

Identifying Indicators of Infant Mortality Using Survival Analysis

Nasrin Khatun

Department of Statistics and Data Science, Jahangirnagar University, Savar, Bangladesh

Email: nasrin@juniv.edu

How to cite this paper: Khatun, N. (2025) Identifying Indicators of Infant Mortality Using Survival Analysis. *Open Journal of Statistics*, 15, 361-370.

<https://doi.org/10.4236/ojs.2025.154019>

Received: February 24, 2025

Accepted: August 1, 2025

Published: August 4, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In practice, when dealing with censored data, survival analysis must be employed. In this case, parametric and non-parametric models are appropriate to analyze survival data to obtain optimal estimates of the parameters of interest. To identify significant determinants of infant mortality in rural Bangladesh, survival data have been extracted from the Bangladesh Demographic and Health Survey (BDHS), 2017-2018. In this study, an event involving an infant death within the past 12 months; otherwise, 0 will be used for censoring purposes. The main aim of this study is to find out the relationship between infant death and demographic factors. Used the Cox proportional hazard model (COXPH) to determine the responsible factors and found that age group, religion, and residence area significantly affected mortality. This study found that urban areas had a higher survival rate than rural areas. On the other hand, the age group 20 - 34 has a higher survival probability than other groups. And also, the "Others" have a higher mortality rate than the Muslim religion. Notably, background factors are effective on health facilities which help to increase the survival rate.

Keywords

Censoring Data, Survival Analysis, Cox Proportional Hazards Model

1. Introduction

In reliability and survival analysis, the time-to-failure data play an important role in the development of the reliability and life characteristics of the products. These kinds of data are sometimes modeled using a competing risks model. Several inferences are frequently derived based on censored data in many practical investigations such as clinical trials, industrial experiments, mortality analysis, material strength, etc. Al-Essa *et al.* described that, when the life of the unit is distributed

using the Gompertz distribution, noting that the units come from two lines of production and two independent causes of failure are activated [1]. There are different types of censored data such as Type-I censored data, Type-II censored data, Hybrid censored data, Joint censored data, etc. Competing risk models are applied to different types of censored data and compute statistical inference.

In this study, censored data have been used to correctly figure out and analyze death rates. When it comes to the topic of child mortality, the term “censored data” refers to circumstances in which the whole survival information of a child is not observed [2]. One of the most important indices of a nation’s socioeconomic development is the mortality rate of its children. This concern is also of the highest priority in developing nations worldwide. Despite Bangladesh’s notable advancements in reducing child mortality in recent years, the rate continues to remain high. Dina A. Ramadan *et al.* described that they take into consideration statistical inferences in competing risk models with Akshaya sub-distributions based on the type-II censoring scheme [3].

In 2017, 85 percent of the 6.3 million deaths that occurred worldwide occurred in the first five years of life. More than eighty percent of the deaths that occurred in children under the age of fifteen in the first five years of their lives, regardless of the mortality rate [4]. Survival rates have risen for all age groups since 2000, but the gains were not spread out evenly. The most significant improvements in survival rates were observed among children between the ages of 1 and 4 years. The mortality rate for individuals in this age bracket experienced a 60% decrease between the years 2000 and 2017. The rate of newborn mortality decreased by 41 percent during this period, while the mortality rate among children aged 1 - 11 months, known as the post-neonatal period decreased by 51 percent. Between 2000 and 2017, there was a 37 percent decrease in mortality rates among children aged 5 to 14 [4].

It is possible to evaluate life potentiality assumptions at various points in time through survival analysis. One of the primary challenges in life data analysis is comparing the lifetime distributions of different groups that have been studied [5]. Regression models are commonly used to model the impact of explanatory variables on the dependent variable of longevity. The models mentioned, specifically in references [6], play a crucial role in survival analysis. Thus, this specific model is referred to as the Cox regression model or the proportional hazards regression model [7]. Alghamdi *et al.* described that they developed statistical inference of competing risks samples collected under a joint Type-II censoring scheme of products with Weibull lifetime distributions [8]. Amulya Kumar Mahto *et al.* described that inference for a competing risks model is studied when latent failure times follow the Kumaraswamy distribution and causes of failure are partially observed [9]. Chandrakant Lodhi *et al.* described that statistical inference for a competing risks model when latent failure times belong to a general family of inverted exponentiated distributions [10]. Chandrakant Lodhi *et al.* described that they study a competing risks model using the Gompertz distribution under progressive

Type-II censoring when probability distributions of failure causes are identically distributed with scale and different shape parameters [11].

The main objective of this study is to use censored data to describe the effect of categorical or quantitative variables on infant mortality by using the Cox proportional hazards regression model. The proportional hazard assumption is met by comparing the survival curves for two different groups of subjects. For example, compare the survival pattern of subjects who live in rural areas. Also, compare the survival pattern of people who practice Religion. Also, find out the determinants of infant mortality.

2. Methodology

The study is an analytical cross-sectional study through secondary data from the Bangladesh Demographic Health Survey (BDHS 2017-2018). To date, eight rounds of BDHS have been conducted in Bangladesh by the National Institute for Population Research and Training under the Ministry of Health and Family Welfare of Bangladesh. The dataset has 12 variables such as Age, ANC visit, Father's Education, Education, Religion, Infant Mortality, Child Mortality, Residence, Wealth Index, Birth Order, Mother age at First Birth, Highest Education level, and 7511 observations.

The data collected from the Bangladesh Demographic and Health Survey (BDHS) was analyzed using the COXPH and survival analysis tool using the R programming language. The Cox model was selected due to its ability to estimate the effect of multiple covariates simultaneously without assuming a specific baseline hazard, making it more suitable than the Kaplan-Meier or Weibull models for this multivariate analysis. While the Cox model included key variables such as maternal education, wealth, and antenatal care, potential interactions or confounding between them were not explicitly tested. Future work could include the interaction terms or stratified analysis to better understand these interdependencies. To apply this COXPH, the proportionality assumption has also been met.

This study utilized the Cox proportional hazard ratios for the factors influencing newborn and child mortality. The proportional hazard model, derived from Cox's (1972) research, postulates that for an individual with a set of Covariates represented by x_i , the hazard rate at time t_i can be expressed as:

$$h_i(t_i; x_i) = h_0(t_i) \exp(\beta^i x_i) \quad (1)$$

$h_i(t_i; x_i)$ is the hazard function for the comparison group at time t_i , $x_i = 1$, β represents a vector containing coefficients that are unknown and represent the effects of covariates, x_i represents a recognized vector of predictor variables linked to the individual. The baseline hazard function, $h_0(t_i)$ represents the danger for an individual with $x_i = 0$ at time t_i . Also, chi-square was used to assess the relationship between all the categorical variables. P-values below 0.05 were considered significant in this study.

3. Analysis and Results

Table 1 represents the age group of 21 - 30 years old, accounting for 59.9% of the population. 19% and 19.8% of the population, respectively, are under 20 years and between 31 to 40 years old. There is a relatively balanced distribution between individuals who received less than 4 ANC visits (34.6%) and those who received more than 4 ANC visits (32.1%). The majority of parents (Fathers (83.4%) and Mothers (93%)) received their education, and most of them belong to the religion of Islam. The observed infant mortality rate (1.9%) is consistent with national estimates from BDHS 2017-18, which reported an infant mortality rate of approximately 28 per 1000 live births (2.8%). In wealth distribution, 20.9% are from the richest family, whereas 19.4% are from poorer families. The birth order number is highest from [1, 2), which is 36.4%. 78.4% of the respondents had their first child during their teenage years (12 - 20 years). About 20.9% of respondents had their first child between the ages of 21 and 30. Only 0.7% of respondents had their first child between the ages of 31 and 41. 17.2% of respondents attained a higher education level, whereas only 7% of respondents had no education.

Table 1. Distribution of the background characteristics of the respondents.

Covariates	n	Percentage
Age		
Less than 20 years	1424.00	19
21 - 30 years	4496.00	59.9
31 - 40 years	1485.00	19.8
41 - 50 years	106	1.4
Antenatal Care		
Less than 4 ANC	2598.00	34.6
More than 4 ANC	2414.00	32.1
Father's Education		
No	1244.00	16.6
Yes	6267.00	83.4
Mother's Education		
No	523	7
Yes	6988.00	93
Religion		
Islam	6843.00	91.1
Others	668	8.9
Infant Mortality		
Alive	7370.00	98.1
Dead	141	1.9

Continued

Child Mortality		
Alive	7355.00	97.9
Dead	156	2.1
Residence		
Rural	4830.00	64.3
Urban	2681.00	35.7
Wealth Index		
Poorer	1458.00	19.4
Poorest	1599.00	21.3
Middle	1357.00	18.1
Richer	1525.00	20.3
Richest	1572.00	20.9
Birth_order		
[1, 2)	2737.00	36.4
[2, 3)	2511.00	33.4
[3, 13)	2263.00	30.1
Respondent Age at 1st Birth		
12 - 20 years	5888.00	78.4
21 - 30 years	1570.00	20.9
31 - 41 years	53	0.7
Highest Education Level		
Higher	1293.00	17.2
No Education	523	7
Primary	2118.00	28.2
Secondary	3577.00	47.6

In **Table 2**, compared to other age groups, mothers between the ages of 21 and 30 had a higher percentage of infant deaths. Mothers who had fewer than four antenatal care visits had a greater death rate (30.8%) than mothers who had more than four ANC visits (46.1%). Compared to infants with educated fathers, infants whose fathers had no formal education appear to have higher death risk, and the mother's schooling is also a similar instance. Mortality rates between rural and urban areas show that rural areas (61.7%) have slightly higher infant mortality rates. Compared to infants from wealthier households, infants from the poorest households appear to have significantly higher mortality rates. Infant death rates according to birth order indicate a higher infant mortality rate for those with higher birth orders. Mothers who gave birth to their first child between the ages of 12 and 20 had a greater mortality rate. Compared to respondents with greater levels of education, infants who had uneducated parents appear to have slightly higher mortality rates.

Table 2. Determinants of infant mortality.

Covariates	N	Infant Mortality	
		Alive %	Dead %
Mother's Age			
Less than 20 years	1424.00	1396 (18.9)	28 (20.2)
21 - 30 years	4496.00	4424 (60)	72 (51)
31 - 40 years	1485.00	1446 (19.6)	39 (27.4)
41 - 50 years	106	104 (1.4)	2 (1.4)
Antenatal Care			
Less than 4 ANC	2598.00	2533 (34.4)	65 (46.1)
More than 4 ANC	2414.00	2371 (32.2)	43 (30.8)
Father's Education			
No	1244.00	1215 (16.5)	29 (20.3)
Yes	6267.00	6154 (83.5)	113 (80)
Mother's Education			
No	523	511 (6.9)	12 (8.5)
Yes	6988.00	6862 (93.1)	126 (89.2)
Religion			
Islam	6843.00	6727 (91.3)	116 (82.5)
Others	668	645 (8.8)	23 (16.1)
Residence			
Rural	4830.00	4743 (64.4)	87 (61.7)
Urban	2681.00	2627 (35.6)	54 (38)
Wealth Index			
Poorer	1458.00	1427 (19.4)	31 (21.7)
Poorest	1599.00	1565 (21.2)	34 (23.8)
Middle	1357.00	1333 (18.1)	24 (17.3)
Richer	1525.00	1498 (20.3)	27 (19.5)
Richest	1572.00	1548 (21)	24 (16.7)
Birth Order			
[1, 2)	2737.00	2696 (36.6)	41 (29.1)
[2, 3)	2511.00	2463 (33.4)	48 (33.8)
[3, 13)	2263.00	2211 (30)	52 (36.9)
Total			
Mother Age at 1st Birth			
12 - 20 years	5888.00	5776 (78.4)	112 (79.3)
21 - 30 years	1570.00	1539 (20.9)	31 (22.3)
31 - 41 years	53	52 (0.7)	1 (0.7)
Highest Education Level			
Higher	1293.00	1279 (17.4)	14 (10.1)
No Education	523	511 (6.9)	12 (8.5)
Primary	2118.00	2065 (28)	53 (37.6)
Secondary	3577.00	3516 (47.7)	61 (43.1)

Survival Model

The Cox proportional hazards model was employed to analyze the impact of various demographic, socioeconomic, and health-related factors on survival rates. The model included age group, religion, antenatal care (ANC) visits, birth order, residence, wealth, and education status as predictors. The results revealed significant findings for religion and residence. Specifically, individuals identifying with religions other than Islam exhibited a significantly higher hazard (estimate = 0.58158, p-value = 0.041), indicating an increased risk of the event occurring compared to those identifying with Islam. Additionally, urban residents had a significantly higher hazard compared to rural residents (estimate = 0.4421, p-value = 0.032). Also found that the 20 - 34 age group (estimate = -0.01003, p-value = 0.009) has lower mortality than the other group (Table 3). The Cox model requires a reference category for each categorical variable. In this study, Islam for religion, Rural for area, less than 20 for age group, less than four for ANC, secondary education for education, more than 1 for birth order, and middle for wealth are selected as the reference group. Hazard ratios (HR) are interpreted relative to the reference group. A hazard ratio greater than 1 indicates increased risk of infant mortality compared to the reference, while a value less than 1 indicates decreased risk.

To provide more intuitive insights, we exponentiated the Cox model coefficients as $\exp(\beta)$ to obtain hazard ratios (HRs), which quantify the relative risk of infant mortality associated with each covariate compared to its reference group. A hazard ratio greater than 1 indicates a higher risk, while a value less than 1 suggests a lower risk relative to the baseline.

Mothers aged 20 - 34 years had a slightly lower risk of infant mortality compared to those under 20, with a hazard ratio of 0.99 (1.0% lower risk). However, mothers aged 35 or more had a 26.3% higher risk (HR = 1.26), though this was not statistically significant ($p = 0.573$). In terms of religion, individuals identifying with religions other than Islam had a 78.9% higher risk of infant mortality (HR = 1.79, $p = 0.041$) compared to Muslims. This indicates that Muslim families, in this sample, experienced better survival outcomes. For antenatal care (ANC), mothers who had 4 or more ANC visits experienced a 26.1% higher hazard (HR = 1.26) than those with fewer visits, though this difference was not statistically significant ($p = 0.306$), possibly due to unobserved confounders or limitations in categorization. Regarding birth order, first-born children had a hazard ratio of 0.98, indicating a 1.7% lower risk compared to children of higher birth orders, which was statistically insignificant ($p = 0.945$). Mothers residing in urban areas had a significantly higher risk of infant mortality than those in rural areas, with a hazard ratio of 1.56 (*i.e.*, 55.6% higher risk, $p = 0.032$). This result is counterintuitive and may be influenced by reporting differences, healthcare access patterns, or environmental exposures. When comparing household wealth levels to the middle-income group, the poorer (HR = 0.70), poorest (HR = 0.65), richer (HR = 0.60), and richest (HR = 0.78) categories all had lower hazards of infant mortality,

though none of these differences reached statistical significance ($p > 0.14$ in all cases). Educational level also showed variation in risk: mothers with no education (HR = 1.20) and primary education (HR = 1.22) had a higher risk of infant mortality compared to those with secondary education (the reference). Meanwhile, those with higher education showed virtually the same risk (HR = 0.99). These differences were not statistically significant ($p > 0.6$), suggesting weak or confounded associations in this sample. These hazard ratios provide a clearer understanding of the relative likelihood of infant mortality across demographic and socioeconomic groups, and future work should further explore interactions and stratified effects to validate and expand these findings.

Table 3. Impact of demographic variables on infant mortality, Hazard model estimates of relative risk.

	Estimate	p-value
Age group		
Less than 20		
20 - 34	-0.01003	0.009
35 or more	0.23333	0.573
Religions		
Islam		
Others	0.58158	0.041
ANC		
less than 4		
greater than or equal to 4	0.23199	0.306
Birth order		
more than 1		
1	-0.01705	0.945
Residence		
Rural		
Urban	0.4421	0.032
Wealth		
middle		
Poorer	-0.36244	0.278
Poorest	-0.4367	0.192
Richer	-0.50957	0.141
Richest	-0.2469	0.502
Education status		
Higher		
No education	0.184514	0.725
Primary	0.196202	0.6047
Higher	-0.00635	0.9859

4. Discussion and Conclusion

In this study, the aim is to find out the causes and factors of infant mortality by using COXPH. Findings that the highest mortality occurred in the age group 20 - 30, and less than 4 ANC had higher mortality than more than or equal to 4 ANC. The Number of rural mortalities is higher than urban, but their percentage is lower. While **Table 2** shows slightly higher mortality among mothers whose first birth occurred between 21 and 30, the difference is minimal. Therefore, the claim that early first birth leads to higher infant mortality requires further statistical testing and may not be conclusive in this dataset. Religion was treated as a categorical variable to capture potential differences in health practices and access to services. However, the results show that non-Muslims had a significantly higher hazard of infant mortality because the hazard rate is high that is 0.58158, indicating better survival outcomes among Muslims in this sample. This finding suggests that non-Islamic religious affiliation is associated with an increased risk of mortality, Muslims were more concise about health.

The model indicates that rural residents have a significantly higher mortality ratio compared to urban residents. A similar result was found in a prior study that urban areas have a higher survival rate than rural areas [3]. This finding is somewhat counterintuitive, as urban areas are typically associated with better access to healthcare facilities and services. Age is one of the best factors, found that 20 - 34-year-old mothers' child mortality is lower than others. Similar results proved that urban areas had a lower.

From this study, we suggest that we need to stop early births and increase the health facilities in rural areas. This study used secondary data from BDHS 2017-2018, which may not reflect more recent trends. While BDHS offers high-quality, nationally representative data, limitations include potential recall bias and lack of control over variable definitions. Future studies could benefit from using updated datasets such as BDHS 2022-2023 if available.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Al-Essa, L.A., Soliman, A.A., Abd-Elmougod, G.A. and Alshanbari, H.M. (2023) Comparative Study with Applications for Gompertz Models under Competing Risks and Generalized Hybrid Censoring Schemes. *Axioms*, **12**, Article 322. <https://doi.org/10.3390/axioms12040322>
- [2] Nareeba, T., Dzabeng, F., Alam, N., Biks, G.A., Thysen, S.M., Akuze, J., *et al.* (2021) Neonatal and Child Mortality Data in Retrospective Population-Based Surveys Compared with Prospective Demographic Surveillance: EN-INDEPTH Study. *Population Health Metrics*, **19**, Article No. 7. <https://doi.org/10.1186/s12963-020-00232-1>
- [3] Ramadan, D.A., Almetwally, E.M. and Tolba, A.H. (2022) Statistical Inference to the Parameter of the Akshaya Distribution under Competing Risks Data with Application HIV Infection to Aids. *Annals of Data Science*, **10**, 1499-1525.

<https://doi.org/10.1007/s40745-022-00382-z>

- [4] UN IGME (2018) Levels & Trends in Child Mortality: Estimates: Report 2018. WHO, UNICEF, World Bank, UN, 1-48.
- [5] Biswas, A., *et al.* (2008) Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics. John Wiley & Sons, Inc.
- [6] Cox, D.R. and Oakes, D. (1984) Analysis of Survival Data. Chapman and Hall.
- [7] Fisher, L.D. and Lin, D.Y. (1999) Time-Dependent Covariates in the Cox Proportional-Hazards Regression Model. *Annual Review of Public Health*, **20**, 145-157. <https://doi.org/10.1146/annurev.publhealth.20.1.145>
- [8] Alghamdi, A.S., Abd-Elmougod, G.A., Kundu, D. and Marin, M. (2022) Statistical Inference of Jointly Type-II Lifetime Samples under Weibull Competing Risks Models. *Symmetry*, **14**, Article 701. <https://doi.org/10.3390/sym14040701>
- [9] Mahto, A.K., Lodhi, C., Tripathi, Y.M. and Wang, L. (2021) Inference for Partially Observed Competing Risks Model for Kumaraswamy Distribution under Generalized Progressive Hybrid Censoring. *Journal of Applied Statistics*, **49**, 2064-2092. <https://doi.org/10.1080/02664763.2021.1889999>
- [10] Lodhi, C., Tripathi, Y.M. and Wang, L. (2021) Inference for a General Family of Inverted Exponentiated Distributions with Partially Observed Competing Risks under Generalized Progressive Hybrid Censoring. *Journal of Statistical Computation and Simulation*, **91**, 2503-2526. <https://doi.org/10.1080/00949655.2021.1901290>
- [11] Lodhi, C., Tripathi, Y.M. and Bhattacharya, R. (2021) On a Progressively Censored Competing Risks Data from Gompertz Distribution. *Communications in Statistics—Simulation and Computation*, **52**, 1278-1299. <https://doi.org/10.1080/03610918.2021.1879141>