ANALYSIS AND COMPARISON OF UNDER FIVE CHILD MORTALITY BETWEEN RURAL AND URBAN AREA IN BANGLADESH

Arfan Raheen AFZAL[1]

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B)

E-mail: nafzal@ucalgary.ca

Sabrina ALAM[2]

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B)

Abstract:

Knowledge of factors that affect the under-five year child mortality is important because it pertains to policy and programs. Causes and differences of under-five mortality between rural and urban area help decision makers to assess programmatic needs and prioritize interventions. This paper investigates the causes and differences of under-five mortality between rural and urban area in Bangladesh using Kaplan-Meier, Cox Proportional Hazard (Cox-PH) and Accelerated Failure Time (AFT) Regression model. Bangladesh Demographic and Health Survey (BDHS)-2007 data are used for the study. The results show that for both the areas, survival probability for children whose mothers have higher education is very high and in urban area the failure rate is very high for children of poor economic status. The Cox-PH analysis reveals that risk of death was lower for children whose mothers were matured and higher educated than younger and no educated mother in rural area. In urban area, children from rich family and the 2nd or 3rd child have lower risk of death compared to poor and 1st child. The AFT analysis shows that for both the areas Weibull distribution better fits the data.

Key words: *child mortality; urban; rural; Bangladesh; ATF; Cox-PH; KM; BDHS*

1 INTRODUCTION AND LITERATURE REVIEW

We are interested in analyzing and comparing child mortality between rural and urban area because identifying these may help the government to correct and formulate its policy to reduce child mortality in Bangladesh. Therefore, the analysis done on child mortality has received considerable attention. This paper provides empirical evidence that, some covariates influence child mortality where the covariates are: sex and birth order of the child, mother's age, education and economic status of the child using Kaplan-Meier, Cox Proportional Hazard (Cox-PH) Regression model and Accelerated failure time (AFT) regression model.

Mozumder et al. (1998) obtained data on a cohort of 21,268 children born during 1983-1991 in three rural Project sites obtained from the longitudinal Sample Registration System (SRS) of the MCH-FP Extension Project (Rural) of the International Centre for Diarrhoeal Disease Research, Bangladesh. The data suggest that there is a significant relationship between childhood immunization and reduced child mortality. Access to tubewell water was also associated with a reduced risk of mortality for young children. Baqui and others have reported on causes of death under age five based on verbal autopsy interviews in the 1993-1994 and 1996-1997 BDHS sample (Baqui et al., 1998; Baqui et al., 2001). The study in the BDHS 1993-1994 revealed that about one-quarter of deaths among children under five years were associated with acute respiratory infections (ARI) and about one-fifth of the deaths were associated with diarrhea (Baqui et al., 1998). Drowning was a major cause of death in children age 1-4 years. Neonatal tetanus and measles were the other important causes of death. The same verbal autopsy instrument and cause of death algorithms were used in the 1996-1997 BDHS. Comparison of the two surveys revealed that deaths due to almost all causes declined. The exceptions were deaths due to neonatal tetanus, diarrhea, and malnutrition (Baqui et al., 2001).

Becher et al. (2004) performed a survival analysis of births under demographic surveillance from a demographic surveillance system in 39 villages around Nouna, western Burkina Faso. All children born alive in the period January 1, 1993 to December 31, 1999 in the study area followed-up until December 31, 1999. Within the observation time, 1340 deaths were recorded. In a Cox regression model a simultaneous estimation of hazard rate ratios showed death of the mother and being a twin as the strongest risk factors for mortality. For both, the risk was most pronounced in infancy. Further factors associated with mortality include age of the mother, birth spacing, season of birth, village, ethnic group, and distance to the nearest health centre. Finally, there was an overall decrease in childhood mortality over the years 1993–99. Kembo and Ginneken (2009) address some important issues in infant and child mortality in Zimbabwe in their study. They found that births of order 6+ with a short preceding interval had the highest risk of infant mortality. The infant mortality risk associated with multiple births was 2.08 times higher relative to singleton births.

It is clear from the review of the literature above that the all of the Kaplan-Meier (K-M), Cox-PH and AFT approach of child mortality analysis are rarely done in Bangladesh where in other countries Cox PH approach of analysis was quite pronounced. In our study, we have used the BDHS-2007 data to analyze the under five child mortality. This study has important application from several aspects. Firstly, we have analyzed the child mortality considering several socioeconomic and maternal factors. Secondly, we have used non-parametric (K-M), semi-parametric (Cox-PH) and also parametric (AFT) approach so that from every perspective we can have an idea about the child mortality. And lastly, we have not only analyzed the child mortality in Bangladesh but also compared it between rural and urban area, which can give us a clue that in which respect under five child mortality differs between these two areas.

The paper is organized as follows: Section two discusses empirical methodology and data, while Section three presents empirical results. In Section four concluding remarks are provided.

2 DATA AND METHODOLOGY

This study is conducted using Bangladesh Demographic and Health Survey (BDHS)-2007 data, the fifth BDHS undertaken in Bangladesh. A two-stage sampling technique was conducted for this survey. We have collected our information about child mortality aged less than five years from the Women's questionnaire where the mother was asked to provide information about her children i.e., birth order of the child, its living status. According to the BDHS-2007 data, the number of children aged five years of less were 6241, out of which 4104 were from rural and 2137 were from urban area. From the total children, 366 were failed (5.9%) of which 260 were from rural and 106 were from urban area. As influential factors for child mortality we considered the variables: Sex (SEX), mother's age (MAGE), mother's education (MEDU), birth order (BORD) and economic status of the family (WEALTH) of each considered child.

At the first step, a univariate approach of survival analysis is done. For this purpose, Kaplan-Meier (K-M) (1958) or Product-Limit survival analysis which is a nonparametric estimate of the survivor function is used. K-M estimate can accommodate missing data such as censoring & truncation and estimates absolute risk. If $t_1 < t_2 < \cdots < t_M$ denote distinct times at which deaths occur, then the K-M estimate of survivor function is given by, $S(t) = \prod_{t_j \le t}(1 - d_j/n_j)$ where d_j is the number of deaths that occur at t_j and n_j is the number at risk (alive & under observation just before t_j).

Next the concentration is extended to multivariate method of survival analysis. Two types of regression models are used for this purpose: Cox Proportional Hazards (Cox-PH) (1972) model and accelerated failure time (AFT) models. The Cox-PH model is the most popular model describing the relationship between risk factors and survival time. This is a semi-parametric model of survival analysis and is given by,

$$h(t|x) = h_0(t)\exp(b_1 x_1 + \cdots + b_p x_p) \quad (1)$$

where x_i's are the risk factors and $h_0(t)$ is the baseline hazard. $\exp(b_i)$ is interpreted as a hazard ratio (or relative risk). PH assumption requires that $\exp(b_i)$ are constant across time, between groups.

Accelerated failure time (AFT) regression models are parametric approach of survival analysis. AFT model is given by the equation,

$$S(t|x) = S_0\{\exp(b_1 x_1 + \cdots + b_p x_p)t\} \quad (2)$$

where $\exp(b_i)$ is interpreted as a time ratio.

In this study, we have analyzed the under five child mortality both for the rural and urban area using the K-M, Cox-PH and AFT approach of survival analysis.

3 RESULTS AND DISCUSSION

Figures (1-10) represent the Kaplan-Meier plots. From the K-M plots we can see that female child mortality is higher than male in rural area where the opposite is true for urban area. For urban area the child whose birth order is four or more has a very high failure rate. For both the areas, survival probability for children whose mothers have higher education is very high compared to the children whose mothers have primary, secondary or no education. In rural area the failure rate is almost similar for children of all economic status but is very high for children of poor economic status in urban area.

Next we employed the Cox-PH analysis. For this purpose, we had to specify the appropriate model first, i.e., selecting covariates to go into the model. We employed the step wise selection for this purpose. In step wise selection, we first, add, one-by-one, best covariate that is excluded from model, secondly, exclude, one-by-one, the worst covariate that is in the model. We define a stopping rule as a condition for inclusion or exclusion of a variable. In our case the stopping rule is defined on p-value and AIC. Firstly, we define two thresholds: $p_E = 0.15$ is a threshold on the p-values for entering a term into the model and $p_R = 0.2 > p_E$ is a threshold for removing terms from the model. We will choose the model with lowest AIC. By this procedure, we choose the covariates for rural area are MAGE, MEDU and BORD and the covariates for urban area are BORD, WEALTH and MEDU.

Figure 1: K-M plot for sex of Child (rural)

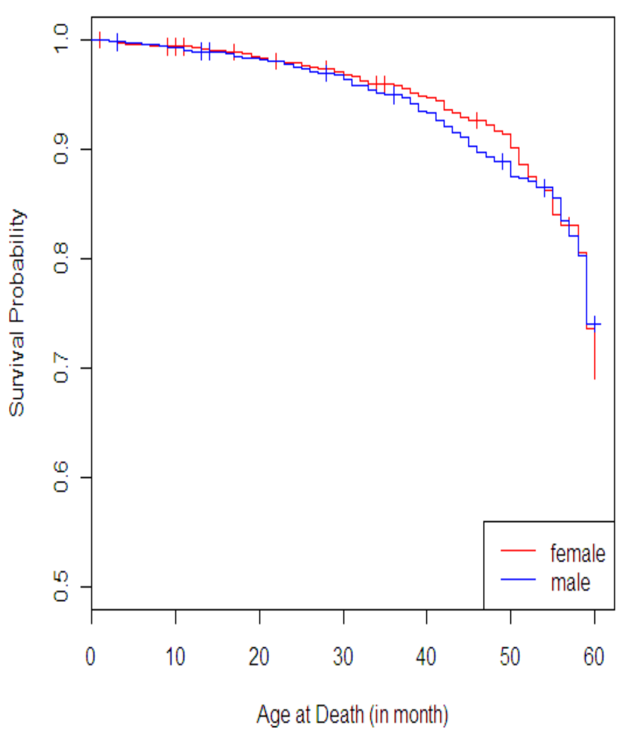


Figure 2: K-M plot for sex of Child (urban)

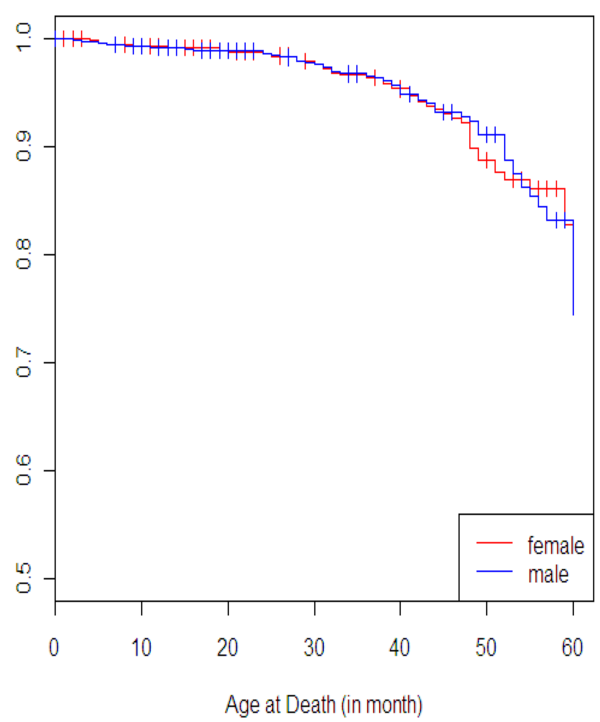


Figure 3: K-M plot for mother's age of Child (rural)

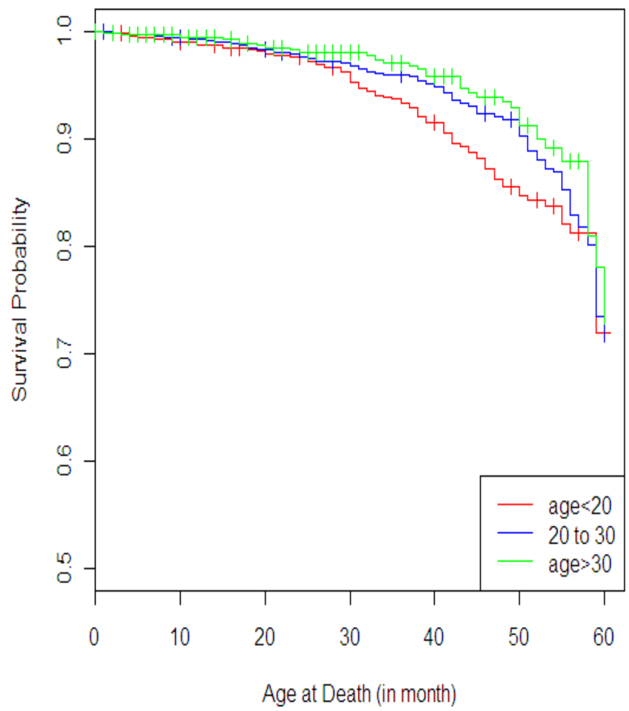


Figure 4: K-M plot for mother's age of Child (urban)

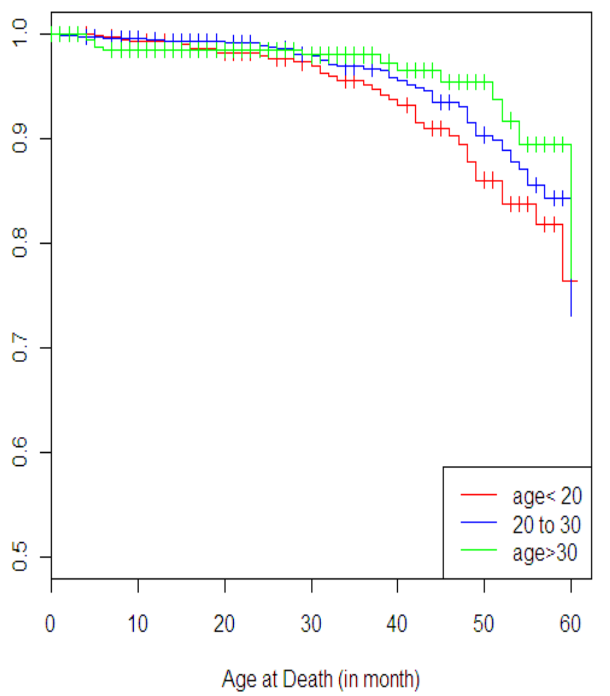


Figure 6: K-M plot for birth order (urban)

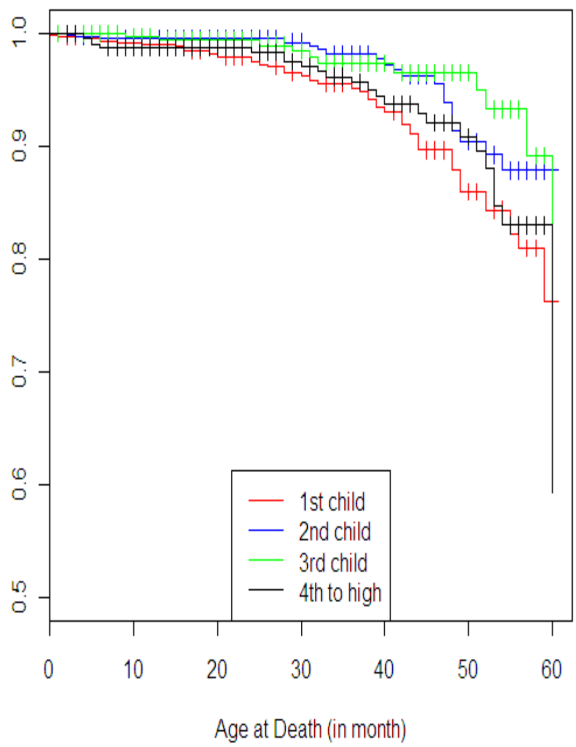


Figure 5: K-M plot for birth order (rural)

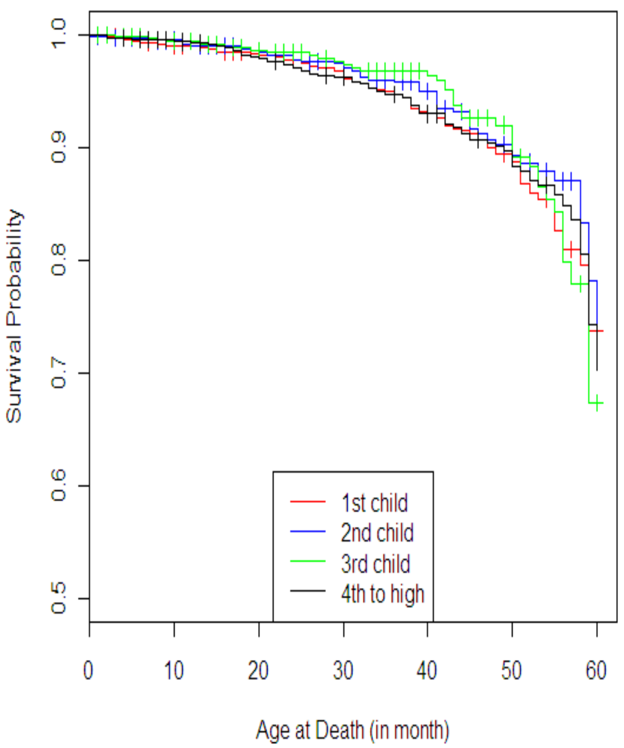


Figure 7: K-M plot for mother's education (rural)

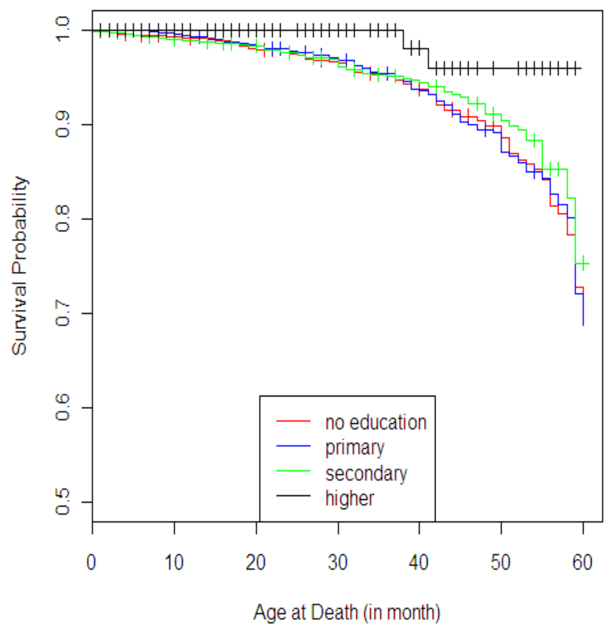


Figure 8: K-M plot for mother's education (urban)

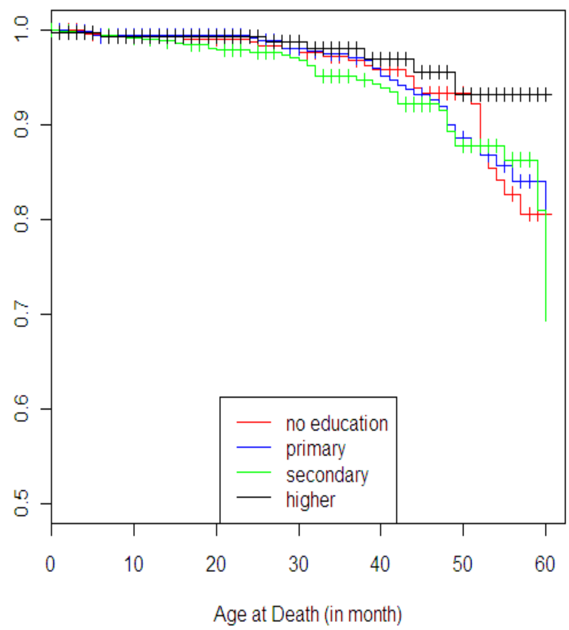


Figure 9: K-M plot for economic status (rural)

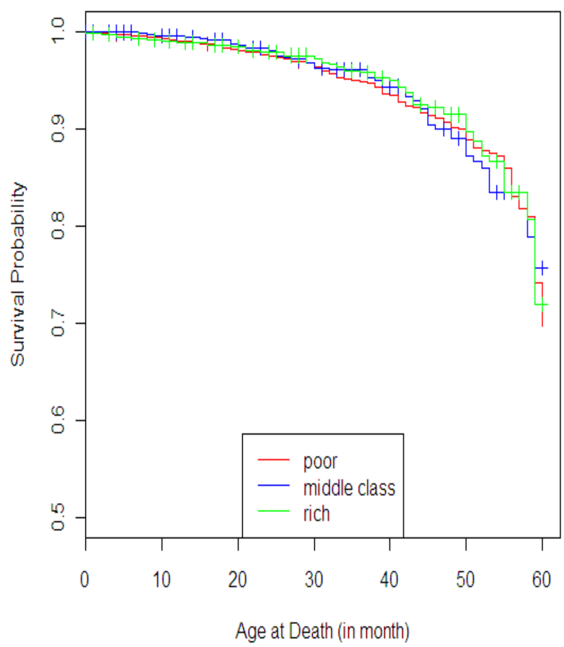
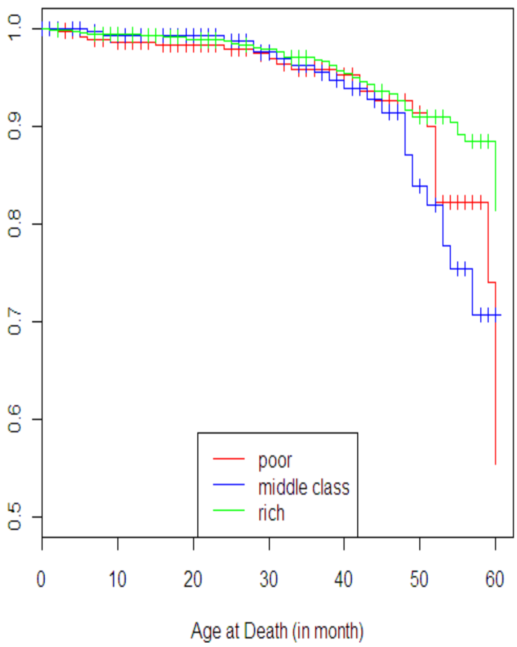


Figure 10: K-M plot for economic status (urban)



Next we were needed to check the proportionality assumption of the selected covariates. PH assumes that the estimates $b_1, b_2, ..., b_p$ do not vary much over time. Table 1 shows that all the variables both for rural and urban area satisfy the PH assumption. These results are further assessed by the Log-minus-log plots and Schoenfeld residuals plot (not presented here). In Log-minus-log plots, the curves of the categories for any predictor is compared after transforming the vertical axis by log (-log(S(t))) and plotted against log (time). If the curves of the different categories are parallel, the proportional hazard assumption is unlikely to be violated. However, when the categories for any predictor are more than two, these graphs are very difficult to assess. The Schoenfeld residual is defined as the covariate value for the individual that failed minus its expected value (yields residuals for each individual who failed, for each covariate). If the impact of an independent variable meets the proportional hazard assumption, the smoothed values of a quantity called scaled Schoenfeld residuals would be roughly horizontal when plotted against survival time. Since all the considered predictors are categorical we have created reference group for each categorical variables. In our analysis for mother's age (MAGE) the reference group is less than 20 years, for WEALTH it's poor, for mother's education (MEDU) it's no education, for birth order of the child its first child and for sex of the child it's female.

Concluding that all the considered variables both for rural and urban area satisfy the proportionality assumption we moved to Cox-PH analysis of the under five child mortality. Table 2 represents the result obtained from the Cox-PH analysis. Within the Cox model, the best interpretation of β for a categorical variable is the hazard ratio. Here, $\exp(\beta)$ is the hazard ratio for being in the considered group versus the reference group. The factor MAGE came significant for both the considered group (mother's age between 20 to 30 years and above 30 years)for child mortality compared to the reference group (mother's age less than 20 years) for rural area. That is,

the estimated hazard ratio (relative risk) of death of children whose mother's age are between 20 to 30 years relative to children whose mother's age are under 20 years is 0.64. In other words, children whose mother's age are in between 20 to 30 years have 36% lower hazard (risk) of death than those children whose mother's age are under 20 years. The other significant variables can also be interpreted in the same manner. BORD appeared to be a significant factor for child

Table 1 Checking Proportionality Assumption

		Rural			Urban		
		rho	chisq	p	rho	chisq	p
MAGE	20 to 30	0.09	0.70	0.40			
	>30	0.12	1.49	0.22			
WEALTH	middle				0.06	0.38	0.54
	rich				-0.10	1.18	0.28
MEDU	primary	0.03	0.21	0.65	-0.02	0.04	0.84
	secondary	-0.02	0.11	0.74	-0.08	0.69	0.41
	higher	-0.01	0.02	0.89	-0.06	0.35	0.55
BORD	2nd child	-0.03	0.27	0.60	0.07	0.48	0.49
	3rd child	0.02	0.15	0.70	0.07	0.53	0.47
	4th to high	-0.11	3.21	0.07	0.02	0.06	0.81

mortality both for the rural and urban area (though for different group) while WEALTH and MEDU is significant for urban and rural area respectively. The Likelihood ratio test (LRT), Wald test and Score test, test the global null hypothesis that $\beta = 0$. The global test is analogous to the overall F-test in an Analysis of Variance (ANOVA) or linear regression. It tests whether all of the covariates have no "influence" on survival time. Since the null hypothesis is rejected in all the three tests, we can say that at least one of the covariates has influence on survival time of under five children both in the rural and urban area and our previous findings are also supported.

Table 2 Cox Proportional Hazard analysis of Child Mortality

		Rural		Urban	
		coef	exp(coef)	Coef	exp(coef)
MAGE	20 to 30	-0.45***	0.64		
	>30	-0.94***	0.39		
WEALTH	middle			0.10	1.11
	rich			-0.47*	0.63
MEDU	primary	-0.06	0.94	0.10	1.11
	secondary	-0.22	0.80	0.32	1.38
	higher	-1.41**	0.24	-0.32	0.72
BORD	2nd child	-0.07	0.93	-0.66***	0.52
	3rd child	0.14	1.15	-0.84***	0.43
	4th to high	0.39*	1.48	-0.17	0.84
Likelihood Ratio Test	on 8 DF	23.37***		21.58***	
Wald Test	on 8 DF	20.29***		20.52***	
Score (logrank) Test	on 8 DF	21.16 ***		21.31***	

Note: ***, ** and * denote significance level at 1%, 5% and 10% respectively.

We move to the parametric analysis of child mortality next. We used the AFT models for this purpose considering four distributions: Weibull, Exponential, Log-Logistic and Log-Normal. AFT models assume a linear relationship between log of completed (latent) survival time t and

Table 3 AFT analysis of child mortality

		Rural				Urban			
		Weibull	Exp	Log logistic	Log normal	Weibull	Exp	Log logistic	Log normal
Intercept		4.71	5.82	4.66	5.15	**4.77*** **	5.94***	4.72***	5.33***
SEX	male	-0.03	-0.06	-0.03	-0.04	-0.03	-0.06	-0.02	-0.02
MAGE	20 to 30	0.20	0.34	0.20	0.24	-0.01	-0.08	-0.01	-0.07
	>30	**0.43***	0.79	0.45*	0.55	0.20	0.34	0.21	0.08
MEDU	primary	0.05	0.14	0.05	0.10	-0.04	-0.04	-0.04	-0.05
	secondary	0.14	0.40	0.14	0.17	-0.13	-0.16	-0.13	-0.20
	higher	0.74	1.70	0.73	0.94	0.18	0.54	0.17	0.17
BORD	1st child	0.04	0.14	0.04	0.05	0.31	0.71	0.32	0.47
	2nd child	-0.50	-0.03	-0.04	-0.00	0.37	0.84	0.38	0.57
	4th to high	-0.17	-0.27	-0.18	-0.21	-0.003	0.08	-0.01	0.08
WEALTH	middle	-0.01	-0.01	-0.01	0.03	-0.05	-0.12	-0.05	0.02
	rich	-0.07	-0.17	-0.07	-0.10	0.18	0.36	0.18	0.28
Scale		0.467	1.00	0.457	1.2	0.469	1.00	0.461	1.27
Log (scale)		**-0.76** ***		-0.78 ***	0.18 **	**-0.76** ***		-0.77 ***	0.24
LogL		-1773.9	-1855.6	-1778	-1812.9	-740.5	-773.6	-742.3	-759.4
LogL (Intercept)		-1786.1	-1867.5	-1790.3	-1824.9	-751.7	-785	-753.3	-769.5
Chisq	on 11 d.f	**24.49** ***	23.9 ***	24.49 ***	23.92 ***	**22.45****	22.69**	21.9***	20.25**
AIC		**3573.8**	3735.1	3582.0	3651.8	**1506.99**	1571.22	1510.66	1544.78

Note: ***, ** and * denote significance level at 1%, 5% and 10% respectively.

covariate x. Now, both for rural and urban area the AIC value is smallest for Weibull distribution indicating Weibull distribution better fits the data than the other distributions. In rural area, only the variable MAGE is significant for greater than 30 years, meaning that for one unit (month) increase in the children's age, the expected survival time increases by $\exp(.43) = 1.54$ or 54% more for children whose mother's age are more than 30 years than children with mother aged less than 20 years. In urban area only the intercept term is significant. The term scale is a time scaling factor, it's greater than 1 means failure is accelerated (survival time shortened) and vice versa. The Log(scale) is statistically significant relative to 0 and scale is smaller than 1 for Weibull distribution in both areas , indicating failure is decelerated (survival time lengthened).

4 CONCLUSIONS

Using non parametric, semi parametric and parametric approach of survival analysis, this paper investigates the factors that affect the child mortality of children aged under five years in Bangladesh and also compares the child mortality between rural and urban area. The analysis has important implications for the government and non-government organizations and policy makers of the country who deal with child affair and health. The non parametric analysis suggests that in urban area the 4th or higher birth ordered child has a very high failure rate. For both the areas, survival probability is very high for children with higher educated mother and in urban area the failure rate is very high for children of poor economic status. The Cox-PH regression analysis indicate that in rural area the covariates MAGE, MEDU and BORD have significant affect on child mortality while the significant covariates for urban area are WEALTH and BORD. The AFT analysis shows that for both the areas Weibull distribution better fits the data and only the covariate MAGE is significant for rural area.

REFERENCES

1. Becher, H., Muller, O., Jahn, A., Gbangou, A., Kynast-Wolf, G. and Kouyate, B., **"Risk factors of infant and child mortality in rural Burkina Faso"**, *Bulletin of the World Health Organization*, 82 (4), 2004.
2. Kembo, J. and Ginneken, J.K.V., **"Determinants of infant and child mortality in Zimbabwe: Results of multivariate hazard analysis"**, *Demographic Research,* 21, 2009, PP 367-384.
3. Kaplan, E.L. & Meier, P., **"Nonparametric estimation from incomplete observations",** *Journal of the American Statistical Association,* 53, 1958, 457-481.
4. Cox, D., **"Regression models and life tables"**, *Journal of the Royal Statistical Society, Series B*, 34, 1972, 187-220.
5. Mozumder, ABM. K. A., Khuda, B., Kane, T. T., **"Determinants of Infant and Child Mortality in Rural Bangladesh"**, International Centre for Diarrhoeal Disease Research, Bangladesh, Working Paper No. 115, 1998.
6. Baqui, A.H., Black, R.E., Arifeen, S. E., Hill, K., Mitra, S.N. and Sabir, A. A., **"Causes of childhood deaths in Bangladesh: results of a nationwide verbal autopsy study"**, *Bulletin of the World Health Organization*, 76 (2), 1998, 161-171.
7. Baqui, A.H., A.A. Sabir, N. Begum, S.E. Arifeen, S.N. Mitra, R.E. Black., **Causes of childhood deaths in Bangladesh: An update**. *Acta Paediatr,* 90, 2001, 682-690.
8. Bangladesh Demographic and Health Survey. **National Institute of Population Research and Training (NIPORT), Mitra and Associates, ORC Macro**. Bangladesh Demographic and Health Survey 2007. Dhaka, Bangladesh and Calverton, Maryland: NIPORT, Mitra and Associates and ORC Macro, 2007.

---

[1]Arfan Raheen AFZAL - Post graduate in Applied Statistics from Institute of Statistical Research and Training (ISRT), University of Dhaka. Currently is a doctoral student in University of Calgary, Alberta, Canada. Afzal is also a GAT in UofC. Previously he worked in the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B).
Scientific Research Interests: Survival Analysis, Biostatistics, Epidemiology.
Publications:
1."An Empirical Analysis of the Relationship between Macroeconomic Variables and Stock Prices in Bangladesh", Bangladesh Development Studies, Vol. XXXIV, December 2011, No. 4, PP 95-105.
2. "Determinants and Status of Vaccination in Bangladesh", Dhaka Univ. J. Sci. 60(1): 47-51, 2012 (January), PP 47-51.

[2]Sabrina ALAM - Post graduate in Applied Statistics from Institute of Statistical Research and Training (ISRT), University of Dhaka. Worked in the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B).
Scientific Research Interests: Survival Analysis, Time Series Analysis, Econometrics.