Survival analysis for under-five child mortality in Uttar Pradesh

Rakesh K. Saroj[1*] & K.N Murty[1] & Mukesh Kumar[2]
[1]Department of Kayachikitsa, IMS, Banaras Hindu University Varanasi-221005
[2]Department of Statistics, MMV, Banaras Hindu University Varanasi-221005

Received: May 21, 2018 Accepted: June 27, 2018

ABSTRACT *The effects of socio-economic and demographic variables play significant role in infant mortality in less developed states in India. The determinants of infant mortality in developing countries can be classified under two major headings: socioeconomic and demographic. In this study we have attempted to find out the impact of socioeconomic, demographic, environmental, health related and nutritional factors in under- five mortality of child. In this attempt, we first analyzed our data using Kaplan-Meier non-parametric method of estimation of survival function. Also we had used Cox proportional hazard model of under-five mortality of child for different with three mainly category covariates. The factors affecting under-five child survival from the data set was selected based on the paper by Mosley. These factors were categorized into the following four categories; social demographic and social economic, environmental and proximate or biological factors but we took three categories after including proximate or biological factors together.*

Keywords: *Cox proportional hazard; Under- five; Child mortality, KM-Curve, Proximate factor*

Introduction: Reducing under-five mortality is now a global concern. In 2001 as part of the Millennium Development goals (MDG) for health, nations pledged to ensure a two-thirds reduction in under-five mortality between 1990 and 2015(1) and at once a series of articles in Lancet by the Bellagio Study Group described various aspects of child survival (2, 3, 4, 5, 6).Although under-five mortality is declining worldwide as a result of socioeconomic development and implementation of child survival interventions, nearly 8.8 million children die every year before their fifth birthday. The National Family Health Survey (NFHS) is a large-scale, multi-round survey conducted in a representative sample of households throughout India. In 2014-2015, India will implement the fourth National Family Health Survey (NFHS-4). Like its predecessors, NFHS-4 will be conducted under the stewardship of the Ministry of Health and Family Welfare, coordinated by the International Institute for Population Sciences, Mumbai, and implemented by a group of survey organizations and Population Research Centre's, following a rigorous selection procedure. Technical assistance for NFHS-4 will again be provided by ICF International, USA with the major financial support from the United States Agency for International Development and Ministry of Health and Family Welfare, Government of India. In addition to the 29 states, NFHS-4 will also include all six union territories for the first time and will also provide estimates of most indicators at the district level for all 640 districts in the country as per the 2011 census. The infant mortality rate in Uttar Pradesh in NFHS-4 is estimated at 64 deaths before the age of one year per 1,000 live births, The under-five mortality rate for Uttar Pradesh is 78 deaths per 1,000 live births, Boys have a higher mortality rate than girl during the neonatal period (in the first month of life). In comparison, girls have higher mortality rates than boys during the postnatal period and between ages one and five years. Children born less than two years after a previous birth are much more likely to die in infancy than children whose mothers waited two or more years between births. First births and those of birth order four or higher are more likely to die during infancy than children of birth order two or three. All the early childhood mortality rates are considerably higher in rural areas than in urban areas (7). The survival analysis is a method to measure the patient's survival that after treatment in any specifics disease. Survival and reliability is some manner is same, reliability use for machine working period or machine reliable in period, while survival analysis use in check the survival status of patients. In the survival analysis; we estimate the Survivor function, or the Hazard function. The hazard is the instantaneous event (death) rate at a particular time point t. In survival analysis doesn't assume the hazard is constant over time. The total hazard is the cumulative hazard experienced up to time t. The survival function is probability an individual survives up to and including time t. It's the probability that the event (e.g., death) hasn't occurred yet. The Kaplan-Meier curve estimates the survival curve of patients representing the proportion of patients who survived over time (8). Survival analysis refers to analyzing a set of data in defined time duration before event occurs. A survival analysis is different from traditional model like regression and classification problems as it models two different parameters. There are various types of distribution of Survival/hazard functions which are frequently assumed while doing a survival analysis; these distributions come from the type of probability distribution of the failure function. There are

the two plots we will refer in each case and these are the important ones to select the distribution which are hazard function and survival function.

1. **Lifetime Distribution Function (F)** : This is the probability of failure happening before a time 'T'.
$$F(t) = \Pr(T \leq t)$$
2. **Lifetime Probability distribution (f)** : A differential of F will give us probability distribution. All the names of distribution function is based on this probability distribution.
$$f(t) = F'(t) = \frac{d}{dt}F(t)$$
3. **Survival Function (S)** : Survival is the inverse of Lifetime. It is one minus Lifetime distribution.
$$S(t) = \Pr(T > t) = \int_t^{\infty} f(u)du = 1 - F(t)$$
4. **Hazard Function (Lambda)** : Hazard function is the rate of event happening. Hazard function can be derived from the Survival function as follows :
$$\lambda(t) = \lim_{dt \to 0} \frac{\Pr(t \leq T < t + dt)}{dt \cdot S(t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}.$$
5. **Cumulative Hazard Function** : This is simply the integral of the hazard function and is given as below :
$$\Lambda(t) = \int_0^t \lambda(u)\, du$$
6. Also, by integrating the hazard function equation we get following equation :
$$\Lambda(t) = -\log S(t)$$

The Cox proportional hazards model: Proportional hazard model is a class of survival models that assesses the relationship between one or more covariates with time. One of the advantages of this model is it does not require strong assumptions on the distribution of data (9). This model consists of two parts: the underlying hazard function, often denoted as $h_0(t)$, describing how the risk of the event per time unit changes over time at baseline levels of covariates; and the effect parameters describing how the hazard varies in response to explanatory covariates (X is a vector of explanatory covariates and β is a vector of unknown regression parameters). The hazard function in Cox's regression model is given by (10).The Cox PH model (for p independent variables $X1, ..., Xp$ is described by the equation:

$$h(t) / h_0(t) = e^{(B_1X_1 + B_2X_2 + ... + B_pX_p)}$$

where $h(t)$ is the hazard function, $h_0(t)$ is called the baseline hazard function (the expected hazard without any effect of the considered factors), e is a base of the natural logarithm,$B_1, ..., Bp$ regression coefficients. The expression $h(t)/h_0(t)$ is called the hazard ratio (HR) and indicates a growth or decrease of hazard caused by an effect of factors $X_1, ..., Xp$.

Studies on determinants of child mortality have mainly used either logistic regression or Cox proportional hazards model assuming that the outcomes are independent. To find more accurate estimates for the determinants of child mortality that has critical implications for resource allocation for improving child survival, sibling structures in child mortality data from demographic surveys have been treated as multivariate failure time data (11, 12, 13, 14). As far as UP is concerned, poor health delivery system, poor maternal and health care services are responsible for low infant and child mortality. Infant mortality rates are very high for some districts of the state. This is a matter of serious concern for the human as well as social development of the state. The study found the Infant Mortality and its Determinates in Uttar Pradesh, India through on logistic regression methods.(15).Another study done on maternal health problem of Uttar Pradesh (16).These studies was restricted to the analysis of mortality risks in children at survival analysis. After review found that few study done on under-five mortality survival status in Uttar Pradesh but none study found on that parametric survival model fitting in the under-five mortality data especially in Uttar Pradesh.

Methodology's: Secondary data has been taken from National Family Health Survey-4 of Uttar Pradesh children data sets. For analysts, the age of the child in months was calculated as follows: age = V008 – B3, where V008 is the century month code (CMC) of the date of interview, and B3 is the CMC for date of birth of the child. DHS calculated age for children by subtracting the month and year of birth from the month and year of interview to give age in month. Total 41751 sample size collected from 71 districts of Uttar Pradesh from January 2015 to August 2016. The factors affecting under-five child survival from the data set was

selected based on the paper by Mosley et al. (17). These factors were categorized into the following four categories; social demographic and social economic, environmental and proximate or biological factors but we took three categories after including proximate or biological factors together. First use the KM-curve for basic analysis and after that we use the Cox-regression analysis for the finding the important factors to play role in child survival status. We have been use the computation through R 3.2 statistical and SPSS software.

Result: The table1 shows status of child survival status of under-five year of Uttar Pradesh. In this table shows that total 93% cases censored as per data because we choose the child death is our event in this study. Total survival of children was 59 months and same which pasted to days (Figure 1). The plot shows that the survival probability and censor cases report of child survival and censored observations are indicated by a + symbol in the graph. In the table 1.1 calculate the Socio-economic and demographic factors to play important role in child survival status through Cox-regression with the help of Statistical package for social science (SPSS). In this analysis include the various factors and find the variable which is playing the significant role in child survival status. In this analysis found that educational level, women's age in years and religion play significant role in the child survival status. In the table 1.2 calculate the environment factors to play important role in child survival status. In this analysis none variable found significant role in the child survival status. In the table 1.3 calculate the proximate and biological factor to play important role in child survival status. In this analysis none variable found that total children ever born, births in last five years, number of living children, currently breastfeeding, smokes cigarettes, desire for more children, size of child at birth, delivery by caesarean section, ANC visits and birth order play significant role in the child survival status. In the table 1.4 calculate the all combine together factor which was found the significant in the individual level in child survival status. This combine Cox regression analysis found that women's age in years, total children ever born, births in last five years, number of living children, currently breastfeeding, smokes cigarettes, desire for more children, size of child at birth, delivery by caesarean section, ANC visits and birth order play significant role in the child survival status.

Conclusions: In this study, we tried to explore and show which factors are associated with the risk of under-five mortality of child in Uttar Pradesh. Using the Kaplan Meier non-parametric and Cox proportional hazard model the difference of child survival among different covariates that significantly influence the survival of child are identified. We first analysis socio-demographic, environment, proximate and biological level factors one by one with the help of Cox-regression and finally took the significant variables from the each model and make combine variable and again apply the Cox-regression for finding the important variable for under-five mortality of child. Cox regression model showed that women's age in years, total children ever born, births in last five years, number of living children, currently breastfeeding, smokes cigarettes, desire for more children, size of child at birth, delivery by caesarean section, ANC visits and birth order are found statistically significant effect on child survival in Uttar Pradesh.

Table 1 Survival status of under-five year of Uttar Pradesh

Case Processing Summary			
		Censored	
Total N	N of Events	N	Percent
41751	2830	38921	93.2%

Table 1.1 Cox-regression analysis Socio-economic and demographic factors of the child survival status

Socio-economic and Demographic factors						
S.No.		B	Sig.	Exp(B)	95.0% CI for Exp(B)	
					Lower	Upper
1.	Educational level	-.173	**.003**	.841	.751	.943
2.	Sex of household head	-.086	.577	.917	.677	1.242
3.	Women's age in years	-.041	**.000**	.959	.939	.981
4.	Current marital status	.206	.100	1.228	.962	1.569
5.	Husband/partner's education level	-.082	.170	.921	.819	1.036
6.	Husband/partner's occupation	.005	.244	1.005	.997	1.012
7.	Respondent currently working	.158	.525	1.171	.720	1.905
8.	Respondent's occupation	-.006	.091	.994	.988	1.001
9.	District	-.002	.301	.998	.993	1.002

10.	Religion	-.310	**.025**	.734	.560	.962
11.	Caste	.081	.108	1.084	.982	1.197

Table 1.2: Cox-regression analysis Environment Factors of the child survival status

Environment Factors						
					95.0% CI for Exp(B)	
S.No.		B	Sig.	Exp(B)	Lower	Upper
1.	Place of residence	-.141	.607	.868	.506	1.489
2.	Wealth index	-.072	.474	.931	.764	1.134
3.	Source of drinking water	.014	.227	1.014	.992	1.036
4.	Slum designation by observation	-.124	.682	.883	.487	1.600
5.	Type of toilet facility	-.016	.148	.984	.963	1.006

Table 1.3 Cox-regression analysis The Proximate and Biological factor of the child survival status

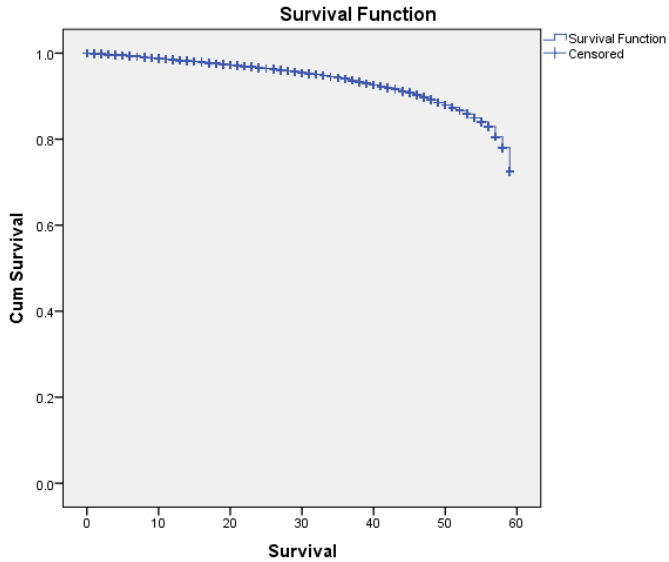
S.No.		B	Sig.	Exp(B)	95.0% CI for Exp(B)	
					Lower	Upper
1.	Total children ever born	0.295	**0.00**	1.343	1.287	1.403
2.	Births in last five years	0.645	**0.00**	1.905	1.796	2.021
3.	Number of living children	-0.747	**0.00**	0.474	0.454	0.494
4.	Currently breastfeeding	-0.622	**0.00**	0.537	0.483	0.596
5.	Anemia level	-0.045	0.102	0.956	0.905	1.009
6.	Smokes cigarettes	2.064	**0.00**	7.881	2.53	24.545
7.	Chewing tobacco	0.167	0.684	1.181	0.53	2.635
8.	Desire for more children	-0.155	**0.00**	0.857	0.833	0.882
9.	Sex of child	0.016	0.737	1.016	0.927	1.113
10.	Size of child at birth	0.151	**0.00**	1.162	1.12	1.207
11.	Birth weight (kg)	0	0.055	1	1	1
12.	Delivery by caesarean section	0.511	**0.00**	1.667	1.395	1.993
13.	ANC Visits	-0.307	**0.001**	0.735	0.61	0.887
14.	Birth Order	0.834	**0.00**	2.302	1.996	2.656
15.	Media Exposure	-0.033	0.51	0.968	0.879	1.066
16.	Birth Interval	0.072	0.153	1.075	0.974	1.187

Table 1.4: Cox-regression analysis of combine factor of the child survival status

S.No.	Covariates	B	Sig.	HR	95.0% CI for Exp(B)	
					Lower	Upper
1.	Women's Age in Years	-.043	**.000**	.958	.947	.969
2.	Education Level	-.020	.336	.980	.941	1.021
3.	Religion	-.009	.845	.991	.903	1.087
4.	Total children ever born	.493	**.000**	1.637	1.571	1.705
5.	Births in last five years	.324	**.000**	1.383	1.313	1.456
6.	Number of living children	-1.000	**0.00**	.368	.354	.383
7.	Currently breastfeeding	-.415	**.000**	.660	.606	.719
8.	Smokes cigarettes	1.979	**.001**	7.238	2.327	22.511
9.	Desire for more children	-.164	**.000**	.849	.830	.868
10.	Size of child at birth	.167	**.000**	1.182	1.146	1.219

11.	Delivery by caesarean section	.143	**.041**	1.154	1.006	1.323
12.	ANC Visits	-.394	**.000**	.674	.580	.784
13.	Birth Order	.583	**.000**	1.792	1.642	1.956

Figure 1: Child survival status of under-five year



References:

1. UN General Assembly, 56th session. Road Map towards the implementation of the United Nations millennium declaration: report of the Secretary-General (UN document no. A/56/326) New York: United Nations; 2001.
2. Black RE, Morris SS, Bryce J. Where and why are 10 million children dying every year? Lancet. 2003;361(9376):2226–2237. [PubMed]
3. Jones G, Steketee RW, Black RE, Bhutta Z, Morris SS, The Bellagio Child Survival Study Group How many child deaths can we prevent this year? Lancet. 2003;362(9377):65–71. [PubMed]
4. Bryce J, el Arifeen S, Pariyo G, Lanata CF, Gwatkin D, Habicht JP, The Multi-Country Evaluation of IMCI Study Group Reducing child mortality: can public health deliver? Lancet. 2003;362(9378):159–164. [PubMed]
5. Victora CG, Wagstaff A, Schellenberg JA, Gwatkin D, Claeson M, Habicht JP. Applying an equity lens to child health and mortality: more of the same is not enough. Lancet. 2003;362(9379):233–241. [PubMed]
6. The Bellagio Study Group on Child Survival. Knowledge into action for child Survival. Lancet. 2003;362(9380):323–327. [PubMed]
7. https://ruralindiaonline.org/resources/national-family-health-survey-nfhs-4-2015-16-uttar-pradesh/
8. Kaplan, E. L., and Meier, P. (1958). Non-parametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53(282), 457 – 481.
9. Cox, D.R. (1972) Regression Models and Life Tables (with Discussion). Journal of the Royal Statistical Society, 34, 187-220.
10. Kleinbaum D, Klein M. Survival analysis: a self-learning text. New York: Springer; 2005.
11. Guo G. Use of sibling data to estimate family mortality effects in Guatemala. Demography. 1993;30(1):15–32. [PubMed]
12. Guo G, Rodríguez G. Estimating a multivariate proportional hazards model for clustered data using the EM algorithm with an application to child survival in Guatemala. Journal of the American Statistical Association. 1992;87(420):969–976.
13. Sastry N. A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. Journal of the American Statistical Association. 1997;92(438):426–435. [PubMed]
14. Hung Wen-Shai, Shu-Hsi Ho. Survival analysis for unobserved heterogeneity on estimated mortality in Taiwan. Economics Bulletin. 2008;9(25):1–10.
15. B. P. Singh, S. Maheshwari: Infant Mortality and its Determinates in Uttar Pradesh, India,J. Stat. Appl. Pro. Lett. 1, No. 3, 47-52 (2014.
16. R.K.Saroj,A.Kumar and M.Kumar, Uttar Pradesh statistical approach for finding the determinants of maternal health services in Uttar Pradesh,IJOAR, Volume 4, Issue 8, August 2016
17. Mosley W, Chen L. An analytical framework for the study of child survival in developing countries. Population and Development Review. 1984; 10(Suppl.): 24-45.