Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

Towards Better Outcomes: The Role of Education and Awareness in Alpha Thalassemia Management

Qamruz Zaman[1*], Nabi Ur Rahman[2], Najma Salahuddin[3], Soofia Iftikhar[4], Sundas Hussain[5], Syed Habib Shah[6]

[1,2] Department of Statistics, University of Peshawar, Pakistan.
[3,4,5] Shaheed Benazir Bhutto Women University Peshawar, Pakistan.
[6]Institute of Numerical Sciences, Kohat University of Science and Technology, Pakistan.

ABSTRACT:

Biostatistics offers the essential procedural and quantitative foundations for medical research. This branch of statistics has wide applications in genetic, bioinformatics', epidemiology, and clinical trials, etc. A disease may be polio, hemophilia, thalassemia, hepatitis, etc. For checking the life expectancy of these diseases, the field of Survival Analysis is used. Survival Analysis is one of the frequently used sets of Statistical methods in medical research and is considered as the backbone of biostatistics. One of the most frequently inherited diseases in Pakistan is thalassemia. It has emerged as one of the major health problems and is one of the most causes of death. The main objective of this study is to model the α-thalassemia patients.

The data of 184 alpha thalassemia patients was obtained from the Fatimid Foundation Peshawar, Pakistan. These 184 were further divided into 137 males and 47 females.

The data was collected with help of patients, their parents and from their records. Out of the selected patients, 87 belonged to the joint family system and 97 were the member of nuclear family system. Further, 58.2% of the patients' parents were first cousins, 19% were relatives and 22.8% were married outside of the family. Majority of patients were illiterate. Nearly, 73.4% parents were not infaour of during pregnancy test. Majority of the patients belonged to the low family income. For the inferential statistics, survival analysis techniques were used. The log-rank and the Tarone-ware tests for the comparisons of different situations, revealed no difference.

For the α-Thalassemia patients, there factors namely i.e. cast of patients, MCHC and WBC were in the model.

Keywords: Thalassemia, Alpha Thalassemia, Genetic Factors, Socioeconomic Status, Consanguinity, Survival Analysis, Patient Education, Healthcare Access, Family Structure, Public Health, Genetic Testing

1. INTRODUCTION OF THE TOPIC

Statistics is a discipline which is the integral part of numerical as well as the social science research. Nowadays, without statistics, research is impossible. Even the role of statistics cannot be ignored in the analysis of sports e.g. cricket. In medical science, we use the form of statistics, called Biostatistics. Biostatistics is one of the main parts of statistics. Biostatistics offers the essential procedural and quantitative foundations for medical research. This branch

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

of statistics has wide applications in genetic, bioinformatics', epidemiology, and clinical trials, etc. In clinical trials, it can be used to investigate a specific disease within a population or in a specific region and also discovers its reasons and consequences. A disease may be polio, hemophilia, thalassemia, hepatitis, etc. [1]. For checking the life expectancy of these diseases, the field of Survival Analysis is used. Survival Analysis is one of the frequently used sets of Statistical methods in medical research and is considered as the backbone of biostatistics. In survival analysis, missing and partial information on variable of interest i.e., time of occurrence of an event are known as censor data. In case of failure time, the researcher waits till occurrence of event to last individual included in the study. The movement of patients from one place to other also is one of the reasons of censoring which may cause the reduction in the information of the patients' variable of interest. Censor data can be of different types. The right censoring where initial information is known but upper information about the variable of interest is missing. This type of censoring is common in real data. The left censoring gives the information about the upper level only. And when the information regarding variable of interest is available in interval and the information is missing at lower and upper end, is said to interval censoring [2].

The concept of censoring is also divided into three parts namely, Type-I, Type-II and Type-III. In type-I censoring, the time for observing the subjects is fixed as well as the number of subjects (Figure-1.1). The number of events e.g. death, recovery from disease, marriage is a random variable i.e. not fixed.

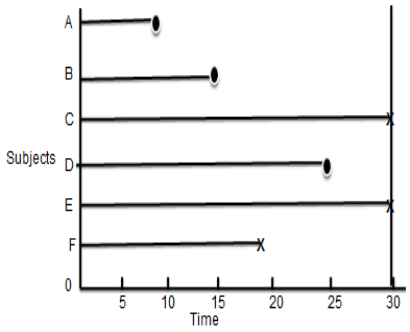

Figure 1.1 Example of type- I censored data

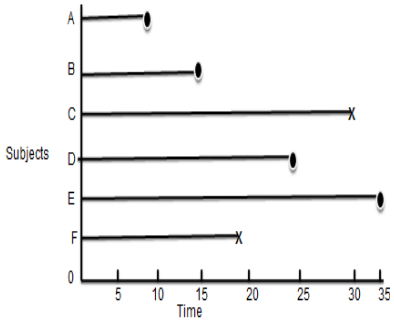

Figure 1.2 Example of type-II censored data

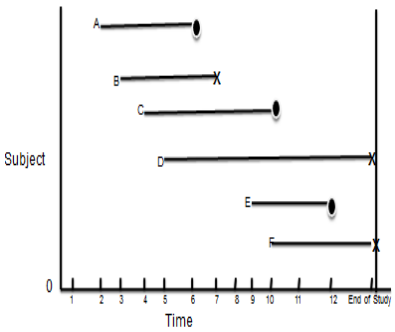

Figure 1.3: Example of type-III censored data

In type-II censoring, the number of events and the subjects are fixed while the waiting time is continued until we observe the fixed number of events (Figure-1.2). Staggered entries are allowed in Type-III censoring. In this the only

the time is fixed. Anyone can enter into and leave the study at any point of time (Figure-1.3). The tools of survival analysis and modeling are now essential aids to the researchers in the medical field. Survival analysis has found wider applications in the field of biostatistics. The techniques of survival analysis is divided into three categories namely; Parametric, Semi-parametric and non-parametric. In parametric procedure of survival analysis, baseline hazard function has specific form and follows the distribution like Exponential, Weibull, and Pareto etc. Cox Regression [3], Stratified Cox Regression and Time Dependent Cox Regression are the three main procedures of semi parametric procedure. In this the base line hazard function is constant and exponential part is the parametric part. The procedure is commonly used for predicting the risk factors under the specific conditions. The most commonly use technique of survival analysis the non-parametric technique. Kaplan-Meier (KM) survival function [4] and Log Rank test are the most commonly use techniques of Survival Analysis. One of the most frequently inherited blood diseases in Pakistan is thalassemia. It has emerged as one of the major health problems and is one of the most causes of death. The origination of the word Thalassemia may have some fifty-thousand years ago in the valley of Greece which refers to the abnormality in one or more of the globin genes [5]. In thalassemia patients, the amount of hemoglobin is less than expected. In case of minor and mild thalassemia, patient might not require the blood. But major form needs proper treatment and regular blood transfusion. Weakness, slow growth, yellowish color and dark color are some of the symptoms of thalassemia. Disease appear due to a mutations in the DNA that make less amount of hemoglobin than the normal. Alpha and beta chains produce the hemoglobin molecules. Thalassemia is of two types, named as α-thalassemia and β-thalassemia.α-thalassemia occurs in patients when hemoglobin does not create enough alpha protein. It has four stages i.e., silent carriers, moderate, medium, and major thalassemia. It is a disorder in blood inherited from family to family lead to abnormal form of hemoglobin. It is produced by inherited transformation from family to family or when the genetic sequence deletes in a human body. If only one of the two partners is carrier of the thalassemic gene, then it may produce a child with “thalassemia minor”. If both are carriers, then there is 25% chance of inheriting a major type of a disease [6]. Apart from inheritance, most of the research shows that thalassemia is also related to the area and frequently it affects the people of Middle East and South Asia including India, Pakistan, and Bangladesh [7].

1.2AREA

Thalassemia is more common among the people of Asia, Middle Eastern, Greek, African descent and Italian.

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

1.3 CAUSES OF THALASSEMIA

Thalassemia is a blood disease which transfers from parents to children. If both the parents are suffering from the minor thalassemia, the children have more chances of inheriting the serious thalassemia disease. If one of the parents suffering from the minor thalassemia, children may have chances of inheriting the minor thalassemia. Iron overload and jaundice problems are the main hurdles in the way of long life expectancy of thalassemia patients. Usually the people with thalassemia die up to the age of 30 [1].

A very few investigations in Pakistan have been attempted regarding thalassemia. The current study is an attempt to determine the significant risk factors associated with the disease. The analysis will draw attention to highlight not only the significant risk factors but will also provide the status of the newly born baby. Therefore this study attempts to determine the survival probabilities by using the Kaplan-Meier Survival function and to obtain the estimated mean and median survival time of thalassemia patients. In addition to this, determining the association and strength of association between the related variables and to obtain an overall and gender-wise Cox proportional hazards regression model for the event i.e., death from Thalassemia.

2. LITERATURE REVIEW

Thalassemia is an integral illness related to an abnormality of globin genes and mostly transmissible disease around the globe. An estimated cases presently in Pakistan are more than 0.1 million which is the about 5% of the total around the world. In Pakistan, every year the number of patients of Thalassemia are increasing by 5 % per 100000 population. In a report, it observed that about 5 to 7 percent of the population being affected by Thalassemia [8]. The statistics of the disease can help to understand and comparison of the disease country wise and the region wise within a country. The statistical importance of a disease can be studied well with the help of survival analysis. Survival analysis consists of statistical methods in which the time to occurrence of an event is variable of interest. Alternatively, the outcome variable in such studies is time until death occurs. Danken et al. [9] analyzed the association between the types of thalassemia and the country of origin of the patients. He observed a significant relation between the type of abnormal genes a thalassemia patients have and its origin. In his study he concluded that the increase in thalassemia patient in Switzerland is due to immigration from Mediterranean countries. Zhu et al. [10] investigated the metamorphosis variety of thalassemia carrier patients

and α- and β-hemoglobin genes for the purpose to launch techniques on prenatal gene analysis. After the analysis it was suggested that the prenatal gene testing and carrier screening in pregnant women should be adopted to avoid thalassemia major. Borgna et al., [11] also studied the information on survival, causes of deaths and complications of patients with thalassemia major from seven centers of Italy and analyzed the relations of gender, birth cohort, barrier and ferritin on survival of the patient. Al-Suliman [12] suggested that in all those countries in which there is a high incidence of hemoglobinopathies, healthcare programs must be initiated for premarital screening to identify and prevent the high risk of marriages and identifying high incidence of attribute in the premarital couples, the future healthcare programs must be started so that to know about the actual incidence of thalassemia. α and β thalassemia types highly found prevalence in Saudi Arabia. Thalassemia is the most challenging disease for almost every country and every country is taking keen interest in eradicating this disease. Olivieriand and Brittenham [13] studied the attitude and perception of people towards thalassemia. He conducted a telephonic interview of the person having age 18 years and above between July and December 2009, in a multi-racial population in Malaysia. In a total of 3723 responded family only 2846 respondents have heard about thalassemia. The study factors were education, employment, age and monthly income. Generally, the response of people who interviewed for screening before marriage was very positive but the ratio who screened before marriage was only 13.6%. A large group of people (63.4%) were not in favor of termination of feotusese who diagnosed with thalassemia major. The finding of this study shows that premarital screening must be more encouraged, and awareness must be created among the people to escape from pregnancies with thalassemia major. Shivalingappa and Parameshwar [14] used various invariants of the logistic model with logit, Probit, and log-log as link function to identify the significant risk factors of periodontal disease. In this paper they find out that religion, status, sources of drinking water, sweet consumption, brushing their teeth and the source are the significant risk for the said disease. Nasir and Zaidi [15] in 2009 conducted a followed-up retrospective survival study for registered Thalassemia patients in Multan of nine years. The study included 120 patients of Thalassemia during 1994-2002 was analyzed by Kaplan-Meier (KM) curve and Nelson-Aalen survival methods to compare the survival function gender wise. Log rank test was used to study the statistical significance and the Cox Proportional Hazard (PH) model was used to study the significant factors related to increase of death both in case of censoring and non-censoring i.e., censored observations ignored. The study concluded that in prognosis female patients found batter than male patients. The survival analysis of real dataset revealed that there found a positive

relationship between the Haemochromatosis level and high risk of death. Constantoulakis et. al. [16] used the longitudinal study procedure and examined the 229 thalassemia patients. Their study revealed that the level of low hemoglobin created the problem of slow growth. Smith and Cauchi, B. [17] determined the association of Mediterranean origin is generally associated with detectable hematological changes. Dincol et. al. [18] determined the common type of thalassemia in Turkey. Wood et. al. [19] conducted a comparative study of Saudi Arabia with African Origin for the Fetal hemoglobin (Hb F) by considering 22 patients. Their research concluded no effect of alpha thalassemia of Fetal Hemoglobin. Beris et. al.[20] examined the effect of globulin chain synthesis by using the case control statistical procedure for high risks group. Meloni et. al.[21] studied the newborn infants for effect of alpha thalassemia and hyperbilirubinemia. Anderson, -B-B determined no association of thalassemia with red-cell metabolism of B6. Cao, et. al. [22] conducted a survey based research based on two different group of patients. They examined the beta thalassemia types by globin chain synthesis analysis. Gajalakshmi et. al. (1984) studied the strong correlation between the blood group and thalassemia based on 330 patients. Vidaud-raphanaud,-D published a paper with a title "Interaction of deletional alpha-thalassemia with sickle cell beta-thalassemia and its influence on foetal hemoglobin expression." In this a rare association of three hemoglobin defects, viz: traits for deletion form of alpha-thalassemia, beta-thalassemia, and sickle cell gene, in a family of French West Indies origin, was studied both at phenotype and genotype levels. In this sickle cell beta-thalassemia, interacting alpha-thalassemia was shown to influence the foetal hemoglobin expression. A reverse relationship between the foetal hemoglobin level and the number of alpha genes was observed. Karim [23] showed that bone marrow density is one of the best tools for improving the life expectancy of thalassemia patients. Their research based on 126 consisting of major and mild thalassemia patients. They also pointed out that the improvement is faster among the patients less than 20 years age group as compared to old age patients. Premawardhena [24] report revealed that a continuous burden on the health as well as the on economy of Asian countries. They have conducted the observational study based on different types thalassemia patients. Their report concluded that the government not only control the disease but also take the steps to prevent malaria and other diseases. Al-Moshary et.al.[25] studied that thalassemia is not an independent disease. They checked and investigated the chances of other diseases like hepatitis B, hepatitis C, and human immunodeficiency virus (HIV) in thalassemia patients. They also concluded that most of the cases included in their study belonged to the poor family.

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

3. PROPOSED METHODOLOGY:

This retrospective cross-sectional study was carried out at Fatimid foundation blood bank Peshawar, district Peshawar. The data from Thalassemia patients was collected through a structure questionnaire. This questionnaire was developed with the consultation of research supervisor and medical officer working in Fatimid Foundation. The sample of study consisted of all those Thalassemia patients who receive regular blood transfusion at Fatimid Foundation. In the data collection process, Type III censoring procedure was used for two months study duration. The personal file of the patient helped us to collect some basic demographic and medical related information.

3.1 DATA SOURCE

Type-III censoring procedure was used for the collection of data. For this two months study period was fixed i.e. from the 1st May 2021 to 30st June 2021. A total of 476 patients' data was collected with the help of the patients, their parents, Doctors and their records from the Fatimid Foundation, Peshawar. Since the treatment of thalassemia is a life lasting process, due to which sometimes patients may leave the study, move from one place to another, change the treatment and may be due to death. This creates the space of censoring. Therefore, the nature of data suggested that the Survival data analysis would be an appropriate statistical technique.

3.2 STATISTICAL METHODOLOGY

The following statistical methodology will be used to achieve the mentioned research objectives

3.2.1 KAPLAN MEIER (KM) PROBABILITY FUNCTION AND CURVE

An estimate of survival function introduced by Kaplan and Meier in 1958[4]. This estimator is also known as product limit estimator. In a situation where individual has similar supplementary characteristics may affect the result of the study. This supplementary information may be age of patient, gender of patient, socio economic status of patient which can be used as a factor as well in developing a hazard model called cox proportional hazard model [3].

The KM probability function is the product of the probability of surviving at the time $t_{i-1}$and the conditional probability of surviving past at time t_i.

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

Symbolically

$$\hat{S}(t_i) = \hat{S}(t_{i-1}) * P(T > t_i | T \geq t_i) \quad (1)$$

In other words

$$\hat{S}(t_{i-1}) = \prod_{j=1}^{i-1} P(T > t_i | T \geq t_i) \quad (2)$$

The Kaplan-Meier product limit estimator is defined by

$$\hat{S}_{t(t)} = \prod \left(1 - \frac{e_i}{n_i}\right) \quad (3)$$

The number of persons at risk in the above equation is denoted by n_i and e_i is used for events.

Greenwood [27] variance estimator for their survival function

$$Var\left(\hat{S}(t)\right) = \hat{S}^{2}_{KM} \sum \left\{ e_i \Big/ \left[n_i (n_i - e_i) \right] \right\} \quad (4)$$

Just like Kaplan-Meier Survival function, also Greenwood variance estimator does not consider the waiting time between events.

3.2.2 PROPORTIONAL HAZARD ASSUMPTION

If the Kaplan Meier survive cures of two diseases do not cross at any point i.e. the Hazard Rate is constant at each point of time. Then the assumption of proportional hazard assumption is said to be satisfied (Figure; 3.1).

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

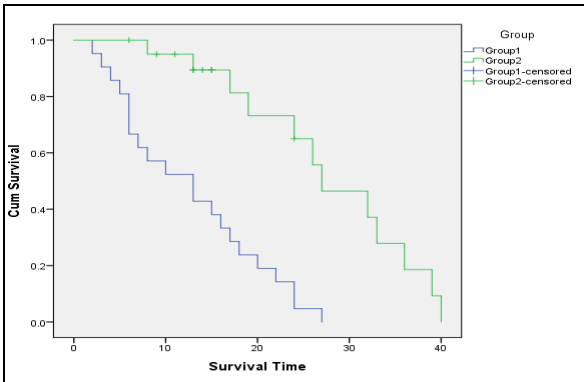

Figure:3.1 Proportional Hazard Assumption is Satisfied

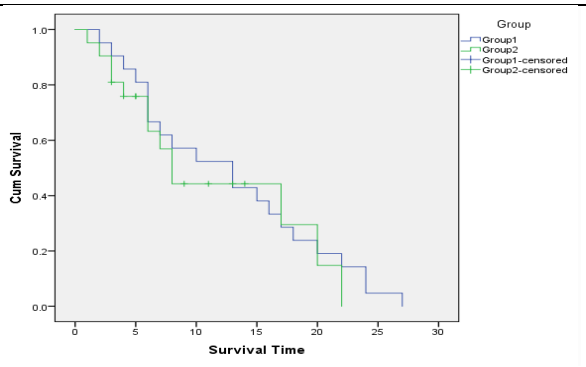

Figure:3.2 Proportional Hazard Assumption is not Satisfied

If the assumption is not satisfied, then the result obtained as illustrated in Figure 3.2.

3.3 LOG-RANK TEST AND WEIGHTED TESTS

For comparing the survival curves of two categories Rank test is commonly used. e.g. Comparing male and female survival, smokers and non-smokers life period etc. Out of the Rank Tests the most commonly use test is the Log-Rank test. The test is more powerful when the proportional hazards assumption is satisfied otherwise weighted tests, Gehan Wilcoxon and Tarone Ware are more popular. The log-rank test statistics is simply called the unweighted rank test considered as the nonparametric technique of survival analysis. Just like the chi-square test of association, it compares the observed frequencies with the expected frequencies. Smaller the difference between the two indicates the two KM cures are equally likely otherwise not.

3.3.1 PROCEDURE OF LOG-RANK TEST

Consider a survival time study and p different event times of two groups I and II ordered as $t_{(1)} < t_{(2)} < \cdots < t_{(p)}$. At time $t_{(i)}$, e_{1i} and e_{2i} be the number of events observed in two groups. Similarly, n_{1i} and n_{2i} be the number of persons at risk in two groups respectively. Therefore, $e_i = e_{1i} + e_{2i}$, $n_i = n_{1i} + n_{2i}$ respectively.

Table 3.1. Rank Tests Layout

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

Group	Number of failures at $t_{(i)}$	Number surviving	Number at risk just before $t_{(i)}$
I	e_{1i}	$n_{1i}-e_{1i}$	n_{1i}
II	e_{2i}	$n_{2i}-e_{2i}$	n_{2i}
Total	e_i	n_i-e_i	n_i

Based under the assumption of null hypothesis that survival is independent of group membership, both the number of failures for two groups and the number of persons surviving can be determined from the value of e_{1i} alone. Once the marginal sums of Table 1 remain fixed, d_{1i} follows a hyper-geometric distribution with mean $E(e_{1i}) = {n_{1i}e_i}/{n_i}$ and variance of e_{1i} is

$$Var(e_{1i}) = \frac{n_{1i}n_{2i}e_i(n_i-e_i)}{n_i^2(n_i-1)}$$

For the overall measure of deviation between the observed and expected failure, sum their differences over the total number of death times to get the statistic.

$$U = \sum_{i=1}^{k}(e_{1i} - E(e_{1i})) \text{ and } Var(U) = Var\left(\sum_{i=1}^{k}(e_{1i} - E(e_{1i}))\right)$$

$$Var(U) = \sum_{i=1}^{k}\frac{n_{1i}n_{2i}e_i(n_i-e_i)}{n_i^2(n_i-1)}$$

Furthermore, according to Collett [20]

$$\frac{U}{\sqrt{Var(U)}} \sim N(0,1)$$

$$\frac{U^2}{Var(U)} \sim \chi^2_{(1)}$$

Therefore, test statistic for the log-rank test is:

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

$$\frac{\left(\sum_{i=1}^{k}\left(e_{1i}-E\left(e_{1i}\right)\right)\right)^{2}}{Var\left(U\right)}\sim\chi_{(1)}^{2} \quad (5)$$

The log-rank test is more suitable, powerful, and consistent when compared to the test statistics used in circumstances where the survival curves remains parallel. In the case of a crossing curve, we use weighted tests. The most frequently used weighted test is Gehan Wilcoxon test, which assigns more weight to initial failure. The test statistic is:

$$\frac{\left(\sum_{i=1}^{k}n_{i}\left(e_{1i}-E\left(e_{1i}\right)\right)\right)^{2}}{Var_{w}\left(U\right)}\sim\chi_{(1)}^{2}$$

Where n_i is the total number of person's prior to time $t_{(i)}$. Where

$$Var_{w}\left(U\right)=\sum_{i=1}^{k}\frac{n_{1i}n_{2i}e_{i}\left(n_{i}-e_{i}\right)}{\left(n_{i}-1\right)}$$

The other commonly used weighted test is the Tarone-Ware, where weight suggested is the square root of the total number of person's prior to time $t_{(i)}$ i.e. $tw=\sqrt{n_i}$.

The Tarone-Ware test is:

$$\frac{\left(\sum_{i=1}^{k}\sqrt{n_{i}}\left(e_{1i}-E\left(e_{1i}\right)\right)\right)^{2}}{Var_{tw}\left(U\right)}\sim\chi_{(1)}^{2} \quad (6)$$

$$Var_{tw}\left(U\right)=\sum_{i=1}^{k}\frac{n_{1i}n_{2i}e_{i}\left(n_{i}-e_{i}\right)}{n_{i}\left(n_{i}-1\right)}$$

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

3.4 COX REGRESSION MODEL

To model the hazard of the thalassemia patient at time t with the set of explanatory variables say Xi, the Cox regression model will be fitted. The general form of the Cox regression model is defined as

$$h_{overall}(t|X) = h_o(t)\exp\left[\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k\right] \quad (7)$$

here $h_{overall}(t|X)$ is the hazard of thalassemia patients at time t, β_i is the ith regression coefficient and x_i is the ith explanatory variable. Similarly, the gender wise hazard of thalassemia patient at a time t will model by the following respective models for male and female patients i.e.,

$$h_{male}(t|X) = h_o(t)\exp\left[\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k\right] \quad (8)$$

And

$$h_{female}(t|X) = h_o(t)\exp\left[\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k\right] \quad (9$$

For finding and comparing the risk factors of alpha and beta thalassemia patients, two separated models are illustrated below:

$$h_{\propto Thalassemia}(t|X) = h_o(t)\exp\left[\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k\right] \quad (10)$$

And

$$h_{\beta Thalassemia}(t|X) = h_o(t)\exp\left[\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k\right] \quad (11)$$

4. RESULTS AND DISCUSSION

4.1 **RESULTS:**

Table-4.1 described that 184 were alpha thalassemia patients data was obtained from the Fatimid Foundation Peshawar Pakistan . These 184 were further divided into 137 males and 47 females. The data was collected with the help of patients, their parents and from their records. Out of the selected patients, 87 belonged to the joint family system and

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

97 were the member of nuclear family system. Further, 58.2% of the patients parents were first cousins, 19% were relatives while 22.8% were married outside of the family.

Table-4.1: Types of family of patients.

		Frequency	Percent	Cumulative Percent
Valid	Nuclear	97	52.7	52.7
	Joint	87	47.3	100.0
	Total	184	100.0	

The pie chart in Figure 4.1 below indicates these facts.

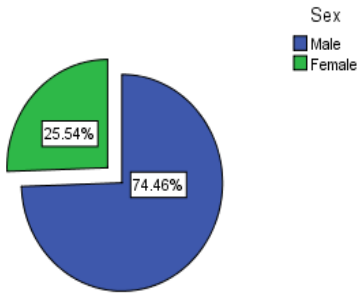

Figure: 4. 1 Gender

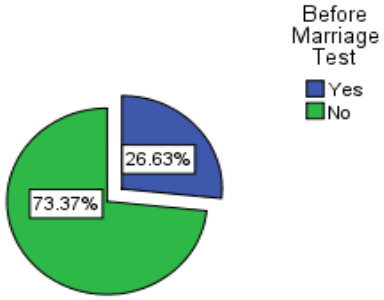

Figure: 4. 2 In favor of Before Marriage test

Figure 4.2 showed the surprising results based on the opinion of the patients parents. 80.46% parents were against the before marriage test for the disease and only 19.54% were in favor of the test.

Table 4.2 gave the detail of the distribution of 184 α-thalassemia patients. 75 patients were censored and 109 were the events.

Table-4.2: Detailed of 184 α-thalassemia Patients

Statistic		Sex		Total
		Male	Female	
Status	Censored	57	18	75
	Event	80	29	109
Total		137	47	184

The Kaplan-Meier Survival function of α-thalassemia Patients are summarized in the Table-4.3. The mean and median survival times of all the patients were 114.502 months and 120 months respectively along with their confidence

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

intervals. 25% of the patients having survival times 156 months and 75% have 79 months respectively. The survival pattern of the patients was described by the Kaplan-Meier Survival Function Figure 4.3.

Table-4.3: Survival Estimates for α-thalassemia Patients

Statistic	Estimate	95% Confidence Interval	
		Lower	Upper
Mean	114.502	107.014	121.991
Median	120.000	104.182	135.818
Q1	156.000	144.169	167.831
Q2	120.000	104.182	135.818
Q3	79.000	70.890	87.111

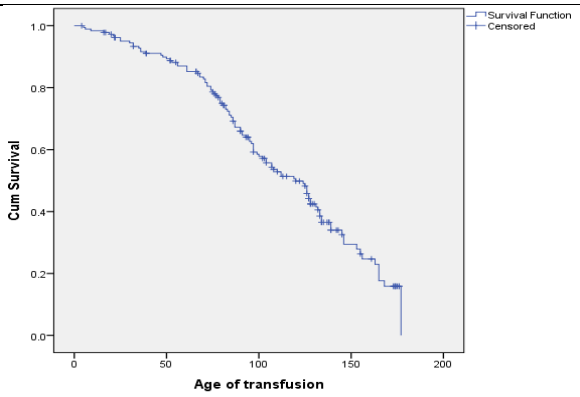

Figure: 4. 3: Kaplan-Meier Survival function of α-thalassemia Patients

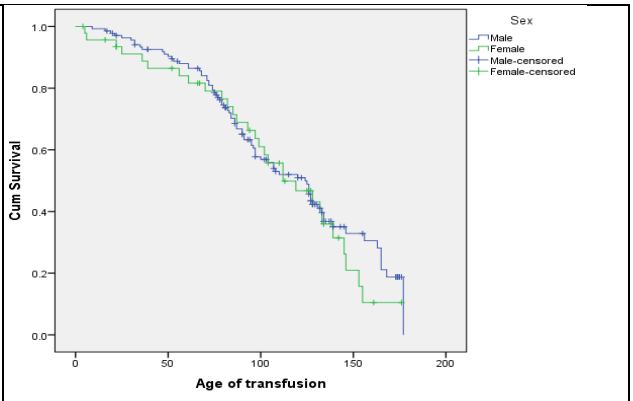

Figure: 4. 4: Genderwise Comparison of α-thalassemia Patients

The graph of survival curves to check the PH assumption is given in Figure 4.4. This indicates that there is no significant difference between the male and female patients survival. The test statistics values of Log Rank test and Tarone-Ware tests are very small i.e. 0.480 and 0.129 indicating no difference between male and female of α-thalassemia Patients (Table- 4.4).

Table-4.4: Comparison of α-thalassemia Patients

Tests	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.480	1	.489
Tarone-Ware	.129	1	.719

The best three models by using forward selection procedure is given in the Table-4.5. Since there are only three variables are included in step-3, so the model is selected as the best model.

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

Table-4.5: Best Three Selected Models of α-thalassemia Patient

-2 Log Likelihood			B	SE	Wald		df	P-Value	Exp(B)
Step 1	780.505	WBC	.000	.000	7.051		1	.008	1.000
Step 2	774.634	C	-.129	.054	5.616		1	.018	.879
		WBC	.000	.000	6.431		1	.011	1.000
Step 3	772.391	C	-.131	.055	5.743		1	.017	.877
		MCHC	.007	.003	3.928		1	.047	1.007
		WBC	.000	.000	6.954		1	.008	1.000

In terms of Cox Proportion Hazard Model the best model is represented as:

$$h(t, X) = h_0(t) \exp [-.131C + 0.007MCHC + 0.000WBC]$$

There appears to be an effect of C i.e. cast of patients, MCHC and WBC in the model. Smaller values of P-values for these variables show that these are the significant factors and play important role in the spread of α-thalassemia disease.

4.2 DISCUSSION:

This study modeled thalassemia, using data from 184 patients sourced from the Fatimid Foundation. Among these, 184 had alpha thalassemia, comprising 137 males and 47 females. Data collection involved patients, their parents, and medical records. Various risk factors were assessed, including transfusion duration, type of thalassemia, parental relations, family structure, education level, household conditions, language, caste, family income, and blood parameters. In alpha thalassemia, significant factors were caste, mean corpuscular hemoglobin concentration (MCHC), and white blood cell count (WBC). The analysis of the patient demographics reveals critical insights into the prevalence and familial patterns of thalassemia.

This finding raises important questions regarding genetic predispositions and environmental factors influencing the disease. The family structure analysis indicates a slight predominance of joint family systems, which may have implications for genetic counseling and support systems for affected families. The strong resistance among parents toward premarital testing underscores a potential gap in awareness and education regarding genetic disorders. This highlights the need for targeted educational campaigns to inform families about the benefits of early detection and intervention. The high percentage of parental relationships among first cousins points to a significant hereditary risk factor, emphasizing the importance of genetic counseling for families with a history of thalassemia. This could inform public health strategies aimed at reducing incidence rates through informed family planning. Educational attainment among patients reveals concerning trends, with a significant portion being illiterate or having incomplete primary education. The cyclical nature of thalassemia—where frequent medical interventions hinder educational pursuits—

demands a multifaceted approach to support affected children, ensuring that medical needs do not eclipse educational opportunities. The linguistic and regional distribution of patients indicates potential cultural and ethnic dimensions to thalassemia, warranting further exploration into community-specific health initiatives. The economic data reflect the challenges faced by low-income families, suggesting that financial constraints may limit access to healthcare resources and interventions. Survival analysis results highlight significant insights into the longevity of patients with alpha thalassemia, with mean survival times suggesting a positive outlook but also indicating substantial variability. The Kaplan-Meier survival curves show no significant gender differences, which could point towards similar treatment responses or disease progression in males and females. The Cox Proportional Hazard Model identifies key factors influencing survival, including cast, mean corpuscular hemoglobin concentration (MCHC), and white blood cell (WBC) count. These variables are pivotal for tailoring individualized treatment plans and may help in stratifying risk among patients.

5. Conclusion: this research underscores the complex interplay between genetic, familial, and socio-economic factors influencing thalassemia outcomes. The findings indicate a pressing need for enhanced awareness and education regarding genetic testing, particularly within communities with high rates of consanguinity. Furthermore, the correlation between economic status and educational attainment among patients necessitates the development of support systems that address both medical and educational needs. Future research should focus on expanding educational outreach and providing resources for families, alongside continued investigation into the biological factors affecting thalassemia progression. Collaborative efforts between healthcare providers, educators, and community leaders will be crucial in mitigating the impact of thalassemia and improving the quality of life for affected individuals.

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

Table-4.4: Family cast/ Place of Birth and Mother tongue

				Count	Column N %
Family Cast/ Place of Birth	Peshawar	Mother Tongue	Urdu	35	27.6%
			Pashto	40	31.5%
			Hindko	52	40.9%
	Total			127	
	Mohmand	Mother Tongue	Urdu	0	0.0%
			Pashto	64	100.0%
			Hindko	0	0.0%
	Total			64	
	Bajawar	Mother Tongue	Urdu	4	4.8%
			Pashto	79	95.2%
			Hindko	0	0.0%
	Total			83	
	Afghan	Mother Tongue	Urdu	0	0.0%
			Pashto	117	100.0%
			Hindko	0	0.0%
	Total			117	
	Migrated From India	Mother Tongue	Urdu	11	84.6%
			Pashto	2	15.4%
			Hindko	0	0.0%
	Total			13	
	Others	Mother Tongue	Urdu	7	9.7%
			Pashto	65	90.3%
			Hindko	0	0.0%
	Total			72	

A question asked from the parents about during pregnancy test to avoid the disease revealed the following results (Figure: 4.3). Majority of the parents (74.58%) were not in favor of test, 19.75% had no knowledge about the test and only 5.67% were in the favor of conducting the test for the prevention of this lifelong disease.

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

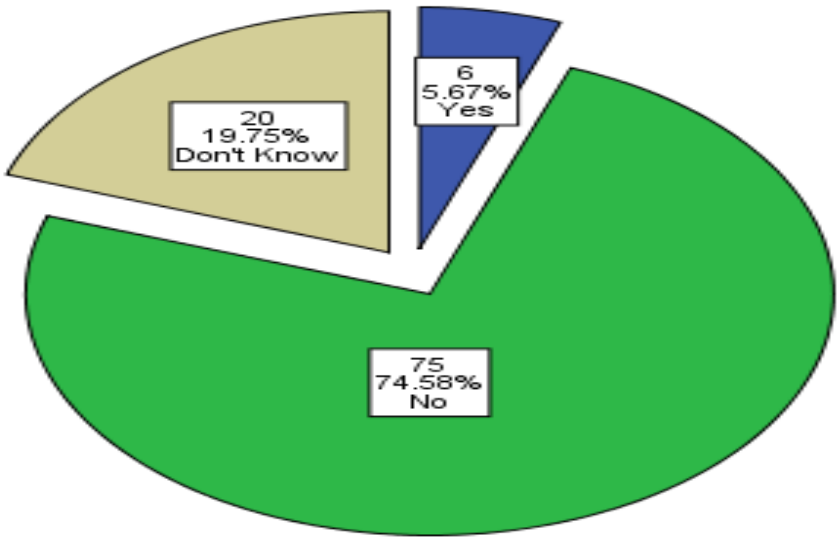

Figure: 4. 3: During pregnancy test

The information's about the family income were summarized in the Table 4.5. Table shows 126 (26.5%) patients were belonged to the family having less than equal to Rs. 10000 income, 169 (35.5%) were belonged to the family having less than equal to Rs.15000 income, 20.6% were belonged to the family less than equal to Rs. 20000 income, 5.0% were belonged to the family less than equal to Rs. 25000 income, 3.2% were belonged to the family less than equal to Rs. 30000 income, 6.9% were belonged to the family less than equal to Rs. 50000 income, 2.1% were belonged to the family less than equal to Rs. 80000 income and only 0.2% families earning greater than 80000. This shows that the majority of the patients belong to low earning families.

Table-4.4: Family cast/ Place of Birth and Mother tongue

Income	Frequency	Percent
<=10000	126	26.5
10000<I<=15000	169	35.5
15000<I<=20000	98	20.6
20000<I<=25000	24	5.0
25000<I<=30000	15	3.2
30000<I<=50000	33	6.9
50000<I<=80000	10	2.1
>80000	1	.2
Total	476	100.0

Since the records consists of both type of thalassemia and information obtained from both male and female patients. Therefore, the inferential analysis was done on the following grounds.

α-thalassemia Patients Survival Analysis

Vol. 10 No. 2 (2024):1363-1387
DOI: https://doi.org/10.17720/qgwm7t89

Table 4.7 gave the detail of the distribution of 184 α-thalassemia patients. 75 patients were censored and 109 were the events.

Table-4.7: Detailed of 184 α-thalassemia Patients

Statistic		Sex		Total
		Male	Female	
Status	Censored	57	18	75
	Event	80	29	109
Total		137	47	184

The Kaplan-Meier Survival function of α-thalassemia Patients are summarized in the Table-4.8.

The mean and median survival times of all the patients were 114.502 months and 120 months respectively along with their confidence intervals. 25% of the patients having survival times 156 months and 75% have 79 months respectively. The survival pattern of the patients was described by the Kaplan-Meier Survival Function Figure 4.5

Table-4.8: Survival Estimates for α-thalassemia Patients

Statistic	Estimate	95% Confidence Interval	
		Lower	Upper
Mean	114.502	107.014	121.991
Median	120.000	104.182	135.818
Q1	156.000	144.169	167.831
Q2	120.000	104.182	135.818
Q3	79.000	70.890	87.111

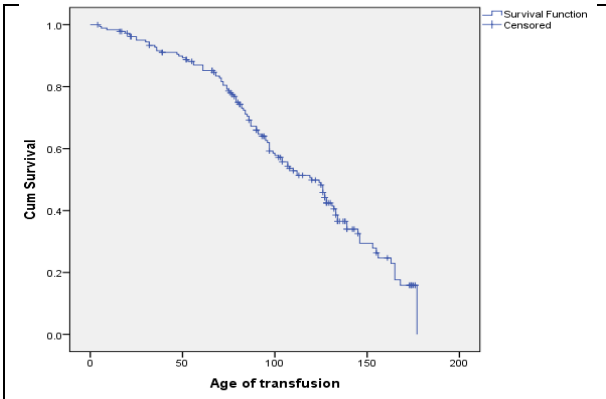

Figure: 4. 5: Kaplan-Meier Survival function of α-thalassemia Patients

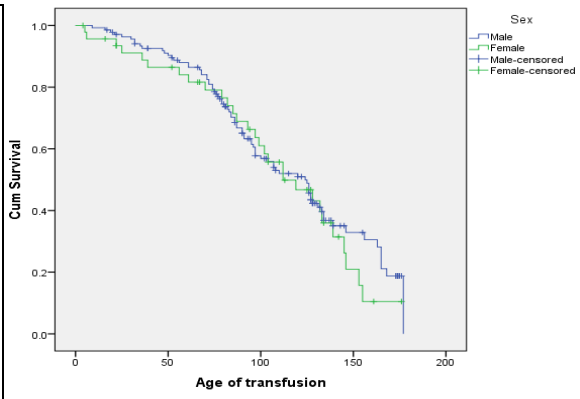

Figure: 4. 6: Genderwise Comparison of α-thalassemia Patients

The graph of survival curves to check the PH assumption is given in Figure 4.6. This indicates that there is no significant difference between the male and female patients survival. The test statistics values of Log Rank test and Tarone-Ware tests are very small i.e. 0.480 and 0.129 indicating no difference between male and female of α-thalassemia Patients (Table- 4.9).

Table-4.9: Comparison of α-thalassemia Patients

Tests	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.480	1	.489
Tarone-Ware	.129	1	.719

The best three models by using forward selection procedure is given in the Table-4.7. Since there are only three variables are included in step-3, so the model is selected as the best model.

Table-4.7: Best Three Selected Models of α-thalassemia Patient

-2 Log Likelihood			B	SE	Wald		df	P-Value	Exp(B)
Step 1	780.505	WBC	.000	.000	7.051		1	.008	1.000
Step 2	774.634	C	-.129	.054	5.616		1	.018	.879
		WBC	.000	.000	6.431		1	.011	1.000
Step 3	772.391	C	-.131	.055	5.743		1	.017	.877
		MCHC	.007	.003	3.928		1	.047	1.007
		WBC	.000	.000	6.954		1	.008	1.000

In terms of Cox Proportion Hazard Model the best model is represented as:

$$h(t, X) = h_0(t) \exp [-.131C + 0.007MCHC + 0.000WBC]$$

There appears to be an effect of C i.e. cast of patients, MCHC and WBC in the model. Smaller values of P-values for these variables show that these are the significant factors and play important role in the spread of α-thalassemia disease.

4.2 DISCUSSION:

This study modeled thalassemia, using data from 476 patients sourced from the Fatimid Foundation. Among these, 184 had alpha thalassemia, comprising 137 males and 47 females, while 292 had beta thalassemia, with 210 males and 82 females. Data collection involved patients, their parents, and medical records. Various risk factors were assessed, including transfusion duration, type of thalassemia, parental relations, family structure, education level, household conditions, language, caste, family income, and blood parameters. The analysis showed that 244 patients came from joint families, while 232 were from nuclear families. Notably, 56.7% of parents were first cousins, suggesting a significant genetic link. Education levels indicated that 87.1% of patients achieved only primary education. Most parents opposed premarital testing (80.46%) and prenatal testing (74.58%) for thalassemia, highlighting a need for greater awareness. Economic data revealed that 62% of patients belonged to families with incomes below Rs. 15,000, indicating a prevalence of low-income households. The survival analysis identified 191 censored cases and 285 events, divided into overall survival, and survival times for alpha and beta thalassemia, as well as male and female patients. The log-rank test revealed no significant gender differences in thalassemia prevalence. Kaplan-Meier estimates provided survival probabilities across the five categories. Using Cox Regression analysis, key

factors influencing survival were identified. For the overall model, significant variables included type of thalassemia, mean corpuscular hemoglobin (MCH), hematocrit (HCT), and household conditions. In alpha thalassemia, significant factors were caste, mean corpuscular hemoglobin concentration (MCHC), and white blood cell count (WBC). For beta thalassemia, significant factors included family type, household conditions, mean corpuscular volume (MCV), and HCT. Among males, significant factors were thalassemia type, MCH, MCHC, and red blood cell (RBC) count, while in females, the significant factors were abortion preference and MCH. The analysis of the patient demographics reveals critical insights into the prevalence and familial patterns of thalassemia. Among the 476 patients, the division between alpha and beta thalassemia shows a notable gender disparity, with a higher incidence of both types in males. This finding raises important questions regarding genetic predispositions and environmental factors influencing the disease. The family structure analysis indicates a slight predominance of joint family systems, which may have implications for genetic counseling and support systems for affected families. The strong resistance among parents toward premarital testing underscores a potential gap in awareness and education regarding genetic disorders. This highlights the need for targeted educational campaigns to inform families about the benefits of early detection and intervention. The high percentage of parental relationships among first cousins points to a significant hereditary risk factor, emphasizing the importance of genetic counseling for families with a history of thalassemia. This could inform public health strategies aimed at reducing incidence rates through informed family planning. Educational attainment among patients reveals concerning trends, with a significant portion being illiterate or having incomplete primary education. The cyclical nature of thalassemia—where frequent medical interventions hinder educational pursuits—demands a multifaceted approach to support affected children, ensuring that medical needs do not eclipse educational opportunities. The linguistic and regional distribution of patients indicates potential cultural and ethnic dimensions to thalassemia, warranting further exploration into community-specific health initiatives. The economic data reflect the challenges faced by low-income families, suggesting that financial constraints may limit access to healthcare resources and interventions. Survival analysis results highlight significant insights into the longevity of patients with alpha thalassemia, with mean survival times suggesting a positive outlook but also indicating substantial variability. The Kaplan-Meier survival curves show no significant gender differences, which could point towards similar treatment responses or disease progression in males and females. The Cox Proportional Hazard Model identifies key factors influencing survival, including cast, mean corpuscular hemoglobin concentration (MCHC), and white blood cell

(WBC) count. These variables are pivotal for tailoring individualized treatment plans and may help in stratifying risk among patients.

5. Conclusion: this research underscores the complex interplay between genetic, familial, and socio-economic factors influencing thalassemia outcomes. The findings indicate a pressing need for enhanced awareness and education regarding genetic testing, particularly within communities with high rates of consanguinity. Furthermore, the correlation between economic status and educational attainment among patients necessitates the development of support systems that address both medical and educational needs. Future research should focus on expanding educational outreach and providing resources for families, alongside continued investigation into the biological factors affecting thalassemia progression. Collaborative efforts between healthcare providers, educators, and community leaders will be crucial in mitigating the impact of thalassemia and improving the quality of life for affected individuals.

REFERENCES

[1] Al-Aidaroos, A.A., Bakar., Othman, Z. Medical Data Classification with Naïve Bayes Approach. Information Technology Journal. 2012,11: 1166-1174.

[2] Ananth, C.V., Kleinbaum, D.G. Regression models for ordinal responses: A review of methods and applications. International Journal Epidemiology. 1997, 26:1323-1333.

[3] Kaplan, E.L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457-481.

[4] Cox, D.R. and Oakes, D. 1984. Analysis of Survival Data. London: Chapman & Hall.

[5] McLaren., R. Implementation of evidence in practice settings: some methodological issues arising from the South Thames evidence-based project. Clinical Effectiveness in Nursing. 2000, 4:2–16.

[6] Thalassemia Federation of Pakistan. Electronic references. Retrieved from http://www.tfp.org.pk/. On 01, June 2021.

[7] Thein, S.L. Genetic insights into the clinical diversity of beta thalassemia. Brazilian Journal of Haematol.2004, 124(3):264-74.

[8] Amin ST. Over 0.1 m suffer from Thalassemia. The Dawn (REVIEW) Aug 16, 2021 (13:23:37 +05:00) [Health-List]

[9] Danken, E., Killer, D., Fischer, S., Marti, H.R. Incidence of thalassemia in Switzerland. Schweizerische Medizinische Wochenschrift. 1975, 105(4):102-105.

[10] Zhu, B.S., He, J., Zhang, J., Zeng, X.H., Su, J., Xu, X.H., Li, S.Y., Chen, H., Zhang, Y.H. A study on gene mutation spectrums of α- and β-thalassemia in populations of Yunnan Province and the prenatal gene diagnosis. Zhonghua Fu Chan KeZaZhi. 2012, 47(2):85-9.

[11] Borgna, Pignatti., Rugolotto, S., DeStefano, P., Zhao, H., Cappellini, Del, G.C., Vecchio, M.A., Romeo, G.L., Forni, M.R., Gamberini, R., Ghilardi, A., Piga, A.C. Survival And Complications In Patients With Thalassemia Major Treated With Transfusion And Deferoxamine. Haematologica. 2004, 89:1187-1193.

[12] Al-Suliman. Prevalence of β-thalassemia trait in premarital screening in Al-Hassa, Saudi Arabia. Annals of Saudi Medicine. 2006, 26(1):14-16.

[13] Olivieri, N. F. & Brittenham, G. M. Iron-chelating therapy and the treatment of thalassemia. Blood.1997, 89(3); 739-61.

[14] Shivalingappa, B. Ja., Parameshwar, V.P. A comparison of ordinal regression models in an analysis of factors associated with periodontal disease. Journal of Indian Society of Periodontology.2010, 14(3):155–159.

[15] Nasir, J. A. & Zaidi, S. A. Modelling survival data of thalassaemia patients in Pakistan. Journal of Ayub Medical College Abbottabad. 2009, 21(1). 142-145

[16] Constantoulakis,-M; panagopoulos,-G; Augoustaki,-O (1975). Stature and longitudinal growth in Thalassaemia major. A study of 229 Greek patients. Clinical Pediatrics. 14(4): 355-7/362-8.

[17] Smith, M., Cauchi, B (1979). Quantitative studies of Hb Bart's levels and red cell indices in alpha thalassaemia trait in Mediterraneans. Pathology. 11(4): 621-7

[18] Dincol,-G; Aksoy,-M; Erdem,-S (1979). Beta-Thalassaemia with increased haemoglobin A2 in Turkey. A study of 164 thalassaemic heterozygotes. Hum-Hered. 29(5): 272-8.

[19] Wood, WG., Pembrey, ME., Serjeant, GR., Perrine, RP., Weatherall, DJ. (1980). Hb F synthesis in sickle cell anaemia: a comparison of Saudi Arab cases with those of African origin. British Journal of Haematology. 45(3): 431-45.

[20] Beris, P., Boreux, G., Klein, D., Miescher, PA. (1980). The detection of beta-Thalassaemia minor. Nouv-Rev-Fr-Hematol. 22(3) : 223-34.

[21] Meloni, T., Erre, S., Gallisai, D., Cutillo, S. (1980). Beta-Thalassaemia trait and hyperbilirubinemia in G-6-PD deficient newborn infants. European Journal of Pediatrics. 134(2): 119-20.

[22] Cao, A., Furbetta, M., Ximenes, A., Angius, A., Rosatelli, C., Tuveri, T., Scalas, MT., Falchi, AM., Maccioni, L., Melis, MA., Galanello, R. (1981).Beta-Thalassaemia types in southern Sardinia. Journal of Medical Genetics. 18(3): 196-9.

[23] Karimi M, Ghiam A F, Hashemi A, Alinejad S, Soweid M, Kashef S (2007). Bone Mineral Density in Beta-Thalassemia Major and Intermedia . Indian Pediatric. 44: 29-32.

[24] Anuja Premawardhena. A., Shanthimala. dS.,, Arambepola. M., Nancy Olivieri, N., Merson, L., , Julia Muraco, J., Angela Allen.A, Fisher, C. , Peto, T., Elliot Vichinsky, E., Weatherall, D., (2004).Thalassemia in Sri Lanka: a progress report. Human Molecular Genetics.(13);203-206.

[25] Al-Moshary M., Al-Mussaed, E., Khan, A., (2019) Prevalence of Transfusion Transmitted Infections and the Quality of Life in β-thalassemia Major Patients.Cures 2019 11(11): e6129. DOI 10.7759/cureus.6129.

[26] Esmaeilzadeh, F., Ahmadi, B., Vahedi, S., Barzegari, S., Rajabi, A. (2021). Major Thalassemia, Screening or Treatment: An Economic Evaluation Study in Iran. International Journal of Health Policy and Management. x(x), 1–8.

[27] Miller, Jr and Rupert G. (1983). What Price Kaplan-Meier. Biometrics, 39, 1077-1081.