MODELING TOTAL NUMBER OF CHILD DEATH UNDER-5 AGE IN HOUSEHOLD OF BANGLADESH USING ZERO INFLATED NEGATIVE BINOMIAL REGRESSION

***Md. Maidul Husain1, Biplab Biswas1, Anjuman Ara1 and Asmaul Husna1**
(1) Department of Statistics, Bangabandhu Sheikh MujiburRahman Science and Technology University, Gopalganj-8100, Bangladesh

IJASR 2019
VOLUME 2
ISSUE 4 JULY – AUGUST **ISSN: 2581-7876**

Abstract – Total number of child death under 5 years is an indicator of child health and overalldevelopment of a country, as it reflects the socio-economic condition of a country where children are growing up. Despite substantial progress in reducing child mortality over the world, remains child death under 5 years urgent concern for many developing country specially in Bangladesh is 32 per 1,000 live births, according to new mortality estimates released by UNICEF, WHO, the UN Population Division, and the World Bank Group. There are many factors contributed to child mortality such as mother education, place of residence, awareness of caring child, maternal age of mother etc. In this paper we have modeled the number of under-5 children deaths experienced by a mother to the associated available factors suggested by the literature. In BDHS 2014 survey, nearly 84.83% of the mothers never experienced any under-5 child death. That means, our response variable is zero-inflated, which motivated us to fit Zero inflated negative Binomial (ZINB) models to the data. ZINB combines of binomial distribution and logit distribution, model has statistical advantage to modeling over dispersion and the excess no. zeros in the data set. The empirical results of this study support the hypothesis about children deaths under 5 years age in a household.

Keywords: Child Mortality, Count Model, ZINB, BDHS.

Introduction

Bangladesh is one of the most densely populated countries with about 1,100people per square kilometer. Globally sociologists and policymakers pay great concentration on the factors affecting under-five child mortality as it is an important indicator for describing standard of living, socio-economic well-being of a country. World-wide several studies have been undertaken that focused on the socio-economic determinants of child mortality. In Bangladesh between 1989 and 2014, under-five child mortality rate reduced by two-third, from 133 to 46 deaths per thousand live births (BDHS 2014). Despite these prominent gains in reducing child mortality, the progress needs to be accelerated to achieve third Sustainable Development Goals (SDG) by 2030, where under-five mortality to at least as low as 25 per 1,000 live births. Bangladesh is one of the developing countries of Asia with high under-five child mortality rate. Various social, demographic and environmental factors had often been studied in relation to under-five child mortality in Bangladesh. Earlier studies shown that less educated mother's experienced higher child mortality rate than those who are educated[3] .Because, it is known that the educated mother well nurtured her children than non-educated mother[5]. Besides these optimal mothers child bearing age 20-34 also strongly associated with child mortality where premature or later pregnancies increase the risk of child mortality. Mortality also varied by different socio economic characteristic such wealth index of household, the administrative region where they lived. This study also intends to analyze of exposure of media and facility of community health care located in the region.

Majority of the studies conducted on child mortality used multivariate logistic regression model or cox proportional hazard model for identifying the risk factor associated with child mortality. For example, hazard model are used for identifying the covariates on the child mortality [7; 11; 1]. Muriithi and Muriithialso employed Cox regression survival analysis to the 2008 - 09 Kenya Demographic Health Surveys [8]. A multivariate Cox regression analysis has been performed by Nisar and Dibley [9] to determine the potential risk factors of neonatal mortality in Pakistan. Moreover, cox proportional hazard model and logistic regression model also used for identifying the risk factors of child mortality [6; 5]. This approach may result in loss of information and this may lower the statistical power of the model as opposed to treating the response variable as a count data (i.e. number of infant death or child death)[2; 10]. In this paper, we used ZINB model, the total number of child death in a household (count) decreases last decade and most of the household exhibit child mortality zero. This data prone to over dispersion and excess no. of zero and they adjust well data from a mixture population, one that has zero

counts and another in which count come from a discrete distribution. The result of the Bernoulli trials is used to determine which of the two process (either zero or count) is used. The primary objective of the current study is to identify socioeconomic and demographic risk factors/predictors of the number of children's death under 5 years for women aged 12-49 from the Bangladesh Health and Demographic Survey (BDHS) administered in 2014 and also compare with usual count model for overly dispersed data.

Study Sites and Participants

This study use the secondary data come from Bangladesh Demographic Health Survey(BDHS) conducted in 2014. The detailed methodology of the survey design, data collection, and data management has been described in BDHS 2014 survey report. Information was taken from birth history section where total 14077 women interviewed aged from 15-49. For this study we exclude the missing subjects by considering the missing observation completely at random, and nationally got total no 7886 women's for further analysis. Women's were asked to provide the total number of son's and total number of child daughter ever died, by adding we get total number of children died in a households, and considered as dependent variable. For independent variable mother's age, mother's education, economic status , occupation , aware of community clinic, no. of living children, and media exposed are related to the under-5 children deaths as a categorical covariates. And only continuous covariate age of mother at first birth of her children with mean 17.47.

Table 1: Summary statistics of data

Variable	Category	Percentage/Mean
Dependent Variable		
Total Child	Count	0.18
Independent Variable		
Mother's Age	"15-24"	12.1
	"25-34"	34.9
	"35-49"	53.0
Economic Status	Poor	41.1
	Middle	19.2
	Rich	39.7
Education level	No Education	34.0
	Primary	32.7
	Secondary	28.2
	Higher	5.1
Occupation	Working	65.8
	Not Working	34.2
Division	Dhaka	17.5
	Chittagong	19.2
	Barishal	11.5
	Khulna	10.9
	Rajshahi	12.2
	Rangpur	12.1
	Sylhet	16.6
Aware of Community Clinic	Yes	30.3
	No	69.7
Age First Birth	Continuous	17.4
Living Child	Child 0	52.4
	Child 1	35.1
	Child 2+	12.5
Media	Exposed = 1	48.3
	Not-Exposed = 0	51.7

Briefly, most of the women's are belong to "35-49" age group, has a no education, most of the womenemployed as household agricultural work, and parity of the women in all division almost same. Underhousehold characteristics, highest percent of households are poor, have a more child live and exposed and non-exposed to

the media are same. Here most important subjects that average age of mother's at first births of her children has increased and reached to 18, which is the one of influential factor reducing under -5 child mortality in Bangladesh

Methodology

The zero inflated negative binomial regression models assume that there are two distinct data generation process. Bernoulli trials are used for determined two processes with probability p_i for zero counts and $1-p_i$ for other counts from negative binomial distribution respectively. Let $Z_i$is a binary response with $\Pr(Z_i = 0) = p_i$ and $\Pr(Z_i = 1) = 1 - p_i$. Also suppose that Y_i is a negative binomial variate with mean μ_i and dispersion parameter $\phi^{-1}$with pmf is given by

$$f(Y_i^*, \mu_i, \phi) = \frac{\Gamma(Y_i^* + \phi)}{\Gamma\phi Y_i^*!}\left(\frac{\phi}{\mu_i + \phi}\right)^{\phi}\left(\frac{\mu_i}{\mu_i + \phi}\right)^{y^*} \quad ; \mu_i > 0 \; and \; \phi \; > 0$$

Following [4] one can define the observed zero inflated random variable $Y_i$as $Y_i = Y_i * Z_i$. Therefore Y_i can take zero if either binary variable Z_i or Y_i takes values zero. So,

$$\begin{aligned} Pr(Y_i = 0) &= Pr(Z_i = 0 \cup \; Y_i^* \; = 0) \\ &= \; Pr(Z_i = 0) + Pr(Y_i^* = 0) - Pr(Z_i = 0)Pr(Y_i^* = 0) \\ &= p_i + \left(\frac{\phi}{\mu_i \; + \phi}\right)^{\phi} - p_i\left(\frac{\phi}{\mu_i + \phi}\right)^{\phi} \\ &= p_i + (1 - p_i)\left(\frac{\phi}{\mu_i \; + \phi}\right)^{\phi} \end{aligned}$$

Again Y_i can take the value k (k=1,2,3...) if and only if the binary variable Z_i takes the value 1 and the count variable Y_i^* take k. So,

$$\begin{aligned} Pr(Y_i = k) &= Pr(Z_i = 1 \cap \; Y_i^* = k) \\ &= \; Pr(Z_i = 0) \; Pr(Y_i^* = k) \\ &= (1 - p_i)\frac{\Gamma(\mathrm{k} + \phi)}{\Gamma\mathrm{k}\phi!}\left(\frac{\phi}{\mu_i \; + \phi}\right)^{\phi}\left(\frac{\mu_i}{\mu_i \; + \phi}\right)^{k} \end{aligned}$$

So, a count response $Y_i$with excess zero has zero inflated negative binomial is given by

$$\begin{aligned} Pr(Y_i = y_i) &= \{p_i + (1 - p_i)\left(\frac{\phi}{\mu_i \; + \phi}\right)^{\phi} \; if \; y_i = 0 \\ &= (1 - p_i)\frac{\Gamma(\mathrm{k} + \phi)}{\Gamma\mathrm{k}\phi!}\left(\frac{\phi}{\mu_i \; + \phi}\right)^{\phi}\left(\frac{\mu_i}{\mu_i \; + \phi}\right)^{k} \; if \; y_i = 1,2, \end{aligned}$$

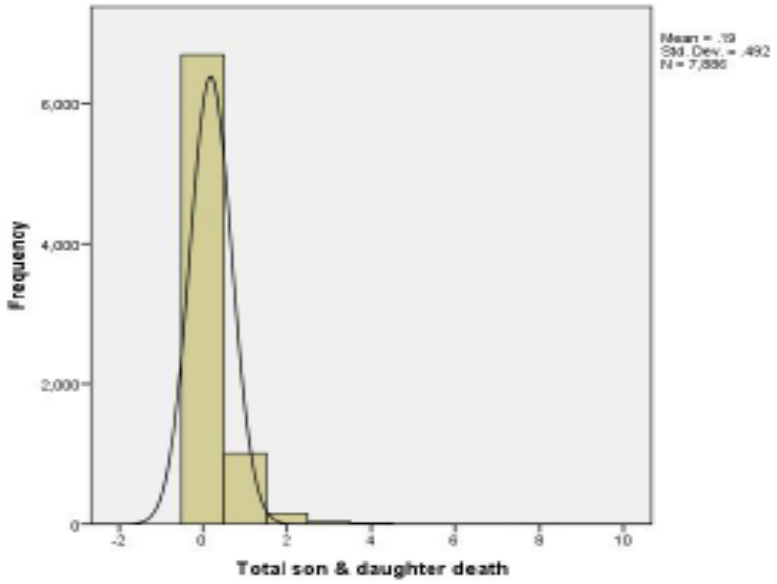
Results and Discussion

At first of the beginning of the analysis we had shown the some descriptive statistics of the dependent variable. From Table 2 it is obtained that the mean number of children deaths under 5 years age is 0.18 and variance is 0.24, which exhibit the over dispersion in the data set. Excess percentage of zeros (88.83) also responsible for over dispersion in the data set (Figure 1 also revealed this). As a count model, in that case, we can applied two regression models, either negative binomial model(NB) or zero inflated negative binomial regression (ZINB), allowed over dispersion to estimate the covariate effect on the response.

Table 2: Descriptive statistics of number of death under 5 year age

	N	Mean	Variance	Zero (%)	Non-Zero (%)
Total Child Death	7886	0.18	0.24	84.83	16.17

Figure 1: Distribution of total number of child death under-5 year age



Result of the negative binomial regression model is shown in Table 3, and we also estimated the incidence risk ratio for covariates. Since excess no. of zero present in the data set zero inflated negative binomial model is more appropriate for estimated parameter, which is supported by the following voung test results reported in Table 4.

Table 3: The result of negative binomial regression approach

Covariates	Category	Risk Ratio(estimated)	Standard (Std) Error
Intercept		0.218	0.237
Mother's Age	"15-24"		
	"25-34"	1.13	0.105
	"35-49"	0.994	0.110
Economic Status	Poor		
	Middle	0.686	0.08
	Rich	0.576	0.08
Education Level	NoEducation		
	Primary	0.996	0.077
	Secondary	0.977	0.087
	Higher	1.196	0.147
Occupation	Not Working		
	Working	1.217	0.06
Division	Barishal		
	Chittagong	1.424	0.116
	Dhaka	1.077	0.124
	Khulna	1.077	0.136
	Rajshahi	1.204	0.127
	Rangpur	0.996	0.130
	Sylhet	1.893	0.113
Awareness of Community Clinic	No		
	Yes	1.033	0.06
Age First Birth	Continuous	0.992	0.011
Living Child	Child 0		
	Child 1	0.845	0.069
	Child 2+	0.999	0.094

Table 4: Comparison Test for NB and ZINB

	Z-statistic	p-value
Voung Test	-4.3107	<0.000

So, we naturally preferred zero inflated negative binomial model (ZINB), since it accommodate over dispersion in child death under-5 years (count) causing excess no. of zeros. Finally, we reported incidence rate ratio (IRR) for child death under-5 years for all groups of categorical variables in these two models as well as the standard error. Comparing estimated parameter and standard error from both the Negative Binomial model and Zero inflated negative binomial models are quite dissimilar (same estimates). Of the list of explanatory variables estimated risk ratio using the ZINB (From Table 5) we can interpret our fi dings. Since day by day modern health facilities for a child improved that's why in age group "15-24" mothers experienced comparatively lower incidence child death rate of than mother's age "25-34", but we can find this statistically insignificant . Education status of a mother strongly associated with nutrition condition of a child, as this known to her importance of child health care. Increasing in the years of education of a mother delay the age at marriage and perhaps get large gap of first childbirth which reduce the risk rate of child death. Mother's having education primary and secondary incidence death rate of is much more significantly lower than the women who don't having any education. Higher socio economic status provides a household greater access to health facilities, adequate nutrition and sanitary conditions and protective factors of child death. So, the middle class and rich class households have 20% and 32% lower incidence total child death rate then the poor community. In Bangladesh, all administrative regions have not achieved a significant equal degree of modern facilities of child caring due to insufficient resources. That's why child mortality also varied across divisions, and among all divisions chittagong and Sylhet are statistically associated with number of under 5 age children's death (Barishal is considered as a reference category). The incidence death ratio of children's death in sylthet and chittagong divisions are 1.74 and 1.33 times higher than the barishal divisions. Media is encompasses watching television, listening radio , reading newspaper and internet has becomes our daily needs and play important rolecreating and shaping public awareness about under 5 year child mortality. The women who have exposed to the media experienced about 25% lower incidence deaths rate of children's death then women's who are unexposed to the media.

Table 5: The result of Zero Inflated Negative Binomial regression approach

		Count Part		Binary Part	
Covariate	Category	Estimate	Std Error	Estimate	Std Error
		(RR)		(OR)	
Intercept		.407	0.292	9.832	1.44
Mother's Age	"15-24"				
	"25-34"	1.234	0.122	1.683	0.636
	"35-49"	0.951	0.126	0.465	0.707
Economic Status*	Poor				
	Middle	0.817	0.103	5.055	0.632
	Rich	0.683	0.112	4.552	0.580
	NoEducation				
Education level*	Primary	0.763	0.105	0.149	0.562
	Secondary	0.634	0.118	0.013	1.973
	Higher	0.800	0.168	4.3E-8	1291
Occupation	Not Working				
	Working	1.009	0.084	0.09	0.659
Division*	Barishal				
	Chittagong	1.33	0.138	.423	0.677
	Dhaka	1.053	0.140	.725	0.722

	Khulna	1.007	0.153	.626	0.825
	Rajshahi	1.128	0.146	0.404	1.099
	Rangpur	1.056	0.154	1.260	0.672
	Sylhet	1.74	0.129	0.376	0.238
Awareness of Community Clinic	No				
	Yes	0.944	0.075	0.346	0.508
Age First Birth	Continuous	0.980	0.012	0.855	0.069
Living Child	Child 0				
	Child 1	1.029	0.098	5.842	0.456
	Child 2+	1.082	0.112	2.266	0.594
Media*	Unexposed				
	Exposed	0.756	0.096	0.523	0.490

Conclusion

The zero inflated negative binomial regression has been shown to adequately model such count data having excess zeros and over-dispersion and hence should be considered as a statistical method of analyzing under 5 years age children death data. Our main focus was, to identify socio-economic and socio-demographic factors on total no. of child deaths of women aged 12-49 in Bangladesh by comparing ZB and ZINB in handling over dispersed count data causing excess no. of zeros. Regional influence, mother's education. Economical position in society, and media exposed are statistically significantly associated with the number of children's death in a family. These findings provide proper insights to policymaker and guidelines for implementing needed intervention to reduce the child mortality under 5 years age in Bangladesh as well as other countries of the world.

References

[1] Bhuiya, A. and Streateld, K. (1992). A hazard logit model analysis of covariates of childhood mortality in matlab, bangladesh. Journal of biosocial science, 24(4):447-462.

[2] Britt, C. L. and Weisburd, D. (2010). Statistical power. In Handbook of quantitative criminology, pages 313-332. Springer.

[3] Cleland, J., Bicego, G., and Fegan, G. (1991). Socio-economic inequalities in childhood mortality: the 1970s compared with the 1980s. [Unpublished] 1991. Presented at the Demographic and Health Surveys World.

[4] Greene, W. H. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models.

[5] Islam, R., Hossain, M., Rahman, M., and Hossain, M. (2013). Impact of socio-demographic factors on child mortality in bangladesh: an multivariate approach. International Journal of Psychology and Behavioral Sciences, 3(1):34-39.

[6] Kwara, K. (2012). Modelling the risk factors of neonatal mortality in Ghana using logistic regression. PhD thesis.

[7] Martin, L. G., Trussell, J., Salvail, F. R., and Shah, N. M. (1983). Co-variates of child mortality in the philippines, indonesia, and pakistan: an analysis based on hazard models. Population Studies, 37(3):417-432.

[8] Muriithi, D. M. and Muriithi, D. K. (2015). Determination of infant and child mortality in kenya using cox-proportional hazard model. American Journal of Theoretical and Applied Statistics, 4(5):404-413.

[9] Nisar, Y. B. and Dibley, M. J. (2014). Determinants of neonatal mortality in pakistan: secondary analysis of pakistan demographic and health survey 2006{07. BMC Public Health, 14(1):663.

[10] Piza, E. L. (2012). Using poisson and negative binomial regression mod-els to measure the in uence of risk on crime incident counts. Rutgers Center on Public Security.

[11] Trussell, J. and Hammerslough, C. (1983). A hazards-model analysis of the covariatesof infant and child mortality insrilanka. Demography, 20(1):1-26.