

MODELLING SURVIVAL DATA OF THALASSAEMIA PATIENTS IN PAKISTAN

Jamal Abdul Nasir, Syed Arif Ahmed Zaidi*

Department of Statistics, The Islamia University of Bahawalpur, *Department of Community Medicine, Quaid-e-Azam Medical College Bahawalpur, Pakistan.

Background: Fatimid Foundation with its centers serves as a charitable organization for millions of blood disease carriers in Pakistan. This retrospective survival study is an analysis of the Thalassaemia patients registered in Multan centre who are followed up to nine years to assess the gender risk of death. **Methods:** Data on 120 patients of Thalassaemia during 1994-2002 was analyzed. The standard Kaplan-Meier and Nelson-Aalen procedures were used to compare the survival function of the male and female patients of Thalassaemia. The statistical significance was also assessed using log rank test. The Cox Proportional Hazards (PH) model using a forward selection procedure was used to identify the potential factors associated with the increase risk of death. **Results:** By ignoring the censoring, the average survival time of males and females were calculated to be 1308.1 and 1574.7 days respectively. The average hazard rate for the female group was 0.00033 and for the male group it was 0.00061. The median survival time was estimated from the Kaplan-Meier survival curves (Males: 1400 days, Females: 1785 days). Under the censoring mechanism, the mean survival time for males was calculated to be 1465.82 days with standard error of 146 days, where as for females it was calculated to be 1855.26 with standard error of 151.82 days. When considering death as event of interest Under the Cox modelling approach, we found a male patient of Thalassaemia was 1.484 times more likely to die when compared to his female counterpart. When considering the iron over load as the event of interest, we found a male patient of Thalassaemia was 1.753 fold more likely to reach excessive Haemochromatosis level compared to female patient. **Conclusion:** The analysis reinforces that female group prognosis is better than the male group. Further, our analysis of real survival data set shows that the high status of Haemochromatosis level is associated with the increased risk of death in patients with Thalassaemia with other possible confounders.

Keywords: Censoring; Cox PH model, Kaplan-Meier estimator, Nelson-Aalen estimator, Thalassaemia

INTRODUCTION

Thalassaemia is an inherent disorder in which there is an abnormality in one or more of the globin genes. It is most widespread hereditary disease in Pakistan; presently there are an estimated 100,000 cases of Thalassaemia in Pakistan which makes up for almost five percent of the world cases.¹ This number is increasing by about 5,000 new births every year. In a recent report from to the Ministry of Health Govt. of Pakistan, it is reported that 5-7% of our population are carrying genes of Thalassaemia and are known as Thalassaemia carriers.² So in dealing with Thalassaemia trait it is important for each country to know statistically its relative position to other countries or even at regional level within the same country. To explore the statistical significance about the life time of the Thalassaemia patients, we use the survival analysis which is a collection of statistical procedures for studying the occurrence and timing of events, for which the outcome variable of interest is time until an event, occurs. Worldwide many authors have used survival analysis techniques to present the results.³⁻⁸ In this study, the primary outcome variable was the time from diagnosis to death.

However, the main objectives of this paper are: (i) to develop the estimator of the survival function by using the standard Kaplan-Meier estimator for the patients of Thalassaemia in Southern Punjab; (ii) to develop the estimator of the cumulative hazard function by using the Nelson-Aalen Procedure; (iii) to estimate the point and interval estimates of mean survival time and median survival time; (iv) Using a forward selection procedure to develop a multiplicative Cox proportional hazards regression model for the event of death from Thalassaemia and for the event of excessive Haemochromatosis level (H-level) from Thalassaemia.

MATERIAL AND METHODS

This retrospective cross-sectional study was carried out at Fatimid foundation blood bank Haematological services Multan centre. Data on the individuals with Thalassaemia who had received regular blood transfusion since January 1, 1994 at Fatimid Foundation in Multan were at risk of having high serum ferritin level was drawn using a simple random sampling design out of a total of 390 patients registered in Multan centre. Demographic data were obtained from the patients files at the centre. The maximum follow up was nine years. January 1, 1994

to December 30, 1994 was the enrolment period. The total study participants consist of 120 individuals, 87 of who found to be in absorbing state at the end of study. The study time for the individuals is measured in days. There were total 57 female participants and 63 male individuals. Several potential risk factors were measures for each individual as patient age (in months) at the time of registration, patient sex, patient H-status, patient region MBD (Multan, Bahawalpur, DG Khan).

A survival function estimator was developed by Kaplan and Meier (1958) and is discussed in many books on survival analysis.⁹⁻¹¹ This estimator is also called the product limit (PL) estimator. When the interest is, the subjects in the group have some additional characteristics that may effect their outcome, for example age, gender, socioeconomic status, education, a popular and widely used multiplicative hazards model due to Cox (1972), often called the proportional hazard model.¹² The Kaplan-Meier in connection with the Nelson-Aalen estimators and the Cox proportional hazard models using stepwise approach were used to assess the gender-based survival for the patients of Thalassaemia. The data was analyzed using SPSS v16 and SAS v9.

We assume that we have a random sample of n patients. We further assume that there is a fixed study Area Southern Punjab to cover. Censoring issue that is considered in this study is the generalized type-I censoring. For more convenience the censoring could be due to the following reasons.

- (i) An individual emigrates out of the study area or loss to follow-up.
- (ii) An individual survives past the end of study period.
- (iii) The censoring is non-informative.

RESULTS

We inspect the survival times given for each gender group. The average survival time were calculated to be 1574.7 and 1308.1 days for females and males respectively. Using the average hazard rate as the other descriptive measure it was seen that the female prognosis (0.000333) was better than their counterparts (0.00061).

We have used the data set by considering the main models labelled model 1 and model 2. The issue of gender risk of death was considered in model 1 by considering the two groups, where $g=1$ shows males group and $g=2$ shows the females group. Towards the model 2 we consider the time to reach excessive H-level as the event of interest and then find the risk factors to asses this dependence. The Product limit estimator provide the summary information about survival. We focus on the survival probabilities of the

two groups of model 1. Figure-1 presents the plot of estimated survival curves for the two groups. We found that the curves end at slight different points as the largest time on study was different for the both groups (3280 days for females, 3235 days for males).

Figure-1 also suggests that the group of females had best and the male group had least favourable prognosis. Even for the matter of convenience of about 760 days, the survival probabilities are 0.6140 with standard error (SE=0.0619) for the male groups and 0.7814 with (SE=0.0559) for the female groups. In this way the corresponding 9 years survival probabilities are 0.0613 with (SE=0.0532) for the males and; 0.1603 with (SE=0.0675) for the females, presents the good female prognosis. The estimated mean death or survival time for the male groups is $\hat{\mu}_{3217} = 1465.82$ days with SE 146 days (Table-1).

A 95% confidence interval (CI) for the mean survival time was (1179.66, 1751.98). The estimated mean survival time for the female group is $\hat{\mu}_{3250} = 1855.26$ with SE 151.82 days. A 95% CI for the mean survival time for female groups is 1557.69, 2152.83 (Table-1). To make comparisons which adjust for the differences in the largest study time, the estimated mean time restricted to the interval 0–3217 days, is computed for each group in Table-2.

When we consider model 2 in which the event of interest is the time to reach the excessive H-level, then the discussion of the significance findings about the gender groups are as follows. The 7-year survival probabilities were 0.0417 with (SE=0.0408) for male group and 0.1250 with (SE=0.0675). The mean time to reach excessive H-level for male group is 1362.5 with SE 197.08, the 95% CI is (976.22, 1748.78) and for female group is 1630.96 with SE 255.96, the 95% CI is (1129.2, 2132.64).

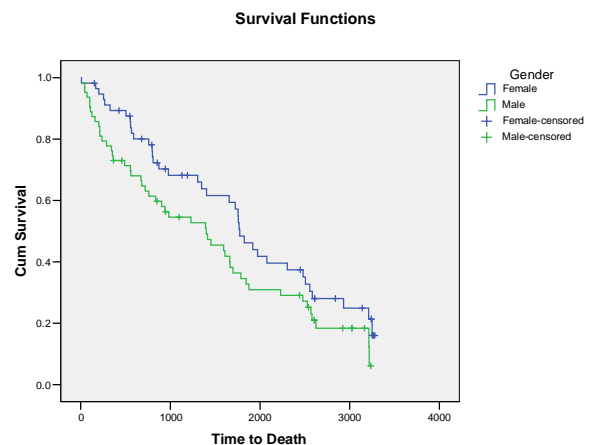


Figure-1: Comparison of Survival function for male and female patients

Table-1 Means and Medians for Survival Time

Sex	Mean				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Female	1855.258	151.822	1557.686	2152.829	1773.000	107.091	1563.101	1982.899
Male	1465.821	146.001	1179.659	1751.984	1400.000	333.906	745.544	2054.456
Overall	1651.230	107.012	1441.486	1860.975	1662.000	154.677	1358.833	1965.167

Table-2: Mean Restricted to 3217 days

Group	Mean	SE	95% CI
Males	1465.82	146	1179.6–1751.9
Female	1845.16	149.85	1551.4–2138.8

Table-3: Regression parameter estimates from Cox model based on Thalassaemia patients of Southern Punjab

Model label	Parameter Estimate	Standard Error (SE)	Wald-Chi Square	p-Value
Variables				
Sex(z ₁)	0.3950	0.2227	3.1471	0.0761
H-level(z ₂)	-1.8182	0.5303	11.7550	0.0006
H-level(z _{2m})	-1.2540	0.2707	21.4525	0.0001
AgexH(z ₅)	0.0185	0.0075	6.1547	0.0131
Variables				
z ₁	0.5614	0.3134	3.2080	0.0733
z ₂	4.4403	1.0754	17.0503	0.0001
Age(z ₄)	0.0840	0.0207	16.4669	0.0001
AgexH(z ₅)	-0.0417	0.0112	13.8120	0.0002

For the Cox regression modelling approach the results of our analysis are summarized in Table-3, where we now propose the following two Cox proportional hazard model. In model 1 we consider the death as the event of interest which is as follows:

$$h(t/z) = h_0(t) \exp[\hat{\beta}_1 z_1 + \hat{\beta}_2 z_2 + \hat{\beta}_3 z_{2m} + \hat{\beta}_5 z_5]^1$$

in model 2 we consider the time to reach excessive H-level as the event of interest, and constructed the following model after using a stepwise forward selection procedure.

$$h(t/z) = h_0(t) \exp[\hat{\beta}_1 z_1 + \hat{\beta}_3 z_{2m} + \hat{\beta}_4 z_4 + \hat{\beta}_5 z_5]^2$$

where Z₁ takes value 0 for female and 1 otherwise, Z₂ shows the status of H-level which is coded as Z₂=1 if 800<H<1400 ng/ml or zero otherwise, and Z_{2m} represents the Excessive H-level as Z_{2m}=1 if H>1400 ng/ml or zero otherwise, Z₄ is the age of the patients measured in months and Z₅ shows the interaction of age and H-level.

From Table-3 under the column heading label 1 refer to the model for the time to death of the patients of Thalassaemia in Southern Punjab given in equation-1 and the model under the label heading 2 shows the model for the same process but the event of interest in this model is time to reach in the state of having excessive H-level given in equation-2. The proposed models confirm the importance of H-level. For the case of model 1 the two gender groups based on proportional hazard model, a male patient of Thalassaemia is $\exp(\beta_1) = 1.409$ times more likely to

die compared to or relative to a female patient of Thalassaemia.

However the evidence shows that there was a potential for H-level to respond differently for males and females, so that the relative risk of males compare to females depend on H-level. So the risk of death for a male patient with H-level was 0.356 relative to a female patient.

As in general linear models, interactions between variables may exist. Generally the interaction effects exist if the relative risk for the two levels of one factor differs for different levels of a second factor. Here in our modelling approach Z₅, the interaction between age and H-level exist, so the exponential of coefficient of Z₅ is, $\exp(0.0185) = 1.0187$ is the excess relative risk of males compared to females. The corresponding 95% CI for $\beta_1, \beta_2, \beta_3$ are give (0.9167, 2.1645), (0.1955, 0.4766), (0.999, 1.008) respectively.

For the case of model 2, a male patient of Thalassaemia is 1.753-fold more likely to reach the excessive H-level compared to female patient. This model clearly shows that time to reach excessive H-level for the gender groups depends on the age of the patient, and interaction (age x H-level). Their respective risks are 84.803, 1.0883 and 0.959. However the relative risk for age 1.088, reflects the excess risk of reaching excessive H-level for each additional year. The corresponding 95% CI are (0.9484, 3.2403), (10.3052, 697.8624), (1.0444, 1.1327), (0.9382, 0.9805). Also from Table 3, the evidence is substantial for each of the regression coefficients we will reject the hypothesis of value of the regression co-efficient to be zero as indicated in p-value.

DISCUSSION

In dealing with Thalassaemia trait it is important for each country to know its relative gender wise position and survival of patients. This study was undertaken not only to evaluate the survival of Thalassaemia patients in connection with the iron overload but also the gender wise survival of these patients. Our results based on gender wise survival were statistically significant where as in another study conducted on the patients of Thalassaemia in Iran the gender wise survival was found to be insignificant.¹³ Further, our results agrees with that of a similar study conducted in seven Italian hospitals, which reported that the female survivorship was

significant than males ($p=0.0003$).¹⁴ Iron over load remained the biggest threat to the health of patients with Thalassaemia. Our study confirms the importance of this factor which adds credibility with the results obtained.

CONCLUSION

Generally patients of Thalassaemia live with more risks than people without Thalassaemia, however, for the patients of Thalassaemia first time in the region of southern Punjab we statistically found that there has been a significant evidence of gender survival that male patients of Thalassaemia are at higher risk than female patients. Our analysis of real data set shows that high status of H-level is strongly associated with increased risk of death in patients with Thalassaemia.

ACKNOWLEDGEMENT

We owe thanks to Dr. Zaheer Ahmed, Administrator Officer blood bank and haematological services Fatimid foundation Multan centre for providing the relevant data for this research.

REFERENCES

1. Amin ST. Over 0.1 m suffer from Thalassaemia. The Dawn (REVIEW) Nov 16, 2001 (13:23:37 +05:00) [Health-List]
2. Haider M. (07/11/2003). Enhanced Yearly Grant in Aid of Rs 15,000 Million from 2004–05 for prevention program of Thalassaemia Disease.
3. Karduan O. Statistical Analysis of Male Larynx-Cancer Patients: A case Study. *Statistica Nederlandica* 1983;37:103–26.
4. Sedmak DD, Meineke TA, Knechtges DS, Anderson J. Prognostic Significance of Cytokeratin-Positive Breast Cancer Metastases. *Modern Pathology* 1989;2:519–20.
5. Mitchell HG. Evaluating serial cancer marker studies in patients at risk of recurrent disease. *Biometrics* 1981;37:67–78.
6. Franks P, Clancy CM, Gold MR. Health insurance and mortality. *JAMA* 1993;270:737–41.
7. Saunders B, Baily S, Phillips M, Allsop S. Women with alcohol problems: Do they relapse for reasons different to their male counterparts? *Addiction* 1993;88:1413–22.
8. Lagakos SW, William JS. Models for censored survival analysis: a cone class of variable-sum models. *Biometrika* 1978;65:181–9.
9. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of American Statistical Association* 1958;53:457–81.
10. Cox DR, Oakes D. *Analysis of Survival Data*. London: Chapman and Hall; 1984.
11. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data*. New York: John Wiley & sons Inc; 2002.
12. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc* 1972;B34:187–220.
13. Kosaryan M, Vahidshahi K, Karami H, Forootan MA, Ahangari M. Survival of Thalassaemia patients referred to the Boo Ali Sina teaching hospital, Sari, Iran. *Haemoglobin* 2007;31:453–62.
14. Borgna-Pignatti C, Rugolotto S, De Stefano P, Zhao H, Cappellini MD, Del Vecchio GC, *et al*. Survival and complications in patients with thalassaemia major treated with transfusion and deferoxamine. *Haematologica* 2004;89(10):1187–93.

Address for Correspondance:

Jamal Abdul Nasir, Department of Statistics, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan.

Email: njamal76@yahoo.com