Science World Publishing

Open Access Journal | Volume: 1.1

CLINICAL BIOSTATISTICS AND BIOMETRICS

Comparison of Count Models and associated risk factors of Neonatal Mortality in Ethiopia

Million Wesenu Demissie

Department of Statistics, College of Computing and Informatics, Haramaya University, Dire Dawa, Ethiopia

Article Information

Article Type: Research
Journal Type: Open Access
Volume: 1 **Issue:** 1
Manuscript ID: CBB-1-102
Publisher: Science World Publishing

Received Date: 25 September 2020
Accepted Date: 30 September 2020
Published Date: 05 October 2020

***Corresponding author:**

Million Wesenu Demissie

Department of Statistics
College of Computing and Informatics
Haramaya University
Dire Dawa
Ethiopia
E-mail: millionwesu@gmail.com

Citation: Million Wesenu Demissie, (2020) Comparison of Count Models and associated risk factors of Neonatal Mortality in Ethiopia. Clin Biostat Biometr. 1(1);1-9

ABSTRACT

The Duration from 0- 28 day of life is a crucial time for survival of any child. Reducing neonatal mortality rate is a serious problem in many low and middle-income countries including Ethiopia. The main objective of this investigation was to find out the associated risk factors for number of neonatal mortality per women and select models that best fit of the neonate dataset. Recorded survey data of Ethiopian Demographic health survey in 2016 were obtained for a retrospective of 9958 women aged 15-49 years and considered in this investigation. Count models of Standard Poisson, zero-inflated Poisson , Negative binomial Poisson and zero-inflated negative binomial regression were used to distinguish the potential risk factors and model comparison have been done using Bayesian Information Criteria (BIC) and Akaike Information Criteria (AIC). Based on this, zero-inflated negative binomial regression model had minimum value of AIC and BIC when compared with other models and best fit for the neonatal mortality dataset. From results of zero-inflated negative binomial regression model parental Residence, Geographical-Region, education level of mother, Religion, mother Age at first time of birth, Source of drinking water, Antenatal care visits, Husband/partner's education level, postnatal care visit and Toilet facility were found to be statistically significant factors of neonatal mortality per mother. It is therefore recommended that women should be attending antenatal care and postnatal care visit during pregnancy and after birth, respectively in addition to improve level of education for the better survival neonates in the first 28th day of life.

KEYWORDS

Count model, Over-dispersion, Neonatal, Akaike information criteria

INTRODUCTION

The Neonatal period is the most vulnerable and high-risk time in life because of the highest mortality and morbidity incidence in human life during the first 28th day of life. It refers to the incidence of death of a live born infant within the first 28th days of life. Neonatal death is a significant public health concern in the world. World health organization (WHO, 2011) reported that, an estimated 40 percent of deaths in children less than five years of age occur during the first 28 days of life. In the Study conducted by Rajaratnam, et al., (2010) suggested that the average daily mortality rate during the neonatal period is close to 30 fold higher than during the postnatal period (one month to one year of age). In addition, an estimated 7.7 million children under five years of age died worldwide. This included 3.1 million neonatal deaths, 2.3 million post neonatal deaths (age one month to one year) and 2.3 million childhood deaths (age 1-5 years).

According to investigation conducted by Black et al. (2010), the most important causes of neonatal mortality were complications of preterm birth (12%), birth asphyxia (9%), tetanus (7%), sepsis (6%) and pneumonia (4%). The causes of neonatal death may also vary within the same country depending upon the socioeconomic status of the regional population and access to health care services. Based on the reports of World Health Organization (WHO, 2011) of the 140 million babies born worldwide, 90% were born in low-income countries and 10% in high-income countries while approximately 99% of neonatal deaths occur in low income countries and 1% in high income countries. Every minute seven newborn babies die worldwide (415 newborn babies every hour). Based on the State of the World's Children Report (UNICEF, 2005), the vast majority of neonatal deaths occur in South Asia and Sub Saharan Africa countries since the socioeconomic status of these countries adversely affects maternal child health as it limits access to adequate nutrition, quality health care, medications, safe water, adequate sanitation, and basic services.

A comparative Study conducted in Ethiopia to find out mortality trends by comparing data from DHS conducted in 2000, 2005, and 2011. The result of these surveys suggested that mortality rates of under-five and infant were continuously decreasing trend in death.

On the one hand, neonatal mortality rate decreased from 49 deaths per 1,000 live births in 2000 to 39 deaths per 1,000 live births in 2005. However, the rate remained stable at 37 deaths per 1,000 live births in 2011. Hence, 42% of the under-5 mortality approximately in Ethiopia is attributable to neonatal deaths (CSA: EDHS 2011). Study conducted by Oestergaard, et al (2011) suggested that Ethiopia experiencing a high neonatal mortality rate at 37 per 1000 live births as comparable to the average rate of 35.9 per 1000 live births for the African region overall using Ethiopian Demographic and Health Survey.

To reduce neonatal death there are two health systems programmes. Maternal health Programmes which intervened pregnancy, childbirth, early neonatal care where as child health programmers extend through infancy into childhood. Considering neonatal mortality requires continuity between these elements of care which is lacking in many settings and Care for the neonate often receives little attention in either maternal or child health programmes. The main problem in coverage of care falls during the crucial first week of life, when the majority of neonatal and maternal deaths occur, often at home and with no contact with the formal health care system. This study reveals that to identifying potential associated risk factors to neonatal mortality and comparison of count data models to select appropriate models for the best fit of neonatal mortality case using Bayesian information criteria (BIC) and Akaike information criteria (AIC).

METHODS AND MATERIALS

Ethiopia Demographic and Health Survey of 2016 data were used which was conducted by Central Statistical Agency (CSA). The main goal of survey is to provide current and reliable data on fertility and family planning behavior, child mortality, adult and maternal mortality, children's nutritional status, use of maternal and child health services, knowledge and prevalence of HIV/AIDS and anemia. This study analyzes responses from each of 9958 women of age 15-49 and only those who have ever born a child, on the counts of the number of deaths of children aged less than one month that the mother experienced in her lifetime. In EDHS, information on child mortality was found from the birth history of women who were included in the survey.

A supposed to the classical linear regression model, the regression models for counts are nonlinear with many properties for the response variable that relate to discreteness, nonlinearity and deal with non-negative values only. The Poisson regression model is a good starting point of count data modeling as it lends itself well with the nature (properties) of count data. However, despite the advantages highlighted, the standard Poisson model still suffers one potential problem. This relates to the assumptions of equality of variance and mean, a property called "equi-dispersion". When this assumption is violated, for instance, the variance of the observed counts exceeds the mean, and an "over-dispersion" will occur. Failure to control for over-dispersion will lead to inconsistent estimates, biased standard errors and inflated test statistics. Hence, in modeling count data, it is a usual practice after the development of Poisson regression model to proceed with analysis of correcting for over-dispersion if it exists. One of the approaches to modeling over-dispersion is to use Quasi Likelihood estimation technique proposed by Wedderburn (1974). Alternatively, one can use the Negative Binomial (NB) regression model which is the generalization and extension of Poisson-gamma regression model (Cameron and Trivedi, 1998; Cameron and Trivedi, 2005; Agresti, 2002). In the NB model, a dispersion parameter is included in the model to cater for over-dispersion by allowing the variance to be greater than the mean and accommodate the unobserved heterogeneity in the count data.

In addition to over-dispersion problems, excessiveness of zero in count data is a very common problem to real life.Basically, the problem of over-dispersion and excessiveness of zeroes overtake by applying regression model of negative binomial model and zero-inflated Poisson model respectively. The characteristic of count data are nonnegative, skewed, over-dispersed and excessiveness of zeroes in really situation. On the top of this, the characteristics enabled the application of various statistical methods and count data regression models. Thus, the study gives realistic approach of modeling count data focusing on data that exhibits over-dispersion and excess zeroes.

METHOD OF DATA ANALYSIS

Regression Model (PR)

The Poisson regression model is often considered as a benchmark model for modeling count data. This model dominates the count data modeling activities as it suits the statistical properties of count data and is flexible to be re-parameterized into other form of distributional functions (Cameron and Trivedi, 1998). Poisson regression assumes a Poisson distribution, characterized by a substantial positive Skewness with variance equals mean. It tends to fit such data better than the linear regression model. However, if the variance is larger than the mean, it induces deflated standard errors and inflated standardized normal (i.e. Z-normal) values, resulting in increased Type I errors that make Poisson regression less adequate.

Suppose Y1, Y2, …,Yk are independent random variables. Let yi denote the value of an event count outcome variable (neonatal death in this study) for ith mother within a given time or exposure periods with mean parameter λi and Xi denote a vector of explanatory variables for the ith mother. Poisson regression is the simplest regression model for count data and assumes that each observed count yi is drawn from a Poisson distribution with the conditional mean λi(Y~Poisson (λi), i=1,2,…,k) on a given vector Xi for case i. Then the Poisson equation of the model with rate parameter λi (Nelder and Wedderburn, 1972) is given by:

$$P(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \lambda > 0 ... and .. y_i = 0,1,2,3 \qquad (1)$$

With mean and variance E(Yi)=Var(Yi)==λi

Equality of the mean with the variance is the equidispersion property of the Poisson model. The mean of the response variable λi is related with the linear predictor through the so called link function. As λi has to be positive, an appropriate choice is the logarithmic function, so that we do not need further restrictions on the parameters. It is well known from the literature (McCullagh and Nelder, 1989) that this is the natural link function for the Poisson distribution.

Let X be nx(k+1) matrix of explanatory variables. The relationship between Y and ith row vector of X, Xi linked by g(λi), is the canonical link function given by;

$$E(Y_i) = \lambda_i = e^{X_i^{'}\beta} \qquad (2)$$

where, Xi = (Xi0,Xi1,…,Xik)′ is the ith row of covariate matrix, (with Xi0=1) and β=(β0,β1, β2,…,βk)′ is unknown (k+1) dimensional vector of regression parameters.

The model comprising equations (1) and (2) is known as the Poisson regression or log-linear model. The log of the mean λi is assumed to be a linear function of the independent variables, that is,

$$\ln(\lambda_i) = \beta_0 + \sum_{i=1}^{k} \beta_i x_{ij} \qquad (3)$$

Hence xij represent the ith value related to the jth factor, k is the number of factors in model and βj is the model parameter.

In a Poisson distribution, the model coefficients are estimated by the maximum likelihood method. The likelihood function L is the product of the terms in equation (1) over all n measured values yi. This function is viewed as a function of the parameters λiand the parameters βi. The parameters are estimated by maximizing the likelihood, or more usually, by maximizing the logarithm of the likelihood. To find the ML of equation (1), we define the likelihood function as follows:

$$L(\beta)=\prod_{i=1}^{n}\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}=\prod_{i=1}^{n}\frac{e^{-e^{X_i'\beta}}(e^{X_i'\beta})^{y_i}}{y_i!} \quad (4)$$

Taking log on both sides, the log-likelihood function is given by:

$$\ln L(\beta)=\sum_{i=1}^{n}\left[y_i\ln\lambda_i-\lambda_i-\ln y_i!\right]=\sum_{i=1}^{n}\left[y_i x_i'\beta-e^{x_i'\beta}-\ln y_i!\right] \quad (5)$$

From equation (4) and (5), the first two partial derivatives of the log-likelihood function exist and β is the solution of the estimating equations obtained by differentiating the likelihood function with respect to β and solving them to zero.

Negative Binomial Regression Model (NBR)

The standard Poisson regression accounts for observed differences among the observations; however negative binomial regression includes a random component that involves unobserved variance among observations. In an over-dispersed data, this random component results in more accurate standard errors and z-statistics for the regression coefficients than using the standard Poisson regression.

Let Y1, Y2 , . . ., Yn be a set of n independent random variables where Yi follows a negative binomial distribution with mean λi and dispersion parameter α (denoted by Yi~NB(λi, α)). Then the negative binomial model takes the form (Cameron and Trivedi, 1986):

$$P(Y_i=y_i)=\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma(y_i+1)\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{\alpha\lambda_i}{1+\alpha\lambda_i}\right)^{y_i}\left(\frac{1}{1+\alpha\lambda_i}\right)^{1/\alpha},\ y_i=0,1,2,.....and..\alpha\geq 0 \quad (6)$$

With mean E(YI)= λi= EXP(x'iβ) and variance Var(Yi)=λi(1+αλi) (McCullagh and Nelder, 1989), where α is the over-dispersion parameter. To see the relationship the introduction of covariates into a regression model using Negative Binomial distribution (Lawless, 1987 and Hinde, 1996).

Based on the Cameron and Trivedi (1986) the negative binomial regression model parameter estimation can be obtained through minimizing the negative of the log-likelihood function and its mathematical equation is given by:

$$\ln(L)=\sum_{i=1}^{n}\left\{y_i\ln\left(\frac{\alpha\lambda_i}{1+\alpha\lambda_i}\right)-\frac{1}{\alpha}\ln(1+\alpha\lambda_i)+\ln\Gamma\left(y_i+\frac{1}{\alpha}\right)-\ln\Gamma(y_i+1)-\ln\Gamma\left(\frac{1}{\alpha}\right)\right\} \quad (7)$$

The maximum likelihood estimates, ( ) , may be obtained by maximizing ln(L) with respect to β and α since $\lambda_i=e^{x_i\beta}$.

Zero-inflated Regression Models

Real life count data frequently exhibit over-dispersion and excess zeros. Although the negative binomial model can solve an over-dispersion problem, it may not be well flexible to handle excess zeros. This motivates the development of Zero-Inflated count model to model excess zeros in addition to over-dispersion. According to Sarkisian and Gerstel (2004) the there is a common problem in the Standard models associated to under-predicting zeros and over-predicting zeros and such complication occurred when the distribution have excessive zeroes. Therefore, in such situations the standard Poison and negative binomial models are not suitable. Due to this, Zero inflated negative binomial and Zero inflated Poisson regression models used to handle excess zeros in analysis.

Zero-inflated Poisson Regression Model (ZIP)

There are two pillars in the zero inflated Poisson regression models which are zero counts and expectation of Poisson distribution that originated from inflated part and parameter λ respectively.

Generally, the zero-inflated probability mass function has the form:

$$P(Y_i=y_i)=\begin{cases}\phi+(1-\phi)p(y=0).....if..y_i=0\\(1-\phi)p(Y=y_i).........if...y_i=1,2,..\end{cases} \quad (8)$$

When the random variables Yi are not dependent possessing zero-inflated Poisson distribution, the zeros are assumed to arise in two ways corresponding to distinct underlying states. Namely; the probability ϕias the first occurrence of zeros that give rises only zeros, whereas the probability (1 −ϕi) as second states that leads to a standard Poisson count with mean λi. The occurrence of zero comes from the first and second state are known as structural zeros and sampling zeros respectively. This two-state process gives a simple two-component mixture distribution with probability mass function:

$$P(Y_i=y_i)=\begin{cases}\phi+(1-\phi)e^{-\lambda_i}.....if..y_i=0\\(1-\phi)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}.........if...y_i=1,2,..\end{cases} \quad (9)$$

This is denoted by Yi ~ ZIP(λi, ϕi) such that 0 ≤ ϕi< 1, where λi is the mean of the non-zero outcomes that can be modeled with the associated explanatory covariates using a natural logarithmic link function as: ln(λi) = $x_i\beta$. Where Xi=(1,Xi1,Xi2, Xi3, ...,Xik)' is a (k+1) x1 vector of explanatory variable of the ith subject and β is (k+1)x1 vector of regression coefficient parameters. Φi(0 <Φi<1) is the probability of an excess zero (being in the zero mortality state) determined by a logit model (Lambert, 1992 and Long, 1997). To predict membership in the "Always Zero" group, we can use the same variables or we can use a smaller subset of the variables or even different variables altogether.

Zero-inflated negative binomial regression model (ZINB)

When there are excess zeros and high variability in the non-zero outcomes, Zero-Inflated Poisson models are less adequate than Zero-Inflated Negative Binomial models. ZINB model is different from zero inflated Poisson models in which the Poisson distribution function changed to negative binomial distribution for count data. In generally, the zero inflated negative binomial distribution have the following probability density function as stated below:

$$p(Y_i=y_i)=\begin{cases}\phi_i+(1-\phi_i)(1+\alpha\lambda_i)^{-\frac{1}{\alpha}}......if..y_i=0\\(1-\phi_i)\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)(\alpha\lambda_i)^{y_i}}{\Gamma(y_i+1)\Gamma\left(\frac{1}{\alpha}\right)(1+\alpha\lambda_i)^{y_i+\frac{1}{\alpha}}}......if..y_i>0\end{cases} \quad (10)$$

Where λiis the mean of the non-zero response that can be modeled with the associated explanatory covariates using a natural logarithm link function as defined in $\lambda_i=e^{x_i\beta}$, and φi is the probability of excess zeros which can be estimated by the logistic regression (Zuuret.a,. 2009).

The Zero-Inflated Negative Binomial model is a special case of a two-class finite mixture model like the ZIP model with mean E(Yi)= λi(1−ϕi) and variance Var(Yi) = λi(1−ϕi)(1+ αλi+ϕiλi) , where the parameters λi and ϕi depend on the covariates and α ≥ 0 is a scalar. Thus we have over-dispersion whenever either ϕior α is greater than 0.

ASSESSING MODEL FIT

Testing Hypotheses for the Significance Model Parameters

Once we have fit a model and obtained estimates for the various parameters of interest, we want to answer questions about the contributions of various factors to the prediction of the response variable. To test whether the entire set of explanatory variables contribute significantly to the prediction of the response variable, we can use the deviance test. The test statistic is given as follows:

$$D=2\left\{\sum y_i\log\left(\frac{y_i}{\hat{\lambda}_i}\right)-y_i+\hat{\lambda}_i\right\} \quad (11)$$

Where $\hat{\lambda}_i$ is the forecasted value from the fitted model. Under the null conjecture

H0: $\beta_1 = \beta_2 = \beta_3 = ... = \beta_k = 0$, D is approximately distributed as chi-square with k degrees of freedom.

Test for Over-dispersion Parameter

The implication of over-dispersion has similar consequences as the failure of the assumption of Homoscedasticity in the linear regression model. The negative binomial regression model reduces to the Poisson regression model when the over-dispersion parameter is not significantly different from zero. To assess the adequacy of the negative binomial model over the Poisson regression model, we can test the hypothesis: H0:α =0 versus Ha: α > 0. The presence of the over-dispersion parameter in the Negative Binomial regression model is justified when the null hypothesis H0: α = 0 is rejected. In order to test the above hypothesis a score test statistic is proposed (Dean and Lawless, 1989). To test the null hypothesis the score test statistic can be given by:

$$s_\alpha = \frac{\left[\sum_{i=1}^{n}\left\{(y_i - \hat{\lambda}_i) - y_i\right\}\right]^2}{2\sum_{i=1}^{n}\hat{\lambda}_i^2} \quad (12)$$

Where, $\hat{\lambda}_i$ represent the forecasted value of Poisson regression model and score statistic have chi-square distribution with one degree of freedom in limiting distribution of the count model under the null conjecture. Alternatively to assess the adequacy of the negative binomial model over the PR model, we test the hypothesis that the mean and the variance are equal (equi-dispersion). Deviance and Pearson Chi-Square divided by the degrees of freedom are used to detect over dispersion or underm-dispersion in the Poisson regression. The deviance of a model m is:

$$D_m = 2(L_f - L_m) \quad (13)$$

Where model "m" is the model under consideration, Lf is the log-likelihood that would be achieved if the model gave a perfect fit (λi = yi for each i, and α= 0) and Lm is the log-likelihood of the model under consideration. If the latter model is correct, Dm is approximately a chi-squared random variable with degrees of freedom equal to the number of observations minus the number of parameters (n-k). Thus, when deviance divided by degrees of freedom ($\frac{D_m}{n-k}$) is larger than 1, over dispersion is indicated (Bauer et al,.2007).

The likelihood ratio test is inappropriate to compare negative binomial or Poisson with the zero inflated negative binomials and Poisson since they are not nested under the other models. In such situation, The best test statistic that used to compare standard non-nested regression models to zero inflated regression models inf count data is Vuong Statistic test (Vuong, 1989). Suppose f1(yi/Xi)and f2(yi /Xi) denote the probability density functions of zero-inflated Poisson, zero-inflated negative binomial and standard Poisson or negative binomial model, respectively. Then mi is given as:

$$m_i = \log\left[\frac{\hat{f}_1(y_i / x_i)}{\hat{f}_2(y_i / x_i)}\right] \quad (14)$$

Where $\hat{f}_1(y_i / x_i)$ and $f_2(y_i / x_i)$ are predicted probabilities of the corresponding models, respectively.

let $\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$ and $S_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})^2}$

announce the average and square root of variance of the mensuration's m. here the mathematical equation for Vuong Statistic is given by:

$$V = \frac{m}{S_m / \sqrt{n}} \quad (15)$$

Under the null hypothesis and big data the two distribution functions are equal.

Goodness of fit-tests

To assess the model appropriateness in-lined with response variable and measure the best fit dataset goodness of fit test were applied. In this study, a likelihood ratio test was used to compare the Poisson with the negative binomial and zero inflated Poisson with zero inflated negative binomial. The likelihood ratio statistic is given by:

$$L = 2(l_1 - l_0) \quad (16)$$

Where, l_1 is the model's log likelihood under the alternative and l0 the model'slog likelihood under null hypothesis.

STATISTICAL RESULTS AND DISCUSSION

Summary of Descriptive Statistics: The results of Summary Statistics in (**Table 1**), indicates that the covariates that associated to neonatal death. Those covariates are number of antenatal visits, husband/partner's education level, religion, Geographical region, mothers' education level, source of drinking water supply, parental residence, place of delivery, postnatal care, Mother's age at first gave birth and presence of Toilet facility.

The Study was considered the women whose age is in category of 15-49 about 9958 out of that 2868(28.80%) women experienced neonatal death. Out of these about 2.80% and 26.0% of neonatal deaths have been happened in resident of Urban and Rural areas, respectively. Concerning Geographical region, Addis Abeba city has the smallest percentage of births in the first 28th day death per women(0.30%) whereas Oromia and Somalia experienced by much had percentage of neonates mortality (4.30% and 4.70%), in the order given. The percentage of neonatal mortality increased with an increment in the number of antenatal care visits. In distinction from others, the death rate of infants in the first 28th day was 8.80% for mothers that not attended antenatal-care-visits and women that attended 1-3 visits had 11.10%, then decreased with an increment in the number of antenatal care visits.

Another important factors associated to neonatal death was source of water and type of toilet facility which is sanitation indicators. The availability of pure water reduces the risk of infection for diarrhea which emerges from an intake of impure water and nutrient that have direct impact on reducing neonatal mortality and indirectly implies that mothers or primary care providers are capable to give greater period to childcare rather than fetching water. An output showed that 1297 (13.0%) respondents had piped water whereas about 4526 (45.5%) households able to drank water from public tap. The percentage of neonatal death per women regarding with source of water supply exposed for public tap was 14.1% of neonatal mortality, while 1.70% of neonatal mortality occurred as a reason of mothers who supplied water from piped water sources. Similarly, the summary statistics of other Socio-demographic variables that expected to affect the number of neonatal death were found in below.

The bar graph of the above indicates that more of the women did not experienced neonatal death which is about 71.2% and the percentage of neonatal death decreased as the number of neonatal to mother death increased.

DATA ANALYSIS OF COUNT MODEL AND MODEL SELECTION

In the data analysis, the initial step is to find out group of independent covariates having latent effect have been enclosed in multivariate regression model of count data of linear component. The selected Poisson model is then tested for over-dispersion. Accordingly, the results of Poisson regression model, chi-square test of deviance indicated that a chi-square value of 1079.68 with p-value p<0.0001, which would suggest that the best fit for the model of Poisson regression model. However, the over-dispersion shall be as a result of dispersion between observations or overabundance zeros. The Pearson chi-square goodness of fit and score test statistic were used to check the presence of over-dispersion. The output of Pearson chi-square value of 10606.35 and test score 13396.34 having significance level< 0.0001 showed that presence of over-variability in

Table 1: Summary Statistics of Explanatory variables and number of neonatal death per women.

Variables	Category /class	Frequency of neonatal death/ women					
		0(%)	1(%)	2(%)	3(%)	4(%)	Total death(%)
Mother Education level	No education	4144(41.60)	1415(14.20)	521(5.20)	228(2.30)	87(0.90)	2251(22.60)
	Primary	2009(20.20)	385(3.90)	77(0.80)	24(0.20)	9(0.10)	495(5.00)
	Secondary	590(5.90)	83(0.80)	19(0.20)	3(0.001)	0(0.00)	105(1.10)
	Higher	347(3.50)	16(0,20)	1(0.0001)	0	0	17(0.20)
Region	Tigray	742(7.50)	159(1.60)	38(0.40)	12(0.12)	9(0.10)	219(2.20)
	Afar	667(6.70)	190(1.90)	100(1.00)	55(0.60)	16(0.20)	361(3.60)
	Amhara	592(5.90)	179(1.80)	74(0.70)	25(0.30)	7(0.10)	285(2.90)
	Oromia	1042(10.50)	309(3.10)	88(0.90)	29(0.30)	3(0.001)	429(4.30)
	Somalia	995(10.00)	339(3.40)	93(0.90)	34(0.30)	6(0.10)	472(4.70)
	BenishangulGumuz	574(5.80)	132(1.30)	56(0.60)	33(0.33)	29(0.30)	250(2.50)
	SNNPR	800(8.00)	250(2.50)	78(0.80)	28(0.30)	18(0.20)	374(3.80)
	Gambela	500(5.00)	124(1.20)	34(0.33)	18(0.20)	0	176(1.80)
	Harari	421(4.20)	102(1.00)	26(0.30)	7(0.10)	2(0.001)	137(1.40)
	Addis Ababa	393(3.90)	25(0.30)	4(0.001)	0	0	29(0.30)
	Dire Dawa	364(3.70)	90(0.90)	27(0.30)	13(0.11)	6(0.10)	136(1.40)
Place of Residence	Urban	1557(15.60)	226(2.30)	37(0.40)	10(0.10)	2(0.001)	275(2.80)
	Rural	5533(55.60)	1673(16.80)	581(5.80)	245(2.50)	94(0.90)	2593(26.00)
Religion	Orthodox	2159(21.70)	456(4.60)	154(1.50)	48(0.50)	21(0.20)	679(6.80)
	Catholic	36(0.40)	21(0.20)	3(0.001)	4(0.001)	0	28(0.30)
	Protestant	1268(12.70)	314(3.20)	89(0.90)	35(0.40)	20(0.20)	458(4.60)
	Muslim	3507(35.20)	1076(10.80)	351(3.50)	162(1.60)	55(0.60)	1644(16.50)
	Traditional	71(070)	14(0.10)	15(0.20)	1(0.001)	0	30(0.30)
	Others	49(0.50)	18(0.20)	6(0.10)	5(0.10)	9	29(0.30)
Drinking water sources	Piped water	1125(12.30)	131(1.30)	22(0.30)	9(0.10)	0	172(1.70)
	Public tap	3125(31.40)	916(9.20)	309(3.10)	121(1.20)	55(0.60)	1401(14.10)
	Spring water	1536(15.40)	479(4.80)	159(1.60)	57(0.60)	17(0.20)	712(7.20)
	Other sources	1204(12.10)	373(3.70)	118(1.20)	68(0.70)	24(0.21)	583(5.90)
Place of delivery	Home	4427(44.50)	1429(14.40)	507(5.10)	203(2.00)	73(0.70)	2212(22.20)
	Health institution	2663(26.70)	470(4.70)	111(1.10)	52(0.50)	23(0.20)	656(6.60)
Husband/partner's education level	No education	3060(30.70)	1026(10.30)	386(3.90)	169(1.70)	64(0.60)	1645(16.50)
	Primary	2370(23.80)	617(6.20)	161(1.60)	75(0.80)	29(0.30)	882(8.90)
	Secondary	988(9.90)	135(1.40)	54(0.50)	7(0.10)	3(0.001)	199(2.00)
	Higher	672(6.70)	121(1.20)	17(0.20)	4(0.001)	0	142(1.40)
PNC	No	4715(47.30)	1132(11.40)	373(3.70)	154(1.50)	53(0.50)	1156(11.60)
	Yes	2375(23.90)	767(7.70)	245(2.50)	101(1.00)	43(0.40)	1712(17.20)
Toilet facility	No	3038(30.50)	803(8.10)	226(2.30)	104(1.00)	36(0.40)	1169(11.70)
	Yes	4052(40.70)	1096(11.00)	392(3.90)	151(1.50)	60(0.60)	1699(17.10)
ANC Visit	No antenatal visits	1771(17.80)	553(5.60)	208(2.10)	82(0.80)	29(0.30)	872(8.80)
	1-3	3194(32.10)	757(7.60)	226(2.30)	86(0.90)	38(0.40)	1107(11.10)
	4-6	1603(16,10)	424(4.30)	127(1.30)	52(0.50)	18(0.20)	621(6.20)
	Greater than 6	522(5.20)	165(1.70)	57(0.60)	35(40)	11(0.10)	268(2.70)

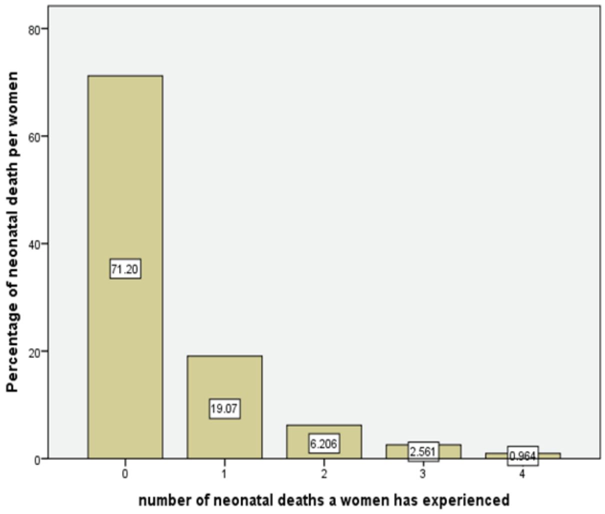

Figure 1: A Diagram for the percentage of neonatal deaths per mother

addition to Poisson regression is not appropriate.

Due to inappropriateness of standard Poisson regression model, we need to fit negative binomial, zero-inflated Poisson or zero-inflated negative binomial regression model to select better model that fit the observed dataset and cop-up the problem of over-dispersion. Model comparison and selection of suitable models that fit the neonatal death dataset have been done using Likelihood ratio test, Bayesian information criterion and Akaike information criterion. As a result the Zero-inflated negative binomial (ZINB) regression model has the smallest AIC(16668.14) and BIC(16917.54) as compared to standard Poisson, negative binomial and zero inflated Poisson regression model which suggested that Zero-inflated negative binomial (ZINB) regression model better fit the dataset to predict the number of neonatal death per women. This implies zero inflated negative binomial regression models give a good prevision of the number of neonatal mortality. In addition, to find out the suitable model in this study Vuong's test was used. In testing the null hypothesis of Vuong test which stated that standard Poisson model as compared to zero inflated Poisson models have equally similar observed distribution. The Value of the test statistic (V=10.31, having significant value <0.0001) indicated that the Zero-inflated Poisson regression model had more statistically significant in testing observed data than Poisson model which implies that excessiveness of zeroes present. Also, comparison of negative binomial with zero inflated negative binomial have been done using Vuong test statistic(v=2.61) that resulted Zero-inflated negative binomial regression model better fit the neonate dataset. than negative binomial regression model at standard significance level.

Goodness of Fit and Test of Over-dispersion

The validity of Poisson regression analysis relies heavily on the assumption of equi-dispersion. Hence, the reason for over-dispersion presence would be either variation among observations or excess zeroes (heterogeneity). The null conjecture (H0): $\alpha=0$ against alternative Ha: $\alpha>0$ implied there is over-dispersion due to data excess zeros H0: $\alpha=0$ rejected and significantly different from zero. The Likelihood-ratio test chi-square test has value 6.39 with p-value 0.0057 suggested that the over-dispersion occurred as reason in presence of excess zeros. Also, the vuong Statistic having the hypothesis H0: $\Phi=0$ versus Ha: $\Phi>0$ were applied(v=2.61 having p-value 0.0045) showed that presence of over-dispersion due to eminent variance in the cardinal responses. The result of Likelihood-ratio chi-square test and Vuong test statistic are failed to reject alternative hypothesis which implies that controlling high variation as a reason of non-zero responses and excess zeros simultaneously regression model of zero inflated negative binomial is best fit to the data. The assessment of model goodness of fit was assessed using Pearson chi-square statistic (=783.65, p-value<0.0001) for the fitted showed that ZINB model have high accuracy than other examinee models.

Potential factors of Neonatal Mortality and Interpretations of Final model

Outcomes of final model parameter estimation with potential covariates associated to neonatal death are found below (Table 3). The results of parameter estimation have regression coefficients, standard error, Z-value, p-value, odds ratio and 95% confidence interval for odds ratio. Based on this result, the potential factors associated to number of neonatal death per mother were mothers' Residence, Education level, Region, Religion, mother Age at first give birth, Source of drinking water, husband/partner education level, Antenatal care visit, Toilet facility and postnatal care visit at 5% significance level. This implies those variables are statistically significant predictors of neonatal mortality per women. Whereas, the place of delivery was not significantly associated to number of neonatal mortality per mother based on the EDHS 2016 dataset. The result of this study showed that children born to mother Resident in rural area significantly associated to neonatal mortality as compared to mother resident in urban area. Based on results in (Table 3)the odds ratio associated to neonates who lived in rural resident attributed to high death as compared to neonate who lived in urban resident. This implies that the hazard of neonate mortality was approximately 1.2992 times higher for a child whose mother lived in rural areas compared to their urban counterparts (OR: 1.2992; 95% CI: 1.0988--1.5361). Basically, it is expected, inhabiting in urban resident should be consociated with a higher standard of living, improved sanitation, and improved health facilities compared rural residence.

Regarding the Regional differences in number of neonatal mortality, that parameter estimation result showed that Children in Amhara (OR: 1.2409; 95% CI: 1.0441-1.4746), Benishangul-Gumuz(OR: 1.4424; 95% CI: 1.1884 -1.7509) and SNNPR (OR:

Table 2: Model Comparison Using Akaike information criterion (AIC) and Bayesian information criterion (BIC).

Model	Ll(null)	Ll(model)	Df	AIC	BIC	LR χ^2/vuong
Standard Poisson regression	-9087.864	-8548.024	33	17162.05	17399.85	χ^2 =1079.68**
Negative Binomial regression	-8700.309	-8313.414	34	16694.83	16939.84	χ^2 =773.79**
Zero-inflated Poisson regression	-8703.369	-8302.264	34	16672.53	16920.35	χ^2 =802.21** Vuong=10.31**
Zero-inflated negative binomial regression	-8690.894	-8299.07	35	16668.14	16917.54	χ^2 =783.65** Vuong=2.61**

**indicates the model is significant at 5% level of significance and alpha=0 is rejected for vuong test.

Table 3: Parameter Estimates of the Zero Inflated Negative Binomial Regression Final Model.

Covariates	Category	β	SE	Z	P-Value	Exp(β)	95% CI for Exp(β)	
							Lower limit	Upper limit
Geographical Region (Tigray)	Afar	0.1815519	0.1059988	1.71	0.087	1.199077	0.9741383	1.475956
	Amhara	0.2158004	0.088062	2.45	0.014	1.240855	1.044149	1.474618
	Oromia	-0.0590291	0.0974191	-0.61	0.545	0.9426793	0.7788265	1.141004
	Somalia	-0.0682116	0.1045434	-0.65	0.514	0.9340628	0.7610069	1.146472
	Benishangul	0.3663416	0.0988604	3.71	0.000	1.442448	1.188365	1.750856
	SNNPR	0.2671159	0.0977446	2.73	0.006	1.306192	1.078466	1.582003
	Gambela	0.1940894	0.1205096	1.61	0.107	1.214205	0.9587691	1.537694
	Harari	-0.0269976	0.1232406	-0.22	0.827	0.9733635	0.7644912	1.239303
	Addis Ababa	-0.3595227	0.2066808	-1.74	0.082	0.6980094	0.4655156	1.046618
	Dire Dawa	0.2022571	0.1216554	1.66	0.096	1.224163	0.9644638	1.55379
Residence (urban)	Rural	0.2617535	0.0854583	3.06	0.002	1.299206	1.098844	1.536103
Mother Education (no Education)	Primary	-0.5697663	0.052012	-10.95	0.000	0.5656576	0.5108354	0.6263633
	Secondary	-0.5620041	0.1094176	-5.14	0.000	0.5700655	0.4600323	0.706417
	Higher	-1.577232	0.2558163	-6.17	0.000	0.206546	0.1251024	0.3410106
Religion (orthodox)	Catholic	0.4652072	0.1966749	2.37	0.018	1.592344	1.082996	2.341245
	Protestant	0.0719209	0.0783503	0.92	0.359	1.07457	0.921601	1.25293
	Muslim	0.1567307	0 .0675948	2.32	0.020	1.169681	1.024544	1.335377
	Traditional	-0.1063152	0.1822436	-0.58	0.560	0.8991412	0.6290741	1.28515
	Others	0.1544941	0.1886057	0.82	0.413	1.167067	0.8064073	1.68903
Age of mother at First Birth (<=20)	Above 20	-0.238247	0.0437258	-5.45	0.000	0.788008	0 .7232878	0 .8585195
Source of water (piped)	Public tap	0.2457449	0.1021825	2.40	0.016	1.278573	1.04652	1.562081
	Spring water	0.1740956	0.1068433	1.63	0.103	1.190169	0.9653028	1.467418
	Other sources	0.2950443	0.1071326	2.75	0.006	1.343186	1.088791	1.657019
ANC Visits (No-ANC check-ups)	1-3 visits	-0.0817457	0.0466719	-1.75	0.080	0.9215063	0.8409517	1.009777
	4-6 visits	-0.1232571	0.0569909	-2.16	0.031	0.8840363	0.7906046	0.9885096
	Greater than 6	0.053123	0.0752255	0.71	0.480	1.054559	0.9099949	1.22209
Husband/partner Education(No Education)	Primary	-0.0972444	0.0433983	-2.24	0.025	0.9073342	0.8333483	0.9878887
	Secondary	-0.3194882	0.0762696	-4.19	0.000	0.7265208	0.6256439	0.8436627
	Higher	-0.2059684	0.0988662	-2.08	0.037	0.8138588	0.6704926	0.9878798
PNC visit (No)	Yes	-0.1734181	0.0432987	-4.01	0.000	0.840786	0.7723774	0.9152536
Toilet facility(No)	Yes	0.1206032	0.036278	3.32	0.001	1.128177	1.050745	1.211315
Intercept		-0.6803529	0.1392113	-4.89	0.000	0.5064382	0.3855046	0.665309
Alpha		0.1492793	0.0730976			0.1492793	0.057173	0.3897694

* Reference categories are in parentheses, Likelihood-ratio test of α=0: χ^2 = 6.39 p-value = 0.0057 and Vuong test of ZINB vs. standard negative binomial: z = 2.61 p-value= 0.0045.

1.3062; 95% CI: 1.0785 -1.5820) region are more exposed to children mortality/death than children in Tigray region. However, statistically insignificant association observed concerning hazardous of neonate mortality among Afar, Oromia, Somalia, Gambela, Harari region and the two city administration (Addis Ababa and Dire Dawa) as compared to Tigray region. Another important factor was education level of mother which significantly associated to neonatal death. The result showed that risk of neonatal mortality whose mothers' have primary and secondary education were consociated with about 43% decrement in hazardous of neonatal mortality in comparable to infants comes from mothers with no education. At the same time, women's having higher education decreased the risk of neonatal death by 79.4% held other condition stable. Additionally, religion of mother is another potential factors associated to neonatal death. As Compared to neonatal of Orthodox mothers, those neonate whose mother's religion is Catholic (OR: 1.592; 95% CI: 1.0830 –2.3412) and Muslim (OR: 1.1697; 95% CI: 1.0245 – 1.3354) have a significantly higher risk of neonatal mortality.

The output in Table 3 showed that children born from mothers whose age at first give birth is greater than 20 have a significantly lower risk of mortality as compared to those born from mothers whose age at first give birth is less than or equal to 20. The risk of infants in the first 28^{th} day of life mortality was approximately 21.2% get down for women whose age was greater than twenty in comparison to children from women less than or equal to twenty years old (OR: 0.7880; CI: 0 .7233-0.8585). Another potential risk factor associated to neonatal death is source of drinking water which is significantly affect neonatal mortality in Ethiopia. The study showed that mothers who use water from public tap and other origins are most likely capable to being wounded/faced neonate mortality than those who use piped drinking water. The jeopardy of births in the first 28^{th} day of life death in-lined to infants having mothers who utilize public tap and other source of drinking water were 27.86% and 34.32% higher as compared to those who use piped water source(OR: 1.2786; 95% CI: 1.0465 – 1.5621) and (OR: 1.3432; 95% CI:1.0888-1.6570) Respectively. The study also reveals the effect of postnatal care on neonatal death, that is, attending health care services after birth has decreased the risk of dying of neonates by 16%.

Finally, there is statistically significant relationship between mother's antenatal care service and death of babies in the first 28^{th} day of life. Babies comes from women take care 4-6 antenatal care check-up had a 11.6% decrement hazard of neonatal death than babies born from women not attending antenatal care check-ups (OR:0.8840; 95% CI: 0.7906-0.9885).The national coverage of antenatal care for public health interventions should be directed at improving the awareness of mothers and family members about the importance of antenatal care checks and to further increase the utilization of these services. Quality, accessibility, and availability of the services should be enhanced to ensure optimal results for neonatal health. The other socio-demographic factors that have an association with neonatal mortality is husband educational level which significantly effect on mortality in this analysis. The husband/parents level of education have significant relationship with births of babies in the first 28th day of life which indicated that having primary, secondary and higher level of education minimized risk of death for neonates relative to husbands has no education. Hence, children of mothers whose husband has secondary education decreased neonatal mortality risk about 27.35% than neonates of mothers whose husband has no education.

CONCLUSIONS

The first twenty eighth day of life is a crucial time for survival of any child, more than four million newborns die yearly within the first four weeks of life with three million of these deaths occurring in the early neonatal death. To key out potential risk factors that associated to neonatal mortality, different count models were applied and compared using Akaike Bayesian Information Criteria (BIC) and Akaike information criteria (AIC). Based on this, zero-inflated negative binomial is better fitted to number of neonatal mortality per women dataset than other count models. These was an over-dispersion effect on the neonatal mortality that arises due to differences of high heterogeneity in the non-zero outcome and qualified by overabundance zeros. This indicates the existence of high variability and necessitates the count models.

The result of zero–inflated negative binomial models showed that parental Residence, Geographical-Region, education level of mother, Religion, mother Age at first give birth, Source drinking of water, Antenatal care visits, Husband/partner education level, postnatal care visit and Toilet facility are the potential risk factors that associated to number of neonatal death per women in Ethiopia. Among these potential predictors education level of mother, mother Age at first give birth, Antenatal care visits, Husband/partner's education level and postnatal care visit after birth decreased the risk of neonatal mortality per women while living in rural areas, being born mother of Ahmara, Benishangul Gumuz and SNNPR Region, using public tap water source and having catholic and Muslim religion increased the risk of neonatal mortality per women. It recommended that awareness have to be given for the society on those potential factors and level of education, attending antenatal care service, check-up on postnatal care visit and early age of give birth should be give attention to improve survival of neonate in the Ethiopia.

BIBLIOGRAPHY

1. WHO (2011). Guidelines on optimal feeding of low birth weight infants in low and middle income countries.
2. Rajaratnam JK, Marcus JR, Flaxman AD, Wang H, Levin-Rector A, et al. Neonatal, post neonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: A systematic analysis of progress towards Millennium Development Goal 4. Lancet. 2010, 375(9730):1988-2008.
3. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, et al. Global, regional, and national causes of child mortalityin 2008: A systematic analysis. Lancet 2010, 375(9730):1969-1987.
4. UNICEF (2009). The state of the world's children. New York.
5. UNICEF (2005). The State of the World's Children.
6. Oestergaard MZ, Inoue M, Yoshida S, Mahanani WR, Gore FM, et al. Neonatal mortality levels for 193 countries in 2009 with trends since 1990: A systematic analysis of progress, projections, and priorities. PLoS Medicine. 8(8),doi:10.1371/journal.pmed.1001080
7. EDHS (2011). Ethiopian Demographic and Health Survey. Addis Ababa
8. Agresti, A. (2002). Categorical Data Analysis. Wiley Interscience, New York.
9. Agresti, A. (1996). An Introduction to Categorical Data Analysis, John Wiley and Sons, Inc New York.
10. Akaike, H. (1973). Information theory and extension of the maximum likelihood principle, Second international Symposium on Information Theory, ed. B.N. Petrov and F. Csaki, Budapest: AkademiaiKiado, pp. 267-281.
11. Wedderburn, R.W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss–Newton method, Biometrika, 61:439-47.
12. Cameron, A.C. &Trivedi, P.K. (2005). Microeconometrics. Methods and Application, Cambridge University Press.
13. Cameron, A.C. &Trivedi P.K. (1998). Regression Analysis of Count Data, Cambridge University Press, Cambridge, UK.
14. Cameron, A. C. and Trivedi P.K. (1986). Econometric models based on count data: Comparisons and applications of some estimators, Journal of Applied Econometrics, 1: 29-53.
15. Mccullagh, P. and Nelder J.A, (1989). Generalized Linear Models, Monographs on Statistics and Applied Probability, 2nd edn, Chapman and Hall, London, UK.
16. Nelder JA, RWM. Wedderburn (1972). Generalized linear models, Journal of the Royal Statistical Society, 135(3):370-384.

17. Lawless, J. F. (1987). Negative binomial and mixed Poisson regression, Canadian Journal of Statistics, 15(3): 209-225.
18. Hinde, J. (1996). Macros for fitting over dispersion models. Glim Newsletters. 26:10-26.
19. Sarkisian, Natalia and Naomi Gerstel. (2004). Explaining the Gender Gap in Help to Parents: The Importance of Employment. Journal of Marriage and the Family, 66:431-451.
20. Lambert D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34:1-14.
21. Long, JS. (1997). Regression Models for Categorical and Limited Dependent Variables, Thousand Oaks, CA: Sage.
22. Zuur FA, Ieno NE, Walker NJ, Saveliev AA. & Smith, MG. (2009). Mixed effects Models and Extensions in Ecology with R. Springer, New York, USA.
23. Dean C. & Lawless, JF. (1989). Tests for detecting over-dispersion in Poisson regression models, Journal of the American Statistical Association, 84:467-472.
24. Bauer L, Greibe P, Hua L & Liang L. (2007): Statistical Models of Accidents on interchange ramps and speed change lines, FHWA-RD – 97-106, U.S. Department of Transportation.
25. Vuong, QH. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica, 57:307-333.