Annals of Healthcare Systems Engineering

www.ahse.reapress.com

Ann. Healthc. Syst. Eng. Vol. 2, No. 4 (2025) 233–238.

Paper Type: Original Article

Survival Analysis of Patients with Beta-Thalassemia Major in Isfahan Province

Ali Azarbar*

Department of Statistics, University of Urmia, Urmia, Iran; a.azarbar@urmia.ac.ir.

Citation:

Received: 11 April 2025 *Revised: 22 June 2025* *Accepted: 26 August 2025*	Azarbar, A. (2025). Survival analysis of patients with Beta-thalassemia major in Isfahan province. *Annals of healthcare systems engineering,* 2(4), 233-238.

Abstract

Beta-thalassemia major is one of the most prevalent forms of anemia that has become increasingly widespread in recent years. If not diagnosed and managed in a timely manner, the disease may lead to patient mortality. The application of appropriate statistical models, such as survival analysis in affected patients, can provide an effective approach for identifying and determining the factors influencing the risk of death among these patients. The present study aimed to identify the best-fitting survival model, using conventional survival analysis methods, to determine the risk factors associated with mortality in patients with beta-thalassemia major. This study utilized data from 190 patients with beta-thalassemia major who attended a treatment center in Isfahan between 2011 and 2021. Using the R statistical software, exponential, Weibull, log-normal, log-logistic, and Gompertz models—with and without frailty terms—as well as the semi-parametric Cox model were fitted to the data. Based on the Akaike Information Criterion (AIC) and a significance level of 0.05, the log-normal model was identified as the best-fitting model, whereas the Gompertz model demonstrated the poorest fit. Accordingly, the parametric log-normal model was selected and recommended for the survival analysis of beta-thalassemia major, and the effects of the relevant variables were subsequently estimated.

Keywords: Beta-thalassemia major, Survival analysis, Log-normal model, Cox model, Akaike information criterion.

1 | Introduction

Beta-thalassemia major is a highly severe disorder and represents the most common hereditary anemia in Iran. The disease is characterized by early clinical manifestations, which typically appear between three and six months of age. To prevent mortality and ensure patient survival, regular blood transfusions administered at scheduled intervals are essential [1]. This condition is associated with physical deformities, skeletal complications, weakness, and growth retardation, all of which have significant psychological and social implications in addition to their physical consequences [2], [3]. Fortunately, recent advances in medical science have enabled timely treatment and continuous care for affected patients, and the disease is no longer

Corresponding Author: a.azarbar@urmia.ac.ir
https://doi.org/10.22105/ahse.v2i4.51

considered as fatal as it once was. In Iran, the major prevalence clusters of thalassemia are located in the coastal and neighboring provinces, including Khuzestan, Bushehr, Hormozgan, Sistan and Baluchestan, Gilan, and Mazandaran [4].

Today, statisticians employ appropriate statistical models to establish relationships between a set of explanatory variables and a response variable. These models enable researchers to explain and predict the response variable based on the explanatory variables [5]. In survival models, also known as time-to-event or failure-time models, the response variable represents the time until the occurrence of a particular event. Accordingly, the identification of causes and determinants of mortality among patients with beta-thalassemia major can be effectively investigated using survival analysis models. One of the most widely used and popular approaches in survival analysis is the semi-parametric Cox proportional hazards model, primarily because it does not require specification of the probability distribution of survival times. Another approach involves the use of parametric survival models, such as the Weibull, log-normal, log-logistic, and Gompertz models. Among these, the Weibull model is considered more general and is more frequently applied than other parametric models because its hazard function is not constant over time. Furthermore, the inclusion of an additional parameter, referred to as the shape parameter, provides greater flexibility to the model. In survival studies, patients with similar risk factors may nevertheless exhibit different survival durations. This variation may be attributable to the influence of one or more unobserved variables not included in the model. In such situations, semi-parametric or parametric frailty models can be employed. In these models, a significant frailty component indicates that individuals with identical explanatory variables and similar characteristics do not necessarily experience the same survival times, suggesting that one or more additional unmeasured covariates may be required to improve model fit. To address this issue, a random effect component is incorporated into the model, and the significance of this component can be used to evaluate the heterogeneous effects of risk factors across patients [5], [6]. Most previous studies in this field have primarily adopted semi-parametric approaches. Therefore, conducting the present study using parametric methods, in addition to identifying mortality risk factors, may contribute to the introduction of a more efficient and better-fitting model. The aim of this study was to fit various parametric survival models alongside the semi-parametric Cox model and to determine the most appropriate model for patients with beta-thalassemia major.

2 | Materials and Methods

In the present study, data from 190 patients with beta-thalassemia major who had attended a treatment center in Isfahan between 2011 and 2021 were analyzed. The explanatory variables included patient age, age at disease onset, age at first Desferal administration, age at initiation of blood transfusion, splenectomy status (splenectomized vs. non-splenectomized), sex, blood transfusion status (regular, irregular, or no transfusion), parental relationship (unrelated, distant relatives, or close relatives), type of Desferal administration (subcutaneous, pump infusion, both methods, or none), maternal educational level (illiterate, below diploma, diploma, or university education), patient age group (born before 1986 vs. born after 1986), and place of birth (coastal vs. non-coastal regions). The response variable was defined as the duration between disease diagnosis and either death or continuation of treatment. For the survival analysis of patients, parametric models including the exponential, Weibull, log-normal, log-logistic, and Gompertz models, as well as the semi-parametric Cox proportional hazards model, were employed. In addition, the significance of the frailty component was assessed for all fitted models. To evaluate the effects of explanatory variables and to identify the optimal model, variables found to be significant in the univariate analyses were initially entered into the corresponding multivariable models. Subsequently, variables with borderline significance were added sequentially according to ascending p-values. If the inclusion of a variable resulted in a higher Akaike Information Criterion (AIC) value compared with the previous step, that variable was removed from the model. In other words, model selection was based on the AIC, whereby smaller AIC values indicate a more appropriate and better-fitting model [7].

2.1 | Semi-Parametric Cox Model

The semi-parametric Cox model is one of the most widely used and general models in survival analysis. In this model, no assumption is made regarding the probability distribution of the baseline hazard function, and the hazard rate for one individual is assumed to maintain a constant proportional relationship with that of another individual [6], [7]. The model assumes that the hazard function is expressed as the product of an unspecified function of time, common to all study units, and a known link function based on a linear combination of explanatory variables [6]. The Cox model is considered a robust model because, regardless of the adequacy of a specific parametric form, it generally provides an acceptable fit to the observed data. In other words, despite the unspecified probability distribution of the baseline hazard function, the estimated regression coefficients and other resulting inferences are often close to those obtained from parametric models. An important assumption of this model is that failure times of different study units are independent of one another, and that the value of the explanatory variables for a given individual does not influence the survival times of other individuals [6], [7]. In this model, the conditional hazard function is defined as follows:

$$h_i\left(t \mid X\right) = h_{\circ}\left(t\right)\exp\left(\beta X_i\right). \tag{1}$$

In this model, t the survival time for the i[th] study unit is denoted by $h_{\circ}(t)$, and the baseline hazard function, $X = 0$, is unspecified and assumed to be identical for all study units. The unknown coefficients β, describe the relationship between the time-to-event variable and the explanatory (covariate) variables.

2.2 | Parametric Models

In parametric survival models, it is assumed that the survival times of individuals follow a specific probability distribution. In these models, the survival function is defined based on the assumed underlying distribution. The exponential, Weibull, and log-normal functions are among the most commonly used distributions in survival analysis, with their survival functions denoted accordingly $S(t) = \exp(-\lambda t)$, $S(t) = \exp\left(-\lambda t^p\right)$, and $S(t) = \left(1 + \lambda t^p\right)^{-1}$. In these distributions, the exponential model has a single scale parameter, whereas the Weibull and log-normal models involve both scale and shape parameters. Other commonly used parametric survival models include the log-logistic and Gompertz distributions [6]. By introducing a random component into the survival functions, a frailty term can be incorporated into parametric models to account for unobserved heterogeneity, representing the effect of unknown or unmeasured covariates. A significant frailty component implies that individuals with identical observed covariates and similar characteristics may not necessarily experience identical survival times, suggesting that the model may require one or more additional explanatory variables for improved fit. For the random effect component, distributions such as the gamma or inverse Gaussian distributions can be considered [6]. In this study, a gamma distribution was assumed for the frailty component.

3 | Results

In this study, data from 190 patients with beta-thalassemia major who had been referred to a treatment center in Isfahan between 2011 and 2021 were collected. The mean age of the patients was 30.71 years, with a standard deviation of 0.52. Among the study participants, 99 (52%) were female. Regarding place of residence, 92 individuals (48%) lived in coastal areas. In terms of maternal education, 59 cases (31%) had mothers with below-diploma education or were illiterate, while this figure for fathers was 58 (30%). A total of 139 patients (73%) were born in or after 1986. Splenectomy had been performed in 80 patients (42%). Regular blood transfusion was reported in 141 cases (74%). Regarding Desferal administration, 61 patients (32%) received subcutaneous treatment. Additionally, 70 patients (37%) had no family relationship or their degree of consanguinity was unknown. The mean survival time estimated using the Kaplan–Meier curve was 40.65 years. Moreover, the mean survival times for males and females were 40.12 and 39.48 years, respectively. The distribution of patients' characteristics is presented in *Table 1*.

Table 1. Distribution of characteristics of patients with Beta-Thalassemia major.

Percentage	Number	Variable	Percentage	Number	Variable
		Parental Consanguinity			**Gender**
34.7	66	Unrelated	46.8	89	Male
4.7	9	Distant relatives	52.1	99	Female
58.4	111	First-degree relatives	1	2	Unknown
2.2	4	Unknown			Age group
		Splenic status	25.3	48	Born before 1986
42.1	80	Removed	73.1	139	Born after 1986
41.6	79	Not removed	1.6	3	Unknown
16.3	31	Unknown			Father's education level
		Blood transfusion status	5.2	10	Illiterate
74.2	141	Regular	25.2	48	Below diploma level
2.8	5	Irregular	22.6	43	Diploma
1	2	No administration	44.2	84	University education
22	42	Unknown	2.8	5	Unknown
		Type of Desferal administration			Mother's education level
32.1	61	Subcutaneous	7.3	14	Illiterate
18.8	36	Pump-based	23.7	45	Below diploma level
7.8	15	Both	21	40	Diploma
4.7	9	None	45.3	86	University education
3.6	69	Unknown	2.7	5	Unknown
		Event status			Place of birth
12.6	24	Death occurrence	48.4	92	Coastal regions
50.5	96	No death occurrence	50	95	Non-coastal regions
36.9	70	Unknown	1.6	3	Unknown

The AIC was used to select the best-fitting model. The model with the smallest AIC value was considered the most appropriate model. The AIC values for each model are presented in the table below.

Table 2. AIC values for different models.

Model	AIC
Weibull (without frailty)	29.34
Weibull (with frailty)	31.12
Log-normal (without frailty)	19.17
Log-normal (without frailty)	*
Log-logistic (without frailty)	26.01
Log-logistic (without frailty)	*
Gompertz (without frailty)	28.36
Model	**AIC**
Gompertz (with frailty)	*
Cox semi-parametric model	25.16

* The model was non-convergent, and the AIC could not be computed.

Among the frailty models, only the Weibull model achieved convergence, whereas the remaining frailty-based models did not converge; therefore, the AIC could not be computed for those models. According to the results presented in *Table 3*, the parametric log-normal model exhibited the lowest AIC value and was thus selected as the best-fitting model. Consequently, a multivariable survival analysis of patients with beta-thalassemia major was conducted using this model, and the statistically significant variables were identified.

The findings indicate that patients' place of birth, age, maternal education level, age at first Desferal administration, and age at initiation of blood transfusion were statistically significant. For instance, the significance of the place of birth suggests that, based on the selected model, patients living in coastal provinces had lower survival compared with other patients.

Table 3. Estimated coefficients of the log-normal model and p-values.

P-Value	Standard Error	Regression Coefficient	Variables
0.72	160.19	1.82	Fixed coefficient
0.02	1.23	0.41	Place of birth
0.00	0.97	0.29	Age at disease onset
0.03	0.12	-0.09	Age at first Desferal administration
0.64	34.12	1.39	Splenectomy status
0.01	0.18	0.24	Mother's education level
0.92	44.12	0.51	Parental consanguinity
0.04	0.24	-0.16	Age at initiation of blood transfusion
0.91	167.12	-0.66	Age group

4 | Conclusion

Beta-thalassemia major is the most prevalent form of anemia in Iran and other parts of the world, and various survival analysis studies have been conducted on patients affected by this disease. The present study aimed to fit standard parametric models, the semi-parametric Cox model, and frailty models in order to identify the most appropriate model for survival analysis of these patients. Ultimately, the log-normal parametric model was selected as the best-fitting model. In a similar study, Borgna-Pignatti et al. [8] concluded that the survival rate of individuals born after 1970 was higher than that of those born before 1970. Rodbari et al. [9] demonstrated that regular blood transfusion, attention to blood screening, and increased family awareness were effective in preventing complications and improving the life expectancy of patients with thalassemia. Ladis et al. [10] estimated the survival probability of 647 patients with beta-thalassemia major using the Kaplan–Meier method and compared the results using the log-rank test. They showed that survival was higher among individuals born after 1975 compared with those born earlier. Latifi and Zandian [1] investigated the survival of patients with beta-thalassemia major in Ahvaz using life tables, Kaplan–Meier curves, and the log-rank test. Their results indicated that ethnicity (Fars, Arab, and Lur) and gender had no significant effect on survival. Yavarian et al. [11] reported the survival of 101 patients with thalassemia using the Kaplan–Meier curve, showing that 68% of patients survived up to the age of 20 years, and only half of the patients were expected to survive to 30 years of age. The findings of the present study indicated that place of birth, age at disease onset, age at first Desferal administration, maternal education level, and age at initiation of blood transfusion had a statistically significant effect on patient survival. Unlike previous studies, birth cohort (year of birth) did not show a significant effect on survival in this study; however, results related to factors such as age at disease onset, maternal education, and age at initiation of blood transfusion were consistent with previous findings. In earlier studies, survival analysis was mainly conducted using the Kaplan–Meier method, log-rank test, and univariate models, while parametric and frailty models were not applied; therefore, direct comparison with these approaches was not possible. One of the limitations of this study was the presence of a considerable amount of missing data among explanatory variables, which led to a reduction in the effective sample size during model fitting. Given that the log-normal parametric model was identified as the best model for survival analysis of beta-thalassemia major in this study, its application is recommended for future research in this field. Furthermore, increasing parental awareness, particularly maternal education, timely blood transfusion, and proper blood screening for early diagnosis may contribute to improved survival outcomes in these patients.

Authors' Contributions

All aspects of the research and manuscript preparation were carried out by the author. The author has read and approved the final version of the manuscript.

Funding

This study did not receive any specific funding from public, commercial, or non-profit funding agencies.

Data Availability

All data are included in the text.

Conflict of Interest

The author declares that he does not have any conflict of interest.

Consent for Publication

The author has given consent for the publication of this manuscript.

Ethics Approval and Consent to Participate

This study does not involve any research conducted on human participants or animals.

References

[1] Latifi, S. M., & Zandian, K. (2010). Survival analysis of β-thalassemia major patients in Khouzestan province referring to Shafa hospital. *Scientific medical journal (AJUMS), 9*(1), 83–92. http://smj.ajums.ac.ir/index.php/MainJournal/issue/archive

[2] Ansari, H., & Tabatabaei, S. (2006). Assessment of survival without cardiac disease of thalassemic patients of Shiraz, Iran, 2005. *Zahedan journal of research in medical sciences, 8*(1), e94917. https://brieflands.com/journals/zjrms/articles/94917

[3] Moayeri, H., & Oloomi, Z. (2006). Prevalence of growth and puberty failure with respect to growth hormone and gonadotropins secretion in beta-thalassemia major. *Archives of Iranian medicine, 9*(4), 329-34. https://pubmed.ncbi.nlm.nih.gov/17061604/

[4] Henderson, R. (1995). Problems and prediction in survival-data analysis. *Statistics in medicine, 14*(2), 161–184. https://doi.org/10.1002/sim.4780140208

[5] Rosthøj, S., & Keiding, N. (2004). Explained variation and predictive accuracy in general parametric statistical models: The role of model misspecification. *Lifetime data analysis*, *10*(4), 461–472. https://doi.org/10.1007/s10985-004-4778-6

[6] Kleinbaum, D. G., & Klein, M. (1996). *Survival analysis a self-learning text*. Springer. https://doi.org/10.1007/0-387-29150-4

[7] Biglarian, A., Bakhsi, E., Rahgozar, M., Karimloo, M. (2011). Comparison of artificial neural network and logistic regression in predicting of binary response for medical data the stage of disease in gastric cancer. *Journal of north Khorasan university of medical sciences, 3*(5), 15–21. https://doi.org/10.29252/jnkums.3.5.S5.15

[8] Borgna-Pignatti, C., Rugolotto, S., De Stefano, P., Zhao, H., Cappellini, M. D., Del Vecchio, G. C., Romeo, M. A., Forni, G. L., Gamberini, M. R., Ghilardi, R., Piga, A., & Cnaan, A. (2004). Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. *Haematologica journal*, *89*(10), 1187–1193. https://haematologica.org/article/view/3248

[9] Roudbari, M., Soltani-Rad, M., & Roudbari, S. (2008). The survival analysis of beta thalassemia major patients in South East of Iran. *Saudi medical journal*, *29*(7), 1031–1035. https://doi.org/10.15537/1658-3175.4401

[10] Ladis, V., Chouliaras, G., Berdousi, H., Kanavakis, E., & Kattamis, C. (2005). Longitudinal study of survival and causes of death in patients with thalassemia major in Greece. *Annals of the new york academy of sciences*, *1054*(1), 445–450. https://doi.org/10.1196/annals.1345.067

[11] Yavarian, M., Farsheedfar, G. R., Karimi, M., Almoazzez, M., Harteveld, C. L., & Giordano, P. C. (2006). Survival analysis of transfusion dependent Î²-Thalassemia major patients. *Journal of research in health sciences*, *6*(2), 8-13. https://pubmed.ncbi.nlm.nih.gov/15477200/