

Analysis of Antenatal Care Visit Data in Bangladesh Using Zero Modified Count Regression Model

Nasiba Maruf Ahmed and Taslim Sazzad Mallick*

Department of Statistics, Dhaka University, Dhaka-1000, Dhaka, Bangladesh

(Received: 20 March 2019; Accepted: 23 June 2019)

Abstract

In medical science, pharmaceutical studies, public health and socio-economic researches we often encounter the situation of excess of zeros in count data. This preponderance of zeros leads to overdispersion. In such cases traditional count data regression models like Poisson and negative binomial (NB) regression may not be pertinent for inference. The two most commonly used types of model that have been developed to adjust for excessive zeros in count data are Hurdle and zero-inflated models. In this study we have analyzed the antenatal care (ANC) visit data of pregnant women in Bangladesh using traditional and zero-modified count models. Based on the model selection criteria, we found that negative binomial hurdle model fits the data best. Through this analysis, we have perceived that the variables age of mother, division, birth order (order a child is born), place of residence, economic condition, media exposure of the mother, main access road to village and education gap between husband and wife have significant impact on the mean number of ANC visits taken.

Keywords: Antenatal care (ANC) visit, count data, zero-inflated, Poisson, negative binomial.

I. Introduction

Antenatal care prepares a woman for delivery and helps to gain insights about warning signs during pregnancy and childbirth. The World Health Organization (WHO) recommends a minimum of four antenatal care visits. But only about 50% of all pregnant women can avail this amount of care. Good care during pregnancy is important for the safety of the mother. Pregnancy is the right time to promote healthy behaviors. Lack of care during this period breaks a critical link in the continuum of care, and affects both mother and baby. However, globally women in the richest quintile and urban region are more likely to receive antenatal care than those in the poorest quintile and rural region respectively¹.

Various researches have been carried out worldwide in order to identify the key determinants affecting the ANC visit. Yasouka et al.² conducted a community-based cross-sectional survey in 62 villages in Cambodia along with GPS (Global Positioning System) data to accurately measure actual travel distance of pregnant women to access health facilities and to examine geographical and environmental barriers. Promoting factors and barriers for ANC use among pregnant women living in remote, agricultural villages were revealed through this study. To assess the quality and utilization of ANC services in rural areas of the Lao PDR, Manithip³ carried out an extensive study on pregnant women between 2008 and 2011 using structured and semi-structured interviews along with the implementation of a low-cost ANC up-grading community-based intervention. Soe et al.⁴ performed a cross-sectional study to identify factors associated with the utilization of ANC among women living in Naypyidaw region of Myanmar. Joshi et al.⁵ obtained data from Nepal Demographic and Health Survey 2011 for 4,079 mothers to explore the determinants associated with the use and quality of antenatal care in Nepal. In all the above mentioned studies, ANC visit has been considered as a binary response and logistic regression

has been used. The common determinants that are found to have significant impact on the ANC visit in these studies includes age, education, husband's education, economic status, place of residence, number of children, presence of health issues during pregnancy, the nearest health center, knowledge about ANC and women's autonomy.

Bangladesh is one of the countries with high maternal mortality ratio (MMR) and neonatal mortality rate (NMR). Although NMR has decreased from 41 per 1000 live births in 2004 to 28 per 1000 live births in 2014, whereas the corresponding MMR has been reduced from 318 to 196 per 1000 live births but both the NMR and MMR needs to be sufficiently emphasized as it is far behind reaching the target⁶⁻⁷. With the Millennium Development Goals (MDG) phasing out and the Sustainable Development Goals (United Nations, 2015) phasing in, the 4th Health, Population and Nutrition Sector Program 2017-2022 has set the target of reaching an NMR of 12 per 1000 live births and an MMR of 105 per 100,000 live births in 2022 (Ministry of Health and Family Welfare⁸, 2017). The decline in NMR and MMR took place (from 2000 to 2014) due to several factors, where antenatal care visit turns out to be a significant one. In Bangladesh, the percentage of pregnant women who made at least one ANC visit increased from 51 to 78 and those who made four or more ANC visits has inflated from 17 percent in 2000 to the current level of 31 percent⁶. A substantial improvement is still needed in order to achieve the SDG goals. An increased number of ANC visit is expected to reduce MMR and NMR and hence it is necessary to identify the factors that are associated with ANC visits. Thereby, the analysis of ANC happens to be a crucial issue that needs to be prioritized. Islam et al.⁹ found that place of residence, financial status, age, academic qualification of the pregnant woman and exposure to mass media play significant role on the number of antenatal visits. A negative binomial regression model was used for the analysis. Mohammad et al.¹⁰ in their research depicted a significant increase of adequate ANC visits among pregnant

* Author for correspondence. e-mail: tsmallick@du.ac.bd

women with the increase in maternal education level and interaction between maternal education and place of residence provides a significant effect on complete ANC visits. Sequential logistic regression model has been used in the above study. Note that, ANC counts for under developed and developing countries often have too many zeros that a standard count data model such as Poisson or negative binomial model cannot accommodate. In recent years, count data with excess zeros are modeled using mixture models. Two common approaches are widely used in the literature. In zero inflated models, the probability of zero class is inflated in conjunction with the standard count data model. On the other hand, probability of observing zero and positives are separately modeled in hurdle approach. A marginalized zero-inflated Poisson (ZIP)¹¹ model, a modified Poisson regression model (that accommodates the excess zeros) has been used for analyzing ANC counts obtained from Bangladesh Demographic and Health Survey (BDHS) data¹². This type of modified count data model has been used widely in the literature for handling zero-inflated ANC data^{12,14-15}. A comparison of the performance of modified Poisson regression models like ZIP and Zero-inflated negative binomial (ZINB), and Hurdle Poisson (HP) and Hurdle negative binomial (HNB)¹³ models in determining the factors associated with the number of ANC visits taken, have been made in such studies. A study in Nigeria conducted by Yusuf et al.¹⁴ revealed that ZINB and HNB fitted the data better than the ZIP or HP. Sekata¹⁵ analyzed the determinants of the barriers in number of antenatal care service visits among pregnant women in rural Ethiopia. She identified that access to modern ANC visits during pregnancy was low in rural Ethiopia than the national value. Count models like Poisson, NB, ZIP, ZINB, HP, and HNB regression models were fitted in this study. Heavy work load was found to be significantly associated with not attending ANC visits along with the other common influential factors. In this paper, we will use zero modified count model to identify the factors associated with the number of ANC visits taken using BDHS (Bangladesh Demographic and Health Survey) data, 2014. To be specific, we will compare standard count models with the modified models in order to depict the variation in results that occurs when the excess zero problem is taken into account. Based on the best fitted model, we will estimate the effects of several factors on ANC visits.

II. Data and Variables

In BDHS 2014, data are collected on 4595 variables containing 17863 observations. Information on ANC was assessed for women who gave birth in the three years preceding the survey. The survey is based on a two-stage stratified sample of households. In the first stage, 600 enumeration areas (EAs) were selected with probability proportional to the EA size, with 207 EAs in urban areas and 393 in rural areas. In the second stage of sampling, a systematic sample of 30 households on average was selected per EA. Among women with two or more live

births during the three-year period, data refer to the most recent live birth only. Among the 17863 women who were surveyed, our study focuses on the antenatal care of 4300 women, who had a recent child birth at the time of the survey and had complete information of all the determinants that are considered in this study. The determinants are chosen on the basis of previous studies, which are thought to be potential factors affecting antenatal care visit. The response variable is the number of antenatal care visits of these 4300 mothers. To identify the influential risk factors that effects antenatal care visit is the main purposes of this study. The variables that have been taken under consideration for this study are as follows: region (Barisal, Chittagong, Dhaka, Khulna, Rajshahi, Rangpur, and Sylhet), age of mother (age below or equal to 20, Age 21-35, and age above 35), women empowerment (yes, no), birth order number (1st, 2nd – 3rd, 4+), place of residence (urban and rural), wealth index (poor, middle, rich), media exposure (exposed, unexposed), NGO membership of mothers (yes, no), availability of maternal and child welfare center MCWC (yes, no), main access road to the village/moholla(all weather road, others), education gap between husband and wife (husband more educated, else) and a variable named ‘mother having son’ has been defined with two categories (if the mother has never given birth to a son or not).

III. Methodology

Zero-modified techniques permit the researcher to answer two questions that pertain to low base rate-dependent variables. In other words, two regression equations are created: one predicting whether the count occurs and a second one predicting the occurrence of the count¹⁶. A count response Y with excess zeros has a zero-inflated distribution, given by

$$\begin{aligned} \Pr(Y_i = 0) &= f_1(0) + (1 - f_1(0))f_2(0) \\ \Pr(Y_i = k) &= (1 - f_1(0))f_2(k), \\ &k = 1, 2, 3 \dots \end{aligned} \quad (1)$$

where $f_1(\cdot)$ models the binary part and $f_2(\cdot)$ models count part of the data. The binary part is commonly modeled by logit model; for the count part, Poisson or negative binomial is used¹⁶. Hurdle regression is also known as two-part model which is originally developed by Mullahy¹². Mullahy states, ‘The idea underlying the hurdle formulations is that a binomial probability model governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is non-zero (positive), the ‘hurdle is crossed’, and the conditional distribution of the positives is governed by a truncated-at-zero count data model.’

A count response Y with excess zeros has a Hurdle distribution, given by

$$\begin{aligned} \Pr(Y_i = 0) &= f_1(0) \\ \Pr(Y_i = k) &= \frac{(1-f_1(0))}{(1-f_2(0))} f_2(k), \quad k = 1, 2, 3 \dots \end{aligned} \quad (2)$$

The log likelihood function for hurdle is given by $\log L = L_1 + L_2$; where,

$$L_1 = \sum_{i=1}^n [(1 - 1[y_i \in J]) \ln Pr[y_i = 0|x_i]]$$

$$+ \sum_{i=1}^n 1[y_i \in J] \ln Pr[y_i = 1|x_i]$$

$$L_2 = \sum_{i=1}^n 1[y_i \in J] \ln Pr[y_i | x_i, y_i > 0] \tag{3}$$

where the indicator function is $1[y_i \in J] = 1$ if $y_i \in J$, $1[y_i \in J] = 0$ if $y_i = 0$ and $J = 1, 2, 3, \dots$ is a positive set¹⁶.

In case of zero-inflated models, the log likelihood function is also given by $\log L = L_1 + L_2$ but the indicator function is reversed and the weights are altered accordingly. The R package ‘pscl’ has been used to estimate the parameters.

IV. Results and Discussion

Bivariate Analysis

Table 1 represents unadjusted effects of several factors on ANC counts using bivariate setup, where the *p*-values are obtained using ANOVA. The variables - age of mother,

division, empowerment status of mother, birth order (order a child is born), place of residence, economic condition, media exposure of the mother, main access road to village, mother having son and education gap between husband and wife were found to have significant association with the number of ANC visits taken. For example, women living in urban area have significantly higher ANC visits (*p*-value<0.001) on an average, in comparison to women living in rural areas. With the increase in the parity of mothers, average number of ANC visits decreases (*p*-value<0.001). The highest number of mean ANC visits taken among the mothers is in Khulna and the lowest number of mean ANC visits taken is in Sylhet. Women having no son, have 6.6% more ANC visits as compared to those who have at least one son (*p*-value=0.026). The mean number of ANC visits happens to be 31.25% more for those women who have all weather road as main access road in comparison to those who have not (*p*-value<0.001).It is also observed that, when wife is more educated than husband or they are equally educated, the mean number of ANC visits tends to be 11% more than the group where husband is more educated (*p*-value<0.001).

Table 1. Mean number of ANC visits by selected Demographic and Socioeconomic variable

Variables	Category	Mean No. of ANC visits (CI)	P-value
Residence	Urban	3.69 (3.54, 3.84)	< 0.001
	Rural	2.36(2.28,2.44)	
Region	Barishal	2.45 (2.24,2.65)	< 0.001
	Chittagong	2.61 (2.44,2.78)	
	Dhaka	3.06(2.87,3.24)	
	Khulna	3.46(3.23,3.70)	
	Rajshahi	2.84 (2.61,3.08)	
	Rangpur	3.20 (2.99,3.41)	
	Sylhet	2.06(1.88,2.24)	
	Birth order	1 st	
	2 nd & 3 rd	2.79 (2.67,2.90)	
	4 th +	1.69 (1.53,1.86)	
Wealth index	Poor	2.02 (1.92,2.13)	< 0.001
	Middle	2.51 (2.36,2.65)	
	Rich	3.65 (3.52,3.78)	
Mother’s age at birth	Age≤20	2.75 (2.62,2.78)	0.005
	21 to 35	2.85(2.75,2.95)	
	35 above	2.16(1.78,2.53)	
NGO membership	No	2.81(2.73,2.90)	0.201
	Yes	2.68(2.50,2.86)	
Media exposure	No	1.80 (1.70,1.90)	< 0.001
	Yes	3.38(3.28,3.48)	
Women empowerment	Yes	2.85(2.77,2.94)	0.001
	No	2.53(2.36,2.69)	
Availability of MCWC	Yes	2.80(2.71,2.88)	0.636
	No	2.73(2.47,2.99)	
Main access road	All weather road	3.04(2.94,3.13)	< 0.001
	Others	2.09(1.96,2.22)	
Mother having son	No son	2.91(2.78,3.04)	0.026
	Else	2.73(2.63,2.82)	
Education gap	Husband more	2.54(2.38,2.70)	0.001
	Else	2.85(2.77,2.94)	

Multivariate Analysis

The frequency distribution of ANC visits shown in Table 2 evidently reveals that the percentage of zero visits is very high. Thereby, it makes more sense to assume a separate source for extra zero and hence it will be justified to use a suitable zero modified count model. Since Poisson and Negative Binomial (NB) are the basic count data model, we commenced our analysis by applying them. We then analyze ANC visit counts by using Zero Inflated Poisson (ZIP), Zero Inflated Negative Binomial (ZINB), Hurdle Poisson (HP), and Hurdle Negative Binomial (HNB). Table 3 depicts the performance of these models on ANC visit data, using Akaike’s information criteria (AIC) and Bayesian information criteria (BIC).

Table 2. Frequency Distribution of ANC Visit data

No. of ANCvisit	0	1	2	3	4	5	6	7	8+
Percentage	21.1	16.1	16.8	13.8	11.0	6.8	4.0	3.9	6.4

Table 3. Model Selection Criteria for the Regression Models

Criterion	Poisson	NB	ZIP	ZINB	HP	HNB
AIC	18902.57	17790.89	18046.69	17606.85	18046.01	17598.41
BIC	19036.27	17930.95	18314.08	17880.60	18313.40	17872.17

From Table 3, we observed that AIC and BIC values for both ZINB and HNB models were similar. Since HNB yielded least AIC/BIC, we have used this to explain the

Table 4. Analysis of ANC visit data using Hurdle NB Model

Variables	Category	Count Part			Binary Part		
		IRR	Std.Error	P-value	OR	Std.Error	P-value
Intercept		2.300	0.086	< 0.001	1.487	0.247	0.108
Residence	Urban	.841	0.031	< 0.001	.769	0.110	0.016
	Rural						
Region	Dhaka						

results. However, if a ZINB was fitted instead of HNB, the same conclusion would have been drawn. For the sake of convenience, the effect of the independent variable on the dependent variable through the incidence rate ratio (IRR) for the count portion and odds ratio (OR) for the binary portion has been reported in table 4. The IRR for the effect of a covariate, x_j (either qualitative or quantitative) can be obtained as $IRR_j = \exp(\beta_j)$; $j = 1, \dots, p$, if the regression model does not involve any interact term involving x_j .

From table 4 it is found that the variables age of mother, division, birth order (order a child is born), place of residence, economic condition, media exposure of the mother, main access road to village and education gap between husband and wife were found to be significant for both count and binary portion. From the binary portion in Table 4, we also found that the variable ‘empowerment status of mother’ was found to have an impact on the odds of taking ANC visits at 10% level of significance. We have observed that women who are aged above 21, have significantly higher odds of taking ANC visits. Women between 21-35 years have 14% higher ANC visits (p -value<0:001) on an average, in comparison to other age groups. Average ANC visits is significantly higher in Khulna (15%, p -value= 0.004) and Rangpur (22%, p -value<0.001) as compared to Dhaka. The study reveals that as the birth order increases, the average number of ANC visit count significantly decreases. As expected, average ANC visits for women from the rural area is significantly lower (16%, p -value<0.001) and rich women have significantly higher number of ANC visits (20%, p -value<0.001). Media exposure tends to have significant impact on the number of ANC visits taken, as the average number of ANC visits taken by the exposed group is 21% higher than that of the unexposed group(p -value<0.001). Women who have an access to all weather road are more prone to taking ANC visits (8% higher, p -value = 0.029) in comparison to others. The results indicate that education gap between husband and wife also turns out to be a significant factor that affects the number of ANC visits. On an average, the number of ANC visits taken by women who are more/equally educated is 10% more in comparison to women whose husbands are more educated (p -value=0.003). Although, the variable mother having son and empowerment status of women had significant association with mean number of ANC visits in the bivariate setup, these did not turn out to be significant in the Hurdle negative binomial model.

	Chittagong	.939	0.044	0.155	.631	0.140	0.001
	Barishal	.976	0.053	0.649	.808	0.156	0.171
	Khulna	1.147	0.048	0.004	1.516	0.185	0.024
	Rajshahi	1.014	0.050	0.764	.780	0.159	0.119
	Rangpur	1.228	0.048	< 0.001	1.105	0.163	0.535
	Sylhet	.922	0.051	0.115	.498	0.142	< 0.001
Birth order	1 st						
	2 nd & 3 rd	.894	0.035	0.001	.516	0.121	<0.001
	4 th +	.725	0.060	< 0.001	.314	0.158	<0.001
Wealth index	Poor						
	Middle	.979	0.042	0.609	1.433	0.109	< 0.001
	Rich	1.202	0.039	< 0.001	2.273	0.117	< 0.001
Mother's age	Age ≤ 20						
	21 to 35	1.142	0.035	< 0.001	1.680	0.119	<0.001
	35 above	1.146	0.095	0.149	1.964	0.242	0.005
NGO membership	No						
	Yes	1.011	0.037	0.748	.906	0.105	0.345
Media exposure	Unexposed						
	Exposed	1.210	0.036	<0.001	2.445	0.093	<0.001
Women empowerment	Empowered	1.040	0.035	0.268	1.201	0.099	0.064
	Not						
Availability of MCWC	Yes	.995	0.048	0.910	1.182	0.143	0.244
	No						
Main access road	All weather	1.082	0.036	0.029	1.190	0.091	0.057
	Others						
Mother having son	No son	.986	0.029	0.642	1.049	0.091	0.595
	Else						
Education gap	Husband more						
	Else	1.105	0.034	0.003	1.428	0.098	< 0.001

V. Conclusion

Although, the current Antenatal care coverage of at least one visit is 78.6% and that of at least four visit is 31.2% (BDHS 2014), it is far beyond reaching the target of 100% and 50% respectively set by the Health, Population, and Nutrition Sector Development Program (HPNSDP) of the Ministry of Health (MoH) of Bangladesh. Hence, it is imperative to identify the determinants that influences the number of ANC visits by pregnant mothers in Bangladesh. The ANC visit data obtained from BDHS 2014, contains a higher percentage of zeros and thereby requires scrutiny. Data with a preponderance of zero count often appear in various fields of research. Several competing zero modified models have been proposed to handle this situation. In this research we have fitted all possible count data models that include hurdle and zero-inflated model. We have found that HNB best fits the ANC data. Mothers who are pregnant for the first time takes more ANC visits than those who are experienced. We could also comprehend that the education gap between husband and wife is also an influential factor; when the wife is equally or more educated than husband, the average count of ANC visit has elevated in comparison to the situation when husband is more educated than wife. The variable "Mother having son" has been found to be associated with the average count of ANC visits in the bivariate analysis, which explains why a mother with no son takes more ANC visits than mothers who already have

one or more son; although, the variable did not turn out to be significant in the model.

Our findings suggest that, women who are aged less than 20, economically disadvantaged, belonging to the class of higher parity should be targeted to improve the current condition. Along with the father, a mother also needs to be equally educated to ensure that she is receiving the mandatory ANC visits during pregnancy. Women from Sylhet, Barishal and Chittagong are found to take least number of ANC visits, so these areas require more attention. Necessary steps should to be taken so that women from rural areas with poor communication system (those who do not have access to all weather road) get the opportunity to avail the minimum number of ANC visits required during pregnancy. Social awareness can be arisen among mass people by strengthening mass-media campaign on the benefits of adequate number of ANC visits.

Acknowledgment

The authors would like to thank the authority of National Institute of Population Research and Training (NIPORT), Bangladesh for allowing them to use the BDHS, 2014 data for their analysis. Moreover, the authors would like to thank Professor Dr. Wasimul Bari and Nasrin Sultana, Department of Statistics, University of Dhaka for their help on the preparation of this manuscript. The authors are

grateful to the editor and the reviewer for their helpful remarks.

References

1. UNICEF Global database, 2016
2. Yasouka, J., K. Nanishi, K. Kikuchi, S. Suzuki, P. Ly, B. Thavrin, T. Omatsu, T. Mizutani, 2018. Barriers for pregnant women living in rural, agricultural villages to accessing antenatal care in Cambodia: A community-based cross-sectional study combined with a geographic information system. *PLoS ONE* **13**, e0194103.
3. Manithip, C., 2012. Quality and utilisation of antenatal care services in Lao PDR. Department of Women's and Children's Health, Karolinska Institutet.
4. Soe, W. M., J. Chompikul and A. Mongkolchat, 2015. Predictors of Quality Utilization of Antenatal Care Services in Naypidaw, Myanmar. *Journal of Public Health and Development*, **13**(3), 3-17
5. Joshi, C., S. Torvaldsen, R. Hodgson, A. Hayen, 2014. Factors associated with the use and quality of antenatal care in Nepal: a population-based study using the demographic and health survey data. *BMC Pregnancy and Childbirth*, **14**, 94.
6. Bangladesh Demographic and Health Survey (BDHS). 2014. NIPORT, Dhaka, Bangladesh; Mitra and Associates, Dhaka, Bangladesh.
7. Bangladesh Progress Report 2015. General Economics Division (GED), Planning Commission, Government of the People's Republic of Bangladesh.
8. Setting Priorities for Data Support to 7th FYP and SDGs: An overview. Bangladesh Bureau of Statistics, 2016.
9. Islam, M.M., M. S. Masud, 2018. Determinants of frequency and contents of antenatal care visits in Bangladesh: Assessing the extent of compliance with the WHO recommendations. *PLoS ONE* **13**(9), e0204752.
10. Mohammad, K. A., Fatima-Tuz-Zahura, M. M. Rahman, 2017. Importance of Maternal Education on Antenatal Care Visits in Bangladesh. *Bangladesh Journal of Scientific Research*, **30**(1&2), 23-33.
11. Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1-14.
12. Islam, M., 2015. Statistical Regression Model for Count Data with many zeros: An approach for Marginal Inference. Department of Statistics, University of Dhaka.
13. Mullahy, J., 1986. Specification and testing of some modified count data models. *Journal of Econometrics*, **33**, 341-365.
14. Yusuf, O. B., R. F. Afolabi, A. S. Agbaje, 2018. Modelling Excess Zeros in Count Data with Application to Antenatal Care Utilisation. *International Journal of Statistics and Probability*, **7**(3), 22-35
15. Sekata, D., 2015. Modeling the Number of Antenatal Care Service Visits Among Pregnant Women in Rural Ethiopia: Zero Inflated and Hurdle Model Specifications. *International Journal of Healthcare Sciences*, **3**(1), 332-355.
16. Cameron, A.C. and P. K. Trivedi, 1986. Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics*, **1**(1), 29-54.