to use antenatal care more while women of higher parity tend to use antenatal care less (Simkhada *et al.* 2008, Awusi *et al.*2009).

Availability and Accessibility: Adequate utilisation of antenatal care cannot be achieved merely by establishing health centres; quality of service also needs to be considered. Quality infrastructure and health facilities should be available and easily accessible. Perceived quality of service is a major factor that influences people's decision to use health care facility (Iyaniwura and Yusuff 2009,Simkhada *et al.* 2008, Awusi *et al.*2009). Lack of proper accessibility of antenatal facility greatly influences the use of antenatal care of pregnant women (Onamade 2014).

Affordability: According to the economic model a perceived lower quality and higher costs of antenatal care, including both time and financial costs of treatment and travel, would reduce the use of antenatal care (Simkhada *et al.*2008),. Studies have shown that effect of distance and/or travel time has a negative effect on antenatal care use (Acharya and Cleland 2000, Magadi, Madise and Rodrigues 2000 and Raghupathy1996).

Region: Place of residence has been found to have a strong influence on the utilisation of ANC. Majority of the studies carried out in developing countries have revealed urban-rural differential in the pattern of antenatal care use (Ibnouf et al., 2007; Rahman et al., 2008). The study carried out by Rahman et al. (2008) in Bangladesh showed strong difference in how antenatal care was utilised. It showed that more than half of urban women utilised antenatal care while about one-quarter of rural women utilised antenatal care services. This has been shown in parts of Nigeria, the study carried out by Dairo and Owoyokun (2010) in Ibadan showed that respondents in urban areas were more likely to attend antenatal care than respondents in rural areas.

2.4 ZERO INFLATED REGRESSION MODEL

Application areas that the zero inflated regression model has been used is diverse and have included the zero-inflated Poisson (ZIP) regression models with an application to defects in manufacturing defects (Lambert, 1992), the Generalized Poisson Regression model has been used to model a household fertility data set (Wang and Famoye, 1997) and to model injury data (Wulu *et al*., 2002), patent applications (Crepon & Duguet, 1997), road safety (Miaou, 1994), species abundance (Welsh *et al*., 1996; Faddy, 1998), medical consultations (Gurmu, 1997), use of recreational facilities (Gurmu & Trivedi, 1996; Shonkwiler & Shaw, 1996) and sexual behaviour (Heilbron, 1994). Lambert (1992) provides a motivation application of these models and discusses the case of zero-inflated Poisson (ZIP) models.
Several models have been proposed to handle count data with too many zeros than expected; The zero-inflated negative binomial (ZINB) regression model with correction for misclassification was applied to dental caries data (Mwalili *et al.* 2008).

Hall (2000) described the zero-inflated binomial (ZIB) regression model and incorporated random effects into ZIP and ZIB models; Other models in the literature include the hurdle model (Mullahy, 1986) and the semi-parametric model.

In a study of the Zero Inflated Regression Models on the effect of air pollutants on hospital admissions, five different models, the Poisson Regression (PR), Negative Binomial Regression (NBR), Generalized Poisson Regression (GPR), Zero Inflated Poisson Regression (ZIPR) and the Zero inflated Negative Binomial Regression (ZINBR) were studied on different type of data and the AIC and BIC were used to check for the model fit. It was discovered that in the data with larger zeros, the AIC and BIC of PR, NBR and GPR was larger than the AIC and BIC values of the zero inflated models (ZINBR and ZIPR) (Cengiz 2011).

In the application of Zero Inflated Regression Model with an application to under-5- deaths, three zero-inflated models, ZIP, ZINB and Hurdle model, were used to observe whether there is any effect of proportion of zeros in the performance of the models. Two classical count data models were also considered, the Poisson and the negative binomial for comparison, it was discovered that the zero inflated models, ZIP, ZINB and Hurdle, are consistent over the changes of the model parameters. It was stated that sometimes over-dispersion 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. But 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. It was discovered that the Zero Inflated models have better performance than the Poisson and Negative binomial models using the AIC statistic for model fit (Al Mamun 2014).

The effect of not using adjusted (inflated or deflated) model when the occurrence of zero differs from what is expected was studied using the zero-adjusted generalized Poisson distribution. The zero-adjusted generalized Poisson model parameters were estimated by the method of maximum likelihood and it showed that more errors are committed for small values of the count if adjustment is ignored. It was noted that the zero-adjusted generalized Poisson distribution fitted very well the fetal movement data it was applied to and this zero-adjusted generalized Poisson regression model also fitted well when it was also applied in a study that looked at the death notice data of London times (Gupta *et al.* 1996).

While comparing the test statistic for zero-inflated Negative Binomial regression model against the zero inflated Poisson model in a study conducted by Muniswamy *et al*, in 2015 the power of three tests (Likelihood ratio test, Wald test and Score test) for different degrees of over-dispersion parameters was tested, it was discovered that when all of the test statistics indicate very strong evidence that the ZIP regression model does not fit the data then ZINB regression model can be used to fit the data. The AIC and BIC can also be used to present additional evidence that indicates that ZINB model better predicts than the ZIP model if its value for the AIC and BIC are small and also that the ZINB model is a better choice than the ZIP model. It was also discovered that the Monte Carlo simulation test indicates that for dataset with small over-dispersion parameter values the ZIP model is more appropriate while ZINB regression model is more appropriate for data that has high over-dispersion values.

Although ZIP and ZINB may be appropriate for different datasets, that is they behave differently for different dataset based on the over-dispersion the dataset exhibits. This point was shown in the example that was used in the study, the AIC and BIC test was used to explain the choice between ZIP and ZINB regression model (Muniswamy *et al.* 2015)

The Generalized Poisson Regression Model (GPR) is used in a situation where there is no Zero-inflation in the data but the data exhibits over-dispersion. When the GPR model was applied to accident data it was discovered that the GPR model performed better than the Poisson Regression (PR) model in identifying demographic factors, driving habits and medication use that is associated with the number of accidents involving elderly drivers. (Famoye *et al*, 2004).

In a study conducted by Ismail and Zamani in 2013 to estimate count data, the Negative Binomial, Generalized Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models were used. The study related the zero-inflated negative binomial and zero-inflated generalized Poisson regression models through the mean-variance relationship and suggests the application of these zero-inflated models for zero-inflated and over-dispersed count data. In the study, variety of models were fitted to two different datasets, involving several forms of NB, GP, ZINB and ZIGP regressions; it was concluded that the selection of the best model depends on many considerations. First, check for over-dispersion, and if the data is slightly over-dispersed than a quasi-Poisson regression can be fitted to the data. Secondly, if the data is largely over-dispersed and it is not caused by excessive zeros but due to variation in the data, then a NB and GP regressions model can be fitted to the data. Thirdly, if the data is both over-dispersed and zero-inflated then a zero-inflated (ZINB, ZIGP) regression model or hurdle (HNB, HGP) regression model can be fitted to the data.

Ismail and Zamani (2013) stated specifically that

> "Zero-inflated models are interpreted as a mix of structural and sampling zeros from two processes; the process that generates excess zeros from a binary distribution which are the structural zeros, and the process that generates both non-negative and zero counts from Poisson or NB distributions which are the sampling zeros. In the other hand, hurdle models assume that all zeros are sampling zeros. Therefore, as a crude guideline, if occurrences of count event do not depend on any condition

and may occur at any time, the hurdle models should be fitted. However, if occurrences of count events depend on specific conditions and/or time, such as the case of deductible or no claim discount in insurance data, the zero-inflated models are more appropriate''.

In studies where the performance of the Poisson Regression and the Zero Inflated Poisson Regression were compared, the Zero Inflated Poisson Regression performed better than the Poisson Regression when excess zeros were suspected. When it was applied to the data on the number of black spots in Corriedale sheep, the performance of the two models was compared under four simulation scenarios and the Deviance Information Criterion was used in accessing model fit. It was discovered that the ZIP model performed well in all situation (Naya *et al.*2008). When the two models were applied to private health insurance data the probability integral transform test and the Vuong's test were used to compare between the two models, it was discovered that the Zero-inflated Poisson regression model performed better than the Poisson regression model this means that the ZIP model fits excess counts data better than the standard Poisson regression (Mouatassim and Ezzahid2012).

ZIP models can be used for data with the occurrence of structural zeros but the count component possesses equality of mean and variance. For data with too many zeros and heavy tails the Zero Inflated Negative Binomial (ZINB), Zero Inflated Generalized Poisson (ZIGP), Zero Inflated Double Poisson (ZIDP), Zero Inflated Inverse Trinomial (ZIIT), Zero Inflated Strict Arcsine (ZISA) models can be used (Phang and Loh2013).

In a study comparing the ZIP, ZINB and the ZIGP to the modelling and testing of claim count data, it was discovered that the value of the AIC for the ZIGP model which was used to test the three different models was the lowest, which suggests that the ZIGP model gives the best results for the dataset (Wolny-Dominiak, 2013).

ZINB regression model sometimes does not converge in fitting a data; Lambert in 1992 observed this problem in fitting ZINB regression model to an observed data set Famoye and Singh 2006 also encountered this problem. This realization led to the development and to application of the ZIGP regression model for modelling over-dispersed data with too many zeros. It was stated that even though the ZIGP regression model is a good competitor of ZINB regression model, the conditions at which it is better is not known, if any, which one will be better. The only observation was the fact that in all dataset fitted to both models; the ZIGP regression model was successfully fitted to all dataset. However, in a few cases, the iterative technique to estimate the parameters of ZINB regression model did not converge (Famoye and Singh 2006)

Zero Inflated Negative Binomial (ZINB) model is useful for analysis of over-dispersed count data with an excess of zeros. In the application of ZINB to model heavy vehicle crash rate the AIC was used to measure the relative goodness of fit of the various models that was tested and the study showed that ZINB is a good fit for data with count outcome (Sharma and Landge2013). When applied to human microbiota sequence data with random effects, it was discovered that the model is useful for analysis of over-dispersed count data with an excess of zeros and repeated measures. The simulation study used also indicated that the method of estimation gave unbiased results for both fixed effects and random effects. The application of the ZINB model to the three selected organism from the microbiota data demonstrated the usefulness of ZINB (Rui Fang 2008).

CHAPTER THREE

METHODOLOGY

3.1 STUDY AREA

Nigeria lies on the west coast of Africa between latitudes 4°16' and 13°53' north and longitudes2°40' and 14°41' east. It occupies approximately 923,768 square kilometres of land stretching from the Gulf of Guinea on the Atlantic coast in the south to the fringes of the Sahara Desert in the north. The territorial boundaries are defined by the republics of Niger and Chad in the north, the Republic of Cameroon on the east, and the Republic of Benin on the west. Nigeria is the most populous country in Africa and the 14th largest in land mass. The country's 2006 Population and Housing Census placed the country's population at 140,431,790.Nigeria has great geographical diversity, with its topography characterised by two main landforms: lowlands and highlands. The uplands stretch from 600 to 1,300 metres in the North Central and the east highlands, with lowlands of less than 20 metres in the coastal areas. Nigeria has a tropical climate with wet and dry seasons associated with the movement of the inter-tropical convergence zone north and south of the equator.

The dry season occurs from October to March with a spell of cool, dry, and dusty Harmattan wind felt mostly in the north in December and January. The wet season occurs from April to September. The temperature in Nigeria oscillates between 25°C and 40°C, and rainfall ranges from 2,650 millimetres in the southeast to less than 600 millimetres in some parts of the north, mainly on the fringes of the Sahara Desert.

Presently, Nigeria is made up of 36 states and a Federal Capital Territory, grouped into six geopolitical zones: North Central, North East, North West, South East, South South, and South West. There are 774 constitutionally recognised local government areas (LGAs) in the country.

3.2 STUDY DESIGN

Data for this study was obtained from the National Demographic and Health Survey (NDHS 2013) and secondary data analysis was conducted to answer the study objectives. The survey made use of a cross-sectional population based study design. This study explores the factors affecting antenatal care visits.

3.3 STUDY POPULATION

For the purpose of this study, women aged 15-49 years are the target population. The study population was gotten as sub samples from the sample of women interviewed in the survey.

3.4 SAMPLING FRAME AND TECHNIQUE

The sample for the 2013 NDHS was nationally representative and covered the entire population residing in non-institutional dwelling units in the country. The survey used as a sampling frame the list of enumeration areas (EAs). Administratively, Nigeria is divided into 36 states and FCT. Each state is subdivided into local government areas (LGAs), and each LGA is divided into localities, then each locality was subdivided into census Enumeration Areas (EAs). A complete list of the EAs served as the sample frame of the survey.

The sampling technique was a stratified sample, selected at random from in three stages from the sampling ~~frame~~. The first stage; each state was stratified into urban and rural areas; this resulted in a list of localities. Second stage; one enumeration area as randomly selected from a selected locality with equal probability selection, the resulting list of household serve as a sampling frame for the selection of households in the third stage. The third stage; fixed number of households were selected in every urban and rural cluster through probability systematic sampling using the household listing.

3.5 SAMPLE SIZE

A nationally representative sample of 40 320 households were selected for the NDHS 2013 survey, of the household occupied 38 522 were successfully interviewed of which 39 902 women aged 15-49 and 18 229 males aged 15-49 were eligible for interview. However 38 948 females and 17 359 males were successfully interviewed. For this study a sample of 31 482 women within the reproductive age of 15-49 who gave birth five years prior to the survey and provided information about antenatal care visits were utilized.

3.6 DATA COLLECTION

Questionnaires were used for data collection. Face to face interviews were conducted and all women within reproductive age 15-49 were selected and interviewed. Information was collected about antenatal visits from women who gave birth five years prior to the survey.

3.7 STUDY VARIABLES

The following are the variables used in the study.

3.7.1 Outcome Variable:

Number of Antenatal care visits. Each woman who had given birth in the last five years preceding the survey were asked about the number of visits they had to an antenatal health care facility

3.7.2 Independent Variables:

1) Respondent's current age
2) Region (based on the six geopolitical zones)
3) Type of place of residence (based on urban or rural)
4) Highest educational level
5) Religion
6) Wealth index
7) Total child ever born (Parity)
8) Husband/partner's educational level
9) Husband/partner's occupation
10) Respondent currently working

3.8 DATA MANAGEMENT AND ANALYSIS

Extraction of relevant data was performed; simple summary statistics (percentage for categorical variables and mean for continuous variables) for all independent variables was also performed. Sample mean and sample variance for the dependent variable (number of antenatal care visits) was calculated to check for the presence of over-dispersion or under-dispersion. Socio-economic factors that affect ANC visits were identified across rural and urban areas. Kolmogorov-Smirnov test was also carried out to check for over-dispersion or under-dispersion. Factors that affect ANC visits were identified using the Zero Inflated

Poisson and the Zero Inflated Negative Binomial model. The zero inflated Poisson regression, zero inflated negative binomial regression parameters were estimated after fitting the independent variables to each model. Significance level was tested at $\alpha = 0.05$. Model fit was assessed using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Log-likelihood ratio test. 95% CI and Incidence Rate Ratio (IRR) were reported. SPSS version 19 and STATA version 12 were used for the analysis.

3.9 OPERATIONAL DEFINITIONS

3.9.1 **Optimum Utilisation of Antenatal care:** Optimum utilisation is measured by the number of visits made to ANC facility which ideally is a minimum of 4 visits, timing of visits and characteristics of users and non-users (WHO, 2002; Ornella et al, 2011), this definition of optimum utilisation of antenatal care by WHO is the same for this study.

3.9.2 **Poor Utilisation of Antenatal care:** Poor utilisation is defined as having less than four visits to made to ANC facility.

3.10 ZERO INFLATED REGRESSION MODEL

In a zero-inflated regression model, for every observation there are two possible data generating processes. For observation *i*, the first data generating process (process 1) generates only zero counts and the second data generating process (process 2) generates counts from either a Poisson or a negative binomial model, where process 1 is chosen with probability φ_i and process 2 is chosen with probability $1 - \varphi_i$. For instance in the number of antenatal care visit data that was used, the two possible processes are that a woman who is pregnant will attend Antenatal care or that she will not. If a woman who is pregnant do not attend ANC then the only possible outcome is zero, if a woman who is pregnant attends ANC; it is then a count process, the count is the number of times she attends ANC. The two parts of a zero-inflated model are a binary model, usually a logit model to model which of the two processes the zero outcome is associated with and a count model, in this case a Poisson model and negative binomial model to model the count process.

In general,

$$y_i \sim \begin{cases} 0 & \text{with probability} \varphi_i \\ g(y_i) & \text{with probability } 1-\varphi_i \end{cases}$$

Therefore, the probability of $\{Y_i = y_i\}$ can be described as

$P(y_i | x_i) = \varphi_i + (1 - \varphi_i)g(y_i)$ for $y_i = 0$

$P(y_i | x_i) = (1 - \varphi_i)g(y_i),$ for $y_i \geq 1$

where $g(y_i)$ follows either the Poisson or the negative binomial distribution.

When the probability φ_i depends on the characteristics of observation i, φ_i is written as a function of $z_i'\gamma$, where z_i' is the $1 \times (q+1)$ vector of zero-inflated covariates to be estimated, associated with the known zero-inflated covariate vector $z_i' = (1, Z_1, \ldots, Z_q)$ and γ is the $(q+1) \times 1$ vector of zero-inflated coefficients to be estimated. The zero-inflated intercept is γ_0, the coefficients for the q zero-inflated covariates are $\gamma_1, \ldots, \gamma_q$ where p is the number of covariates not including the intercept, and q is the number of the covariates Z's not including the intercept. The parameter φ_i, which is often referred to as the zero-inflation factor, is the probability of zero counts from the binary process. For common choice and simplicity, φ_i is characterized in terms of a logistic regression model by writing as *logit*$(\varphi_i) = Zi'\gamma$.

The function F relating the product $z_i'\gamma$(which is a scalar) to the probability φ_i is called the zero-inflated link function,

$$\varphi_i = F_i = F(z_i'\gamma)$$

The zero-inflated link function F can be specified as either the logistic function,

$$F\left(z_i'\gamma\right) = \Lambda\left(z_i'\gamma\right) = \frac{exp(z_i'\gamma)}{1 + exp(z_i'\gamma)}$$

It can also be specified as the standard normal cumulative distribution function (also called the probit function),

$$F(z_i'\gamma) = \phi(z_i'\gamma) = \int_0^{z_i'\gamma} \frac{1}{\sqrt{2\pi}} exp(-u^2/2)du$$

3.10.1 ZERO INFLATED POISSON REGRESSION MODEL

The second data generating process (process 2) is

$$g(y_i) = \frac{exp(-\mu_i)\mu_i^{y_i}}{y_i!}$$

where $\mu_i = e^{x_i'\beta}$.

Thus the ZIP model is defined as

$$P(y_i|x_i, z_i) = \begin{cases} F_i + (1 - F_i)\exp(-\mu_i)\, for y_i = 0 \\ (1 - F_i)\dfrac{exp(-\mu_i)\mu_i^{y_i}}{y_i!}\, for y_i \geq 1 \end{cases}$$

The conditional mean of y_i is given by,

$$E(y_i|x_i, z_i) = \mu_i(1 - F_i)$$

while the conditional variance of y_i is given by

$$V(y_i|x_i, z_i) = E(y_i|x_i, z_i)(1 + \mu_i F_i)$$

The ZIP model exhibits over-dispersion since

$$V(y_i|x_i, z_i) > E(y_i|x_i, z_i)$$

The Log-Likelihood function of the ZIP model is given by

$$L = \sum_{i=1}^{N} \ln[P(y_i|x_i, z_i)]$$

3.10.1.1 ZIP Model with Logistic Link Function

The logit link function is used to model the likelihood of structural zeros. Thus, the presence of structural zeros gives rise not only to a more complex distribution, but also creates an additional link function for modelling the effect of explanatory variables for the occurrence of such zeros. In other words, the ZIP model enables us to better understand the effect of covariates by distinguishing the effects of each specific covariate on structural zeros (likelihood for being non-risk) and on the count response (mean of Poisson model for the at-risk subgroup).

The ZIP model in which the probability φ_i is expressed with a logistic link function, namely

$$\varphi_i = \frac{exp(z_i^{'}\gamma)}{1 + exp(z_i^{'}\gamma)}$$

The log-likelihood function is

$$L = \sum_{\{i:y_i=0\}} \ln[exp(z_i'\gamma) + exp(-\exp(x_i'\beta))$$

$$+ \sum_{\{i:y_i>0\}} [y_i x_i'\beta - \exp(x_i'\beta) - \sum_{k=2}^{n} \ln(k)]$$

$$- \sum_{i=1}^{N} \ln[1 + \exp(z_i'\gamma)]$$

3.10.2 ZERO-INFLATED NEGATIVE BINOMIAL REGRESSION MODEL

This is obtained by specifying a negative binomial distribution for the data generation process 2,

$$g(y_i) = \frac{\Gamma\left(\frac{1+\alpha\mu_i}{\alpha}\right)}{\Gamma\left(\frac{1}{a}\right)y_i!}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}$$

Thus the ZINB model is defined to be

$$p(y_i|x_i, z_i) = \begin{cases} F_i + (1-F_i)\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} for y_i = 0 \\ (1-F_i)\frac{\Gamma\left(\frac{1+\alpha\mu_i}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)y_i!}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \quad \text{for } y_i > 0 \end{cases}$$

Where α is the over-dispersion parameter.

In this case, the conditional mean of y_i is,

$$E(y_i|x_i, z_i) = \mu_i(1-F_i)$$

while the conditional variance of y_i is given by

$$V(y_i|x_i, z_i) = E(y_i|x_i, z_i)[1 + \mu_i(F_i + \alpha)]$$

As with the ZIP model, the ZINB model exhibits over-dispersion because the conditional variance exceeds the conditional mean.

3.10.2.1 ZINB Model with Logistic Link Function

The logit link function is used to model the likelihood of structural zeros. Thus, the presence of structural zeros gives rise not only to a more complex distribution, but also creates an additional link function for modeling the effect of explanatory variables for the occurrence of such zeros. In other words, the ZINB model enables us to better understand the effect of covariates by distinguishing the effects of each specific covariate on structural zeros (likelihood for being non-risk) and on the count response (mean of Negative Binomial for the at-risk subgroup).The ZINB model is obtained by specifying a negative binomial distribution for the data generation process. Using the logit model to model which of the two processesthe zero outcome is associated with and a count model, in this case a negative binomial model, to model the count process. The expected count is expressed as a combination of the two processes.

When there is still dispersion in the at-risk subgroup, we may use the ZINB model, which is identical to ZIP, except that the NB replaces the Poisson to account for over-dispersion for modelling the count response from the at-risk subpopulation. Thus a ZINB regression model has one logistic regression for structural zeros and one NB log-linear for the count response for the at-risk subgroup, with the additional dispersion parameter α from the NB to account for over-dispersion.

The ZINB model in which the probability φ_i is expressed with a logistic link function, namely

$$\varphi_i = \frac{exp(Z_i'\gamma)}{1 + exp(Z_i'\gamma)}$$

The log-likelihood function is

$$\mathcal{L} = \sum_{i:y_i=0} \ln\left[\boldsymbol{exp}(z_i'\gamma) + (\mathbf{1} + \boldsymbol{\alpha exp}(x_i'\beta))^{-\alpha^{-1}}\right] + \sum_{i:y_i>0} \sum_{j=0}^{y_i-1} \ln(j + \alpha^{-1})$$

$$+ \sum_{i:y_i>0} \left\{-\ln(y_i!) - (y_i + \alpha^{-1}) \ln\left(1 + \alpha \boldsymbol{exp}(x_i'\beta)\right) + y_i \ln(\alpha)\right.$$

$$\left. + \, y_i x_i'\beta\right\} \sum_{i=1}^{N} \ln[1 + \boldsymbol{exp}(z_i'\gamma)]$$

3.11 PARAMETER ESTIMATION METHOD

Estimation of the parameters of the Zero Inflated Regression Model is fairly straightforward. Lambert suggested the *EM* algorithm for a Zero Inflated Poisson Model; a straightforward gradient approach will be used for this research.

3.11.1 Maximum Likelihood Estimation

The parameters will be estimated by the MLE method. By taking the partial derivatives of the log-likelihood and setting them equal to zero, the likelihood equations for estimating γ and β are obtained.

3.11.2 EM Algorithm

When the likelihood function has a complicated structure and maximizing it by numerical methods is difficult, a simple alternative procedure is the EM-algorithm developed by Dempster*et al*. The E- and M-steps can be obtained by rewriting the likelihood function so as to accommodate missing data.

3.12 KOLMOGOROV-SMIRNOV TEST

The Kolmogorov-Smirnov test (KS test) is a nonparametric test of the equality of continuous probability distribution that can be used to compare a sample with a reference probability distribution. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution.It compares the empirical distribution function of a random sample to a hypothesized cumulative distribution. The null hypothesis $H_0$is rejected in favor of the alternative hypothesis H_1at a level of significance.

H_0: Antenatal care data follows a normal distribution

H_1: Antenatal care data do not follow a normal distribution

3.13 MODEL SELECTION METHODS

The model selection methods such as Akaike information criterion (AIC), Bayesian information criterion (BIC) and the likelihood ratio test (LRT) are commonly used for model comparison especially among non-nested models. These are model selection methods based on the log-likelihood.

3.13.1 Akaike Information Criterion

Akaike Information Criterion (AIC)was developed by Akaike (1994). AIC is based on information theory and is commonly used in research because of its philosophical and computational advantages. It is derived under the assumption that the operating models belong to the approximating family.

It provides an estimate of the information that would be lost if a particular model were to be employed, the AIC shows how well the model "fits" the data or set of observations.
The criterion takes into account the statistical goodness of fit and the number of parameters that have to be estimated. In general, the smaller the AIC, the better is the model

AIC = -2 log likelihood + 2k

AIC = -2logL(θ) + 2k

where,

k = number of estimated parameters included in the model

and $L(\theta)$ is the maximum likelihood function.

Akaike Information Criterion uses the log likelihood which is the probability of obtaining the chosen data under the given model; hence it makes sense to choose a model that makes the probability as large as possible. The logarithm does not affect the value but the negative sign does, it means minimizing the value of the statistics.

3.13.2 Bayesian Information Criterion

Bayesian Information Criterion is based on information theory and it is commonly used in research because of its philosophical and computational advantages. It is a criterion for model selection among a finite set of models. It is based, in part on the likelihood function and it is closely related to the AIC; the difference is in the second term which depends on the sample size n. The BIC was developed by Gideon E. Schwarz.

Also the smaller the BIC, the better is the model

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

$$BIC = -2 \ln(L) + k \ln(n)$$

where,

L = log likelihood,

k = number of parameters and

n = number of observations

The goal of BIC is to find the best model for prediction using highest posterior probability while the goal of AIC is to identify the model that most plausibly generated the data. Both AIC and BIC can be used whether the models are nested or not.

Information criterion supply information on the strength of evidence for each model, it does not make use of a significant level rather it is based on maximum likelihood. It has a high potential of selecting the best model as it is independent of the order in which the models are computed.

3.13.3 Log-Likelihood

Likelihood Ratio Test is a statistical test used to compare the goodness of fit of two models. The test is based on the likelihood ratio and it has become one of the most popularrnethods for testing restrictions on a statistical model. The maximized likelihood L, for a given model is the value of the likelihood function when the parameters are substituted with their maximum likelihood estimates and the statistic -2logL which is used to compare models. It is a measure of agreement between the model and the data, the larger the maximum likelihood the better the agreement between the model and the observed data, however the smaller the value of -2logL, the better the model.

CHAPTER FOUR

RESULTS

4.1 SOCIO DEMOGRAPHIC CHARACTERISTICS OF RESPONDENTS

The mean age of women was 29.4 years (SD = 7.0). Median number of antenatal care visits was 4 visits (Range = 30).

The age distribution showed that the age groups 25-39 years had the highest proportion of women (65.3%) compared to the age groups 15-24 years (24.6%) and 40-49 years (10.1%). The North West geo political zone had the highest proportion of women(37.0%) compared to other zones: North East (17.5%), North Central (13.6%), South West (13.7%), South South(9.2%) and South East (8.9%).A total of 20,702, (65.0%)women lived in the rural areas while 11,126, (35.0%) lived in urban areas.

Less than half of the women(49.2%) had no education,25.8%had secondary, 19.3% had primary and5.8% had higher education. Similarly, about half of these women practiced Islam (62.3%), 36.8% were Christians while 0.9% practiced traditional religion. The Wealth Index distributions showed that a larger proportion(23.5%) of these women were in the poorest wealth quintile while 23.1%, 18.9%, 17.8% and 16.7% were in the poorer, middle, richer and richest quintiles respectively. Majority of the women (48.1%)had given birth to 2-4 children, 40.4%had given birth to more than 4 children and 11.5% had one child.

A high proportion (39.8%) of the women's partner/husband had no education, 29.1% had secondary, 19.0% had primary education and only 12.1% had higher education. Majority of the women's Husband/Partner were employed with proportion of 99.2% and only 0.8%was unemployed. Table 4.1 shows the socio demographic characteristics of respondents.

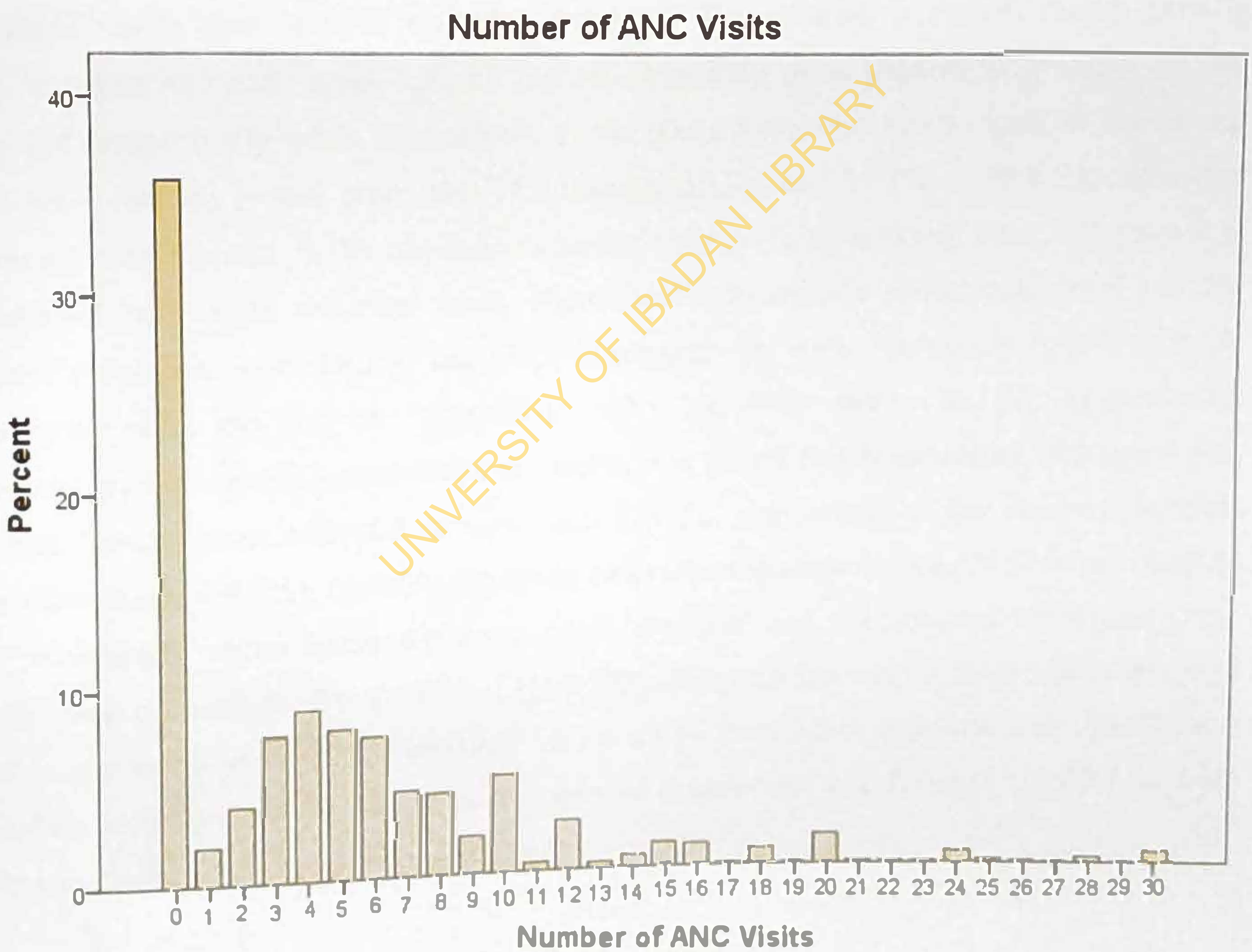
Table 4.1: Socio-demographic characteristics of the respondent

Variables	N (%)
Age group(years)	
15 – 24	7834 (24.6)
25 – 39	20793 (65.3)
40 – 49	3201 (10.3)
Region	
North Central	4340 (13.6)
North East	5578(17.5)
North West	11775 (31.5)
South East	2840 (8.9)
South South	2935 (9.2)
South West	4360 (13.7)
Residence	
Urban	11126 (35.0)
Rural	20702 (65.0)
Educational level	
No education	15657 (49.2)
Primary	6127 (19.3)
Secondary	8211 (25.8)
Higher	1834 (5.8)
Religion	
Christian	11647 (36.8)
Islam	19689 (62.3)
Traditional	291 (0.9)
Wealth index	
Poorest	7496 (23.5)
Poorer	7355 (23.1)
Middle	6001 (18.9)
Richer	5656 (17.8)
Richest	5320 (16.7)
Parity	
1	3670 (11.5)
2 – 4	15294 (48.1)
>4	12864 (40.4)
Husband's/Partner's educational level	
No education	12334 (39.8)
Primary	5884 (19.0)
Secondary	9035 (29.1)
Higher	3767 (12.1)
Husband's/Partner's employment status	
Unemployed	237 (0.8)
Employed	31092 (99.2)
Respondent's employment status	
Unemployed	9823 (31.0)
Employed	21865 (69.0)

4.1.1 Number of ANC Visits

The chart below (Fig. 4.1) illustrates the number of ANC visits. The mean number of ANC visits was 5.15and the variance was36.861. Minimum number of visits was zero and maximum number of visits was 30. Number of zeros present in the data was 6,990 (35.1%) while non-zeros observations were 12,931 (64.9%). The percentage of zero found in the data showed that there is excess zeros in the data.

Figure 4.1: Distribution of the number of ANC visits



4.2 SOCIO-ECONOMIC FACTORS AFFECTING ANC VISITS ACROSS RURAL AND URBAN AREAS

Socio economic factors which affect optimum utilisation of ANC visits are presented in table 4.2. Respondents living in urban area within the age group 25-39 years had the highest proportion of optimum utilisation (80.0%), followed by respondents aged 40-49 years (74.2%) and respondents aged 15-24 years (71.1%),($p<0.001$) while respondents in rural area within the age group 25-39 years had the highest proportion of optimum utilisation (40.7%), followed by respondents aged 40-49 years (37.9%) and respondents aged 15-24 years (35.9%), ($p<0.001$). Antenatal care visits increased with increased respondents level of education; respondents with higher educational level had the highest proportion of ANC visits 94.0% and 91.0% both in urban and rural areas respectively ($p<0.001$). Similarly, with increased wealth quintile ANC visit also increased. Respondents in richest wealth quintile both in urban and rural areas had the highest proportion of utilisation 91.2% and 83.5%, ($p<0.001$) respectively while respondents in the poorest wealth quintile both in urban and rural areas had the lowest proportion of utilisation 39.1% and 17.1%, ($p<0.001$).Utilisation increased with increase in the respondents husband/partner's educational level; the pattern is similar for both urban and rural areas, higher husband/partner's educational level had the highest proportion with (88.3%) and (73.0%) respectively while husband/partner's with no education had the lowest proportion with (42.3%) and (20.7%) respectively. Husband/partner's employment status in rural area revealed that respondents with employed husband/partner's use antenatal care less than (38.4%) respondents whose husband/partners are unemployed (59.5%), ($p<0.001$).In urban areas the proportions were (77.3%) and (82.9%) for respondents whose husband/partners were employed and unemployed respectively, but these were not statistically significant ($p=0.273$). Respondents employment status revealed that employed respondents use antenatal care more for both urban and rural area (80.6%) and (43.4%) respectively ($p<0.001$) than unemployed respondents (68.7%) and (30.5%) for both urban and rural area respectively, ($p<0.001$).

Table 4.2: Socio-economic factors affecting ANC visits across rural and urban areas

Variables	Urban Area			Rural Area		
	Poor utilisation	Optimum utilisation	P-Value	Poor utilisation	Optimum utilisation	P-Value
Age group(years)						
15 – 24	409 (28.9)	1006 (71.1)	<0.001	2445 (64.1)	1370 (35.9)	<0.001
25 – 39	961 (20.0)	3846 (80.0)		4513 (59.3)	3094 (40.7)	
40 – 49	197 (25.8)	568 (74.2)		939 (62.1)	573 (37.9)	
Educational level						
No education	802 (52.7)	721 (47.3)	<0.001	6161 (75.7)	1978 (24.3)	<0.001
Primary	331 (22.8)	1121 (77.2)		1082 (46.2)	1260 (53.8)	
Secondary	377 (12.3)	2688 (87.7)		629 (28.9)	1547 (71.1)	
Higher	57 (6.0)	890 (94.0)		25 (9.0)	252 (91.0)	
Wealth index						
Poorest	142 (60.9)	91 (39.1)	<0.001	3661 (82.9)	753 (17.1)	<0.001
Poorer	232(47.3)	258(52.7)		2677 (66.6)	1343 (33.4)	
Middle	431(38.6)	685 (61.4)		1138 (42.1)	1563 (57.9)	
Richer	502 (23.0)	1682(77.0)		346 (25.8)	997 (74.2)	
Richest	260 (8.8)	2704 (91.2)		75 (16.5)	380 (83.5)	
Parity						
1	241 (17.3)	1150 (82.7)	<0.001	1235 (56.3)	959 (43.7)	<0.001
2 - 4	668 (19.5)	2749 (80.5)		3249 (59.6)	2205 (40.4)	
>4	658 (30.2)	1521 (69.8)		3413 (64.6)	1873 (35.4)	
Husband's/Partner's educational level						
No education	644 (57.7)	473(42.3)	<0.001	5211 (79.3)	1358 (20.7)	<0.001
Primary	256 (21.0)	961(79.0)		1188 (50.8)	1151 (49.2)	
Secondary	441 (15.2)	2454(84.8)		1062 (39.1)	1652 (60.9)	
Higher	178 (11.7)	1347 (88.3)		249 (27.0)	674 (73.0)	
Husband's/Partner's employment status						
Unemployed	13(17.1)	63 (82.9)	0.273	32(40.5)	47(59.5)	<0.001
Employed	1524(22.7)	5202(77.3)		7761(61.6)	4841(38.4)	
Respondent's employment status						
Unemployed	559 (31.3)	1227 (68.7)	<0.001	3084 (69.5)	1475 (33.3)	<0.001
Employed	1004 (19.4)	4170 (80.6)		4777 (56.6)	3657 (43.4)	

4.3 MODEL COMPARISON AND SELECTION

Table 4.3 shows the comparison of the two models; Zero Inflated Poisson regression and Zero Inflated Negative Binomial regression using the AIC, BIC and -2LogL.
The values of the log-likelihood, AIC and BIC for ZINB model were 83,289.14, 83,391.14, and 83,790.99 respectively.

In general, the values of -2logLL, AIC and BIC for ZINB are smaller than the values of -2logL, AIC and BIC for ZIP. Table 4.5 shows the model comparison using the AIC, BIC and -2logL.

Table 4.3: Model comparison using the AIC, BIC and -2logLL

Model	-2LogLL	AIC	BIC
ZIP	90,075.06	90,175.06	90,567.08
ZINB	83,289.14	83,391.14	83,790.99

4.4 ZERO INFLATED NEGATIVE BINOMIAL REGRESSION ANALYSIS OF FACTORS AFFECTING ANC VISITS

Zero Inflated Negative Binomial Regression was performed to investigate the factors that affect the number of ANC visits. Table 4.4 shows the parameter estimates using Zero Inflated Negative Binomial regression analysis.

ANC visit was approximately higher by 5.9% among respondents aged 25-39years compared to respondents aged less than 24 years (IRR = 1.059; 95% CI: 1.027, 1.091).ANC visit was higher by 9.5% among respondents aged 40-49years compared to respondents aged less than 24 years (IRR = 1.095; 95% CI: 1.048, 1.115).

ANC visits among respondents from the south south, south east, and south west were higher by 14.8% (IRR = 1.148, 95% CI: 1.103, 1.194), 30.5% (IRR = 1.305; 95% CI: 1.254, 1.358) and 88.6% (IRR = 1.886; 95% CI: 1.825, 1.950) respectively compared to respondents from the north central zone. However respondents from the north east and north west zone had lower ANC visits 25.4% (IRR = 0.746; 95% CI: 0.719, 0.774) and 26.8% (IRR = 0.732; 95% CI: 0.705, 0.760). ANC visits among respondents who had primary, secondary and higher education were higher by 3.9% (IRR = 1.039; 95% CI: 1.004, 1.075), 10.7% (IRR = 1.107; 95% CI: 1.067, 1.148) and 15.8% (IRR = 1.158; 95% CI: 1.101, 1.218) respectively compared to respondents with no education. Respondents who practised Islam had a 4.5% increase in ANC visit (IRR = 1.045; 95% CI: 1.016, 1.075) compared to respondents who practised Christianity. Respondents in the middle, richer and richest wealth quintiles had a higher chance of attending ANC; 11.5% (IRR = 1.115; 95% CI: 1.066, 1.167), 20.4% (IRR = 1.204; 95% CI: 1.147, 1.264) and 25.5% (IRR = 1.255; 95% CI: 1.190, 1.325) respectively compared to respondents in the poorest wealth quintiles. Respondents who had between 2-4children and more than 4 children had 5.3% increase in ANC visit (IRR = 0.947; 95% CI: 0.919, 0.976) and 5.6% increase in ANC visit (IRR = 0.944; 95% CI: 0.910, 0.979) compared to respondents who had 1 child. Respondents whose husband's/partner's had primary, secondary and higher education were higher by 4.1% (IRR = 1.041; 95% CI: 1.004, 1.080), 8.4% (IRR = 1.084; 95% CI: 1.045, 1.124) and 9.6% (IRR = 1.096; 95% CI: 1.051, 1.143) respectively compared to respondents whose husband's/partner's had no education.

For the absolute zero group, the odds of not attending antenatal care among respondents from the north east, south east, and south west decreased by 33.7% (IRR = 0.663; 95% CI: 0.571, 0.769), 63.0% (IRR = 0.370; 95% CI: 0.274, 0.501) and 30.7% (IRR = 0.693; 95% CI: 0.561, 0.855) respectively compared to respondents from north central zone. However the odds of not attending ANC increased by 84.4% (IRR = 1.844; 95% CI: 1.602, 2.121) and 311.3% (IRR = 4.116; 95% CI: 3.469, 4.888) among respondents from the north west and south south zone compared to respondents from north central zone. The odds of not attending ANC among women who resides in rural areas increased by 70.2% (IRR = 1.702; 95% CI: 1.504, 1.927) compared to women who resides in urban areas. The odds of not attending ANC among respondents who had primary, secondary and higher education decreased by 51.1% (IRR = 0.489; 95% CI: 0.432, 0.552), 69.6% (IRR = 0.304; 95% CI: 0.259, 0.358) and 92.6% (IRR = 0.074; 95% CI: 0.037, 0.150) respectively compared to respondents with no education. Respondents who practised Islam had a decreased odds of 14.4% (IRR = 0.856; 95% CI: 0.741, 0.987) compared to respondents who practised Christianity while respondents who practised traditional religion had a 100.2% increased odds of not attending ANC (IRR = 2.002; 95% CI: 1.385, 2.891) compared to respondents who practise Christianity. The odds of not attending ANC decreased with increased wealth index, respondents in the poorer, middle, richer and richest wealth quintiles had decreased odds of not attending ANC: 45.9% (IRR = 0.541; 95% CI: 0.489, 0.599), 70.0% (IRR = 0.300; 95% CI: 0.264, 0.341), 80.0% (IRR = 0.200; 95% CI: 0.169, 0.237) and 89.2% (IRR = 0.108; 95% CI: 0.082, 0.143) respectively compared to respondents in poorest wealth quintile. Women who had between 2-4 and more than 4 children had increased odds of not attending ANC: 20.1% (IRR = 1.201; 95% CI: 1.050, 1.372), 26.4% (IRR = 1.264; 95% CI: 1.079, 1.480) respectively compared to women who had 1 child. The odds of not attending ANC decreased for respondents whose husband's/partner's had primary, secondary and higher education 51.5% (IRR = 0.485; 95% CI: 0.432, 0.544), 48.9% (IRR = 0.511; 95% CI: 0.451, 0.580) and 61.3% (IRR = 0.387; 95% CI: 0.317, 0.473) respectively compared to respondents whose husband's/partner's had no education. Women who were employed had a 31.1% decreased odds of not attending ANC (IRR = 0.689; 95% CI: 0.631, 0.751) compared to women who were unemployed.

0.855) respectively compared to respondents from north central zone. However the odds of not attending ANC increased by 84.4% (IRR = 1.844; 95% CI: 1.602, 2.121) and 311.3% (IRR = 4.116; 95% CI: 3.469, 4.888) among respondents from the north west and south south zone compared to respondents from north central zone. The odds of not attending ANC among women who resides in rural areas increased by 70.2% (IRR = 1.702; 95% CI: 1.504, 1.927) compared to women who resides in urban areas. The odds of not attending ANC among respondents who had primary, secondary and higher education decreased by 51.1% (IRR = 0.489; 95% CI: 0.432, 0.552), 69.6% (IRR = 0.304; 95% CI: 0.259, 0.358) and 92.6% (IRR = 0.074; 95% CI: 0.037, 0.150) respectively compared to respondents with no education. Respondents who practised Islam had a decreased odds of 14.4% (IRR = 0.856; 95% CI: 0.741, 0.987) compared to respondents who practised Christianity while respondents who practised traditional religion had a 100.2% increased odds of not attending ANC (IRR = 2.002; 95% CI: 1.385, 2.891) compared to respondents who practise Christianity. The odds of not attending ANC decreased with increased wealth index, respondents in the poorer, middle, richer and richest wealth quintiles had decreased odds of not attending ANC: 45.9% (IRR = 0.541; 95% CI: 0.489, 0.599), 70.0% (IRR = 0.300; 95% CI: 0.264, 0.341), 80.0% (IRR = 0.200; 95% CI: 0.169, 0.237) and 89.2% (IRR = 0.108; 95% CI: 0.082, 0.143) respectively compared to respondents in poorest wealth quintile. Women who had between 2-4 and more than 4 children had increased odds of not attending ANC: 20.1% (IRR = 1.201; 95% CI: 1.050, 1.372), 26.4% (IRR = 1.264; 95% CI: 1.079, 1.480) respectively compared to women who had 1 child. The odds of not attending ANC decreased for respondents whose husband's/partner's had primary, secondary and higher education 51.5% (IRR = 0.485; 95% CI: 0.432, 0.544), 48.9% (IRR = 0.511; 95% CI: 0.451, 0.580) and 61.3% (IRR = 0.387; 95% CI: 0.317, 0.473) respectively compared to respondents whose husband's/partner's had no education. Women who were employed had a 31.1% decreased odds of not attending ANC (IRR = 0.689; 95% CI: 0.631, 0.751) compared to women who were unemployed.

Table 4.4: Parameter estimates of factors affecting antenatal care utilisation using the Zero Inflated Negative Binomial regression.

Parameters	β	IRR	Standard error	P- Value	95% CI for IRR	
					Lower Bound	Upper Bound
Constant	1.706	5.505	0.056	<0.001	4.935	6.141
Age group						
15 – 24*						
25 – 39	0.057	1.059	0.015	<0.001	1.027	1.091
40 – 49	0.091	1.095	0.023	<0.001	1.048	1.115
Region						
North Central*						
North East	-0.293	0.746	0.019	<0.001	0.719	0.774
North West	-0.312	0.732	0.019	<0.001	0.705	0.760
South East	0.266	1.305	0.020	<0.001	1.254	1.358
South South	0.138	1.148	0.020	<0.001	1.103	1.194
South West	0.635	1.886	0.017	<0.001	1.825	1.950
Residence						
Urban*						
Rural	-0.004	0.996	0.013	0.752	0.972	1.021
Educational level						
No education*						
Primary	0.038	1.039	0.017	0.027	1.004	1.075
Secondary	0.101	1.107	0.019	<0.001	1.067	1.148
Higher	0.146	1.158	0.026	<0.001	1.101	1.218
Religion						
Christian*						
Islam	0.044	1.045	0.014	0.002	1.016	1.075
Traditional	-0.004	0.996	0.061	0.950	0.884	1.123
Wealth index						
Poorest*						
Poorer	0.021	1.021	0.023	0.357	0.977	1.067
Middle	0.109	1.115	0.023	<0.001	1.066	1.167
Richer	0.186	1.204	0.025	<0.001	1.147	1.264
Richest	0.227	1.255	0.027	<0.001	1.190	1.325
Parity						
1*						
2 - 4	-0.054	0.947	0.015	<0.001	0.919	0.976
>4	-0.058	0.944	0.019	0.002	0.910	0.979
Husband's/Partner's educational level						
No education*						
Primary	0.040	1.041	0.019	0.031	1.004	1.080
Secondary	0.081	1.084	0.019	<0.001	1.045	1.124
Higher	0.092	1.096	0.022	<0.001	1.051	1.143
Husband's/Partner's						

employment status						
Unemployed*						
Employed	-0.068	0.935	0.047	0.148	0. 853	1.024
Respondent's employment status						
Unemployed*						
Employed	0.023	1.023	0.013	0.072	0.998	1.049
Inflation Variable						
Constant	0.869	2.385	0.257	0.001	1.441	3.945
Age group						
15 – 24*						
25 – 39	-0.101	0.903	0.059	0.088	0.804	1.015
40 – 49	-0.149	0.862	0.086	0.083	0.728	1.028
Region						
North Central*						
North East	-0.411	0.663	0.076	<0.001	0.571	0.769
North West	0.612	1.844	0.072	<0.001	1.602	2.121
South East	-0.994	0.370	0.154	<0.001	0.274	0.501
South South	1.415	4.116	0.087	<0.001	3.469	4.888
South West	-0.367	0.693	0.107	0.001	0.561	0.855
Residence						
Urban*						
Rural	0.532	1.702	0.063	<0.001	1.504	1.927
Educational level						
No education*						
Primary	-0.716	0.489	0.062	<0.001	0.432	0.552
Secondary	-1.190	0.304	0.083	<0.001	0.259	0.358
Higher	-2.602	0.074	0.358	<0.001	0.037	0.150
Religion						
Christian*						
Islam	-0.156	0.856	0.073	0.032	0.741	0.987
Traditional	0.694	2.002	0.188	<0.001	1.385	2.891
Wealth index						
Poorest*						
Poorer	-0.614	0.541	0.520	<0.001	0.489	0.599
Middle	-1.205	0.300	0.065	<0.001	0.264	0.341
Richer	-1.607	0.200	0.086	<0.001	0.169	0.237
Richest	-2.222	0.108	0.140	<0.001	0.082	0.143
Parity						
1*						
2 - 4	0.183	1.201	0.068	0.008	1.050	1.372
>4	0.234	1.264	0.081	0.004	1.079	1.480
Husband's/Partner's educational level						
No education*						
Primary	-0.724	0.485	0.059	<0.001	0.432	0.544
Secondary	-0.671	0.511	0.064	<0.001	0.451	0.580
Higher	-0.950	0.387	0.102	<0.001	0.317	0.473
Husband's/Partner's employment status						

Unemployed*						
Employed	-0.285	0.752	0.227	0.209	0.481	1.174
Respondent's employment status						
Unemployed*						
Employed	-0.373	0.689	0.045	<0.001	0.631	0.751
lnalpha	-1.753	0.173	0.024	<0.001	0.165	0.182
alpha	0.173	1.189	0.004		1.179	1.200
-2LogL	83289.14					
AIC	83391.14					
BIC	83790.99					

*** Reference category**

Unemployed*						
Employed	-0.285	0.752	0.227	0.209	0.481	1.174
Respondent's employment status						
Unemployed*						
Employed	-0.373	0.689	0.045	<0.001	0.631	0.751
lnalpha	-1.753	0.173	0.024	<0.001	0.165	0.182
alpha	0.173	1.189	0.004		1.179	1.200
-2LogL	83289.14					
AIC	83391.14					
BIC	83790.99					

*** Reference category**

CHAPTER FIVE

DISCUSSION

5.1 SOCIO ECONOMIC FACTORS AFFECTING ANC UTILISATION ACROSS URBAN AND RURAL AREAS

Antenatal care utilisation has been found to vary across rural areas and urban areas, with women living in rural areas using antenatal care less than women in urban areas. The observed difference between rural and urban areas in the use of antenatal care services is likely to be due to differences in the availability and access to antenatal care services in these areas (Ibnouf et al., 2007).This study showed a difference in the pattern of antenatal care utilisation between urban and rural respondents.

There was a higher percentage of antenatal care use among the urban women than among rural women. This study showed that respondents living in urban areas use antenatal care more than respondents in rural areas and this agrees with the finding by Dairo and Owoyokun(2010); Adamu(2011). This is probably due to the fact that women in urban areas are better informed than rural women and that rural women havelimited access to ANC serviceswhile urban women have a lot of options to choose from (Dairo and Owoyokun, 2010). This difference in utilisation among rural and urban areas may also be explained by higher socio economic factors in urban areas than rural areas which ultimately affects utilisation of ANC visits. Respondents from urban areas had higher proportion of educated women, more respondents belonged to the richest wealth quintile and also had higher proportion of employed respondents compared to respondents from rural areas. All these factors strongly influence utilisation of ANC.

Wealth Index is also a factor affecting ANC utilisation, respondents in higher wealth index category use ANC more; this might be because they have more resources to access ANC services than respondents with lower wealth index. It is obvious that as the wealth index becomes high, number of ANC visits is also increased in both urban and rural areas. This agrees with the findings of Shrestha and Shrestha that women that belongs to the highest wealth quintile more often have resources and ability to buy health services (Shrestha and Shrestha 2011).Other factors such as husband/partner's educational level also affects ANC utilisation, Rahman et al. (2008) found a positive association between husband/partner's education and antenatal care use. Results revealed that the higher the educational attainment of husband/partner's, the more women use antenatal care. This might be because respondents whose husband/partners are educated have the tendency to influence their wives decision in attending ANC. This agrees with Caldwell's opinion that men who had higher levels of education make significant contributions to the decision about child care than men without any education (Caldwell, 1990).

FACTORS AFFECTING ANTENATAL CARE VISITS

The result of this study showed that age was a significant predictor for number of ANC visit, older respondents use ANC more than younger respondents, this might be as a result of the fact that older respondents are more knowledgeable and have more information than younger respondents.This study agrees with a study carried by Dairo and Owoyokun (2010), which found that older women were more likely to attend ANC, this however disagrees with the findings by Awusi et al., (2009)who found that older women use ANC less than younger women.

Geopolitical zones were also found to strongly determine number of antenatal care visits with women in the southern region using antenatal care more compared to women in the northern region. Place of residence was not statistically significant, this result however disagrees with Dairo and Owoyokun (2010); Adamu (2011) who found that urban women attend antenatal care more than rural women, this may be because urban women are more likely informed than rural women about the importance of ANC (Dairo and Owoyokun 2010). Parity was found to have a significant negative effect of antenatal care visits; respondents with higher parity tend to use antenatal care less than women with lower parity and this agrees with the study carried out by Simkhada et al. (2008) and Awusi et al. (2009) which found that parity had a significant negative effect on adequate attendance of antenatal care and stated that women of lower parity tend to use antenatal care more while women of higher parity tend to use it less. This could be due to over-confidence and being ignorant of increased pregnancy-related complication(s) risk with increased parity (Awusi et al., 2009).

Increased level of maternal education and husband's/partner's education also had a significant positive effect on number of antenatal care visits with respondents that has higher education and husband's/partner's that has higher education having the highest number of antenatal care visits this agrees with the study carried out by Simkhada et al. (2008), Awusi et al. (2009), Nisar and White (2003).

of husband/partner's, the more women use antenatal care. This might be because respondents whose husband/partners are educated have the tendency to influence their wives decision in attending ANC. This agrees with Caldwell's opinion that men who had higher levels of education make significant contributions to the decision about child care than men without any education (Caldwell, 1990).

FACTORS AFFECTING ANTENATAL CARE VISITS

The result of this study showed that age was a significant predictor for number of ANC visit, older respondents use ANC more than younger respondents, this might be as a result of the fact that older respondents are more knowledgeable and have more information than younger respondents.This study agrees with a study carried by Dairo and Owoyokun (2010), which found that older women were more likely to attend ANC, this however disagrees with the findings by Awusi et al., (2009)who found that older women use ANC less than younger women.

Geopolitical zones were also found to strongly determine number of antenatal care visits with women in the southern region using antenatal care more compared to women in the northern region. Place of residence was not statistically significant, this result however disagrees with Dairo and Owoyokun (2010); Adamu (2011) who found that urban women attend antenatal care more than rural women, this may be because urban women are more likely informed than rural women about the importance of ANC (Dairo and Owoyokun 2010). Parity was found to have a significant negative effect of antenatal care visits; respondents with higher parity tend to use antenatal care less than women with lower parity and this agrees with the study carried out by Simkhada et al. (2008) and Awusi et al. (2009) which found that parity had a significant negative effect on adequate attendance of antenatal care and stated that women of lower parity tend to use antenatal care more while women of higher parity tend to use it less. This could be due to over-confidence and being ignorant of increased pregnancy-related complication(s) risk with increased parity (Awusi et al., 2009).

Increased level of maternal education and husband's/partner's education also had a significant positive effect on number of antenatal care visits with respondents that has higher education and husband's/partner's that has higher education having the highest number of antenatal care visits this agrees with the study carried out by Simkhada et al. (2008), Awusi et al. (2009), Nisar and White (2003).

Wealth index and religion were also found to have an effect on antenatal care use with respondents in the richest category being the highest user of antenatal care and respondents who practised Islam also being the highest user of ANC.This agrees with the study carried out by Dairo and Owoyokun (2010). This might be because Christianorganizations have spiritual houses that offer care for pregnant women (Dairo and Owoyokun2010).

Respondent's employment status was found to have a significant positive effect on number of antenatal care visits using Zero Inflated Poisson model but it wasn't significant with Zero Inflated Negative Binomial model. The ZIP result agrees with the study carried out byBabalola et al., (2014) which revealed that employed women were more likely to use antenatal care than unemployed women. This might be because employed respondents have more resources to access ANC services than unemployed respondents.

5.2 COMPARISON OF THE PARAMETER ESTIMATES OF ZERO-INFLATED POISSON REGRESSION AND ZERO-INFLATED NEGATIVE BINOMIAL MODELS

If the dispersion parameter is zero and lnalpha is -infinity then a ZINB model will not be necessary, but since it is not zero then a ZINB model is more appropriate than a ZIP regression model. Furthermore, the values of the standard error of the ZIP model are smaller than the values of the standard error of the ZINB model, the larger values of the standard error of ZINB led to some regression parameters being insignificant. This shows that in the presence of over-dispersion the ZIP regression model underestimates the standard error and biases the parameter estimates by overstating the significance of the regression parameter. This result agrees with the study carried out by previous researchers (Al Mamum 2014, Muniswamy et al. 2015, Ismail and Zamani 2013 and Rui Fang 2008).They discovered that in the presence of over-dispersion and excess zerosa ZINB model fits better than a ZIP model.

This studyshows that the values of the log-likelihood, AIC and BIC for ZINB is lower than that of the ZIP model. This shows that the ZINB model fits the data better than the ZIP model which agrees with the findings by Muniswamy et al 2015 that used three tests(Likelihood

Wealth index and religion were also found to have an effect on antenatal care use with respondents in the richest category being the highest user of antenatal care and respondents who practised Islam also being the highest user of ANC.This agrees with the study carried out by Dairo and Owoyokun (2010). This might be because Christianorganizations have spiritual houses that offer care for pregnant women (Dairo and Owoyokun2010).

Respondent's employment status was found to have a significant positive effect on number of antenatal care visits using Zero Inflated Poisson model but it wasn't significant with Zero Inflated Negative Binomial model. The ZIP result agrees with the study carried out byBabalola et al., (2014) which revealed that employed women were more likely to use antenatal care than unemployed women. This might be because employed respondents have more resources to access ANC services than unemployed respondents.

5.2 COMPARISON OF THE PARAMETER ESTIMATES OF ZERO-INFLATED POISSON REGRESSION AND ZERO-INFLATED NEGATIVE BINOMIAL MODELS

If the dispersion parameter is zero and lnalpha is -infinity then a ZINB model will not be necessary, but since it is not zero then a ZINB model is more appropriate than a ZIP regression model. Furthermore, the values of the standard error of the ZIP model are smaller than the values of the standard error of the ZINB model, the larger values of the standard error of ZINB led to some regression parameters being insignificant. This shows that in the presence of over-dispersion the ZIP regression model underestimates the standard error and biases the parameter estimates by overstating the significance of the regression parameter. This result agrees with the study carried out by previous researchers (Al Mamum 2014, Muniswamy et al. 2015, Ismail and Zamani 2013 and Rui Fang 2008).They discovered that in the presence of over-dispersion and excess zerosa ZINB model fits better than a ZIP model.

This studyshows that the values of the log-likelihood, AIC and BIC for ZINB is lower than that of the ZIP model. This shows that the ZINB model fits the data better than the ZIP model which agrees with the findings by Muniswamy et al 2015 that used three tests(Likelihood ratio test, Wald test and Score test) to compare ZIP and ZINB and discovered that ZINB is a better model.

5.3 LIMITATIONS

The NDHS is a cross sectional survey, the information provided on the number of ANC visits was subjective; the exact number of visits reported may not be precise.

CONCLUSION

Factors that affect utilisation of ANC visits were found to be age, region, residence, education, religion, wealth index, parity, husband/partners educational level and respondent's employment status.

Zero Inflated Negative Binomial regression model was found to fit the data on utilisation of Antenatal Care visits in the presence of over-dispersion more than the Zero Inflated Poisson Regression model.

5.4 RECOMMENDATION

Measures to improve utilisation of ANC in rural areas should be put in place. To achieve improved use of antenatal care in rural areas, education of women living in rural areas should be looked into and programs that will make women gainfully employed should be put in place. Job opportunities should be increased in rural areas and northern regions of the country to ensure increased use of antenatal care in rural areas because employment status of respondents was found to be a significant predictor of utilisation of ANC.

In the analysis of count data in the presence of over-dispersion, Zero Inflated Negative Binomial regression model should be used, but if the data does not exhibit over-dispersion then the Zero Inflated Poisson regression model can be used.

REFRENCES

Acharya, L.B. and Cleland, J. 2000.Maternal and child health services in rural Nepal: does access or quality matter more?*Health Policy and Planning,* 15: 223 -240

Ahmed, S. and Mosley, W. H. 2002.Simultaneity in the use of Maternal-Child Health care and contraceptives: evidence from developing countries,*Demography* 39: 7 5 -93

Akaike H. 1974.A new look at the statistical model identification. *IEEE Trans on Automatic Control*; 19:716–723.

Al Mamun, M.A., 2014. Zero-inflated regression models for count data: an application to under-5 deaths. *Ball State University Muncie, Indiana.*

Awusi, V.O., et al., 2009. Determinants of antenatal care services in Emevor village, Nigeria.*Benin Journal of Postgraduate Medicine Vol. 11 Supplemental*

Babalola, B., 2014. Determinants of urban-rural differentials in antenatal care utilization in Nigeria. *African Population Studies* Vol. 28, No. 3.

Bŏhning, D. et al., 1995. Zero-Inflated Count Models and their Applications in Public Health and Social Science, *Department of Epidemiology, Institute for Social Medicine, Berlin,* 333:344.

Carrel, M. et al., 2010. Protection from annual flooding is correlated with increased cholera prevalence in Bangladesh: a zero-inflated regression analysis, *Environmental Health,* 9:13.

Cengiz, M.A., 2012. Zero-Inflated Regression Models for modelling the effect of air pollutants on Hospital Admissions, *Pol. J. Environ. Stud.* 21(3): 565-568.

Crepon, B., Duguet E. 1997. Research and development, competition and innovation {pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity. *Journal of Econometrics*, 79, 355:378.

Dairo, M. D. &Owoyokun, K.E., 2010. Factors affecting the utilization of antenatal care services in Ibadan, Nigeria, *Benin Journal of postgraduate Medicine, Vol. 12 No. 1*

Emelumadu O.F., Onyeonoro U.U., Ukegbu A.U., Ezeama N.N., Ifeadike C.O., Okezie O.K.,2014. Perception of quality of maternal healthcare services among women utilising antenatal services in selected primary health facilities in Anambra State, Southeast Nigeria, *Nigerian Medical Journal,* Vol. 55, Issue 2

Faddy, M. 1998. Stochastic models for analysis of species abundance data. In D. J. Fletcher, L. Kavalieris, and B. F. Manly (Eds.), Statistics inEcology and Environmental Monitoring 2: Decision Making and RiskAssessment in Biology, pp. 3340. *University of Otago Press*.

Famoye, F., Singh, K.P., 2006. Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data, *Journal of Data Science*, 4:117-130.

Famoye, F.,Wang W., 2004. Censored generalized Poisson regression model, *Computational Statistics and Data Analysis* 46 547-560.

Gideon, D.B, Raymond, M., 2008. Deal with excess zeros in the discrete dependent variable, the number of homicide in Chicago census Tract, *Social Statistics Section*

Gupta, P. L., Gupta, R. C., and Tripathi, R. C. 1996.Analysis of zero-adjusted count data.*Computational Statistics and Data Analysis* 23, 207-218.

Gurmu, S. 1997. Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization. *Journal of Applied Econo-metrics*, 12, 225:242.

Gurmu, S. and P. Trivedi 1996. Excess zeros in count models for recreational trips. *Journal of Business & Economic Statistics*, 14, 469:477.

Heilbron, D. 1994. Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, 36, 531:547.

Hinde, J. and Demetrio, C. 1998. Overdispersion: models and estimation. *Computational Statistics and Data Analysis*, 27, 151-170.

Hossain, K et al., 2015. Contribution of Socio-Demographic factors on antenatal care in Bangladesh: Modelling approach, *Public Health Research,* 5(4): 95-102

Hugo, N. et al., 2008. A comparison between Poisson and zero-inflated Poisson regression models with an application to number of black spots in Corriedale sheep, *INRA, EDP Sciences*, 40:379-394

Ibnouf, A.H et al., 2007. Utilization of antenatal care services by Sudanese women in their reproductive age, *Saudi Med J,* Vol. 28 (5): 737-743

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

Iyaniwura, C.A., Yussuf Q. 2009. Utilization of Antenatal care and Delivery services in Sagamu, South Western Nigeria.*African Journal of Reproductive Health* Vol 13 No 3

Kawungezi, P.C., et al. 2015. Attendance and Utilization of Antenatal Care (ANC) Services: Multi-Center Study in Upcountry Areas of Uganda, *Open J Prev Med.*

Lambert, D. 1992.Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing,*Technometrics*, 34, 1, 1992, pp. 1-14.

Magadi, M. A., Madise, N. J. and Rodrigues, R. N 2000. Frequency and timing of antenatal care in Kenya: explaining the variations between women of different communities.*SocSci Med*, 2000; 51: 551–561.*March 1999.* Accra: Ghana Statistical Service.

Martin, R. et al., 1998. Models for count data with many zeros, *International Biometric conference.*

Md. Mosiurrahman, md.Rafiqulislam and ahmedzohirulislam. 2008. Rural-urban differentials of utilization of ante-natal Health-care services in Bangladesh

Miaou, S.-P. 1994. The relationship between truck accidents and geometric design of road sections. Poisson versus negative binomial regressions.*Accident Analysis & Prevention*, 26, 471:482.

Mouatassim Y., Ezzahid. E.H. 2012.Poisson regression and Zero-inflated Poisson regression: application to private health insurance data. *Eur. Actuar. J.* DOI 10.1007/s13385-012-0056-2

Muniswamy, B., Molla D.T and Reddy N.K.2015.Comparison of Test Statistic for Zero-Inflated Negative Binomial against Zero-Inflated Poisson Model, *Indian Journal of Science and Technology*, Vol 8(4), 349–357.

Mwalili SM, Lesaffre E, Declerck D. 2008. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Stat Methods Med Res.*17:123–139. [PubMed: 17698937]

Nanjundan1, G.T., Naika R. 2012.Asymptotic Comparison of Method of Moments Estimators and Maximum Likelihood Estimators of Parameters in Zero-Inflated Poisson Model *Applied Mathematics*. 3, 610-616

National Population Commission, 2013. *Federal Republic of Nigeria: final report on Nigeria Demographic and Health Survey 2013,* Carlverton, MD, USA: ORC Macro.

Ngatchou-Wandji, J. & Paris, C., 2011. On the Zero-Inflated Count Models with Application to Modelling Annual Trends in Incidences of some occupational Allergic Diseases in France, *Journal of Data Science*, 9:639-659

Nisar, N. & White, F., 2003. Factors affecting utilization of Antenatal Care among reproductive age group Women (15-49 years) in an urban squatter settlement of Karachi, *JPMA (Journal Of Pakistan Medical Association)* Vol 53 ,No.2.

Onamade, A., 2009. Access and utilization of antenatal care services in Funtua local government area, Katsina State, Nigeria.*Post graduate school,* Thesis Submited to Ahmadu Bello University, Zaria.

Phang Y.N., Loh E.F. 2013. Zero Inflated Models for Overdispersed Count Data.*International Journal of Mathematical, Computational, Statistical, Natural and Physical Engineering*Vol:7, No:8.

Raghupathy, S. 1996. Education and the use of maternal health care in Thailand.*Social Science and Medicine* 43: 459-471

Rui Fang 2008, Zero-Inflated Negative Binomial (ZINB) Regression Model For Over-Dispersed Count Data With Excess Zeros And Repeated Measures, An Application To Human Microbiota Sequence Data, *North China Coal Mining Medical College*

Rui Fang, Brandie D. Wagner J. Kirk Harris, Sophie A. Fillon.2014. Application of zero-inflated negative binomial mixed 1 model to human 2 microbiota sequence data.*PeerJPrePrints*

Sharma A.K., Landge V.S. 2013. Zero Inflated Negative Binomial For Modelling Heavy Vehicle Crash Rate On Indian Rural Highway,*International Journal of Advances in Engineering & Technology,* Vol. 5, Issue 2, pp. 292-301

Shonkwiler, J. and Shaw W. 1996. Hurdle count-data models in recreation demand analysis. *Journal of Agricultural and Resource Economics*, 21, 210:219.

Shrestha, G & Shrestha, G., 2011.Statistical Analysis of Factors Affecting Utilization of Antenatal Care in Nepal, *Nepal Journal of Science and Technology 12:* 268-275

Simkhada B, Teijlingen ER, Porter M, Simkhada P. 2008. Factors Affecting the Utilization of Antenatal Care in Developing Countries: Systematic Review of the Literature. *Journal of Advanced Nursing.* 61:244–260. [PubMed: 18197860]

SPSS Inc. SPSS for windows, version 16.0. Chicago.

Stata 12.0, college station, Texas 77845 USA

Tsegay, Y. et al., 2013. Determinants of antenatal and delivery care utilization in Tigray region, Ethiopia: a cross-sectional study, *International Journal for Equity in Health 2013, 12:30*

Wang, W. and Famoye, F. 1997.Modeling household fertility decisions with generalized Poisson regression.*Journal of Population Economics* 10, 273-283.

Wolny-DominiakA. 2013 Zero-Inflated Claim Count Modeling And Testing – A Case Study, *Econometrics* 1(39).

Wulu, J.T., Singh, K.P. Famoye, F. and McGwin, G. 2002.Regression analysis of count data.*Journal of the Indian Society of Agricultural Statistics* 55, 220-231.

Yusuf O.B., Ugalahi L.O. 2015. On the performance of the Poisson, Negative Binomial and Generalized Poisson Regression Models in the prediction of Antenatal care Visits in Nigeria, *American Journal of Mathematics and Statistics,* Vol. 5 No 3, pp. 128-136. doi: 10.5923/j.ajms.20150503.04.