Electronic Journal of Applied Statistical Analysis
EJASA (2012), Electron. J. App. Stat. Anal., Vol. 5, Issue 2, 271 – 288
e-ISSN 2070-5948, DOI 10.1285/i20705948v5n2p271

SURVIVAL ANALYSIS OF DIALYSIS PATIENTS UNDER PARAMETRIC AND NON-PARAMETRIC APPROACHES

Maryam Siddiqa[1*], Mueen - ud -Din Azad[1], Muhammad Khalid Pervaiz[2], Muhammad Ghias[1*], Gulzar H. Shah[3], Uzma Hafeez[1]

[(1)]Department of Statistics, GC, University, Lahore, Pakistan
[(2)]Hajvery University, Lahore, Pakistan
[(3)]Jiann-Ping Hsu College of Public Health, Georgia Southern University, USA

Received 24 March 2011; Accepted 02 April 2012
Available online 14 October 2012

Abstract: *Dialysis is a life saving procedure and recommended way of treatment for end stage kidney diseases. Transplantation can also be useful source but it is restricted by financial limitations especially in developing countries like Pakistan. A little work has been done in literature regarding the estimated survival time of dialysis patients in Pakistan. So, this study has estimated the median survival time of male/females patients separately by parametric and non-parametric approaches. Moreover, comparison of survival time to patients (≤ 50 years and >50 years) was also compared. As censoring is an important part of the survival data which causes insensitivity to the usual procedures of analysis. Frequently, in modeling the survival data, most of the time we have no prior information about the theoretical distribution of survival time, that's why, non-parametric methods are commonly used. The significance of this study is the fitting of probability distribution of real life time data of dialysis patients which is not done before. It is very laborious job to fit an appropriate distribution of the data and estimation of parameters. We find that the probability distribution of our real life time data is weibull distribution.*
Finding suggested that the Kaplan-Meier method and weibull model based on Anderson-Darling test provided a very close estimate of the survival function in both genders and age groups. On the average survival time in males is relatively high but not statistically different from females.

Keywords: *Survival data, dialysis, parametric tests, non-parametric tests, Weibull distribution, Kaplan-Meier method.*

[*] E-mail: m.siddiqa85@gmail.com, ghiasgcu@yahoo.com

1. Introduction

The Chronic Kidney Disease (CKD) becomes a common disorder and there are growing burden of CKD patients with worldwide increasing prevalence. As documented by [6-8, 12] the high prevalence of CKD is confirmed by European, Australian and Asian studies. Hence it is obvious that globally there are a large number of patients with Chronic Kidney Disease (CKD). In America alone about 370,000 people undergo dialysis and among these around 50,000 people die each year due to kidney failure [3]. Many of the studies shown that one out of ten people, slightly or seriously, suffering any kind of kidney disease [4]. It has been revealed that out of every three Pakistani's; one is suffering from kidney disease [13]. High prevalence of CKD and the high cost of Renal Replacement Therapy (RRT) are connected with end stage renal disease (ESRD). Dialysis is still the most common treatment of CKD, due to insufficient kidney donation to meet the requirements for renal transplantation. So it is apparent that CKD is a vital public health problem world widely. Particularly in Pakistan due to the high prevalence rate of Diabetes and Hypertension which are the major risk factors of CKD, Chronic Kidney disease is becoming a critical issue [13]. The increasing number of patients who have kidney disease and need dialysis is a sign of importance of research in the field of nephrology. Dialysis is in itself an invasive technique with its own setup of complications and morbidity/mortality. Up till now, physicians have done few descriptive researches related to causes of Kidney failure and quality of life of Dialysis patients, but there is rare chance of any documented statistical inferential work about survival of these patients, done by statisticians in Pakistan.

Pakistan is an under-developed country and chronic renal failure has emerged as a crucial medical, social and economic problem for the people suffering from this disease. Caused by unawareness of common people or else poor symptomatic medical practice, chronic kidney disease developed to ESRD and patients have to experience the costly renal replacement therapy. Therefore, it's very difficult for the people and government as well to afford all the expenses of dialysis modality. In Pakistan, cost of dialysis is eight times the average annual per capita income i.e. almost $3000 per year. Hemodialysis was introduced in Pakistan in 1970 and Eighty five percent patients maintained on hemodialysis [32]. Currently 175 hemodialysis centers are in running condition in Pakistan. These dialysis centers are positioned in four provinces of the country and more than hundred centers are located in Sindh and Punjab province's major cities. The Dialysis registry of Pakistan (2008) reported 6351 patients in 175 Centers are being dialyzed in Pakistan [23].

Analysis of the data that measures lifetime or the length of time until the occurrence of the event, generally focuses on estimating the probability about individual who will survive for a given length of time. Several approaches have been proposed in the literature for analyzing the survival data analysis [19, 20].The nonparametric technique, Kaplan-Meier Product Limit Method (Kaplan-Meier, 1958) is applied to estimate the survival functions and hazard rates of the survival data [14]. In biomedicine, prior knowledge about the distribution of survival function is rarely available to enable an accurate hypothesis. Thus, nonparametric method is more appropriate to compare the survival functions in such situations. The Log-Rank Test proposed by Mantel (1966) is one of the tests available for comparison [22]. Several parametric methods of estimations are available to model survival data for which knowledge of the theoretical distribution ought to exist.

There are several theoretical distributions commonly considered for survival data, including Exponential, Generalized Gamma, Log Logistic and Weibull. Statistical tests are also available to compare two survival curves which follow the same distribution. For instance, Cox (1953) suggested an F-test for comparison of two exponential distributions without any restriction of censored observations [5]. To compare two survival times which follow Gamma distributions, Rao (1952) proposed a test which assumes that the ratio between the averages of both survival times follows F-distribution [26]. For uncensored survival times, Thoman, Bain and Antle (1969) and Thoman and Bain (1969) proposed a test by using maximum likelihood estimator [30, 29].

The main aim of this research is to study the survival curves of Dialysis Patients, by estimating and interpreting survivor functions from survival data by using the Kaplan-Meier estimates. And finally to fit an appropriate parametric model based on Anderson-Darling goodness of fit test. Survival time of Gender and patients greater than 50 years was also compared with less than 50 years in order to observe any significant difference in survival time of both groups.

2. Material and Methods

"Survival analysis is concerned with statistical modals and methods for analyzing data, representing life times, waiting times or more generally times to the occurrence of some specified event. Outcome variable of interest is time until an event occurs. Such data denoted as survival data, can arise in various scientific fields including medicine, engineering, and demography" [17]. Moreover survival data can attain from laboratory studies of animals or from clinical studies of humans who experience acute diseases. Survival data can comprises on survival time, response given treatment, and patient distinctiveness, allied with response and survival [16].

2.1 Censoring or Censored Data

A sample or observation is supposed to be censored, while by accident or design, the measurement of a random variable under study remains unobserved for several items in the sample. e.g., in a study, leukemia patients observed until they quit from state of the remission, if the period of study completed whereas the patient is still in remission (with no event) in that case the patient's survival time is referred as censored. And in this situation the survival time of individual is equal to the follow up period of that individual. However if the patient quit from remission, later than the study bring to an end, the total survival time of patient can not be identified [16].

When time-related variables take on in research, for example survival and recurrence, the researchers do not know the results for all patients after the time of study is fulfilled, consequently these results are known as censored [9]. For those individuals who remain alive at the end of study or whose survival status was unidentified the Survival time can not be determined and this sort of observations are referred as censored observations. The individuals who remain alive at the end of study are called as withdrawn alive and whose status could not be evaluated because they moved away or refuse to become a part of experiment, described as lost to follow-up [10].

Patients do not usually commence treatment or come in the study simultaneously. When the entrance of patients in the study is not at the same time and numerous patients are remaining in the study after the period of follow up, the data is considered as progressively censored or doubly censored [9].When the end point of interest has not so far occur and personnel's accurate survival time partially known at the right side of the follow-up period and study come to an end then this sort of data is known as Right-censored data. And the entire survival time which is incomplete, has been cut off (i.e. censored) from the right side [24].

Moreover, data may be left-censored. In left-censored data the survival time of a person is partially known at the left side of the follow-up period for that person. For instance, in a follow up study of HIV infection, if a person has already positive for the HIV virus before the start of follow-up duration, but the exact time of the first exposure to the virus is unknown. Since there is unidentified follow-up time from the time of primary exposure to the time of first positive HIV test, so the survival time is considered primarily censored on the left side [16].

The relevant data of the dialysis patients was acquired from Institute of Nephrology Sheikh Zaid Hospital Lahore, Pakistan. This hospital is a tertiary care centre and patients are referred not only from Punjab but also from other provinces of Pakistan. All sort of difficult and complicated cases are treated here. At the moment Components of Shaikh Zayed Medical Complex are Shaikh Zayed Hospital, Federal Postgraduate medical Institute, National health research Complex, Institute of Nursing and Health Sciences, Kidney Institute. Addition of the Institute of kidney Diseases as the 5th component of Shaikh Zayed Medical Complex is the Latest Achievement. Establishment of the Kidney Institute was approved in 1998. This is 220 beds faculty that will provide treatment to critically sick kidney patients including Kidney transplantation. The Institute will also provide opportunity for training of doctors at the Postgraduate level in the disciplines of Nephrology and Urology.

The dataset consisted of a total of 138 patient records, including both censored (84 patients) and uncensored (54 patients) observations. Both gender categories were fairly well represented in this population as 67 of all patients were males and 71 females. These patients were followed at monthly basis for up to 36 months or to the date of death which ever was earlier. In medical researches we have no prior information about the distribution of data. Hence, in our study, we obtained survival and hazard curves for gender and age groups by using Product-Limit Method. It enabled us to estimate survival and hazard functions in the existence of censored cases.

We performed our analysis into two steps and draw the conclusion after comparing the results. In the first step nonparametric technique, Product Limit Method is applied to estimate the survival functions and hazard rates and log rank test gives the statistical difference between the two survival functions. While in the second step, parametric approach starts with P-P Plot and the Anderson-Darling test to confirm the distribution of survival data. To find an appropriate distribution among the different available distributions is a painstaking job. After the immense effort it was found that the survival and hazard rates follow the Weibull Distribution in our data. The parameters of Weibull distribution are then estimated by Maximum Likelihood estimation method. Afterward survival functions are thus obtained by using the Weibull Distribution. Comparison of results from parametric and non-parametric tests suggested that the survival functions of Weibull Distribution and Product Limit Method supported the same conclusions. However, difference between the survival times of men and women were not statistically significant (at 5% level of significance), nor were the differences between patients in two age groups, ≤ 50 years and > 50 years.

2.2 Nonparametric approach

The Survivor function S(t) gives the probability that a person survives longer than some specified time t: that is, S(t) gives the probability that the random variable T exceeds the specified time t. In practice Survivor function is estimated as:

$$\hat{S}(t) = \frac{\text{Number of individuals in the sample who survive longer than t}}{\text{Total number of individuals in the sample}} \quad (1)$$

Survivor or Survivorship function is non increasing function of time t, denoted by S(t), is defined as:

$$S(t) = p(T \geq t) = 1 - F(t) \quad (2)$$

Moreover, KM estimator was also used to estimate the cumulative hazard function:

$$\hat{h}(t) = -\ln[\hat{s}(t)] \quad (3)$$

The Hazard function denoted by h(t), also called the instantaneous failure rate, conditional mortality rate, force of mortality and age specific failure rate. Hazard function give the risk of failure per unit time during the aging process or instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t. It is defined as the probability of failure during a very small time interval i to i+1, assuming that the individual has survived until time i. In practice the hazard function is estimated as the proportion of patients dying in an interval per unit time, given that they have survived to the beginning of the interval:

$$\hat{h}(t) = \frac{\text{Number of patients dying per unit time in the interval}}{\text{Number of patients surviving at t}} \quad (4)$$

The hazard function can be expressed in terms of the cumulative distribution function F(t) and the probability density function f(t):

$$h(t) = \frac{f(t)}{\{1 - F(t)\}} \quad (5)$$

The description of the diversity in two or more groups of the estimated survival time distributions and the plots of the survival rates are simply start of the survival analysis. Besides these descriptive guidelines, researchers require a statistical test to conclude that, these differences are statistically significant or caused by "chance variation". As, a short dissimilarity perhaps statistically significant due to the large sample size and a large dissimilarity possibly will not, in the small sample size. So it is expected to execute a statistical test to calculate the amount of differences between two survival curves. The Log-Rank Test is used to compare the difference between the two survival curves.

2.3 Parametric Approach

For choosing a basic survival distribution, simplest and efficient tool is graph which has long been used for display and interpretation of data. When data consists of censored and uncensored observations, then appropriate graphical technique is 'Hazard Plotting'. Hazard Plotting' is a subjective method based on visual examination in which observations are plotted against cumulative hazard values. To determine the appropriate probability distribution of the survival data amongst the available distributions we started with the graphical approach of P-P plot technique for modeling the hazard rates [18]. Subsequently, The Anderson-Darling Test (Stephens, 1974) is employed to test for Distributional Adequacy that a sample of data comes from a population with a specific distribution [28]. This is test which is more sensitive to deviations in the tails of the distribution and utilizes the specific distribution being tested, in computing critical values.

H_0: The data follows a specified distribution.
H_a: The data do not follow a specified distribution.
Test Statistic: $A^2 = -N - S$

Where:

$$S = \sum_{i=1}^{N} \frac{(2i-1)}{N} \left[Log\, F(Y_i) + Log(1 - F(Y_{N+1-i})) \right]$$

Where N is the sample size, F is the cumulative distribution function of specified distribution and Y_i are the ordered data. For the Weibull distribution cumulative distribution function is:

$$F(y_i) = 1 - [\exp - (\lambda y)^{\gamma}]$$

and λ, γ are scale and shape parameters respectively.

2.4 The Weibull Distribution

There are numerous commonly used survival distributions for instance, normal, log normal, exponential, weibull, gamma, logistic, extreme value type 1 distribution and fatigue life distribution. The entire distributions have been used extensively in recent years to deal with reliability of components, lifetimes, learning times, duration of epidemics, traveling times, material strengths, particle dimensions, radioactive intensities, rainfall amounts, and costs of industrial accidents. Weibull distribution is one of these lifetime distributions, named after the Swedish physicist Waloddi Weibull (1939) [32]. The weibull distribution with two parameters has hypothetical confirmations in technical as well as biomedical applications as a purely empirical model. Also it is a very flexible life distribution model and widely used parametric survival model for modeling failure times [11].
The Weibull distribution is defined as:
The continuous random variable t has a weibull distribution, with parameters λ and γ if its density function is given by:

$$f(t)=\gamma\lambda^{\gamma}t^{\gamma-1}\exp[-(\lambda t)^{\gamma}] \quad t \geq 0, \lambda > 0, \gamma > 0 \tag{6}$$

And cumulative distributions function has the form:

$$F(t)=1-[\exp-(\lambda t)^{\gamma}] \tag{7}$$

Survivorship function and Hazard function is, therefore:

$$S(t) = \exp[-(\lambda t)^{\gamma}] \tag{8}$$

$$h(t) = \gamma\lambda(\lambda t)^{\gamma-1} \tag{9}$$

Weibull distribution has two parameters, where γ is the shape parameter and λ is scale parameter. At γ <1, the failure rate decreases over time, at γ =1 failure rate remain constant over time, and at γ > 1 failure rate increases over time [18, 31].

2.4.1 Estimation of Parameters of fitted Distribution with Censored Observations

Analytical procedures for estimating the most commonly used survival distributions are concerned with maximum likelihood estimates of the parameters of the exponential, Weibull, Lognormal, and Gamma distribution. In this study, we used the method of Maximum likelihood estimation to estimate parameters of the Weibull theoretical distribution.

Density function of weibull distribution is:

$$f(t) = \gamma\lambda^{\gamma}t^{\gamma-1}\exp[-(\lambda t)^{\gamma}] \quad t \geq 0, \gamma > 0, \lambda > 0 \tag{10}$$

Cohen (1965) and some other writers write the density function as:

$$f(t) = \frac{\gamma}{\theta}\, t^{\gamma-1}\exp[-(\frac{t^{\gamma}}{\theta})] \quad t \geq 0, \gamma > 0, \theta > 0 \tag{11}$$

As θ in equation (11) is equal to $\frac{1}{\lambda^{\gamma}}$. Estimation of λ and γ in equation (10) is equivalent to estimation of θ and γ in equation (11), and λ can be obtained from θ and γ.

An experiment in which subjects are entered at different time and the experiment terminated after the fixed period of time. In this case the collected data is progressively censored data.

The progressively censored ordered survival data are:

$$t_{(1)} \leq t_{(2)} \leq \ldots\ldots \leq t_{(r)}, t^{+}_{(r+1)}, \ldots\ldots, t^{+}_{(n)}$$

If the survival distribution is Weibull with density of equation (11), the MLE of θ and γ can determined by solving the following two equations simultaneously.

$$\frac{r}{\hat{\gamma}} + \sum_{i=1}^{r} \log_e t_{(i)} - \frac{1}{\hat{\theta}} \left[\sum_{i=1}^{r} t_{(i)}^{\hat{\gamma}} \log_e t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+\hat{\gamma}} \log_e t_{(i)}^{+} \right] = 0 \tag{12}$$

and

$$-\frac{\mathrm{r}}{\hat{\theta}} + \frac{1}{\hat{\theta}^2} \left[\sum_{i=1}^{r} t_{(i)}^{\hat{\gamma}} + \sum_{i=r+1}^{n} t_{(i)}^{+\hat{\gamma}} \right] = 0 \tag{13}$$

By eliminating $\hat{\theta}$ between above two equations and simplifying we have:

$$W(\hat{\gamma}) = \frac{\sum_{i=1}^{r} t_{(i)}^{\hat{\gamma}} \log_e t_{(i)} + \sum_{i=r+1}^{n} t_{(i)}^{+\hat{\gamma}} \log_e t_{(i)}^{+}}{\sum_{i=1}^{r} t_{(i)}^{\hat{\gamma}} + \sum_{i=r+1}^{n} t_{(i)}^{+\hat{\gamma}}} - \frac{1}{r} \sum_{i=1}^{r} \log_e t_{(i)} - \frac{1}{\hat{\gamma}} = 0 \tag{14}$$

These equations can be solved iteratively for the MLE of $\hat{\gamma}$ by using the Newton-Raphson method. The Newton-Raphson method is a numerical iterative procedure that can be used to solve nonlinear equations [25, 1]. Iterative procedure is a technique of successive approximations and each approximation is called iteration. If the successive approximations approach the solution very closely than we say that iterations converge.

Now an initial estimate of γ is required to solve equation (14) iteratively. If the initial estimate is close to the real value of γ than small number of iterations are needed. Cohen (1965) recommends a graphical method for initially estimating of γ, by using the fact that the coefficient of variation (CV) of the weibull distribution is a function of parameter γ [2]. The sample coefficient of variation is determined by:

$$\mathrm{CV} = \frac{s}{\bar{t}} \tag{15}$$

Where s and $\bar{t}$ are the sample standard deviation and mean. CV can be plotted against different values of γ. By using sample coefficient of variation initial estimate of γ can be read from the graph. The obtained value should provide a satisfactory initial estimate for starting the iterative procedure for the MLE [18].

An initial estimate of γ for solving equation (14) iteratively can be obtained by considering all censored observations as exact. The sample coefficient of variation calculated by using all of the n observations, and value of γ can be read from the graph.

Similarly after $\hat{\gamma}$ determined, $\hat{\theta}$ can be estimated by:

$$\hat{\theta} = \frac{1}{r} \left[\sum_{i=1}^{r} t_{(i)}^{\hat{\gamma}} + \sum_{i=r+1}^{n} t_{(i)}^{+\hat{\gamma}} \right] \tag{16}$$

Therefore λ in equation (10) may be estimated by:

$$\hat{\lambda} = \exp(-\frac{\log_e \hat{\theta}}{\hat{\gamma}}) \qquad (17)$$

2.4.2 Comparison of two Theoretical Survival Functions

For comparing two theoretical survival functions, following a particular model with their parameters, two tests can compare the distributions i.e. the likelihood ratio test and an F-test suggested by Cox (1952) .These two tests can test the hypothesis that, whether the two distributions are equal or not. These tests are also applicable for censored observations. Thoman and Bain (1969) proposed a test for uncensored samples, used for comparing two weibull distributions, with Maximum Likelihood Estimator [29]. To test the equality of two weibull distributions, the null and alternative hypotheses are: H_0: $\gamma_1 = \gamma_2$ Vs. H_1: $\gamma_1 > \gamma_2$

In case of rejection of null hypothesis, it is clear that the two weibull distributions are not come to similar. Conversely, if no proper evidence provided to reject the null hypothesis, then we also have to check the equality of two scale parameters λ_1 and λ_2, under the null and alternative hypothesis: H_0: $\lambda_1 = \lambda_2$ against H_1: $\lambda_1 < \lambda_2$ [18].

3. Results & Discussion

The death rate (of 60.9%) among the dialysis patients was notable in this population. The survival rates by using the Kaplan-Meier estimates for males and females are given in Appendix as Table 1. Overall comparison of the males and females is done by using Kaplan-Meier estimates. For comparison in mean ages and survival time in both gender t-test for two independent samples was applied and results depicts no statistical differences as p-values were greater than 0.05. Mean survival time for males and females were noted as with ±SD (18.58±12.59; 16.01±11.66) respectively. Survival time is relatively high in case of males but not statistically different from females. Figure 1 presents the Survival Curves comparing male and female survival rates.

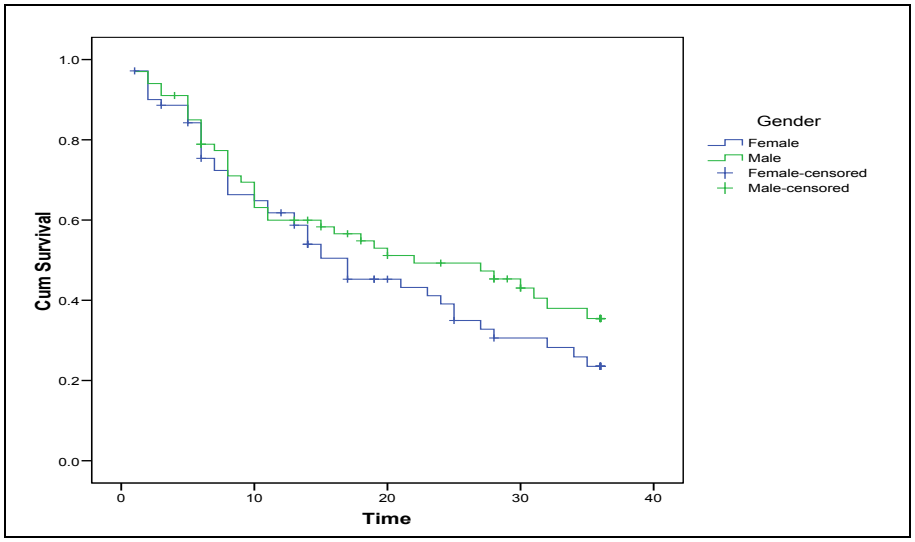


Figure 1. Survival Curves for Gender Group .

Table1. Survival and Hazard Rates by using Kaplan-Meier Method and Survival Rates by using Weibull Distribution.

Females				Males			
Survival Time (months)	Survival Rate K – M	Hazard Rate K – M	Survival Rate Weibull	Survival Time (months)	Survival Rate K – M	Hazard Rate K – M	Survival Rate Weibull
1	0.9718	0.0285	0.9697	1	0.9701	0.0303	0.9684
1	0.9718	0.0285	0.9697	1	0.9701	0.0303	0.9684
2	0.9003	0.1049	0.9367	2	0.9403	0.0615	0.9384
2	0.9003	0.1049	0.9367	2	0.9403	0.0615	0.9384
2	0.9003	0.1049	0.9367	3	0.9104	0.0938	0.9096
2	0.9003	0.1049	0.9367	3	0.9104	0.0938	0.9096
2	0.9003	0.1049	0.9367	5	0.8497	0.1628	0.8551
3	0.8860	0.1209	0.9032	5	0.8497	0.1628	0.8551
5	0.8425	0.1713	0.8373	5	0.8497	0.1628	0.8551
5	0.8425	0.1713	0.8373	5	0.8497	0.1628	0.8551
5	0.8425	0.1713	0.8373	6	0.7890	0.2369	0.8292
6	0.7538	0.2826	0.8053	6	0.7890	0.2369	0.8292
6	0.7538	0.2826	0.8053	6	0.7890	0.2369	0.8292
6	0.7538	0.2826	0.8053	6	0.7890	0.2369	0.8292
6	0.7538	0.2826	0.8053	7	0.7732	0.2571	0.8042
6	0.7538	0.2826	0.8053	8	0.7101	0.3422	0.7800
6	0.7538	0.2826	0.8053	8	0.7101	0.3422	0.7800
7	0.7236	0.3234	0.7740	8	0.7101	0.3422	0.7800
7	0.7236	0.3234	0.7740	8	0.7101	0.3422	0.7800
8	0.6633	0.4104	0.7435	9	0.6943	0.3647	0.7565
8	0.6633	0.4104	0.7435	10	0.6312	0.4600	0.7338
8	0.6633	0.4104	0.7435	10	0.6312	0.4600	0.7338
8	0.6633	0.4104	0.7435	10	0.6312	0.4600	0.7338
10	0.6482	0.4334	0.6853	10	0.6312	0.4600	0.7338
11	0.6181	0.4810	0.6575	11	0.5996	0.5113	0.7118
11	0.6181	0.4810	0.6575	11	0.5996	0.5113	0.7118
13	0.5872	0.5323	0.6047	15	0.5830	0.5395	0.6305
13	0.5872	0.5323	0.6047	16	0.5658	0.5693	0.6117
14	0.5396	0.6169	0.5797	18	0.5481	0.6011	0.5758
14	0.5396	0.6169	0.5797	19	0.5299	0.6350	0.5587
14	0.5396	0.6169	0.5797	20	0.5116	0.6701	0.5422
15	0.5048	0.6836	0.5555	22	0.4927	0.7078	0.5105
15	0.5048	0.6836	0.5555	27	0.4729	0.7486	0.4393
17	0.4525	0.7928	0.5097	28	0.4532	0.7912	0.4264
17	0.4525	0.7928	0.5097	30	0.4306	0.8425	0.4016
17	0.4525	0.7928	0.5097	31	0.4052	0.9031	0.3897
21	0.4320	0.8393	0.4281	32	0.3799	0.3799	0.3783
23	0.4114	0.8881	0.3919	35	0.3546	0.3546	0.3458
24	0.3908	0.9394	0.3748				
25	0.3497	1.0506	0.3585				
25	0.3497	1.0506	0.3585				
27	0.3278	1.1151	0.3277				
28	0.3060	1.1841	0.3132				
32	0.2824	1.2642	0.2611				

Hazard function gives the risk of failure per unit time since the date of first dialysis. Graph of hazard function of two gender groups is given in Figure 2.

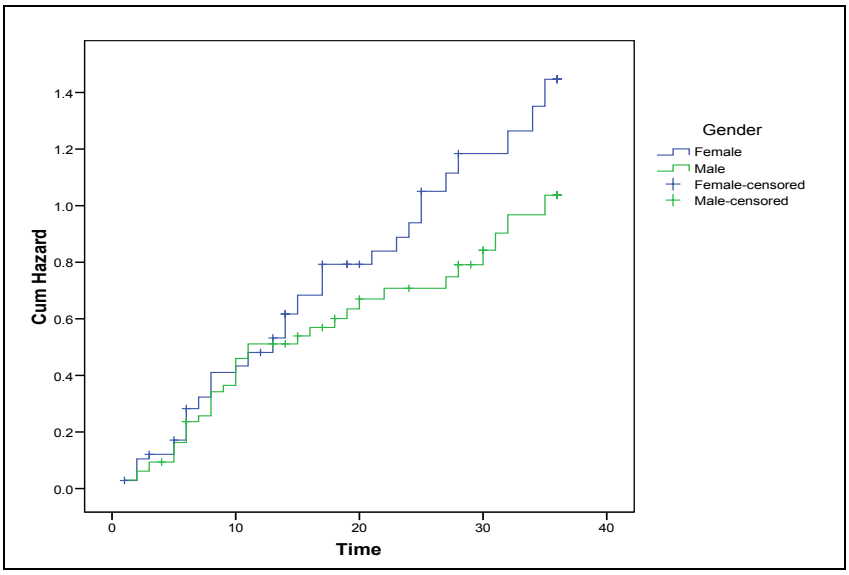
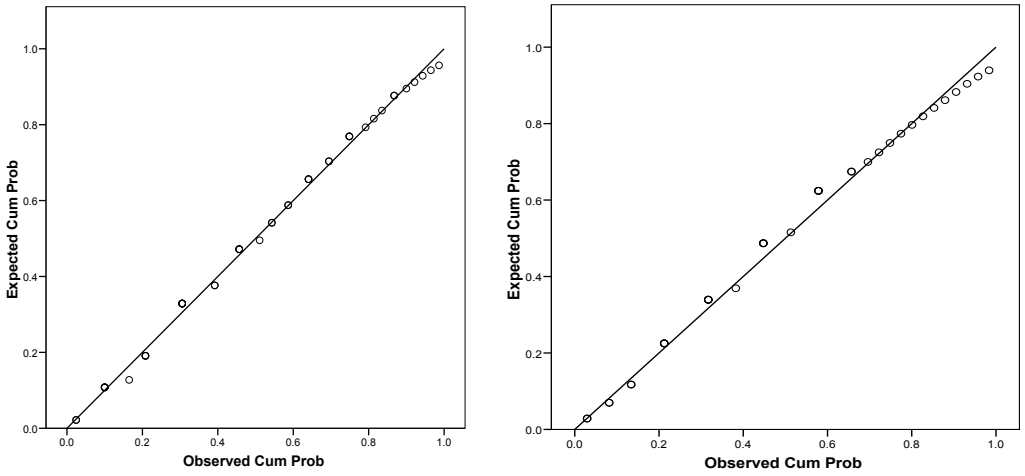


Figure 2. Hazard Curves of for Gender Group.

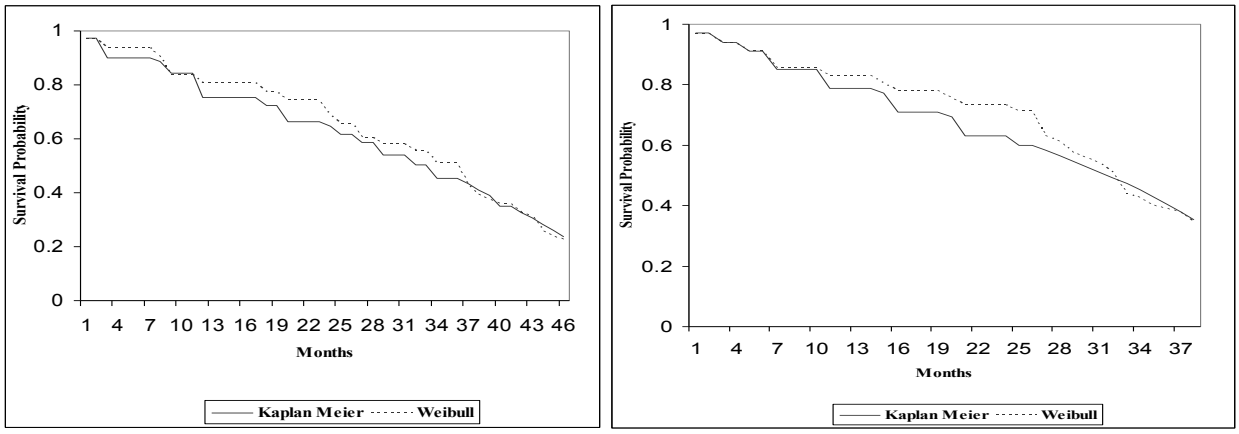
Figure 1 shows that for about 13 months of the follow up, the probability of survival for members of both genders remains same but after that time, the survival time improves in males compared to females. Similarly, Figure 2 demonstrates that after passing the same months of disease, the hazard rate increases in females .Consequently, the widening gap between the both curves suggests that the survival of Males group improved later during follow-up than it is early on and risk of disease is elevated in females than males. Overall, by Log-Rank test the P-value (0.223) indicates that the survival difference between two gender groups is not statistically significant.



Females Males

Figure 3. Weibull P-P Plot for Females / Males.

The P-P plot indicates that the hazard rates of males and females follow the Weibull distribution. Weibull P-P Plot of Females and Males are presented in Figure 3. The Weibull P-P Plot of Hazard Functions for both gender categories showed linearity and seemed a better fit as compared to other distributions. The P-Values for Anderson-Darling test are greater than 0.05 for both gender categories, indicating that the survival functions follow the Weibull Distribution.
The scale and shape parameters of Weibull distribution are estimated by the method of Maximum Likelihood.



Females Males

Figure 4. Survival Curves of Kaplan-Meier and Weibull for Females/Males.

For the Females group these values are $\gamma_1 = 1.09$ and $\lambda_1 = 0.040$. For the Males group these values are $\gamma_2 = 0.984$ and $\lambda_2 = 0.030$. Figure 4 graphically display the estimated survival distribution of Kaplan-Meier compared with Weibull distribution. Figure supports the claim that a Weibull distribution is an effective description to model the data of concern to our study.
The Figure 4 shows that the median survival time for females when calculated from Kaplan-Meier is 17 months and from Weibull curve, it is 36 months.

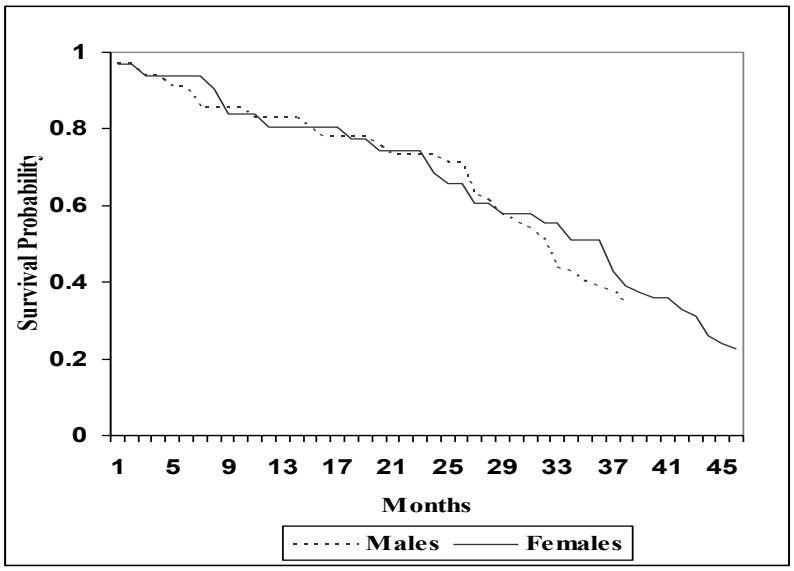


Figure 5. Comparison of Survival Functions of Theoretical Distribution for Gender Group.

In contrast, the median survival time for males, when calculated from Kaplan-Meier is 22 months and from Weibull it is 32 months. Graphical comparison of survival functions for Gender groups calculated by theoretical distribution is shown in Figure 5.
For testing the equality of two Weibull distributions $f_1(t)$ and $f_2(t)$, respectively for Males and Females, the null and alternative hypotheses are:

H_0: $\gamma_1 = \gamma_2$
H_1: $\gamma_1 > \gamma_2$.

The test statistic is $\hat{\gamma}_1 / \hat{\gamma}_2$ = 1.1122, which is insignificant at the 0.05 level ($\upsilon_{0.05}$ = 1.332 for n= 42). It provided no support to null hypothesis. Therefore, we had to check the equality of two scale parameters λ_1 and λ_2, under the hypothesis: H_0: $\lambda_1 = \lambda_2$ against H_1: $\lambda_1 < \lambda_2$. The decision rule is to reject H_0 if,

$$G = \frac{\hat{\gamma}_1 + \hat{\gamma}_2}{2}(\log_e \hat{\lambda}_2 - \log_e \hat{\lambda}_1) > Z_\alpha$$

Where G = -0.2983, which is also insignificant at the 0.05 level ($\upsilon_{0.05}$ = 0.394 for n= 42).

3.1 Analysis for comparing Age groups

Now the same procedure is repeated for age groups (≤50 & >50) to see if significant differences exist between these groups. The survival rates by using the Kaplan-Meier estimates for the age groups "less than or equal to 50" and "greater than 50" are given in the Appendix as Table 2[TABLE 2 HERE]. The cutoff point for dividing the patients into two age groups was taken as 50 years because the average age of the patients was 50 years. The Survival Curves of both Age groups, using Kaplan-Meier estimates are given in Figure 6.

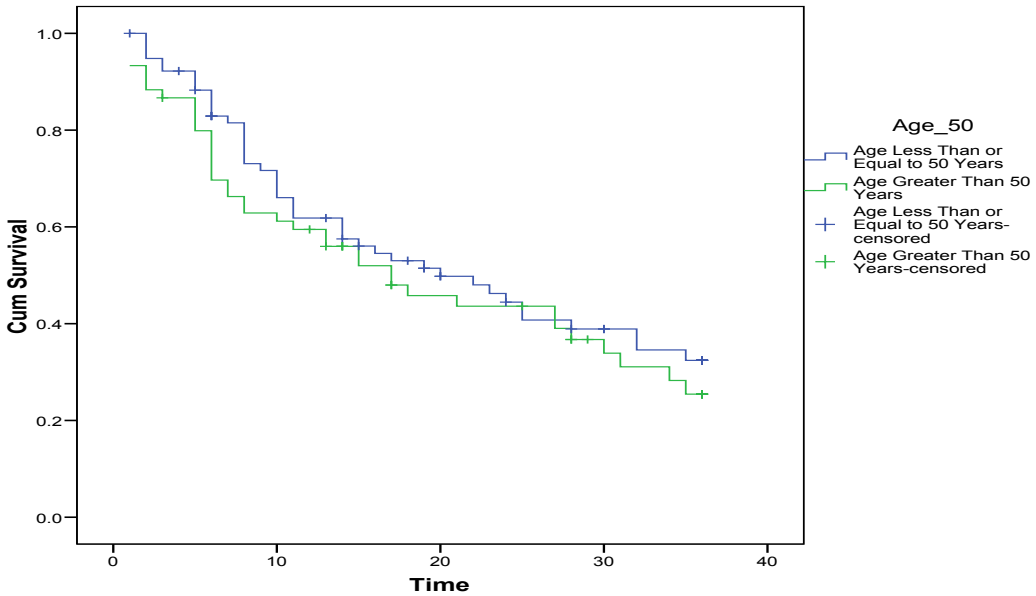
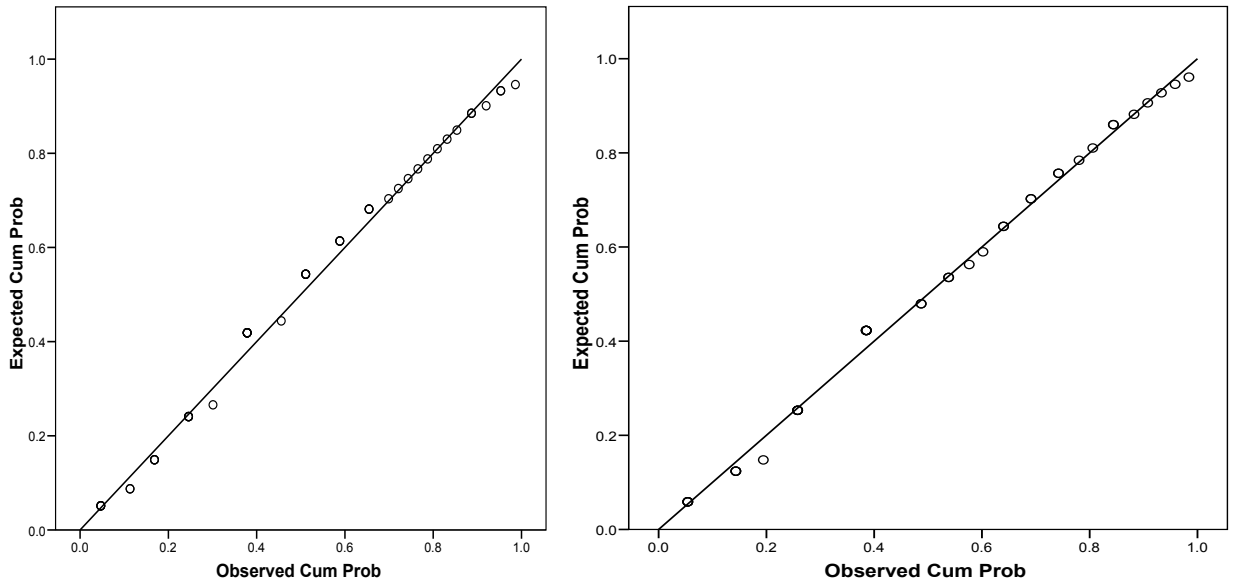


Figure 6. Survival Curves of Patients of Age Groups.

Table 2. Survival Rates by Kaplan-Meier Method and Survival Rates by using Weibull Distribution for Age Groups -50.

Age Group -Less than 50				Age Group -Greater than 50			
Survival Time (months)	Survival Rate K – M	Hazard Rate K – M	Survival Rate Weibull	Survival Time (months)	Survival Rate K – M	Hazard Rate K – M	Survival Rate Weibull
2	0.9481	0.0534	0.9525	1	0.9333	0.0690	0.9555
2	0.9481	0.0534	0.9525	1	0.9333	0.0690	0.9555
2	0.9481	0.0811	0.9525	1	0.9333	0.1241	0.9555
2	0.9481	0.1249	0.9525	1	0.9333	0.1241	0.9555
3	0.9221	0.1875	0.9262	2	0.8833	0.1241	0.9159
3	0.9221	0.2045	0.9262	2	0.8833	0.2248	0.9159
5	0.8826	0.3137	0.8731	2	0.8833	0.2248	0.9159
5	0.8826	0.3137	0.8731	3	0.8667	0.3614	0.8788
5	0.8826	0.3137	0.8731	5	0.7987	0.3614	0.8106
6	0.8291	0.4148	0.8467	5	0.7987	0.3614	0.8106
6	0.8291	0.4808	0.8467	5	0.7987	0.3614	0.8106
6	0.8291	0.5531	0.8467	5	0.7987	0.3614	0.8106
6	0.8291	0.5531	0.8467	6	0.6967	0.4114	0.779
7	0.815	0.5531	0.8206	6	0.6967	0.4640	0.779
8	0.7307	0.5791	0.7949	6	0.6967	0.5196	0.779
8	0.7307	0.6347	0.7949	6	0.6967	0.5802	0.779
8	0.7307	0.7714	0.7949	6	0.6967	0.5802	0.779
8	0.7307	0.8106	0.7949	6	0.6967	0.6543	0.779
8	0.7307	0.8977	0.7949	7	0.6627	0.7344	0.7489
8	0.7307	0.8977	0.7949	7	0.6627	0.7344	0.7489
9	0.7167	1.0620	0.7696	8	0.6288	0.8297	0.7202
10	0.6604	1.1265	0.7448	8	0.6288	0.9409	0.7202
10	0.6604	0.0534	0.7448	10	0.6118	1.0015	0.6664
10	0.6604	0.0534	0.7448	11	0.5948	1.2639	0.6412
10	0.6604	0.0811	0.7448	13	0.5598	0.0690	0.594
11	0.6183	0.1249	0.7205	13	0.5598	0.0690	0.594
11	0.6183	0.1249	0.7205	15	0.5198	0.1431	0.5506
11	0.6183	0.3137	0.7205	15	0.5198	0.2248	0.5506
14	0.5752	0.3137	0.6509	17	0.4798	0.2248	0.5106
14	0.5752	0.3137	0.6509	17	0.4798	0.3614	0.5106
14	0.5752	0.3332	0.6509	18	0.458	0.4114	0.4918
15	0.5604	0.4148	0.6288	21	0.4362	0.4640	0.4396
16	0.5453	0.4148	0.6074	27	0.3903	0.4914	0.3522
17	0.5301	0.4148	0.5865	27	0.3903	0.6543	0.3522
19	0.5145	0.4808	0.5464	28	0.3673	0.7809	0.3395
20	0.4979	0.4808	0.5273	30	0.3391	0.9409	0.3155
22	0.4801	0.6065	0.4906	31	0.3108	1.0816	0.3042
23	0.4624	0.6645	0.4731	34	0.2826	1.1686	0.2727
24	0.4446	0.6973	0.4562	35	0.2543	1.3692	0.2629
25	0.4075	0.7337	0.4397				
25	0.4075	0.9442	0.4397				
28	0.389	0.1875	0.3935				
32	0.3458	0.1875	0.3386				
32	0.3458	0.1875	0.3386				

The Log-Rank test was applied to check whether the hazard difference in two groups was significant. The P-value (0.382) indicates that the difference between two groups is insignificant. Our results indicate that the hazard rates of both age groups again followed the Weibull distribution. Weibull P-P Plot of Age group less than or equal to 50, and greater than 50 years are shown in Figure 7.



Age group less than 50 Age group greater than 50

Figure 7. Weibull P-P Plot for Age Group (<= 50 and >50).

The Weibull P-P Plot of Hazard Functions for both groups shows linearity and seems a better estimator as compared to other distributions. The P-value (0.200) for age group less than or equal to 50 years and the P-value (0.250) for age group greater than 50 years indicates that the survival functions follow the Weibull Distribution.

The scale and shape parameters of Weibull distribution are estimated by the method of Maximum Likelihood estimation. For the Age group less than or equal to 50 years, these values are $\gamma_1 = 1.119$ and $\lambda_1 = 0.033$. For the Age Group-Greater than 50 years these values are $\gamma_2 = 0.951$ and $\lambda_2 = 0.038$. Figure 8 displays the estimated survival distribution of Kaplan-Meier compare with Weibull distribution.

From the figure 8 it is clear that the median survival time for age group less than 50 years calculated from Kaplan-Meier is 20 months and from Weibull it is 36 months. Also figure 8 shows that the median survival time for age group greater than 50-years calculated from Kaplan-Meier is 17 months and from Weibull is 29 months.

To test the equality of two Weibull distributions $f_1(t)$ and $f_2(t)$, of both Age-50 groups, it is adequate to test the hypotheses: H_0: $\gamma_1 = \gamma_2$ Vs. H_1: $\gamma_1 > \gamma_2$. Test statistic is $\hat{\gamma}_1 / \hat{\gamma}_{21} = 1.1766$, which is insignificant at the 0.05 level ($\upsilon_{0.05} = 1.332$ for n= 42). So this provided no proper evidence to reject the null hypothesis, then have to check the equality of two scale parameters λ_1 and λ_2, under the hypothesis: H_0: $\lambda_1 = \lambda_2$ against H_1: $\lambda_1 < \lambda_2$ and will reject H_0 if,

$$G = \frac{\hat{\gamma}_1 + \hat{\gamma}_2}{2}(\log_e \hat{\lambda}_2 - \log_e \hat{\lambda}_1) > Z_\alpha$$

Where G = 0.1510, which is also insignificant at the 0.05 level ($\iota_{0.05}$= 0.394 for n= 42).

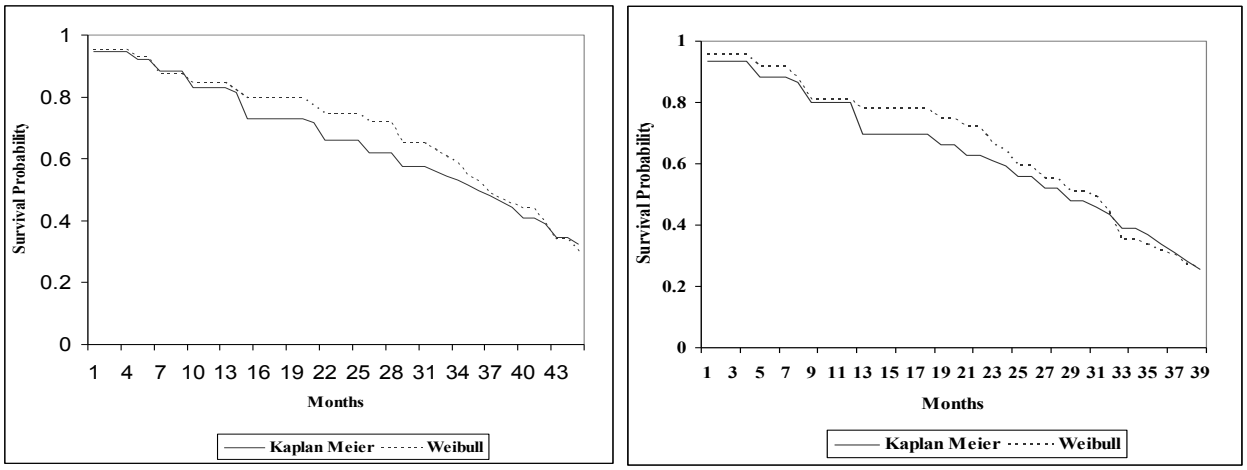


Figure 8. Survival Curves of Kaplan-Meier Method and Weibull for Age Group- Less than 50 and greater than 50 years.

People of this region become victim of ESRD at younger age because detection is not made in time and preventive measures are not adapted. The median age of ESRD population is calculated to be 44 years [27, 15]. Mostly people go to nephrologists late in the course of their disease.73% of patients consult nephrologists almost when they have developed ESRD [21]. Generally ESRD prevalent population opt the hemodialysis therapy and hemodialysis facilities are going better in numerous areas of the country. Still, the majority of patients expired or discontinues the therapy within the first three months due to economic limitations.

In general cohort studies follow individuals for long period of time, which is time consuming and costly to perform. In this study three year time period to follow up Dialysis patients is carried out. A cohort of patients of single hospital with the same clinical conditions in a dialysis unit of ION- Sheikh Zayed hospital Lahore was taken out, and followed maintenance hemodialysis patients, to see how many patients experienced death. Results of almost same survival are arrived at, despite differentiation of age and gender due to limited time. As late-diagnosed dialysis treatment, infections and malnutrition are mostly causes of high mortality and leads to less survival chances of patients. Survival time is almost same in patients of different Age groups and Gender group provided that general health, nutritional status and quality of life of patients is not much fluctuate in this data set.

For further study of greater than three years time periods and consulting more than one hospital, describing the individuals with different clinical conditions and different modalities are required for this population. Result from such a cohort study would have more accurately reflect the outcomes that would be seen in clinical practice, and long follow-up period might have resulted in more occurrences of events and reduced number of censored observations in the data.

4. Conclusion

Survival functions for Males, Females Group, Age Group cut-of point at 50 were estimated by Kaplan-Meier estimates, and the equality of two survival functions in each group is tested by the Log-Rank test. The testing result suggests that each two survival functions are statistically insignificant at 5% level of significance. After selecting the theoretical Weibull Distribution we estimated the survival functions again. The Thoman and Bain test and Log rank test lead to same conclusions. The median survival time of both Gender and Age group are different for the current sample data but these results can not generalize to all the population. Overall it can be concluded that survival time of both gender and age (cut-off point at 50 years) are approximately identical.

References

[1]. Carnahan, B., Luther, H. A., and Wilkes, J. O. (1969). *Applied numerical Methods*. New York: John Wiley & Sons.

[2]. Cohen, A.C. (1965). Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples. *Technometrics*, 7, 579-588.

[3]. Cooper, J. (1999). *US incidence of kidney failure is the highest in the world, The Medical Reporter*. URL: http://medicalreporter.health.org/tmr0799/kidney.html.

[4]. Coresh, J., E. Selvin, et al. (2007). Prevalence of chronic kidney disease in the United States. *JAMA: the journal of the American Medical Association,* 298, 2038- 2047.

[5]. Cox D.R. (1953). Some Simple Tests for Poisson Variates. *Biometrika*, 40, 354–360.

[6]. Chadban, S.J., Briganti, E.M., Kerr, P.G., Dunstan, D.W., Welborn, T.A., Zimmet, P.Z., & Atkins, R.C. (2003). Prevalence of kidney damage in Australian adults: The AusDiab kidney study. *Journal of the American Society of Nephrology, 14*, 131-138.

[7]. Chen, J., Wildman, R. P., Gu, D., Kusek, J. W., Spruill, M., Reynolds, K., Liu,D., Hamm, L.L., Whelton, P.K., & He, J. (2005). Prevalence of decreased kidney function in Chinese adults aged 35 to 74 years. *Kidney international, 68*, 2837-2845.

[8]. DE ZEEUW, D., Hillege, H. L., & DE JONG, P. E. (2005). The kidney, a cardiovascular risk marker, and a new target for therapy. *Kidney international, 68*, 25-29.

[9]. Dawson, B., Trapp, R.G. (2004). *Basic & clinical biostatistics*. 4rth edition. New York: McGraw-Hill.

[10]. Forthofer, R.N., Lee, E.S. (2006). *Introduction to biostatistics*: *Aguid to Design, Analysis and Discovery*. California, USA: Academic Press.

[11]. Gleser, L.J. (1980). *Probability models and applications*. New York: Macmillon Publishing.

[12]. Hallan, S. I., Coresh, J., et al. (2006). International comparison of the relationship of chronic kidney disease prevalence and ESRD risk. *Journal of the American Society of Nephrology*, 17, 2275-2284.

[13]. Jafar, T. H. (2006). The growing burden of chronic kidney disease in Pakistan. *New England Journal of Medicine,* 354, 995-997.

[14]. Kaplan, E.L., Meier, P. (1958). Nonparametric Estimation from incomplete Observations. *Journal of the American Statistical Association*, 53, 457–481.

[15]. Kher, V. (2002). End stage renal disease in developing countries. *Kidney Int*, 62, 350–362.
[16]. Kleinbaum, D.G. (1996). *Survival analysis*. New York: Springer-Verlag.
[17]. Read, C. B., Kotz, S., & Johnson, N. L. (1988). *Encyclopedia of statistical sciences*. Canada: A Wiley-Interscience Publication.
[18]. Lee, E. T. (2003). *Statistical Methods for Survival Data Analysis*. New York: John Wiley & Sons.
[19]. Leung, K. M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual Review of Public Health, 18*, 83-104.
[20]. Mani, M.K. (1998). The management of end-stage renal disease in India. *Artificial Organs*, 22, 182–186.
[21]. Mantel, N. (1966). Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration. *Cancer Chemotherapy Reports*. 50, 163–170.
[22]. Naqvi, S.A.J. (2009). Renal Diseases in Pakistan-'Time to act'. *Journal of Nephrology and Renal Transplantation, 2*, 133-135.
[23]. Petrie, A., & Sabin, C. (2009). *Medical statistics at a glance*. Vol. 29: Wiley-Blackwell.
[24]. Ralston, A., Wilf, H. (1967). *Mathematical Methods for Digital computers*. New York: John Willey.
[25]. Rao C.R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: John Wiely & Sons.
[26]. Rizvi, S. (1998). Present state of dialysis and trans-plantation in Pakistan. *merican journal of kidney diseases: the official journal of the National Kidney Foundation,* 31, xiv–xiviii.
[27]. Stephens, M.A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, 69, 730-737.
[28]. Thoman, D.R., Bain, L.J. (1969). Two sample tests in the Weibull Distribution. *Technometrics*, 11, 805–815.
[29]. Thoman, D.R., Bain, L.J., and Antle, C.E. (1969). Inferences on the Parameters of the Weibull Distribution. *Technometrics*. 11, 445–460.
[30]. Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K. (2005). *Probability statistics for engineers & statistics*. Seventh edition. New Delhi: Pearson Education.
[31]. Weibull W. (1939). A statistical Theory of the Strength of Materials. *Ingeniors Vetenskaps Akademiens*, Stockholm 151.
[32]. Rao, M., Juneja, R., Shirly, R.B., Jacob, C.K. (1998). Haemodialysis for endstage renal disease in southern India—a perspective from a tertiary referral care center. *Nephrol Dial Transplant*, 13, 2494–2500.