American Academic Scientific Research Journal for Engineering, Technology, and Sciences
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402
https://asrjetsjournal.org/index.php/American_Scientific_Journal/index

Model Selection and Inference in Competing Risk Regression Model to Determine the Potential factors of under-5 Child Mortality in Bangladesh

A H M Musfiqur Rahman Nabeen[a], Kakoli Khatun[b] , Mosammat Sonia Khatun[c]*

[a,c] *Department of Statistics, Pabna University of Science and Technology, Rajapur, Pabna, Bangladesh, 6600*
[b] *Department of Statistics, Gopalgonj Science and Technology University, Gopalgonj, Bangladesh, 8100*
[a]*Email: musfiqnabeen@gmail.com*
[b]*Email: kakolikh@buffalo.edu*
[c]*Email: soniakhatun.du.29@gmail.com*

Abstract

Under-5 child mortality is always a critical term for the developing countries like Bangladesh. The primary objective of this study is to investigate under-5 child mortality using Fine and Gray (1999) competing risk regression. The secondary objective is to decide among many other covariates which covariates should be included in Fine and Gray model and which should not. For this purpose, data are extracted from Bangladesh Demographic and Health Survey (BDHS) 2011, where the event of interest, under-5 child mortality may occur due to any of the three causes: Disease, Non-disease and Other. It is found that for "Disease Cause" 4 covariates (wealth index, size of child at birth, gender of child, availability of maternal and child welfare center) are selected. For "Non-disease Cause" 6 covariates (mother's education, place of delivery, size of child at birth, NGO membership of mother, gender of child, birth order number) are selected and for "Other Cause" 6 covariates(mother's education, availability of MCWC, NGO membership of mother, father's education, birth order number, main access road to village) are selected. Finally, for the selected covariates, the Fine and Gary competing risk regression models are fitted to identify the potential factors of under-5 child mortality due to the three different causes of deaths. The identified factors may help to take decision by the health policy makers to increase under-5 child survival in Bangladesh.

Keywords: Competing risk; Fine and Gray model; Under-five mortality; Model selection; Forward and Backward method; AIC.

Received: 6/15/2025
Accepted: 8/15/2025
Published: 8/25/2025

---

* *Corresponding author.*

I. Introduction

Children are vital for a nation's present and future advancement and ensuring their healthy growth and development should be the primary concern of any society. The level of socioeconomic development of a country is revealed by its under-5 mortality rate. As a developing country in Bangladesh under-5 child mortality is always an important phenomenon. Thus, many studies have already been conducted and many are ongoing to identify the causes and determinants of under-5 child mortality in Bangladesh. However most of studies are in non-competing setup (considering one cause of death at a time) and they did not consider competing risk circumstance [1,2,3,4]. In the analysis of survival data, the circumstance of competing risk may occur when an individual experiences an event due to one and only one type of cause, out of more than one types of causes. In practice, under-5 child mortality may occur out of different types of causes. That is, under-5 child mortality is a competing risk phenomenon (considering more than one causes of death at a time).So, it is needed to fit competing risk regression. This is the central motivation behind this article. There are mainly two approaches to fit competing risk regression: Cox proportional hazard (PH) model approach and Fine and Gray model approach [5,6]. In competing Cox PH model, the events occurred due to other than the cause of interest are treated as censored. There is no direct interpretation of cause specific hazard function here. Fine and Gray (1999) proposed an improved way for competing risk regression through Cumulative Incidence Function (CIF), which does not need to treat the events occurred due to other than the cause of interest as censored. Moreover, it has direct interpretation in terms of survival probability [6]. Thus Fine and Gray competing risk regression model is used in this study to find the causes and determinants of under-5 child mortality. But before going to fit this model it is to be decided among many other covariates which should be included in the model and which should not. Primarily, on the basis of literature review [1,2,4,7,8,9] 17 covariates are considered, which are : mother's age at 1st birth, region, aware of community clinic, mother's education level, wealth index, birth order number, gender of child, type of place of residence, place of delivery, sex of the household head, exposure to media, NGO membership of mother, size of child at birth, main access road to village, availability of Maternal and Child Welfare Center (MCWC) in district, women empowerment and father's education level. Among the above 17 covariates, it is needed to decide which covariates should be included in the final competing risk regression model. For this purpose, Bangladesh Demographic and Health Survey-11 [10] is used. To know what are the causes of under-5 death, Verbal Autopsy (VA) Dataset is needed, where the description of the causes of death is given. Though in Bangladesh BDHS-2014 is available, unfortunately there is no VA questionnaire. But, BDHS-2011 includes VA questionnaire and there are 17 causes of death. From which, by recoding 3 categories: Disease, Non-disease and Other causes are obtained.

II. Data and Mathodology

A. *Data and Variables*

Secondary data extracted from nationally representative Bangladesh Demographic and Health Survey (BDHS), 2011 is used for this study which was conducted by National Institute of Population Research and Training (NIPORT), Mitra and Associates of Dhaka. For conducting the survey, Bangladesh was stratified into 20 strata where each stratum is made up of enumeration areas (EAs). It followed a two stage stratified sampling procedure. At the first stage, the survey randomly selected 600 EAs (among them 207 in urban areas and 393 in rural areas)

with probability proportional to the EA size. In second stage of sampling, a systematic sample of 30 households was selected on average from each enumeration area. Finally, from the selected households information of the interviews of 17842 ever-married women aged 12 to 49 years was collected. BDHS-2011 includes their socioeconomic, biodemographic, and maternal and child health related information. It has Verbal Autopsy data (to know the cause of death) and Community data. Both are used in this study.

In BDHS-2011 data, there are 17842 observations and 4007 variables are collected. Moreover, in Verbal Autopsy dataset, there are 490 observations. One of the main purposes of this study is to determine important risk factors that have a substantial impact on the child mortality through a competing risk setup. To achieve this purpose on the basis of previous studies [1,2,4,7,8,9] 17 covariates are selected: mother's age at 1st birth (Age below 20, Age 20-30, Age above 30), region (Barisal, Chittagong, Dhaka, Khulna, Rajshahi, Rangpur and Sylhet), aware of community clinic (Yes, No), mother's education level (No education, Primary, Secondary, Higher), wealth index (Poor, Middle, Rich), birth order number (1st birth, 2nd-3rd birth, Others), gender of child (Male, Female), type of place of residence (Rural, Urban), place of delivery (Home, Hospital/Others), sex of the household head (Male, Female), exposure to media (Yes, No), NGO membership of mother (Yes, No), size of child at birth (Small, Average/large), main access road to village (All weather road, Seasoned/Other), availability of MCWC in district (Yes, No), women empowerment (Yes, No) and father's education level (No education, Primary, Secondary, Higher). Note that information on exposure to media, NGO membership of mother, Availability of MCWC in the district and Women Empowerment are not directly collected in BDHS, 2011 survey.The event of interest is under-5 child mortality. If a child dies before reaching his/her fifth birthday then under-5 mortality occurs. Last five births of women are considered for this study. Information from 8588 children was collected who were born preceding 5 years of the survey. There are 400 events (under-5 deaths) among 8588 observations. So, child mortality rate becomes 46.58 per thousand children. There are 17 causes of death in the Verbal Autopsy dataset, in BDHS-2011. Out of those 17 causes, by recoding 3 causes of death are obtained.

Table 1

Category	Causes of death
Disease (**194**)	Neonatal Tetanus, Measles, Diarrhoea, Meningitis, Neonatal Jaundice, Pneumonia, Respiratory Distress, Serious Infection
Non-disease (**108**)	Congenital Abnormality, Drowning, Birth Asphyxia, Birth Injury, Premature Birth, Malnutrition
Other cause (**98**)	Other, Undetermined, Unspecific categories
Total = **400**	

B. Fine and Gray Model

For competing risks data, Fine and Gray (1999) proposed a model, based on cumulative incidence function [6]. This model has obtained popularity for its convenience in directly assessing the effect of covariates on the

cumulative incidence function (CIF) or sub-distribution function. Fine and Gray proposed to fit the sub-distribution hazard with the following model:

$$\gamma_j(t|x) = \gamma_{0j}(t)e^{\beta_j' x}; j = 1,2,3, \dots, p$$

Where $\gamma_{0j}(t)$ is the baseline sub-distribution hazard due to j^{th} cause, which is left as an arbitrary function. Moreover, $\gamma_j(t|x)$ is the sub-distribution hazard function in the presence of covariates for the j^{th} cause and $e^{\beta_j' x}$ is the relative risk associated with covariates x.

Cumulative incidence function (CIF)

The sub-distribution or cumulative incidence function (CIF) for the event due to j^{th} risk factor, is defined as the joint probability $\tilde{F_j}(t) = P(T \leq t, C = j)$.That is, cumulative incidence function or CIF is the probability that an event occurs because of risk j at or before time t.

C. Variable Selection Methods and Criteria

Variable selection is an important part of regression analysis. 'Which variables are important to the outcome variable and which are not?' is an important question. In this article, the main interest is the variable selection of Fine and Gray competing risk regression model using Backward and Forward model selection method based on Akaike information criteria (AIC). In this regard, a freely available R package called *crrstep* is used.

Backward elimination procedure

Backward elimination procedure is an improvement of the "All Possible Regression". It attempts to examine not for all regression but only for the "Best" regression containing certain number of variables [11]. At first it takes all the covariates (P) in a regression model and calculate its AIC value.

$$Y = \beta_1 X_1 + \beta_2 X_2 + \cdots \dots \dots \dots + \beta_{P-1} X_{P-1} + \beta_P X_P (AIC = AIC_{All})(a)$$

Now, backward elimination method removes covariates one by one from the above model (a) and creates P number of models and then calculates AIC value for each of them. Need to choose the model with minimum value of AIC. Then the process should be repeated for the new model with (P -1) number of covariates. The procedure will stop and will select a model when no covariate can be removed from the new model.

Forward selection procedure

The direction of forward selection procedure is opposite to that of backward elimination process. This is an attempt to achieve a suitable regression equation by inserting variable one by one until regression equation is satisfactory.

Akaike information criterion (AIC)

The partial likelihood of the data for a given Fine and Gray model is a measure of the goodness of fit. However, the partial likelihood is increased when the number of parameters in the model is also increased. This heads to over fitting. To escape this Akaike Information Criterion proposed by Hirotugu Akaike [12] can be used. It is defined as $AIC = -2logL_P + 2P$,here$P$ is the total number of parameters and $logL_P$ is the maximized value of the log partial likelihood for the Fine and Gray model.

III. Results and Discussion

A. Unadjusted Effect

Before going to find out the adjusted effects of the covariates on under-5 child mortality in competing risk regression model, it is needed to see an overview of the unadjusted effects. The Kaplan-Meier (KM) method [13] may be the most common technique for this purpose. But in the presence of competing risk, Kaplan-Meier estimates cannot interpret the probabilities. However, their complement (1-KM) can interpret the probability of an event of interest in an ideal world where the other types of events do not exist. That is, in the presence of competing risks, (1-KM) does not estimate the probability of the occurrence of an event. So, using KM approach to assess the unadjusted association between the covariate of interest and under-5 child mortality is found under non-competing setup actually. The significant covariates for under-5 child mortality due to the different causes of death in KM approach are shown in Table 2.

Table2: Summary table of unadjusted effect in KM approach for under-5 child mortality due to three different causes of death

Covariates	Disease	Non-disease	Other cause
Mother's age at first birth	×	×	×
Region	✓ (p-value=0.05)	×	×
Aware of community clinic	×	×	×
Mother's education level	✓ (p-value=0.02)	✓ (p-value=0.05)	✓ (p-value=0.00)
Wealth index	✓(p-value=0.01)	×	×
Birth order number	×	✓ (p-value=0.00)	×
Sex of child	×	✓ (p-value=0.04)	×
Place of residence	×	×	×
Place of delivery	×	✓ (p-value=0.00)	×
Sex of household head	×	×	×
Exposure to media	✓ (p-value=0.04)	✓ (p-value=0.06)	✓ (p-value=0.02)
NGO membership	×	✓ (p-value=0.05)	✓ (p-value=0.04)
Size of child	✓ (p-value=0.00)	✓ (p-value=0.05)	✓ (p-value=0.10)
Main access road to village	×	×	×
Availability of MCWC	✓ (p-value=0.04)	×	✓ (p-value=0.00)
Women empowerment	×	×	
Father's education level	×	×	✓ (p-value=0.00)

✓: Significant in unadjusted association

×: Insignificant in unadjusted association

B. Selected Variables

The event of interest in this study is under-5 child mortality. Now, the event may occur due to one of the three causes: Disease, Non-disease, Other cause. Note that occurrence of event due to one of the above 3 causes

precludes from observing the occurrence of event due to any other causes. That is why they are termed as competing risk. As there are three competing causes, there will be three Fine and Gray models. Bigger models have less bias but more variance. Smaller models have less variance but more bias [14]. Thus, it is needed some scientific methods that can select model with balance bias and variance. Backward and Forward model selection methods are used in this study using the model selection criteria AIC. The selected covariates for the three Fine and Gary models are given in Table 3.

Table 3 : Selected covariates in Fine and Gray model due to three different causes of death

Method	Disease	Non-disease	Other
Backward	Wealth index Size of child Gender of child Availability of MCWC	Gender of child Mother's education Size of child Birth order number NGO membership Place of delivery	Main access road to village NGO membership Mother's education Availability of MCWC Father's education Birth order number
Forward	Size of child Wealth index Availability of MCWC Gender of child	Birth order number NGO membership Place of delivery Size of child Mother's education Gender of child	Father's education Availability of MCWC Birth order number Mother's education NGO membership Main access road to village

C. Adjusted Effect

The main objective of this study is to find the important factors and determinants of under-5 child mortality through a competing risk regression setup and to evaluate how much influence the factors have on under-5 mortality. The covariates selected for Fine and Gray models are being regressed to know the influence of the factors on under-5 child morality.

Table 4 : Estimated coefficient, p-value, Sub-distribution hazard Ratio (SHR) and 95% confidence interval in Fine and Gray model for the selected covariates due to Disease Cause

Variables	Category	β	p-value	SHR(=e^{β})	95% CI for SHR
Wealth index	Poor (RC)	-	-	-	-
	Middle	0.014	0.940	1.015	(0.708,1.453)
	Rich	-0.468	0.005	0.626	(0.448,0.874)
Size of child at birth	Small (RC)	-	-	-	-
	Average/Large	-0.513	0.001	0.598	(0.435,0.824)
Gender of child	Female (RC)	-	-	-	-
	Male	0.197	0.170	1.219	(0.918,1.618)
Availability of MCWC	No (RC)	-	-	-	-
	Yes	-0.288	0.080	0.753	(0.548,1.035)

Table 3 shows that for the disease cause, under-5 child mortality among the rich people is lowest. This is because; the rich people can afford more medical and other economical supports for their children than the poor people in Bangladesh. Size of child at birth is an important factor of under-5 child mortality due to disease cause. The under weight babies are more likely to die before their fifth birthday.

Table 5 : Estimated coefficient, p-value, Sub-distribution hazard Ratio (SHR) and 95% confidence interval in Fine and Gray model for selected covariates due to Non-disease Cause.

Variables	Category	β	p-value	SHR(=e^{β})	95% CI for SHR
Gender of child	Female (RC) Male	- 0.410	- 0.037	- 1.507	- (1.026,2.215)
Mother's education	No education (RC) Primary Secondary Higher	- 0.435 0.216 -0.885	- 0.210 0.480 0.151	- 1.544 1.279 0.413	- (0.781,3.052) (0.646,2.533) (0.123,1.390)
Size of child at birth	Small (RC) Average/Large	- -0.575	- 0.008	- 0.563	- (0.367,0.863)
Birth order number	1st birth (RC) 2nd/3rd birth Other birth	- -0.241 -1.080	- 0.240 0.010	- 0.786 0.340	- (0.525,1.177) (0.149,0,776)
NGO membership	No (RC) Yes	- 0.570	- 0.005	- 1.769	- (1.189,2.263)
Place of delivery	Home (RC) Hospital/Other	- 0.680	- 0.001	- 1.974	- (1.301,2.994)

In Table 5 it is observed that the male children are more likely to die due to non-disease cause than the female children. Size of child at birth is an important factor for both disease and non-disease causes of death. It reveals that birth order number is an important factor for the under-5 child mortality due to non-disease cause. The "2nd/3rd" or ""other" birth children are less likely to die than the "1st birth" children. Thus, there is a special phenomenon for the "1st birth" children. These findings will give a new window for the doctors and medical science researchers. In Bangladesh a mother involved in NGO is usually poor. As a result, she cannot take proper care of her children and has to face many difficulties in the time of her pregnancy. This is also found in Fine and Gray model for non-disease cause – under-5 mortality is higher for the children whose mothers have NGO membership. Education is always a crucial factor for under-five child mortality. For the non-disease cause, mother's education shows it's importance in unadjusted association (KM approach). But surprisingly the categories do not show significance in adjusted association.

Table 6 : Estimated coefficient, p-value, Sub-distribution hazard Ratio (HR) and 95% confidence interval in Fine and Gray model for selected covariates due to Other Cause.

Variables	Category	β	p-value	SHR(=e^{β})	95% CI for SHR
Main access road to village	All weather road (RC) Seasoned/Other	- -0.344	- 0.140	- 0.709	- (0.447,1.122)
NGO membership	No (RC) Yes	- 0.344	- 0.120	- 1.396	- (0.920,2.119)
Mother's education	No education (RC) Primary Secondary Higher	- -0.567 -0.786 -0.896	- 0.045 0.017 0.380	- 0.567 0.456 0.408	- (0.326,0.986) (0.239,0.868) (0.055,2.985)
Availability of MCWC	No (RC) Yes	- -0.521	- 0.020	- 0.594	- (0.382,0.923)
Father's education	No education (RC) Primary Secondary Higher	- -0.051 -0.586 -1.943	- 0.840 0.070 0.054	- 0.950 0.557 0.143	- (0.576,1.566) (0.295,1.048) (0.019,1.031)
Birth order number	1st birth (RC) 2nd/3rd birth Other birth	- -0.677 -0.844	- 0.004 0.006	- 0.508 0.430	- (0.319,0.809) (0.236,0.784)

From Table 6 it is seen that besides mother's education the father's education is an important factor for the child mortality due to “"other cause" of under-5 child mortality. Like the "disease cause", due to "other cause" of under-5 child mortality the children with the facility of maternal and child welfare center (MCWC) are more likely to survive than the children who do not have the facility of MCWC in their district. Like the Fine and Gray model for non-disease cause, birth order number is an important factor in Fine and Gray model for other cause. Recall that the "other cause" group consists of three kinds of disease: Other, Undetermined, Unspecific. So, to reduce under-5 child mortality due to “other cause", appropriate home care and timely treatment of complications for newborns are mandatory.

IV. Conclusion

In this article, Fine and Gray competing risk survival regression model, has been used to determine the potential determinants of under-5 child mortality in Bangladesh using BDHS, 2011 data. Findings of the study suggest that

to reduce under-5 child mortality due to disease cause, government should provide some social security and medical support freely for the poor people of the society and should arrange family and social awareness programs as well as health related programs more and more for women to be conscious about the ominous consequences of underweight babies.

To reduce under-5 child mortality due to non-disease cause, special care is necessary for the male children and "1st birth" babies. Besides, female education participation needs to be increased. It will play an important role to increase awareness not to give birth of underweight babies.

Not only the level of female education but also the level of male education should be increased to reduce under-5 child mortality due to "other cause". Number of maternal and child welfare center (MCWC) should be increased by the policy makers to reduce under-5 child mortality due to "other cause".

Acknowledgement

We are greatly indebted to NIPORT for its open access of using BDHS-11 data. The authors are grateful to the anonymous referee for a through checking of the details and for valuable comments that helped to improve this article.

References

[1]. N. I. Mondol, K. Hossain, and K. Ali, "Factors influencing infant and child mortality: A case study of Rajshahi District, Bangladesh," J. Hum. Ecol., vol. 26, no. 1, pp. 31–39, 2009.

[2]. K. A. Mohammad and T. Tabassum, "The impact of socio-economic and demographic factors on under-five child mortality in Bangladesh," Imp. J. Interdiscip. Res., vol. 2, pp. 626–631, 2016.

[3]. A. H. M. M. R. Nabeen and M. A. S. Akanda, "Determinants and predictive power of fitted model for under-five child mortality in Bangladesh," South Asian J. Popul. Health, vol. 10, no. 1 & 2, pp. 109–121, 2017. [Online]. Available: https://ssrn.com/abstract=3121951

[4]. E. W. Kanmiki, A. A. Bawah, I. Agorinya, S. F. Achana, R. A. Oduro, F. J. Phillips, and J. Akazili, "Socio-economic and demographic determinants of under-five mortality in rural northern Ghana," Int. Health Hum. Rights, vol. 14, p. 24, 2014.

[5]. D. R. Cox, "Regression models and life-tables (with discussion)," J. R. Stat. Soc., Ser. B, vol. 34, pp. 187–220, 1972.

[6]. J. P. Fine and R. J. Gray, "A proportional hazards model for the sub-distribution of a competing risk," J. Amer. Stat. Assoc., vol. 94, pp. 496-509, 1999.

[7]. Q. H. Chowdhury, R. Islam, and K. Hossain, "Socio-economic determinants of neonatal, post neonatal, infant and child mortality," Int. J. Sociol. Anthropol., vol. 27, no. 12, pp. 1357–1368, 2010.

[8]. A. H. M. M. R. Nabeen, M. S. Khatun, M. A. Islam, M. A. S. Akanda, and J. L. Sharp,

[9]. "Potential factors associated with undernutrition in children under 5 years and its patterns in Bangladesh: Insights from four cross-sectional studies," Dr. Sulaiman Al Habib Med. J., vol. 6, no. 4, pp. 194–209, 2024, doi: 10.4103/DSHMJ.DSHMJ_43_24.

[10]. A. H. M. M. R. Nabeen and M. A. S. Akanda, "Association between economic inequality and under-five child malnutrition: Evidence from Bangladesh Demographic and Health Survey," Dhaka Univ. J. Sci., vol. 66, no. 1, pp. 73–78, 2018, doi: 10.3329/dujs.v66i1.54548.

[11]. NIPORT, Mitra and Associates, and ICF International, Bangladesh Demographic and Health Survey (BDHS) 2011, Dhaka, Bangladesh, and Calverton, MD, USA: NIPORT, Mitra and Associates, and ICF International, 2013.

[12]. N. R. Draper and H. Smith, Applied Regression Analysis, 3rd ed. New York, NY, USA: John Wiley & Sons, 1998.

[13]. H. Akaike, "Information theory and an extension of the maximum likelihood principle,". 2nd Int. Symp. on Information Theory*, B. N. Petrov and F. Csaki, Eds., Tsahkadsor, Armenia, USSR, Sept. 2–8, 1971, Budapest, Hungary: Akademiai Kiado, 1973, pp. 267–281.

[14]. E. L. Kaplan and P. Meier, "Non-parametric estimation from incomplete observations," J. Amer. Stat. Assoc., vol. 53, pp. 457–481, 1958.

[15]. D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, 4th ed. New York, NY, USA: John Wiley & Sons, 2003.