بسم الله الرحمن الرحيم

Sudan University of Science and Technology

College of Graduate Studies

Estimating and Analyses the hemodialysis Patients from Date of Diagnosis until Death Using Parametric Models -The Governmental Hospitals in Khartoum State (2015 -2005)

استخدام النماذج المعلمية لتقدير وتحليل مرضى غسيل الكلى من تاريخ التشخيص وحتي الوفاة -المستشفيات الحكومية بولاية الخرطوم 2015- 2005

A Thesis Submitted for Ph.D in Statistics

Prepared by:

Reem Yousif Mekki Daladom

Supervisor:

Prof. Mohamed Hassan Mudawi
Professor of Statistics

Co. Supervisor:

Dr.Manahil Saidahmed Mustafa
Assistant Professor of Statistic

Date: 8/2022

﴿وَمَا أُوتِيتُم مِّنَ الْعِلْمِ إِلَّا قَلِيلًا﴾

(سورة الإسراء –الآية 85)

DEDICATION

To my parents

To my daughter

To my brothers and sisters

To my friends

To my colleagues

To all whom I love them

.

ACKNOWLEDGEMENTS

- Praise is to ALLAH who helped me to complete this search.
- My special thanks and appreciation to Dr. Mohmmed Hassan Mudawi which the kindly supervising this research.
- Thanks to Dr.Manahil Sid Ahmed Mustafa and Dr.Altaiyb Omer Ahmed Mohmmed.
- Thanks to all teachers in Department of Applied Statistics.
- Thanks to everyone who contributed to the success of this study
- Thanks to the patients and the staff of all the hemodialysis centers in Khartoum State/Sudan, for contribution in this study.
- I am grateful for all who stood beside me and gave me the support to accomplish this research.

ABSTRACT

This research aimed at comparing different models of parametric Proportional Hazards (PH) models mainly (Weibull, Exponential, Gompertz) and Accelerated Failure Time (AFT) models mainly (Lognormal and log logistic) in patients with hemodialysis to determine the best model for assessing the survival of patient and identify significant risk factors for mortality. Recently in Sudan the end –stage renal disease (ESRD) has become a major health problem .The Study consists of 325 hemodialysis patients who were collected from the records at governmental hospitals in Khartoum State in the period from December 2005 to December 2015. Data was used to estimate the survival function with view to identify risk factors such as (age (date of diagnosis of the disease) , Sex , Marital Status, Education status, occupation, Address, regular , Dialysis frequency per week, Hospitals , Diabetes Mellitus, Hypertension , polycystic kidney disease, Renal obstructions, Shrunken kidneys, Uncertain, Other) influencing among the end-stage renal disease (ESRD) population. The result show that the univariate and multivariate analysis, According to hazard ratio and time ratio, the variables including age, diabetes mellitus, diabetes mellitus +hypertension, urea and serum creatinine were considered to be highly significant factors and increased the risk of death in patients (shorter survival) so that they could influence survival in hemodialysis patients in the five models used in this research. Whereas other factors, such as regular, hospital, hypertension, shrunk kidneys, dialysis frequency per week, other have decreased the risk of death (longer survival) and have a direct effect on the survival of the hemodialysis patient. The median overall survival time was estimated at 84 months. Based on the log rank test, the variables considered to be important with p-value < 0.05 were regular, dialysis frequency per week, hospitals, diabetes mellitus , hypertension, diabetes mellitus and hypertension, shrunken kidneys, other. The

Gompertz model, which is based on Cox-Snell Residuals, Akaike Information criterion (AIC) and the Bayesian Information Criterion (BIC),is useful among others models. Furthermore, hypertension (HR=0.612, p-value=0.039), regular (HR=0.485, p-value=0.003), urea (HR=1.004, p-value=0.045), and hospitals (HR=0.842, p-value=0.003) were found to have a significant impact on survival ($P<0.05$). The Gompertz model was found to have the smallest BIC values in multivariate analysis, so it was chosen as the best model for hemodialysis patients.

المستخلص

يهدف هذا البحث إلى مقارنة نماذج مختلفة من نماذج الأخطار النسبية (PH) بشكل رئيسي (Weibull، Exponential، Gompertz) ونماذج وقت الفشل المتسارع (AFT) بشكل رئيسي (and Log logistic Lognormal) للمرضى الذين يعانون من غسيل الكلى لتحديد أفضل نموذج لتقييم بقاء المريض وتحديد عوامل الخطر الهامة للوفاة. في الآونة الأخيرة في السودان أصبح مرض الكلى في نهاية المرحلة (الداء الكلوي بمراحله الأخيرة) مشكلة صحية كبيرة . تتكون الدراسة من 325 مريض غسيل كلى تم جمعهم من السجلات في المستشفيات الحكومية في ولاية الخرطوم في الفترة من ديسمبر 2005 إلى ديسمبر 2015. البيانات استخدمت لتقدير دالة البقاء على قيد الحياة بهدف تحديد عوامل الخطر مثل (العمر (تاريخ تشخيص المرض) ، الجنس ، الحالة الاجتماعية ، الحالة التعليمية ، المهنة ، العنوان ، المنتظم ، تكرار غسيل الكلى في الأسبوع ، المستشفى ، السكري ، ارتفاع ضغط الدم ، تعدد تكيسات أمراض الكلى ، انسداد الكلى ، الكلى المنكمشة ، غير معروف ، وأخرى) التي تؤثر على مرضى الكلى في نهاية المرحلة. أظهرت النتيجة أن التحليل أحادي المتغير ومتعدد المتغيرات ، وفقًا لنسبة الخطر ونسبة الوقت ، فإن المتغيرات العمر ، وداء السكري ، وداء السكري + ارتفاع ضغط الدم ، واليوريا ، و مصل الكرياتينين ، تعتبر من العوامل المهمة للغاية وتزيد من خطر الوفاة لدى المرضى (بقاء أقصر) ولها تأثير على لبقاء على قيد الحياة لمرضى غسيل الكلى في النماذج الخمسة المستخدمة في هذا البحث. في حين أن هناك عوامل أخرى ، مثل منتظم ، المستشفيات ، ارتفاع ضغط الدم ، الكلى المنكمشة ، تكرار غسيل الكلى في الأسبوع ، وأخرى، قد قللت من خطر الوفاة (البقاء على قيد الحياة لفترة أطول) ولها تأثير مباشر على بقاء مريض غسيل الكلى. قُدِّر متوسط وقت البقاء الإجمالي بنحو 84 شهرًا. بناءً على اختبارلوغريثم تساوي الرتب ، المتغيرات المهمة عندما تكون القيمة الاحتمالية ($p\text{-value} < 0.05$) منتظم ، تكرار

غسيل الكلى في الأسبوع ، المستشفيات ، داء السكري ، ارتفاع ضغط الدم ، داء السكري وارتفاع ضغط الدم ، الكلى المنكمشة ، وأخرى. يعد نموذج Gompertz ، الذي يعتمد على Cox–Snell Residuals ، ومعيار معلومات Akaike (AIC) ومعيار المعلومات Bayesian (BIC) ، مفيدًا من بين النماذج الأخرى. علاوة على ذلك ، ارتفاع ضغط الدم (HR=0.612, p–value=0.039), ، العادي (HR=0.485, p–value=0.003) ، اليوريا (HR=1.004, p–value=0.045) ، والمستشفيات (HR=0.842, p–value=0.003) لها تأثير كبير على البقاء ($P < 0.05$). تم العثور على نموذج Gompertz يحتوي على أصغر قيم BIC في التحليل متعدد المتغيرات ، لذلك تم اختياره كأفضل نموذج لمرضى غسيل الكلى.

Index of Contents

List of Tables:

List of Figures:

Chapter One

Introduction

1-0. Preface:

Statistical methods for survival data analysis have continued to flourish in the last two decades. Survival analysis focuses on estimating the probability about individual who exposed to hazard for a given length of time until death. Survival analysis is particularly useful when the probability of occurrence of the event under study changes with time (Kleinbaum & Klein, 2012; Collett, 2003; Lee & Wang, 2003). Parametric models are always used for the analysis of survival data to assess the risk of death or chance of survival in a chronic hemodialysis patient. For physicians, making scores is really helpful, and the predictive method of choice for making such scores will be vital to producing accurate outcomes. In other words, the use of a mathematical model strengthens these approaches by allowing for simultaneous measurement of survival in relation to many factors and, moreover, produces estimates of the intensity of effect for each constituent element.

In the last decade, Sudan has seen the expansion of its renal facilities both in the capital and in provincial hospitals. However, the shortage of needs remains strong and the demand for transplantation facilities is even higher due to the effects of the international embargo, as well as the epidemiological transformation marked by a resilient burden of infectious diseases, which has resulted in chronic diseases such as chronic renal failure. End Stage Renal Disease (ESRD) is one of the world's most important causes of morbidity and death observed last year. There is a drastic rise in renal failure among Sudanese citizens in Khartoum. Sudan's frequency-reported rate of new cases (ESRD) is 70-140 per million inhabitants/year (Elamin, et al., 2012; Banaga, et al., 2015) .

The research focuses on one of the most common health conditions in both developed and developing countries, particularly ESRD, a term used when the kidney approaches a total or nearly complete inability to function; kidneys can no longer expel waste, regulate and concentrate urine**.** Dialysis therapy is a treatment intended to remove excrement and toxic compounds in the body to compensate for the lack of kidney function. One class of dialysis is hemodialysis (PAGE, 2007)**.** It has been has been

estimated that over 1.1 million patients are estimated to have ESRD globally, with an addition of percent each year. Incidence and prevalence rates in the United States, for example, are expected to grow by 44 percent and 85 percent, respectively, from 2000 to 2015, and incidence and prevalence rates by 32 percent and 70 percent per million people. Tre comparable patterns in the progress of ESRD patients in developed countries. (Elsharif & Elsharif, 2011; Elamin, et al., 2012; Gilbertson, et al., 2005)

Scientific studies have uncovered major causes of end stage renal disease in survival time. These causes are affecting in the survival of hemodialysis patients for a live long time.

1-1. Research Problem

Survival analysis is one of the most important applications in Biostatistics. 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 until death. Scientific studies have uncovered major causes of end stage renal disease in survival time. These causes are affecting in the survival of hemodialysis patients for a live long time. Millions of people are being affected with outbreak of kidney disease around the world. As consequences many of those patients lost their life. Due to lack of community awareness about the related risk factors such as age, gender and other disease might increase the potential risk that lead to kidney failure, these factors reduce the hazard ratio. Thus has an effect in the survival hemodialysis patient for a live long time. When parametric models to estimate the hazard ratio we know the hazard ratio for each factor. And the application of these models helps to identify the characteristics that lead to an increased probability of survival.

1-2. Research Importance

This research is a few of the research applications in the field of bio-statistics, and also compares the performance of various parametric models of the survival of hemodialysis patients. Parametric models were selected to estimate the survival probability and determine the most influential factors in the survival for hemodialysis patient.

1-3. Research Objectives:

The main objective of this study is to estimate the best survival model for analyzing the factors of the hemodialysis patients.

Other objectives:

1. To estimate survival and hazard functions using Kaplan-Meier estimator.
2. To estimate of median and quartiles for survival time of hemodialysis patients.
3. To estimate the median of survival time for age.
4. To estimate the median of survival time for sex.
5. To compare between survival functions for male/female patients.
6. To compare between survival functions for dialysis frequency by week.
7. To compare between survival functions for diabetes mellitus.
8. To compare between survival functions for hypertension.
9. To compare between survival function for both diabetes mellitus and hypertension.
10. To Estimate the risk factors affecting the hemodialysis patient using parametric models.

1-4. Research Hypotheses:

1. There is no difference between parametric models when compare the fitness of the models based on AIC, and Cox-Snell residual
2. There is no differences between the hazard ratio in terms of regular

3. There is no differences between the hazard ratio in terms of dialysis frequency by week
4. There is no differences between the hazard ratio in terms of age
5. There is no differences between the hazard ratio in terms of both diabetes mellitus and hypertension

1-6. Limits of the research:

The limits of research are within the limits of the applied side .this study is limited to patients with hemodialysis patients who receive treatment in government hospitals (Ahmed Gasim, Ibn Sienna, Omdurman, Selma center, Ribat ,Bahri) who were diagnosed with the disease from December 2005 to December 2010 and followed up until December 2015.

1-7. Target Population:

Data of the study were collected from the biggest and well-known governmental hospitals (Ahmed Gasim, Ibn Sienna, Omdurman, Selma center, Ribat, Bahri) in these three cities named (Khartoum, Khartoum north and Omdurman). In the period December (2005-2015).Data of all patients were collected based on patients' medical records, and the collected data included the following information: age (date of diagnosis of the disease) , Sex , Marital Status, Education status, occupation, Address, regular , Dialysis frequency per week, Hospital , Diabetes Mellitus, Hypertension , polycystic kidney disease, Renal obstructions, Shrunken kidneys, Uncertain, Other, survival status even death or the date of last follow-up per months, was defined as the interval between the date of diagnosis of the disease and the date of death or date of follow- up time in December 2015(the end of the study).

1-7-1. Inclusion Criteria

Data were obtained from records for hemodialysis patients who regularly visited centers in the six governmental hospitals from December 2005 to December 2015, 1 to 90 years of aged patients were included.

1-7-2. Exclusion Criteria

Hemodialysis patients with acute renal disease, insufficient medical records, hemodialysis patients who have only stayed for a short period and people in emergency cases were also disqualified.

1-5. Research Methodology:

A descriptive method was used to describe research data using tables , figures which helps to arrive the general characteristics of the study data The analytical and inferential method is also used to study survival analysis and it is use in finding survival function , hazard function, building parametric models for hemodialysis patients and knowing the importance of variable the risk death .In addition to reaching the results through the applied side and conclusions thereof and then making recommendations and conclusions. To analyze the study data we will use the statistical software EXCEL, SPSS and STATA packages.

1-8.Ethical Considerations:

The study protocol was authorized by the ethics and research committees of the Ministry of Health of Khartoum (serial number: KMOH-REC-1-2020). Hospitals received informed consent.

1-9.Previous Studies:

1- In 2015 the researchers J D Urrutia, W S Gayo, L A Bautista and E B Baccay Publishing a scientific paper in Journal of Physics entitled "Survival Analysis of Patients with End Stage Renal Disease "This study aimed to determine what variables

affect the probability of survival of Chronic Kidney Disease patients diagnosed with ESRD and to assess the effectiveness of the developed survival model to provider reasonably accurate failure analysis and failure forecasts of the conditional status of patient. Conclusion According to the results, the researchers conclude that having End Stage Renal Disease (ESRD) with complications increases the probability of death. In addition, The developed probability risk of death in ESRD patients generated by Cox regression and Weibull Distribution at certain age will help medical experts to improve the quality of medications and technologies that will lessen the risk of death in ESRD and to patients with or without ESRD as a base of healthy lifestyle living.

2- In 14 October 2012 the researchers Maryam Siddiqa, Mueen - ud -Din Azad, Muhammad Khalid Pervaiz Muhammad Ghias, Gulzar H. Shah, Uzma Hafeez Publish a scientific paper in Electronic Journal of Applied Statistical Analysis entitled "Survival Analysis of Dialysis Patients under Parametric and Non-Parametric approaches (in Pakistan)". The objective of this study has estimated the median survival time of male/females patients separately by parametric and nonparametric approaches. Moreover, comparison of survival time to patients ($\leq$ 50 years and >50 years) was also compared. 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.

3- 1n 2007 Asian Pacific J Cancer Prev, 8, 412-416. the researchers Mohamad Amin Pourhoseingholi, Ebrahim Hajizadeh, Bijan Moghimi Dehkordi, Azadeh Safaee, Alireza Abadi, Mohammad Reza Zali Publish a scientific paper in Asian Pacific J Cancer entitle “Comparing Cox Regression and Parametric Models for Survival of Patients with Gastric Carcinoma” .The objective of this study was to compare two survival regression methods – Cox regression and parametric models - in patients

with gastric a denocarcinomas who registered at Taleghani hospital, Tehran. Results: The survival results from both Cox and Parametric models showed that patients who were older than 45 years at diagnosis had an increased risk for death, followed by greater tumor size and presence of pathologic distant metastasis.

4- In Pakistan Vet. J., 2007, 27(4): 194-198. the researchers M. AKRAM, M. AMAN ULLAH AND R. TAJ Publish a scientific paper in Pakistan. Entitled "SURVIVAL ANALYSIS OF CANCER PATIENTS USING PARAMETRICAND NON-PARAMETRICAPPROACHES "In this study, a retrospective simple random sample design was used; the lifetime data on 202 male and 145 female patients of cancer belonging to different classes was selected. These 347 patients of cancer were treated in Nishtar Hospital Multan during January, 1997 to December, 2001 .Using the non-parametric and parametric modeling strategies. The Kaplan-Meier method and Weibull model based on Anderson-Darling test were applied to the real life time data. Findings suggested different sex-superiority of survival pattern among different groups of cancer patients. Interestingly, Kaplan-Meier and Weibull model provided a very close estimate of the survival function and other.

5- In 2015 the researchers Deepapriya. S and Ravanan. R Publish a scientific paper in International Journal of Scientific and Research Publications entitled SURVIVAL ANALYSIS OF UIS PAIENTS UNDER PARAMETRIC NON-PARAMETRIC APPROACHES USING R SOTWARE. The study deals with the survival analysis of University of Massachusetts AIDS Research Unit (UMARU) Impact Study (UIS) data under non Parametric and parametric method. Conclusion of study there is no significance difference between the estimates of survivorship function obtained by parametric and non-parametric method. On a comparison with the parametric distributions log normal is found to be a better fit in accordance with the AIC value for the UIS data, moreover the scale parameter obtained by Weibull is nearly 1 which can be reduced to exponential survival function so obtained for the UIS data.

6- InJ Ayub Med Coll Abbottabad 2015; 27(1):205–7 Zahid Ahmad, Isaac Shahzad Publish a scientific paper entitled SURVIVAL ANALYSIS OF DIALYSIS PATIENTS IN SELECTED HOSPITALS OF LAHORE CITY This study was done to analyses the survival rate of ESRD patients in Lahore city, and to evaluate the influence of various risk factors and prognostic factors on survival of these patients. Methods: A sample of 40 patients was taken from the Jinnah Hospital Lahore and Lahore General Hospital by using the convenience sampling technique. The Log Rank Test was used to determine the significant difference between the categories of qualitative variables of ESRD patients. Multivariate Cox Regression Analysis was used to analyses the effect of different clinical and socio-economic variables on the survival time of these patients. Results: Different qualitative variables like: age, marital status, BMI, comorbid factors, diabetes type, gender, income level, place, risk factor like diabetes, ischemic heart disease, hypertension and Hepatitis status were analyzed on the basis of Log Rank Test. While age and comorbid factors were found to be statistically significant which showed that the distribution of age and comorbid factors were different. By using the Cox Regression analysis the coefficient of Mass, serum albumin and family history of diabetes were found to be significant.

7- 2012 the researchers HSIN-HUNG LIN, CHING-WEI TSAI, PAO-HSUAN LIN, KUANG-FU CHENG, HONG-DAR WU, I-KWAN WANG, CHING-YUANG LIN, WALTER CHEN and CHIU-CHING HUANG Publish a scientific paper in Nephrology 17 (2012) 621–627 entitled "Survival analysis of pediatric dialysis patients in Taiwan "Aim: The long-term survival of Taiwanese children with end-stage renal disease (ESRD) has not been reported before. This study aimed to determine the long-term survival, mortality hazards and causes of death in pediatric patients receiving dialysis. Conclusion: We conclude that there was no significant difference of pediatric ESRD patient survival between HD and PD treatment in Taiwan. The older pediatric ESRD patients had better survival than younger patients.

8- In 1993 the researchers KuNIT0sHI ISEKI, NOBUYUKI KAWAZOE, AKI1t OSAWA, and KOSHIRO FUKIYAMA Publish a scientific paper in Kidney International entitled Survival analysis of dialysis patients in Okinawa, Japan (1971—1990) the objectives of the present study were to describe the characteristics of patients on chronic dialysis therapy over the last two decades in Okinawa and to provide a longitudinal analysis of survival using the Cox proportional hazard analysis. Maintenance hemodialysis was initiated from 1971 in Okinawa. Before this, uremic patients were occasionally treated by intermittent peritoneal dialysis. Currently, there are 27 dialysis units and six of them are freestanding.

9- In (2016), the researchers Mohsen Vahedi, Mahmood Mahmoodi1, Kazem Mohammad, Sharzad Ossareh2 &Hojjat Zeraati Publish a scientific paper in Global Journal of Health Science entitled "What Is the Best Parametric Survival Models for Analyzing Hemodialysis Data? " .Exponential, Weibull, Gompertz, lognormal and log-logistic were used for analyzing survival of hemodialysis patient . Results: According to the both criteria (AIC and Cox-Snell residual), Weibull survival model manifested better results as compared with other models. In our analysis Gompertz distribution, which had the lowest AIC value, was selected as the most suitable.

10- In (2017), the researchers Enayatollah Bakhshi,Reza Ali Akbari Khoei, Azita Azarkeivan, Maryam Kooshesh, Akbar Biglarian Publish a scientific paper in Medical Journal of the Islamic Republic of Iran (MJIRI) entitled “Survival analysis of thalassemia major patients using Cox, Gompertz proportional hazard and Weibull accelerated failure time models"The present study was conducted to apply the semi-parametric Cox PH model and use parametric proportional hazards (PH) and accelerated failure time (AFT) models to identify the risk factors related to survival of TM patients. Results: the Gompertz model, birthplace and age at onset of the disease were significant factors (p= 0.035, and p= 0.005) in survival time. The Akaike Information Criterion (AIC) for Weibull model was 158.51, which was lower than other

parametric models. Conclusion: According to the results, the Weibull AFT model was found to be a better model for identifying the risk factors related to survival of patients with TM disease.

1-10. Research Organization:

The research includes five chapters; the first chapter contains problem, importance, objectives, hypotheses, methodology, research limits, target population, ethical consideration and previous studies. The second chapter contains: Theoretical framework for research contains definition of survival analysis, nonparametric, semi-parametric approaches and parametric approach. The chapter third contains: definition of kidney, kidney's function, renal diseases, causes of kidney disease, kidney failure, chronic kidney disease and end-stage renal failure diagnosis, risk factors, treatment for end stage kidney disease. The fourth chapter contains: the practical side of the research. The fifth chapter includes the conclusions and recommendations, and then references and appendices.

Chapter Two
Survival Analysis Models

2-0. Preface:

In this chapter will be addressed to all the statistical techniques to be used in the practical side of this research. The main part of this chapter is to find a survival function; hazard function, parametric models of relative risk, best model and the comparison between two or more groups of data survive.

2-1. Definition of survival analysis

Survival analysis is a collection of statistical methods that are used to describe, explain, or predict the occurrence and timing of events. The name survival analysis stems from the fact that these methods were originally developed by biostatisticians to analyze the occurrence of deaths. However, these same methods are perfectly appropriate for a vast array of social phenomena including births, marriages, divorces, job terminations, promotions, arrests, migrations, and revolutions. Other names for survival analysis include event history analysis, failure time analysis, hazard analysis, transition analysis, and duration analysis. Although some methods of survival analysis are purely descriptive (e.g., Kaplan-Meier estimation of survival functions), most applications involve estimation of regression models, which come in a wide variety of forms. A key feature of all methods of survival analysis is the ability to handle right censoring, a phenomenon that is almost always present in longitudinal data. Right censoring occurs when some individuals do not experience any events, implying that an event time cannot be measured. Introductory treatments of survival analysis for social scientists can be found in Teachman (1983), Allison (1984, 1995), Tuma and Hannan (1984), Kiefer (1988), Blossfeld and Rohwer (2001), and Box-Steffensmeier and Jones (2004) (Vallinayagam, et al., 2014).

2-2 Survival data:

The Expression describe the data that measure the time until the event referred to as survival data , and the outcome variable is the time until the event that take a positive real values, which is known as time to stay (Kalbfleisch & L.Prentice, 2002) .

2-2-1. Survival time

Survival time can be defined broadly as the time to the occurrence of a given event. This event can be the development of a disease, response to a treatment, relapse, or death. Therefore, survival time can be tumor-free time, the time from the start of treatment to response, length of remission, and time to death (Lee & Wang, 2003).

2-2-2. Special features of survival data

We must first consider the reasons why survival data are not amenable to standard statistical procedures used in data analysis. One reason is that data are generally not symmetrically distributed. Typically, a histogram constructed from the survival times of a group of similar individuals will tend to be positive skewed, that is the histogram will have a longer "tail" to the right of the interval that contains the largest number of observations. As a consequence, it will not be reasonable to assume that data of this type have a normal distribution. This difficulty could be resolved by first transforming the data to give a more symmetric distribution (collett, 2003).Second reason is that the main features of survival data that renders standard methods in appropriate is that survival times are frequently censored Cox and Oakes (1984).

The survival time of an individual is said to be censored when the end point of interest has not been observed for that individual. This may be because the data from a study are to be analyzed at a point in time when some individuals are still alive. Alternatively, the survival status of an individual at the time of the analysis might not be known because that individual has been lost follow-up. The only information available on the survival experience of that patient is the last date on which he or she was known to be alive. This date may well be the last time that patient reported to a clinic for regular check-up. An actual survival time can also be regarded as censored when death is from a cause that is known to be unrelated to the treatment. However, it can be difficult to be sure that death is not related to a particular treatment that the patient is receiving. In each of these situations, a patient who entered a study at time t_0 dies at time $t_0 + t$, however, t is unknown, either because the individual is still alive or because he or she

has been lost to follow-up. If the individual was last known to be alive at time $t_0 + c$, the time c is called a censored time. This censoring occurs after the individual has been entered into a study that is to the right of the last known survival time, and is therefore known as right censoring. The right-censored survival time is then less than the actual, but unknown survival time. Another form of censoring is left censoring, which is encountered when the actual survival time of an individual is less than that observed .Yet another type of censoring is interval censoring. Here individuals are known to have experienced an event within an interval of time (collett, 2003) .

2-2-3. Time origin

Most methods of survival analysis (e.g., Cox regression) require that the event time be measured with respect to some origin time. The choice of origin time is substantively important because it implies that the risk of the event varies as a function of time since that origin. In many cases, the choice of origin is obvious. In demography studies if the event is a divorce, the natural origin time is the date of the marriage.

In drug trials original time for the disease is the beginning of taking the treatment history (Allison, 2010).

2-2-4. Event

The first step in any application of survival analysis is to define the event. In medical research this event can be the development of a disease, response to a treatment, relapse from remission, or death, and when the end point not fatal, the event can be relief of pain or the recurrence or any designated experience of interest that may happen to an individual (Kleinbaum & Klein, 2012; collett, 2003).

2-2-5. Patient time and study time

In atypical study, patients are not all recruited at exactly the same time. But accrue over a period of months or even years. After recruitment, patients are followed up until they die, or until a point in calendar time that marks the end of the study. When the data are

analyzed, after recruitment some patients may be lost to follow up, while others will still be alive at the end of the study. The calendar time period in which an individual is in the study is known as the study time. The period of time that patient spends in the study, measured from that patient's time origin is often referred to as patient time (Collett (2003)).

2-3. Survival time distribution:

Survival time data measure the time to a certain event, this time are subject to random variations, and like any random variables, form a distribution. The distribution of survival times is usually described or characterized by three functions: the survivorship function, the probability Density function and the hazard function. These three functions are mathematically equivalent if one of them is given, the other two can be derived (Lee & Wang, 2003).

2-3-1. Survival function and hazard function

The actual survival time of an individual t, can be regarded as the value of a variable T, which can take any non-negative value. The different values that T can take have a probability distribution and we call the random variable associated with the survival time. Now suppose that the random variable T has a probability distribution with underlying probability density function$f(t)$.The distribution function of T Continuous variable is then given by Cumulative Distribution function is

$$F(t) = P(T < t) \ = \int_0^t f(u)du,$$

It represents the probability that the survival time is less than some value t.

The survival function, $S(t)$ is defined to be the probability that the survival time is greater than or equal to t and so

$$S(t) = \ P(T \geq t) = \ 1 \ - F(t) \qquad (2.1)$$

(collett, 2003) .Here $S(t)$ is a nonincreasing function of time t with the properties

$S(t) = 1$ for $t = 0$

$S(t) = 0$ for $t = \infty$

That is, the probability of surviving at least at the time zero is 1 and that of surviving an infinite time is zero (Lee & Wang, 2003).

Sometimes, for example, when lifetimes are grouped or measured as a number of cycles of some sort, T may be treated as a discrete random variable. Suppose T can take on values $t_{1,} t_2 \ \ldots,$ with $0 \leq t_{1,} < t_2 \ldots,$ and let the probability Density function be

$$f(t_j) = \Pr (T = t_j) \qquad j = 1, 2$$

The survivor function is then

$$S(t_j) = \Pr (T \geq t) = \sum_{j:t_j > t} f(t_j)$$

2-3-2. Probability density function (Density function)

Like any other continuous random variable, the survival time T has a probability density function defined as the limit of the probability that an individual fails in the short interval t to t+Δt per unit width Δt, or simply the probability of failure in a small interval per unit time. It can be expressed as:

$$f(t) = \frac{\lim_{\Delta t \longrightarrow 0} P[\ (t, t + \Delta t)]}{\Delta t} \qquad (2.2)$$

The density function has the following two properties:

1. $f(t)$ is a nonnegative function:

$$f(t) \geq 0 \quad \text{Forall} t \geq 0$$

$$= 0 \quad \text{For} \, t < 0$$

2. The area between the density curve and the t axis is equal to 1.

The density function is also known as the unconditional failure rate (Lawless & , 2003). The hazard function is widely used to express the risk or the hazard of death at some time t, and is obtain from the probability that the individual dies at time t, conditional on he or she having survived to that time. For formal definition of the hazard function consider the probability that the random variable associated interval with an individual survival time T lies between $t \text{and } t + \Delta t$, conditional on T being greater than or equal to t, written:

$P(t \leq T < t + \Delta t | T \geq t)$ (Lawless & , 2003).

This conditional probability is then expressed as a probability per unit time by dividing by the time interval Δt to give a rate. The hazard function $h(t)$is then the limiting value of this quantity, as Δt tends to zero, so that

$$h(t) = \lim_{\Delta t \to 0} \left\{ \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \right\} \qquad (2.3)$$

The function $h(t)$ is also known as the hazard rate, the instantaneous death rate, the intensity rate, or the force of mortality.

From equation (2.3) $h(t)\Delta t$ is the appropriate probability that an individual dies in the interval $(t, t + \Delta t)$, conditional on that person having survived to time t.

From the definition of the hazard function in equation (2.3) we can obtain some useful relationships between the survival and hazard functions. According to a standard result from probability theory, the probability of an event A, conditional on the occurrence of an event B, is given by

$$P(A|B) = \frac{P(AB)}{P(B)}$$

where $P(AB)$ is the probability of joint occurrence of A and B (collett, 2003) .Using this result, the conditional probability is the definition of the hazard function in equation (2.3) is

$$\frac{P(t \leq T < t + \Delta t)}{P(T \geq t)}$$

This is equal to

$$\frac{F(t + \Delta t) - F(t)}{S(t)}$$

Where $F(t)$ is the distribution function of T

Then

$$h(t) = \Delta t \xrightarrow{\lim} 0 \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)}$$

Now

$$\Delta t \overset{\lim}{\longrightarrow} 0 \left\{ \frac{F(t+\Delta t) - F(t)}{\Delta t} \right\}$$

Is the definition of the derivative of $F(t)$ with respect to t, which is $f(t)$, and so

$$h(t) = \frac{f(t)}{s(t)} \qquad (2.4)$$

It then follows that:

$$h(t) = -\frac{d}{dt}\{\log S(t)\} \qquad (2.5)$$

And so

$$S(t) = \exp\{-H(t)\} \qquad (2.6)$$

Where

$$H(t) = \int_0^t h(u)du. \qquad (2.7)$$

The function $H(t)$ features widely in survival analysis, and is called the Integrated of cumulative hazard. From equation (2.6), the cumulative hazard can be obtained from the survival function since:

$$H(t) = -\log S(t) \qquad (2.8)$$

In the analysis of survival data, the survival function and hazard function are estimated from the observed survival times (collett, 2003).

2-4.Nonparametric and semi parametric approaches:

2-4-1. Estimate of the survival function

Suppose that we have a single sample of survival times, where none of observations are censored. The survivor function $S(t)$, defined in equation (1, 1) is the probability that an individual survives for time greater than or equal to t .This function can be estimated by empirical survivor function given by (Lawless & , 2003):

$$\hat{S}(t) = \frac{\text{number of individuals with survival times} \geq t}{\text{number of individuals in the dataset}} \qquad (2.9)$$

Equivalently, $\hat{S}(t) = 1 - \hat{F}(t)$, where $\hat{F}(t)$ is empirical distribution that is, the ratio of the total number of individuals at time to the total number of individuals in study .Notice that the empirical survivor function is equal to unity for values of t before the first death time, and zero after the final death time .The estimated survivor function $\hat{S}(t)$ is assumed to be constant between two adjacent death time, and so a lot of $\hat{S}(t)$ against t is step function. It decreases immediately after each observed survival time.

2-4-2.The Kaplan-Meier estimate of the survival function

Non-parametric method for estimating survivor function which can be used in the presence of censored times are life - table and Kaplan- Meier estimate (collett, 2003) . The only difference is that the Kaplan- Meier estimate is based on individual survival times, whereas in the life-table method, survival times are grouped into intervals (Lee & Wang, 2003) but the grouping of the survival times does result in some loss of information. This particularly so when the number of patients is small, less than about 30, says (collett, 2003).

For this reason Kaplan- Meier best method for estimate survival data analysis.

To obtain the Kaplan-Meier estimate, series of time intervals is constructed, as for the life- table estimate and each interval includes the one death time, although there could be more than one individual who dies at any particular death time.

We suppose that there are n individuals with observed survival times $t_1, t_{2,} \dots, t_n$ some of this observation may be right censored, and there may also be more than one individual with the same observed survival times. We therefore suppose that there are r death times amongst the individuals. Where $(r \leq n)$ .After arranging these death times in ascending order the jth is denoted $t_j, for j = 1, 2, \dots, r$, and so the r ordered death times are $t_{(1)} < t_{(2)} < t_{(3)} < \cdots < t_{(r)}$. The number of individuals who are alive just before timet_j, including those who are about to die at this time, will be denoted $n_j, for j = 1, 2 \dots r$ and d_j will denote the number who die at this time. We count the total number of individuals alive at the start of the interval $(n_j;\ j = 1, 2 \dots r)$ and the number of individuals who died (d_j) in the time interval. The Kaplan-Meier estimate of the survival function is given by:

$$\hat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right) \qquad (2.10)$$

The Kaplan-Meier estimate is formed as a product of a series of estimated of probabilities .in fact Kaplan-Meier estimate is the limiting value of the life- table estimate and also when the number of intervals tends infinity and their width tends to zero is known as the product- limit estimate of survivor function (collett, 2003).

2-4-3.Nelson -Aalen estimate of survivor function

An alternative estimate of the survivor function, which is based on the individual event times, is the Nelson- Aalen estimate given by:

$$\hat{S}(t) = \prod_{j=1}^{k} \exp(-d_j/n_j)$$

For $j = 1, 2 \ldots r$ and d_j will denote the number who die at this time and so r ordered the death time.

n_j the number of individuals who are alive just before time t_j the number of individuals at risk at t_j (Kalbfleisch & L. Prentice, 2002; collett, 2003) .

2-4-4.Standard error of the estimated survivor function

The Kaplan-Meier estimate of the survival function for any value of t in the interval from $t_{(k)}\ to\ t_{(k+1)}$ can be written as:

$$\hat{S}(t) = \prod_{j=1}^{k} \hat{p}_j$$

$$\text{For } k = 1, 2 \ldots, r \text{ where } \hat{p}_j = \left(\frac{n_j - d_j}{n_j}\right)$$

Is the estimated probability that an individual survives through the time interval that begins at $t_j, j = 1, 2 \ldots, r$. taking logarithms.

$$\log \hat{S}(t) = \sum_{j=1}^{k} \log \hat{p}_j$$

And so the variance of $\log \hat{S}(t)$ is given by:

$$\text{var}\{\log \hat{S}(t)\} = \sum_{j=1}^{k} \text{var}\{\log \hat{p}_j\}. \qquad (2.11)$$

Now the number of individuals who survive through the interval beginning at t_j can be assumed to have a binomial distribution with parameters n_j and p_j, where p_j the true probability of survival through that interval is. The observed number who survives is $n_j - d_j$, and using the result that the variance of a binomial random variable with parameters n, p is:

$np(1-p)$, the variance of $n_j - d_j$ is given by

$$\text{var}(n_j - d_j) = n_j p_j (1 - p_j)$$

Since $\hat{p}_j = \left(\frac{n_j - d_j}{n_j}\right)$

The variance of $\hat{p_j}$ is

$$\text{var}\left(\frac{n_j - d_j}{{n_j}^2}\right) = \frac{1}{{n_j}^2}\text{var}(n_j - d_j) = \frac{1}{{n_j}^2} n_j p_j (1 - p_j)$$

The variance of $\hat{p}_j$ Maythen estimated by

$$\frac{\hat{p}(1-\hat{p}_j)}{n_j} \qquad (2.12)$$

In order to obtain the variance of $\log \hat{p}_j$, we make use of a general result for the approximate variance of a function of a random variable. According to this result the variance of a function $g(X)$ of the random variable x is given by:

$$\text{var}\{g(X)\} \approx \left\{\frac{dg(X)^2}{dX}\right\}\text{var}(X) \qquad (2.13)$$

This is known as the Taylor series approximation to the variance of a function of a random. Using equation (2.13), the approximate variance of $\log \hat{p}_j$ is $\text{var}\left(\hat{p}_j\right)/\hat{p}_j^{\,2}$ and using (2.12), the approximate estimate variance of $\log \hat{p}_j$ is

$$\frac{1-\hat{p}_j}{\left(n_j \hat{p}_j\right)}$$

which on substitution for $\hat{p}_j$, reduces to:

$$\frac{d_j}{n_j\left(n_j - d_j\right)} \ . \qquad (2.14)$$

From equation (2.11),

$$\text{var}\left\{\log \hat{S}(t)\right\} \approx \sum_{j=1}^{k} \frac{d_j}{n_j\left(n_j - d_j\right)} \ , \qquad (2.15)$$

And a further application of the result in equation (2.13) gives

$$\text{var}\left\{\log \hat{S}(t)\right\} \approx \frac{1}{\left[\hat{S}(t)\right]^2} \text{var}\left\{\hat{S}(t)\right\} \ ,$$

So that

$$\text{var}\left\{\hat{S}(t)\right\} \approx \left[\hat{S}(t)\right]^2 \sum_{j=1}^{k} \frac{d_j}{n_j\left(n_j - d_j\right)} \ . \qquad (2.16)$$

Finally the standard error of the Kaplan-Meier estimate of the survivor function, defined to be the square root of the estimated variance of the estimate is given by:

$$se\{\hat{S}(t)\} \approx \hat{S}(t)\left\{\sum_{j=1}^{k}\frac{d_j}{n_j(n_j-d_j)}\right\}^{\frac{1}{2}}, \qquad (2.17)$$

For $t_k \leq t < t_{k+1}$.This result is known as Greenwood's formula.

If there are no censored survival times $n_j - d_j = n_j + 1,$ and expression (2.14) becomes

$(n_j - n_{j+1})/n_j n_{j+1}$. Now,

$$\sum_{j=1}^{k}\frac{n_j - n_{j+1}}{n_j n_{j+1}} = \sum_{j=1}^{k}\left[\frac{1}{n_{j+1}} - \frac{1}{n_j}\right] = \frac{n_j - n_{j+1}}{n_j n_{j+1}},$$

This can be written as

$$\frac{1-\hat{S}(t)}{n_1\hat{S}(t)}$$

(Kalbfleisch & L. Prentice, 2002; collett, 2003)

2-4-5. Confidence intervals for the survival function

Once of standard error of an estimate of the survival function has been calculated a confidence interval for the corresponding value of the survival function at a given time t can be found. A confidence interval is an interval estimate of survival function, and is the interval which is such that there is a prescribed probability that the value of the true survivor function is included within it.

A confidence interval for the true value of the survival function at a given time t is obtained by assuming that the estimated value of the survival function at t is normally distributed with mean $S(t)$ and estimated variance given by equation (2.16).

The interval computed from percentage points of the standard normal distribution. Thus, if Z is a random variable that has a standard normal distribution, the upper (one – sided) $\alpha/2$-point, or the (two sided) á-point, of this distribution is that value $z_{\alpha/2}$ which is such that Pr $(Z > z_{\alpha/2}) = \alpha/2$. This probability is the area under the standard normal curve to the right of $z_{\alpha/2}$, for example the two-sided 5% and 1% points of the standard normal distribution $z_{0.025}$ and $z_{0.005}$, are 1.96 and 2.58, respectively.

A $100(1-\alpha)\%$ confidence interval for $S(t)$, for a given value of t is the interval from

$\hat{S}(t) - z_{\alpha/2}se\{\hat{S}(t)\}to\hat{S}(t) + z_{\alpha/2}se\{\hat{S}(t)\}$, Wherese$\{\hat{S}(t)\}$ is found from equation

(2.17)(Kalbfleisch & L. Prentice, 2002; collett, 2003).

2-5. Estimate of the hazard function:

There are many ways to estimate the hazard function, but the most commonly used is a Kaplan –Meier estimate.

2-5-1. Kaplan –Meier of estimate the hazard function

A natural way of estimating the hazard function for unground survival data is to take ratio of the number of death at a given death time to the number of individuals at risk at that time. If the hazard function is assumed to be constant between successive death times, the hazard per unit time can be found by further dividing by the time interval. Thus if there are d_j deaths at the jth death time,t_j , $j = 1,2, \dots, r$, and n_j at risk at timet_j, the hazard function in the interval from t_j to t_{j+1} can be estimated by:

$$\hat{h}(t) = \frac{d_j}{n_j \tau_J} \qquad (2.18)$$

For $t_j \leq t < t_{j+1}$

Where

$\tau_J = t_{j+1} - t_j$. (Kalbfleisch & L. Prentice, 2002; collett, 2003).

Notice that is not possible to use equation (2.18) to estimate the hazard in the interval that begins at the final death time, since this interval is open ended.

The estimate in equation (2.18) is referred to as a Kaplan-Meier type estimate, because the estimated survival function derived from it is the Kaplan-Meier estimate. To show this, note that since $\hat{h}(t)$, $t_j \leq t < t_{j+1}$, is an estimate of the risk of death per unit time in the jth interval, the probability of death in that interval is $\hat{h}(t)_{\tau_J}$, that is $\frac{d_j}{n_j}$. Hence an estimate of the corresponding survival probability in that interval is $1 - (d_j/n_j)$ and the estimated survival function is as given by equation (2.10).

The approximate standard error of $\hat{h}(t)$ can be found from the variance of d_j, which may be assumed to have a binomial distribution with parameters

n_j and p_j, where p_j is the probability of the death in the interval of length T. Consequently, $\text{var}(d_j) = n_j p_j (1 - p_j)$, and estimating p_j by d_j/n_j

Gives

$$\text{se}\{\hat{h}(t)\} = \hat{h}(t)\sqrt{\frac{n_j - d_j}{n_j p_j}} . \qquad (2.19)$$

(Collett, 2003).

2-5-2. Estimate the cumulative hazard function

The cumulative hazard function is important in the identification of models for survival

data .In addition since the derivative of cumulative hazard function is hazard function. The cumulative hazard at time$t, H(t)$ was defined in equation (2.7) to be the integral of

the hazard function, but is more conveniently found using equation (2.8). According to this result.

$$H(t) = -\log S(t),$$

and so if $\hat{S}(t)$ is used the Kaplan-Meier estimate or Nelson -Aalen estimate of survival function.

$$\hat{H}(t) = -\log \hat{S}(t)$$

If the Nelson -Aalen estimate of survival function is used, the estimate cumulative hazard function,

$\hat{H}(t) = -\log \hat{S}(t)$ is given by:

$$\hat{H}(t) = \sum_{j=1}^{k} \frac{d_j}{n_j}$$

This is the cumulative sum of the estimated probabilities of death from the first to the kth time interval, $k = 1,2, \ldots r$.

(Kalbfleisch & L. Prentice, 2002; collett, 2003).

2-6. Median and Percentiles of survival times:

Since the distribution of survival times tends to be positively skew, the median is the preferred summary measure of the location of the distribution. Once the survival function has been estimated it is straight forward to obtain an estimate of the median survival times. This is the time beyond which 50% of the individuals in the population under study are expected to survive, and is given by that value $t(50)$ which is such that $S\{t(50)\} = 0.5$.

Because the non-parametric estimates of $S(t)$ are step- functions, it will not usually be possible to realise an estimated survival time that makes the survival function exactly equal to 0.5.instead, the estimated Median Survival Time, $\hat{t}(50)$ is defined to be the smallest observed survival times for which the value of the estimated survival function is less than 0.5.In mathematical terms.

$$\hat{t}(50) = \min\{t_i | \hat{S}(t_i) < 0.5\},$$

where t_i is the observed survival time for ith individual, $i = 1,2, \dots, n$.

Since the estimated survival function only changes at a death time, this is equivalent to the definition

$$\hat{t}(50) = \min\{t_j | \hat{S}(t_j) < 0.5\},$$

Where t_j is the jth ordered time, $j = 1,2, \dots, r$.

In the particular case where the estimated survival function is exactly equal to 0.5 for values of t in the interval from t_j to t_{j+1}, the median is taken to be the half-way point in this interval, that is

$$(t_j + t_{j+1})/2.$$

In the situation where there are no censored survival times, the estimated median survival time will be the smallest time beyond which 50% of the individuals in the sample survive.

A similar procedure to that described above can be used to estimate other percentiles of the distribution of survival times. The pth percentile of the distribution of survival times is defined to be the value $t(p)$ which is such that:

$$\mathrm{F\{t(p)\}} = \frac{\mathrm{p}}{100}.$$

In terms of the survival function $\mathrm{t(p)}$ is such that:

$$\mathrm{s\{t(p)\}} = 1 - \left[\frac{\mathrm{p}}{100}\right].$$

Respectively. Using the estimated survivor function, the estimatedpth percentile is the smallest observed survival time $,\hat{\mathrm{t}}(\mathrm{p}),$ for which

$$\hat{\hat{\mathrm{s}}}\{\hat{\mathrm{t}}(\mathrm{p})\} < 1 - \left[\frac{\mathrm{p}}{100}\right].$$

(collett, 2003).

2-6-1. Confidence interval for the median and percentiles

Approximate confidence intervals for the median and percentiles of a distribution of survival times can be found once the variance of the estimated percentile has been obtained. An expression for the approximate variance of a percentile can be derived from a direct application of the general result for the variance of a function of a random variable in equation (2.13). Using this result

$$\text{var}\{t(p)\} = \left(\frac{1}{\hat{f}\{t(p)\}}\right)^2 \text{var}[\hat{s}\{t(p)\}] \quad . \quad (2.20)$$

Where $t(p)$ is the pth percentile of the distribution, $\hat{s}\{t(p)\}$ is the Kaplan- Meier estimate of the survival function at $t(p)$ and $\hat{f}\{t(p)\}$ is estimate the probability density function of the survival times at $t(p)$

The standard error of $\hat{t}(p)$ the estimated pth percentile is therefore given by.

$$\text{se}\{\hat{t}(p)\} = \frac{1}{\hat{f}\{\hat{t}(p)\}} \text{se}[\hat{s}\{\hat{t}(p)\}] \quad . \quad (2.21)$$

the standard error of $\hat{s}\{\hat{t}(p)\}$ is found using Greenwood's formula for the the standard error of Kaplan-Meier estimate of survivor function ,given in equation(2.17),while an estimate of the probability density function at $\hat{t}(p)$ is

$$\hat{f}\{t(p)\} = \frac{\hat{s}\{\hat{u}(p)\} - \hat{s}\{\hat{l}(p)\}}{\hat{l}(p) - \hat{u}(p)},$$

Where

$$\hat{u}(p) = \max\left\{t_j | \hat{s}(t_j) \geq 1 - \frac{p}{100} + \epsilon\right\},$$

And

$$\hat{L}(p) = \min\left\{t_j | \hat{s}(t_j) \leq 1 - \frac{p}{100} + \epsilon\right\},$$

for $j = 1, 2, \ldots, r$,

ϵ is small values in many cases taking 0.05

In particular for equation (2.21), the standard error the median survival time is given by

$$\text{se}\{\hat{t}(50)\} = \frac{1}{\hat{f}\{\hat{t}(50)\}}\text{se}[\hat{s}\{\hat{t}(50)\}] \qquad (2.22)$$

Where $\hat{f}\{\hat{t}(50)\}$ can found from

$$\hat{f}\{\hat{t}(50)\} == \frac{\hat{s}\{\hat{u}(50)\} - \hat{s}\{\hat{l}(50)\}}{\hat{l}(50) - \hat{u}(50)} \quad . \qquad (2.23)$$

$\hat{u}(50)$ is the largest survival time which the Kaplan –Meier estimate of the survivor function exceed 0.55.

$\hat{l}(50)$ is the smallest survival time for which survivor function is less than or to 0.45.

Once the standard error of the estimated pth percentile has been found a $100(1-\alpha)\%$ Confidence interval for $t(p)$ has limit of

$$\hat{t}(p) \pm z_{\alpha/2}\,\text{se}\{\hat{t}(p)\} \qquad (2,24)$$

Where

$z_{\alpha/2}$ is the upper $(\text{one}-\text{sided})$ $\alpha/2$ -point of the standard normal distribution (collett, 2003).

2-7. Comparison of two or more groups of survival data:

The simplest way of comparing the survival times obtained from two or more groups of individuals is to plot the corresponding estimates of the two or more survival functions on the same axes.In the comparison of two groups of survival data, there are a number of methods that can be used to quantify the extent of between group differences. Two non-parametric procedures will now be considered, namely the log-rank test and the Wilcoxon test (Mantel, 1966; collett, 2003).

2-7-1. Log-rank test for comparison of two groups of survival data

In order to construct the long-rank test, we begin by considering separately each death time in two groups of survival data. These groups will be labeled Group I and, Group II. Suppose that there are r distinct death times $t_1 < t_2 < \ldots < t_r$, across the two groups and that at time t_j, d_{1j} individuals in Group I and d_{2j} individuals in Group II die for $j = 1,2,\ldots,r$. Unless two individuals in a group have the same recorded death time, the value of , d_{1j} and d_{2j} will either be zero or unity. Suppose further that there are n_{1j} individuals at risk of death in Group I just before time t_j, and that there are n_{2j} at risk in Group II. Consequently at time t_j, there are $d_j = d_{1j} + n_{2j}$ deaths in total out of $n_j = n_{1j} + n_{2j}$ individuals at risk. We can therefore regard d_{1j} as a random variable, which can take any value in the range from zero to the minimum of $d_j n_{1j}$. In fact d_{1j} has hypergeometric distribution, according to which probability that the random variable associated with the number of deaths in the Group I takes the value d_{1j} is (Mantel, 1966; collett, 2003).

$$\frac{\binom{d_j}{d_{1j}}\binom{n_{j-d_j}}{n_{1j}-d_{1j}}}{\binom{n_j}{n_{1j}}} \qquad (2.25)$$

The expression

$$\binom{d_j}{d_{1j}}$$

represents the number of different ways in which d_{1j} times can be chosen from d_j times and is read as "$d_{1j}Cd_j$ ". It is given by:

$$\binom{d_j}{d_{1j}} = \frac{d_j!}{d_{1j}!\left(d_j - d_{1j}\right)!}$$

The mean of the hypergeometric random variable d_{1j} is given by:

$$e_{1j} = n_{1j}d_j/n_j$$

So that e_{1j} is the expected number of individuals who die at time t_j in Group I. The most straight forward way of doing this is to sum the differences $d_{1j} - e_{1j}$ over the total number of death time r in the two groups is given by:

$$U_L = \sum_{j=1}^{r}\left(d_{1j} - e_{1j}\right). \qquad (2.26)$$

Notice that is $d_{1j} - e_{1j}$ which the difference between the total observed and expected numbers of death in Group I. this statistic will have zero mean, since $E\left(d_{1j}\right) = e_{1j}$. The variance of U_L is simply the sum of the variances of the d_{1j}. The variance of d_{1j} is given by:

$$\mathcal{V}_{1j} = \frac{n_{1j}n_{2j}d_j\left(n_j - d_j\right)}{{n_j}^2\left(n_j - 1\right)}, \qquad (2.27)$$

So that the variance of U_L is:

$$\mathrm{var}(U_L) = \sum_{j=1}^{r}\mathcal{V}_{1j} = V_L, \qquad (2.28)$$

It can be shown that U_L has an approximate normal distribution when the number of death times is not too small. It then follows that $U_L/\sqrt{V_L}$ has a normal distribution with zero mean and unit variance N (0, 1).

$$U_L/\sqrt{V_L} \backsim N(0,1),$$

(Mantel, 1966; collett, 2003).

The square of a standard normal random variable has a chi square distribution on one degree of freedom, denote χ^2_1 (collett, 2003) so we have that:

$$\frac{U_L^{\,2}}{V_L} \sim \chi^2_1 \ . \qquad (2.29)$$

2-7-2.comparison of three or more groups of survival data

The long-rank test can be extended to enable three or more groups of survival data to be compared. Suppose that the survival distribution of g groups of survival, for g ≥2.we then define analogues U - Statistic for comparing the observed numbers of death in groups $1, 2 \ldots, g-1$ with expected values given by:

$$U_{wK} = \sum_{j=1}^{r} n_j \left(d_{kj} - \frac{n_{kj} d_j}{n_j} \right),$$

for $k = 1, 2 \ldots g-1$ these quantities are then expressed in the form of a vector $(g-1)$ components, which we denote by U_L and U_W .

We also need expression for the variances of the $, U_{LK}\ and\ U_{WK}$, and for the covariance between pairs of values. The covariance between U_{LK} for and $U_{Lk'}$ is given by:

$$V_{Lkk'} = \sum_{j=1}^{r} \frac{n_{kj} d_j (n_j - d_j)}{n_j (n_j - 1)} \left(\delta_{kk'} - \frac{n_{k'j}}{n_j} \right),$$

For $k, k' = 1,2, \ldots, g-1$, where $\delta_{kk'}$ is such that:

$$\delta_{kk'} = \begin{cases} 1 \text{ if } k = k' \\ 0 \text{ othermise} \end{cases}$$

These terms are assembled in the form of a variance –covariance matrix,V_L which is asymmetric matrix that has the variance of U_{LK} .for example in the comparison of three groups of survival data, this matrix would be given by:

$$V_L = \begin{pmatrix} V_{L11} & V_{L12} \\ V_{L12} & V_{L22} \end{pmatrix},$$

where V_{L11} and V_{L22} are the variances of U_{L1} and U_{L2} , respectively, and V_{L12} is their covariance.

Finally, in order to test the null hypothesis of no group differences, we make use of the result that the test statistic $U'_L V^{-1}_L U_L$ has a chi- squared distribution $(g-1)$ degrees of freedom, when the null hypothesis is true (collett, 2003).

2-8. Cox's regression model:

In the analysis of extent data, interest centers on the risk or hazard of death at any time after the time origin of the study. As a consequence, the hazard function is modeled directly in survival analysis. There are two broad reasons for modeling survival data. One objective of modeling process is to determine which combinations of potential explanatory variables affect the form of the hazard function. In particular, the effect that the treatment has on the hazard of death can be studied, as can the extent to which other explanatory variables affect the hazard function. Another reason for modeling the hazard function is to obtain an estimate of the hazard function itself for individual

(collett, 2003).The basic mode for survival data to be considered is proportional hazards model .this model was proposed by Cox (1972) and has also come to be known as the Cox regression model. Although the model is based on the assumption of proportional hazards, no particular form of probability distribution is assumed for survival time .The model is therefore referred to as a semi-parametric model (collett, 2003) is usually written in terms of the hazard model formula

$$h(t,X) = h_o(t)e^{\sum_{i=1}^{p}\beta_i X_i} \qquad (2.30)$$

Where

$$X = X_1, X_2, \dots X_p$$

This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables denoted by X. That is X represents a collection of predictor variables that is being modeled to predict an individual's hazard**.** The Cox model formula says that the hazard at time t is the product of two quantities .the first of these,$h_o(t)$, is called the baseline hazard function. The second quantity is the exponential expression e to the linear sum of $\beta_i X_i$, where the sum is over the p explanatory X variables.

An important feature of this formula, which concerns the proportional hazards (PH) assumption, is that the baseline hazard is function of t , does not involve the $X's$ .In contrast ,the exponential expression involves the $X's$, but does not involve t .the $X's$ here are called time –independent $X's$ (Kleinbaum & Klein, 2012).

2-8-1.Assumption of proportional hazards

The PH assumption requires that the hazard ratio (HR) is constant over time, or equivalently, that the hazard for one individual is proportional to the hazard for any

other individual, where the proportionality constant is independent of time. The formula for HR is.

$$HR(t,x) = \frac{\hat{h}(t,x^*)}{\hat{h}(t,x)} = \frac{\hat{h}_o(t)\left[e^{\sum_{i=1}^{p}\hat{\beta}_i x^*{}_i}\right]}{\hat{h}_o(t)\left[e^{\sum_{i=1}^{p}\hat{\beta}_i X_i}\right]}$$

$$= exp\left[\sum_{i=0}^{p}\hat{\beta}_i(x^*{}_i - x_i)\right] \qquad (2.31)$$

$h_o(t)$ is baseline hazard function appears in both the numerator and denominator of the hazard ratio and cancels out of formula, the final expression does not involve time t.

Let $x^* = (x^*{}_1, x^*{}_2 \ldots x^*{}_p)$ and $x = (x_1, x_2, \ldots x_p)$

(Kleinbaum & Klein, 2012)Taking $x^*{}_i = x + 1$, the hazard ratio reduces to $HR = \exp(\beta)$ and corresponds to the effect of one unit increase in the explanatory variable X on the risk of event.

Hence the survival function from hazard function using function (2.30) is

$$s(t,x) = e^{-H(t,x)}$$

Since the cumulative proportional hazards model from equation (2.7) is:

$$H(t,x) = \int_0^t h(t,x)$$

$$= e^{\sum_{i=1}^{p}\beta_i x_i}\int_0^t h_o(t)\,dt$$

$$= e^{\sum_{i=1}^{p} \beta_i X_i} H_o(t)$$

$$S(t,x) = e^{-H_o(t)}\, e^{\sum_{i=1}^{p} \beta_i X_i} = \left(e^{-H_o(t)}\right) e^{\sum_{i=1}^{p} \beta_i X_i}$$

$$S_o(t) = e^{-H_o(t)}$$

$S_o(t)$ is baseline survival function

$S(t,x)$ is Cox PH model survival function

$$= S_o(t) e^{\sum_{i=1}^{p} \beta_i X_i}$$

(collett, 2003).

2-9.Fitting the proportional hazards model:

2-9-1. Partial Likelihood function for Survival times:

To estimate the coefficients,$b_1, \ldots b_p$, Cox (1972) proposes a partial likelihood function based on a conditional probability of failure, assuming that there are no tied values in the survival times. However, in practice, tied survival times are commonly observed and Cox's partial likelihood function was modified to handle ties (Kalbfleisch and Prentice, 1980; Breslow, 1974; Efron, 1977). In the following we describe the estimation procedure without and with ties.

2-9-1-1 .Estimation procedures without tied survival times

Suppose that k of the survival times from n individuals are uncensored and distinct, and n-k are right-censored.$t_{(1)} < t_{(2)} < \cdots t_{(k)}$Be the ordered k distinct failure times with corresponding covariates $x_{(1)}, x_{(2)}, \ldots x_{(k)}$. Let $R_{(t_{(i)})}$ be the risk set at time$t_{(i)}$. $R_{(t_{(i)})}$ consists of all persons whose survival times are at least $t_{(i)}$. For the particular failure at

timet$_{(i)}$, conditionally on the risk set $R_{(t_{(i)})}$, the probability that the failure is on the individual as observed is:

$$\frac{\exp\left(\sum_{j=1}^{p} b_j x_{j(i)}\right)}{\sum_{l\epsilon R_{(t_{(i)})}} \exp\left(\sum_{j=1}^{p} b_j x_{jl}\right)}\left(= \frac{\exp(\acute{b}x_{(i)})}{\sum_{l\epsilon R_{(t_{(i)})}} \exp(\acute{b}x_{(l)})}\right)$$

Each failure contributes a factor and hence the partial likelihood function is

$$l(b) = \prod_{i=1}^{k} \frac{exp(\sum_{j=1}^{p} b_j x_{j(i)})}{\sum_{l\epsilon R_{(t_{(i)})}} exp\left(\sum_{j=1}^{p} b_j x_{jl}\right)}\left(= \prod_{i=1}^{k} \frac{exp(\acute{b}x_{(i)})}{\sum_{l\epsilon R_{(t_{(i)})}} exp(\acute{b}x_{(l)})}\right) \quad (2.32)$$

And the log-partial likelihood is

$$l(b) = logl(b) = \sum_{i=1}^{k}\sum_{j=1}^{p} b_j x_{ji} - \sum_{i=1}^{k} \log\left[\sum_{l\epsilon R_{(t_{(i)})}} \exp\left(\sum_{j=1}^{p} b_j x_{jl}\right)\right]$$

$$=\sum_{i=1}^{k}\left\{\acute{b}x_{(i)} - \log\left[\sum_{l\epsilon R_{(t_{(i)})}} \exp(\acute{b}x_{(l)})\right]\right\} \quad (2.33)$$

The maximum partial likelihood estimator (MLE) $\hat{b}$ of b .That is, $\hat{b}_1, \ldots, \hat{b}_p$ are obtained by solving the following simultaneous equations, which are obtained by taking the derivative of l(b)

$$\frac{\partial l(b)}{\partial b} = 0$$

Or

$$\frac{\partial l(b)}{\partial b_u} = \sum_{i=1}^{k}\left[x_{u(i)} - A_{ui}(b)\right] = 0\, u = 1,2,\dots,p \qquad (2.34)$$

Where

$$A_{ui}(b) = \frac{\sum_{l\epsilon R_{(t_{(i)})}} x_{ul}\exp\left(\sum_{j=1}^{p} b_j x_{jl}\right)}{\sum_{l\epsilon R_{(t_{(i)})}} \exp\left(\sum_{j=1}^{p} b_j x_{jl}\right)} = \frac{\sum_{l\epsilon R_{(t_{(i)})}} x_{ul}\exp\left(\acute{b}x_{(l)}\right)}{\sum_{l\epsilon R_{(t_{(i)})}} \exp\left(\acute{b}x_{(l)}\right)} \qquad (2.35)$$

By applying the Newton-Raphson iterated procedure. The second partial derivatives of $l(b)$respective to b_u and b_v, $u, v = 1,2,\dots,p$,in the Newton-Raphson iterative procedure is:

$$I_{uv} = \frac{\partial^2 l(b)}{\partial b_u\, \partial b_v} = -\sum_{i=1}^{k} C_{(uvi)}\left(b_1,\dots,b_p\right) = -\sum_{i=1}^{k} C_{(uvi)}(b)\; u, v = 1,2,\dots,p \quad (2.36)$$

$$C_{uvi}(b) = \frac{\sum_{l\epsilon R_{(t_{(i)})}} x_{ul}x_{vl}\exp\left(\sum_{j=1}^{p} b_j x_{jl}\right)}{\sum_{l\epsilon R_{(t_{(i)})}} \exp\left(\sum_{j=1}^{p} b_j x_{jl}\right)} - A_{ui}(b)A_{vi}(b) \qquad (2.37)$$

The covariance matrix of the MPLE $\hat{b}$, defined similarly as $\widehat{V}(b)$

$$\widehat{V}\left(\hat{b}\right) = \widehat{cov}\left(\hat{b}\right) = \left[-\frac{\partial^2 l\left(\hat{b}\right)}{\partial b\, \partial \acute{b}}\right]^{-1} \qquad (2.38)$$

Where

$$-\frac{\partial^2 l(\hat{b})}{\partial b\, \partial \acute{b}}$$

is called the observed information matrix with $-l_{uv}(\hat{b})$ as its (u, v)

Element and where $l_{uv}(b)$ is defined in (2.36). Let the(i, j) element of $\hat{V}(\hat{b})$ in (2.38) be v_{ij} Then the $100(1-\alpha)\%$ confidence interval for $\hat{b}_i$ is, according to

$$\left(\hat{b}_i - Z_{\propto/2}\sqrt{vii}\,, \hat{b}_i + Z_{\propto/2}\sqrt{vii}\right) \qquad (2.39)$$

2-9-1-2. Estimation procedure with tied survival times

Suppose that among the n observed survival times there are k distinct uncensored times $t_{(1)} < t_{(2)} < \cdots t_{(k)}$ .Let $m_{(i)}$ denote the number of people at $t_{(i)}$ .let $R(t_{(i)})$denote the set of people at risk at time$t_{(i)}$ and $r_i$be the number of such persons. To approximate the exact partial likelihood function, the following two likelihood functions can be used when each$m_{(i)}$is small compared tor_i Breslow (1974) provided the following approximation

$$L_B(b) = \prod_{i=1}^{k} \frac{\exp(\acute{z}_{u^*(i)} b)}{\left[\sum_{l \in R(t_{(i)})} \exp(\acute{x}_l b)\right]^{m_{(i)}}} \qquad (2.40)$$

An alternative approximation was provided by Efron (1977)

$$L_E(b) = \prod_{i=1}^{k} \frac{\exp\left((\acute{z}_{u^*(i)} b)\right)}{\prod_{j=1}^{m_{(i)}}\left[\sum_{l \in R(t_{(i)})} \exp(\acute{x}_l b) - [(j-1)/m_{(i)}] \sum_{l \in u^*_{(i)}} \exp(\acute{x}_l b)\right]} \qquad (2.41)$$

(Lee & Wang, 2003).

2-9-2. Tests of hypotheses:

Two approaches to testing hypotheses about $\hat{\beta}$

2-9-2-1. Likelihood ratio test

This test is denoted by G and is calculated by means of double difference between the logarithm of partial maximum likelihood of the model containing variables (full) and the logarithm of the partial that does not contain variables (reduced) and the test formula is given as follows:

$$G = -2\ln\left[L_p(\hat{\beta}) - L_p(0)\right] \sim \chi^2{}_{r,\alpha} \qquad (2.42)$$

The value of this statistic G follows the distribution of the Chi-Squares (χ^2) with one degree of freedom. Thus, it can be used to test the estimated model.

2-9-2-2. A Wald test

A Wald test is known as Z statistic (standard normal distribution), which is one of two test statistics typically used with ML estimates. For unvaried the null hypothesis that $\beta = 0$ Can be tested by calculating the value of statistic

$$Z = \frac{\hat{\beta}}{se(\hat{\beta})}, \qquad \text{or} \quad \left[\chi^2 = \frac{\hat{\beta}}{se(\hat{\beta})}\right]^2 \qquad (2.43)$$

where Z is standard normal distribution, χ^2 is chi-square distribution on one degree of freedom and $se(\beta)$ is standard error of $\hat{\beta}$

When given a set of covariates X and a corresponding set of coefficient B,

$\hat{B}_j = 0$ in presence of all other terms that in the model. Can be tested by calculating the value of statistic

$Z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$ or $\left[X^2 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}\right]^2$ (collett, 2003; Lemeshow, et al., 2008).

2-10. Interpretation of parameter estimates:

The proportional hazards model used in the analysis of survival data, the coefficient of explanatory variables in the model can be interpreted as logarithms of the ratio of hazard of death to the baseline hazard.

2-10-1. Models with a variable

The proportional hazards model contains a single continuous variable X, so that the hazard function for the ith of on n individuals, for whom X, takes valuesx_i, is

$$h_i(t) = e^{\beta x_i} h_0(t),$$

The coefficient of x_i in this model can then be interpreted as logarithm of a hazard ratio, the ratio of hazard of death for whom the value $x + 1$ is recorded on X, relative to one for whom the value x obtained is. This is

$$\frac{exp[\beta(x+1)]}{exp(\beta x)} = e^{\beta},$$

And so $\hat{\beta}$ in the fitted proportional hazards model is the estimated change in the logarithm of the hazard ratio when the value of X is increased by one unit.when a continuous variable $X$is included in a proportional hazards model, using a similar argument, the estimated change in log hazard ratio when the value of the variable$X$ is increased by r units is $r\hat{\beta}$, and the corresponding estimate of the hazard ratio is $exp(r\hat{\beta})$,from which confidence intervals for true hazard ratio can be derived (collett, 2003).

2-10-2. Models with a factor

When individuals fall into of m groups, $m \geq 2$, which correspond to categories of an explanatory variable, the groups can be indexed by the levels of a factor .under a proportional hazards model, the hazard function for an individual in the group, $j = 1,2, \ldots m$, is given by:

$$h_j(t) = \exp(\gamma_j) h_0(t),$$

Where γ_j is the effect due to the j level of the factor, and $h_0(t)$ is the baseline hazard function, we take $\gamma_1 = 0$.

The baseline hazard function then corresponds to the hazard of death at time t for an individual in the first group. The ratio of the hazards at time is t for an individual in the jth group, ≥ 2 , relative to an individual in the first group , is then $exp(\gamma_j)$. Consequently, the parameter γ_j is the logarithm of this relative hazard is,

$$\gamma_j = \log\left[{h_j(t)} \Big/ {h_0(t)} \right]$$

The estimated logarithm of relative hazard for an individual in group j , relative to an individual in group 1, is then $\hat{\gamma}_j$, and find $A100(1-\alpha)\%$ confidence interval for the ture log-hazard ratio is the interval from

$$\hat{\gamma}_j - z_\alpha/2 \ se(\hat{\gamma}_j) \text{ to } \hat{\gamma}_j + z_\alpha/2 \ se(\hat{\gamma}_j),$$

where $z_\alpha/2$ is the upper $\alpha/2$ -point of the standard normal distribution.

In some applications, the hazard ratio relative to the level of a factor other than the first is required.

The hazard functions for individuals at level j and j' of the factor are

$\exp(\alpha_j)\, h_0(t)$ and $\exp(\alpha_{j'})\, h_0(t)$, respectively, and so the hazard ratio for an individual at level j, relative to one at level j' , is $\exp(\alpha_j - \alpha_{j'})$. The hazard ratio is then $\alpha_j - \alpha_{j'}$, which is estimated by $\hat{\alpha}_j - \hat{\alpha}_{j'}$. To obtain the standard error of estimate, we use the result that the variance of the difference $\hat{\alpha}_j - \hat{\alpha}_{j'}$ is given by

$$\mathrm{var}(\hat{a}_j - \hat{\alpha}_{j'}) = \mathrm{var}(\hat{\alpha}_j) + \mathrm{var}(\hat{\alpha}_{j'}) - 2\mathrm{cov}(\hat{\alpha}_j \hat{\alpha}_{j'})$$

then estimate the covariance between $\hat{\alpha}_j \; and \; \hat{\alpha}_{j'}$,as well as estimates of their variance and compute $se(\hat{\alpha}_j - \hat{\alpha}_{j'})$ (collett, 2003).

2-10-3 .Model with combination of terms

A fitted model will contain terms corresponding to number of variates factors or combination of two, with suitable coding of indicator variables corresponding to factor in model, the parameter estimates can again be interpreted as logarithm of hazard ratios. When model contain more than one variable, the parameter estimate associated with a particular effect is said be adjusted for other variables in model, and so estimates are log-hazard, adjusted for the other terms in the model. The proportional hazards can therefore be used to estimate hazard ratios, taking account of other variables included in the model.

When interactions between factor or mixed terms involving factors and variates, are fitted, the estimated log-hazard ratios for particular factor will differ according to the level of any factor, the value of variate with which it interacts (collett, 2003).

2-11. Model checking in the Cox regression model:

After a model has been fitted to an observed set of data, the adequacy of fitted model needs to be assessed. Indeed, the use of diagnostic procedures for model checking is an essential part of the modeling process. Many model checking procedures are based on quantities known as residuals. These are values that can be calculated for each individual in the study, and have the feature that their behavior is .known, at least

approximately, when the fitted model is satisfactory .A number of residuals have been proposed for use in connection with the Cox regression Model.

2-11-1. Cox- Snell residuals

The residual that is most widely in the analysis of survival data is the Cox- Snell residuals, so called because it is a particular example of the general definition of residuals given by Cox and Snell (1968).

The Cox- Snell residuals for the ith individual $i = 1,2, \dots, n,$ is given by

$$\mathrm{r_{Ci}} = \exp\left(\hat{\beta}\mathrm{x_i}\right)\hat{\mathrm{H}}_{\mathrm{o}}(\mathrm{t_i}) \qquad (2.44)$$

Where $\hat{H}_o(t_i)$ is an estimate of the baseline cumulative hazard function at time t_i , the observed survival time of that individual Nelson –Aalen estimate generally used, The Cox-Snell residual, $r_{Ci}$is the value of

$$\hat{\mathrm{H}}_{\mathrm{i}}(\mathrm{t_i}) = -\log\hat{\mathrm{S}}_{\mathrm{i}}(\mathrm{t_i}),$$

Where$\hat{H}_i(t_i)$ and $\hat{S}_i(t_i)$ are the estimated values of cumulative hazards and survivor function of the ith individual at t_i (collett, 2003).

2-11-2. Martingale residuals

The modified residuals $\acute{r}_{Ci}$a mean of unity for uncensored observations accordingly, these residuals might be further refined by relocating the $\acute{r}_{Ci}$ so that they have a mean zero when an observations is uncensored .if addition the resulting values are multiplied by -1, we obtain the residuals.

$$r_{Mi} = \delta_i - r_{Ci} \qquad (2.45)$$

Martingale residuals take values between $-\infty$ and unity, with residuals for censored observations, where $\delta_i = 0$,being negative. It can also be shown that these residuals sum to zero and, in large samples, the Martingale value of zero.

2-11-3.Deviance residuals

Although martingale residuals share many of the properties possessed by residuals encountered in other situations, such as in linear regression analysis. They are not symmetrically distributed about zero, even when the fitted model is correct. Residuals, which were introduce by T Herneau et al. (1990), are

$$r_{D_i} = \text{sgn}(r_{Mi})[-2\{r_{Mi} + \delta_i \log(\delta_i - r_{Mi})\}]^{1/2}, \qquad (2.46)$$

Where r_{Mi} is the martingale residual for the ith individual, and $sgn(.)$ is the sign function.

This is the function that takes the value $+1$ if its argument is positive and -1if negative (collett, 2003).

2-11-4.Schoenfeld Residuals

These residual differs from those considered previously in one other important respect .this is that there is not a single value of the residual for each individual, but a set of values ,one for each explanatory variable included in the fitted Cox regression model**.** Theith partial or schoenfeld residual for X_j, the jth explanatory variable in the model is given by:

$$\boldsymbol{r_{Pji}} = \delta_i\{x_{ji} - \hat{a}_{ji}\}, \quad (2.47)$$

Where

x_{ji} is the value of the jth explanatory variable, $j = 1,2,\dots,p$, for the ith individual in the study

$$\hat{a}_{ji} = \frac{\sum_{l \epsilon R_{(t_i)}} x_{ji} \exp\left(\hat{\acute{\beta}} x_l\right)}{\sum_{l \epsilon R_{(t_i)}} \exp\left(\hat{\acute{\beta}} x_l\right)}, \qquad (2.48)$$

And$R_{(t_i)}$ is the set of all individuals at risk at timet_i .The ith schoenfeld residual, for the explanatory variable X_j is an estimate of the ith component of the derivative of the logarithm likelihood function with respect to B_j,which, from equation (2.34) is given by:

$$\frac{\partial \log L(b)}{\partial b_j} = \sum_{i=1}^{n} \delta_i \{x_{ji} - a_{ji}\}, \qquad (2.49)$$

Where

$$a_{ji} = \frac{\sum_l x_{ji} \exp(\acute{\beta} x_l)}{\sum_l \exp(\acute{\beta} x_l)} \qquad (2.50)$$

The ith term in this summation evaluated at $\hat{\beta}$, is the schoenfeld residual for X_j, given in (2.47).since the estimates of the $\beta's$ are such that

$$\frac{\partial \log L(b)}{\partial b_j} \Big| \hat{\beta} = 0$$

The schoenfeld residual must sum to zero. These residuals also have the property that, in large sample, the expected value of $r_{Pji}$is zero, and they are uncorrelated with one another. Grambsch and Therneau (1994) proposed a scale version of the schoenfeld residual, is more effective in detecting departures from the assumed model. Let the vector of schoenfeld residual.For the ith individual be denoted

$r_{Pi} = (r_{P1i}, r_{P2i}, \dots, r_{Ppi})'$. The scale, or weighted, schoenfeld residual r^*_{Pi}, are then the components of vector

$$\mathrm{r^*_{Pi}} = \mathrm{rvar}(\hat{\beta})\mathrm{r_{Pi}},$$

Where r is the number of deaths among the n individuals, and $var(\hat{\beta})$ is the variance -covariance matrix of the parameter estimates in the fitted Cox regression model. These scaled schoenfeld residual are therefore quite straight-forward to compute (collett, 2003).

2-12.Assessment of model fit:

A number of plots based on residuals can be used in the graphical assessment of the adequacy of fitted model. Many graphical procedures that are analogues of residual plots used in linear regression analysis have not proved to be very helpful this is because plots of residuals a giants quantities such as the observed survival times, or rank order of these times, often exhibit a define pattern ,even when correct model has been fitted.

2-12-1 .Plots based the Cox-Snell residuals

After computing the Cox-Snell residuals,r_{Ci}, the Kaplan Meier estimate of the survivor function of these values is found. This estimate is computed in similar manner to the Kaplan Meier estimate of survivor function of survival times, except that the data on which the estimate is based are now the residuals$\mathrm{r_{Ci}}$. Residuals obtained from censored survival times are themselves taken to be censored .denoting the estimate by $\hat{S}(r_{Ci})$, the values of $\hat{H}(r_{Ci}) = -log\,\hat{S}(r_{Ci})$are plotted against$r_{Ci}$. This gives a cumulative hazard plot of residuals. A straight line with unit slope and zero intercept will then indicate that the fitted survival model is satisfactory (collett, 2003).

2-12-2. Plots based on martingale and deviance residuals

An index plot of the martingale residuals will highlight individuals whose survival time is not well fitted by the model. Such observations may be termed outliers. The data from individuals for whom the residual is unusually large in absolute value, will need to be subject of further scrutiny. The plots theses residuals against the survival time, the rank order of the survival times, or explanatory variable, may indicate whether there are particular times, or values of variables, where the model does not fit well.

The deviance residuals are symmetrically distributed than the martingale residuals, plots based on these residuals tend to be easier to interpret .consequently, an index plot of the deviance residuals may also be used to identify individuals whose survival times are out of line. By reconciling information about individuals whose survival times are out line, with the values of their risk score, useful information can be obtained about the characteristics of observations that are not well fitted by the model. Plot of deviance residuals against the risk score is particularly helpful diagnostic (collett, 2003).

2-13.Testing the assumption of proportional hazards:

A crucial assumption made when using the Cox regression model is that of proportional hazards. If hazard are not proportional, this means that the linear component of the model varies with time in some manner. We must therefore consider how the validity of this assumption can be examined. This is followed by description of how diagnostics derived from a fitted model can be used in examining the proportional hazards assumption.

2-13-1.Log –cumulative hazard plot

According to Cox regression model, the hazard of death at any time t for the ith individual is given by

$$h_i(t) = \exp(\beta' x_i) h_0(t) \ , \qquad (2.51)$$

$wherex_i$is the vector of values of explanatory variables for that individual

âIs the corresponding vector of coefficients, and $h_0(t)$ is the baseline hazard function .integrating both sides of this equation over t give

$$\int_0^t h_i(u)\,du = \exp(\beta' x_i)\int_0^t h_0(u)\,du.$$

And so, using equation (3.6),

$$H_i(t) = \exp(\beta' x_i)H_0(t),$$

Where$H_i(t)$ and $H_0(t)$ are the cumulative hazard function .taking logarithms of each side of this equation, we get

$$\log H_i(t) = \beta' x_i + \log H_0(t)$$

The differences in the log-cumulative hazard do not depend on time .this means that if the log-cumulative hazard functions for individuals with different values of their explanatory variables are plotted against time, the curves so formed will parallel if proportional hazards model (2.50) is valid. This provides the basis of a widely used diagnostic for assessing the validity of the proportional hazards assumption. To use this plot, the survival data are first grouped according to the levels of one or more factors. If continuous variable are to be feature in this analysis, their values are will first need be grouped in some way to give a categorical variable (collett, 2003).

2-13-2. Schoenfeld residuals

The schoenfeld residuals, defined in section (2.11.4) are be particularly useful in evaluating the assumption of proportional hazard after fitting a Cox regression model. Grambsch and Therneau (1994) have shown that the expected value of the ith scale schoenfeld residuals, for thejth explanatory variable, Xj in the model,r^*_{Pi}, is given by

$E(r^*_{Pi}) \approx \beta_j(t_i) - \hat{\beta}_j,$

Where $B_j(t)$ is taken to be a time-varying coefficient of , $Xj\beta_j(t_i)$ is the value of the coefficient at ith death time,$t_i$and $\hat{\beta}_j$ is the estimated value of $\hat{\beta}_j$ in the fitted Cox regression model. Consequently, a plot of the values of $r^*_{Pi} + \hat{\beta}_j$ against the death times should give information about the form of the time –dependent coefficient of Xj, $\beta_j(t)$. In particular, a horizontal line will suggest that the coefficient of Xj is constant, and the proportional hazards assumption is satisfied. A smoothed curve can be superimposed on this plot to aid interpretation .this plot can also be supplemented by fitted a straight lines, and testing if the slope of this is zero (collett, 2003).

2-13-3.Adding a time - dependent variable

To examine the assumption of proportional hazard in the regression model a time – dependent variable can be added the model. Consider a survival study in which each patient has been allocated to one of two groups, corresponding to standard and a new treatment. Interest may then Centre on whether the ratio of hazard of death at time t in one treatment group, relative to the other, is independent of survival time.

$$h_i(t) = \exp(\beta_1 x_{1i}) h_0(t) \qquad (2.52)$$

Wherex_{1i} is the value of an indicator variable X_1 that is zero for the standard treatment and unity for the new treatment .the relative hazard of death at any time for a patient on the new treatment, relative to one on the standard is then e^{β_1}, which is independent of survival time.

Define a time-dependent explanatory variable X_2 where $X_2 = X_1 t$. if this variable is added to model in equation (2.51), the hazard of death at time t for the ith individual becomes.

$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i}) h_0(t),$ (2.53)

Where

$x_{2i} = x_{1i}t$ is the value of $X_1 t$ for the ithindividual.

The relative hazard at t is

$$\exp(\beta_1 + \beta_2 t), \qquad (2.54)$$

Since $X_2 = t$ under the new treatment, and zero otherwise.

logt Might use in place t in definition of time dependent variable X_2 equation (2.53) a test of hypothesis is that $\beta_2 = 0$ is that a test of proportional hazards, alternative hypothesis is that the hazard ratio is dependent on the logarithm of time.

The test can be carried out using either a Wald a statistic or likelihood ratio statistic .in either case, the test statistic has chi – square distribution with one degree of freedom under the null hypothesis (collett, 2003; Kleinbaum & Klein, 2012).

2-14.Parametric approach:

Parametric approaches are used either when a suitable model or distribution is fitted to the data or when a distribution can be assumed for the population from which the sample is drawn. Commonly used survival distributions are the exponential, Weibull, lognormal, log logistic and gamma (Ravanan, 2015).

2-14-1.Parametric model

A parametric survival model is one in which survival time (the outcome) is assumed to follow a known distribution ,do make assumptions about the distribution of failure times and the relationship between covariates and survival experience but semi-parametric models make no assumption about the distribution of failure times . Parametric models fully specify the distribution of the baseline hazard/survival function according to some (defined) probability distribution. Parametric models are useful when we want to predict survival rather than identify factors that influence survival. Parametric models can be expressed in: (1) proportional hazard form, where a one unit change in an explanatory

variable causes proportional changes in hazard; and (2) accelerated failure time (AFT) form, where a one unit change in an explanatory variable causes a proportional change in survival time. Examples of distributions that are commonly used for survival time are: the Weibull, exponential (a special case of the Weibull), Gompertz, log-logistic, lognormal. For parametric survival models, time is assumed to follow some distribution whose probability density function f (t) can be expressed in terms of unknown parameters. Once a probability density function is specified for survival time, the corresponding survival and hazard functions can be determined (Kleinbaum & Klein, 2012; Abdelaal & Zakria, 2015; George, et al., 2014) **.**

2-14-2.Parametric PH models:

Parametric PH models similar in concept and interpretation to the Cox (PH) model. The key difference between the two is that the hazard is assumed to follow a specific statistical distribution when a fully parametric PH model is fitted to the data, whereas the Cox model enforces no such constraint. Other than this, the two model types are equivalent. Hazard ratios have the same interpretation, whether derived from a Cox or a fully parametric regression model, and the proportionality of hazards is still assumed. A number of different parametric PH models may be derived by choosing different hazard functions. The commonly used models are exponential, Weibull, or Gompertz models. (Bradburn, et al., 2003; Lawless & , 2003).

2-14-2-1. Models for the hazard function

Once distribution model for survival times has been specified in terms of probability density function, the corresponding survivor and hazard functions can be obtained from the relations

Once

$$S(t) = 1 - F(t)$$

So

$$F(t) = P(T < t) \ = \int_0^t f(u)du,$$

Then

$$S(t) = \ 1 - \int_0^t f(u)\, dy \qquad (2.55)$$

And

$$h(t) = \ \frac{f(t)}{s(t)} \ = -\frac{d}{dt}[\log S(t)], \qquad (2.56)$$

Where $f(t)$is the probability density function of the survival times these relationships were derived in 3.1.alternative approach is to specify functional form for the hazard function, from which the survivor function and probability density function can be determined from the equations

$$S(t) = \exp\{-H(t)\}, \qquad (2.57)$$

And

$$f(t) = h(t)S(t) = -\frac{dS(t)}{dt}, \qquad (2.58)$$

where

$$H(t) \ = \int_0^t h(u)\, du$$

is integrated hazard function (collett, 2003).

2-14-2-2. Exponential distribution

The simplest and most important distribution in survival studies is the exponential distribution. In the late 1940s, researchers began to choose the exponential distribution to describe the life pattern of electronic systems. In addition is often referred to as a purely random failure pattern. It is famous for its unique ‘‘lack of memory,’’ which requires that the age of the animal or person does not affect future survival.

The exponential distribution is characterized by a constant hazard rate λ, its only parameter. A high λ value indicates high risk and short survival; a low λ value indicates low risk and long survival. When λ =1, the distribution is often referred to as the unit exponential distribution.

When the survival time T follows the exponential distribution with a parameter λ, the probability density function is referred as

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \ , \lambda > 0 \\ 0 & t < 0 \end{cases} \qquad (2.59)$$

Survivorship function is then

$$S(t) = \exp[-\lambda t] \qquad t \geq 0 \qquad (2.60)$$

The hazard function is

$$h(t) = \lambda. \qquad t \geq 0$$

a constant, independent of t.

The mean and variance of the exponential distribution with parameter are λ,
Respectively;

$\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$

The coefficient of variation is 1 ((Lee & Wang, 2003; Lee & Wang, 2003).
The median of exponential distribution $t(50)$, is such that

$s\{t(50)\} = 0.5$

That is $exp\{-\lambda t(50)\} = 0.5$,

So that $t(50) = {}^{1}/_{\lambda}\ log2$.

More generally, the pth percentile of the survival time distribution is the value $t(p)$ such That

$S\{t(p)\} = 1 - \frac{p}{100}$,

and using equation (3. 60), this is

$t(p) = \frac{1}{\lambda} log\left(\frac{100}{100-p}\right)$. (collett, 2003).

2-14-2-3 .Weibull distribution

The Weibull distribution is a generalization of the exponential distribution. However, unlike the exponential distribution, it does not assume a constant hazard rate and therefore has broader application. The distribution was proposed by Weibull (1939) and its applicability to various failure situations discussed again by Weibull (1951). It has then been used in many studies of reliability and human disease mortality.

The Weibull distribution is characterized by two parameters γ and λ .The value of γ determines the shape of the distribution curve and the value of λ determines its scaling (Lee & Wang, 2003) .

Amore general form of hazard function is such that

$$h(t) = \lambda\gamma t^{\gamma-1}, \qquad (2.61)$$

for $0 \leq t < \infty$, function that depends on two parameters λ and γ , which are both greater than zero.In the particular case where γ = 1, the hazard function takes a constant value λ , and the survival times have an exponential distribution. For other values of γ, the hazard function increases or decreases monotonically, that is, it does not change direction. The shape of the hazard function depends critically on the value of γ , and so γ is known as the shape parameter, while the parameter λ is scale parameter.

The Survivor function is given by

$$S(t) = \left\{ -\int_0^t \lambda\gamma u^{\gamma-1}\, du \right\} = \exp(-\lambda t^{\gamma}) \qquad (2.62)$$

The density function is then

$$f(t) = \lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma}), \qquad (2.63)$$

For $0 \leq t < \infty$, is the density of a random variable that has a Weibull distribution with scale parameter λ and shape parameter γ This distribution will be denoted $W(\lambda, \gamma)$ .the right hand -tail of this distribution is longer than the left -hand one, and so the distribution is positively skewed

The mean, or expected value, of random variable T that has a $w(\lambda, \gamma)$ distribution can be shown $(\text{collett}, 2003)$ to be given by

$$E(T) = \lambda^{-1/\gamma}\Gamma(\gamma^{-1} + 1)$$

and the variance is

$$\sigma^2 = \frac{1}{\lambda^2}\left[\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma^2\left(1 + \frac{1}{\gamma}\right)\right]$$

Where $\Gamma(\gamma)$ is the well-known gamma function defined is

$$\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x} dx = (\gamma - 1)!$$

When γ is positive integer

Values of $\Gamma(\gamma)$ can be found in Abramowitz and Stegun (1964). The coefficient of variation is then

$$CV = \left[\frac{\Gamma(1+2/\gamma)}{\Gamma^2(1+1/\gamma)} - 1\right]^{1/2}$$ (Lee & Wang, 2003)

However, since the Weibull distribution is skewed, a more appropriate and more tractable, summary of the location of the distribution is the median survival time. This is the value $t(50)$ such that

$S\{t(50)\} = 0.5$, so that

$$\exp\{-\lambda[t(50)]^\gamma\} = 0.5,$$

And

$$t(50) = \left\{\frac{1}{\lambda}\log 2\right\}^{1/\gamma} \qquad (2.64) \quad .$$

More generally, the pth percentile of the weibull distribution, $t(p)$ is such that

$$t(p) = \left\{\frac{1}{\lambda}\log\left(\frac{100}{100-p}\right)\right\}^{1/\gamma} \qquad (2.65)$$

The median and other percentiles of the Weibull are therefore much simpler to compute than the mean of the distribution.

Since the Weibull hazard function can take a variety of forms, depending on the value of the shape parameter, γ , and appropriate summary statistics can be easily obtained, this distribution is widely used in the parametric analysis of survival data.

2-14-2-4.Gompertz distribution

The Gompertz model has found application in demography and the biological sciences. Indeed the distribution was introduced by Gompertz in 1825, as a model for human mortality. The hazard function of the Gompertz distribution is given by

$$h(t) = \lambda e^{\gamma t} \qquad (2.66)$$

for $0 \leq t < \infty$ and $\lambda > 0$ .in the particular case where $\gamma = 0$, the hazard function has a constant value λ, and the survival times then have an exponential distribution. The parameter γ determines the shape of hazard function, positive $\gamma > 0$ values leading to a hazard function that increases with time, decreases if $\gamma < 0$. The hazard function can also be expressed as

$h(t) = \exp(\lambda + \gamma t)$, (collett, 2003)

which shows that the log-hazard function is linear in t .

The Gompertz hazard increase or decreases monotonically. The survival function of the Gompertz distribution is given by

$$S(t) = \exp\left[-\frac{e^{\lambda}}{\gamma}(e^{\gamma t} - 1)\right], \qquad (2.67)$$

and the corresponding density function is

$$f(t) = \exp\left[(\lambda + \gamma t) - \frac{1}{\gamma}\left(e^{\lambda+\gamma t - e^{\lambda}}\right)\right], \qquad (2.68)$$

The *p*th percentile is such that

$$t(p) = \frac{1}{\gamma}\log\left[1 - \frac{\gamma}{\lambda}\log\left(\frac{100 - p}{100}\right)\right] \qquad (2.69)$$

From which the median survival time is

$$t(50) = \frac{1}{\gamma}\log\left[1 + \frac{\gamma}{\lambda}\log 2\right] \qquad (2.70)$$

(collett, 2003; Lee & Wang, 2003).

2-14-3.Parametric accelerated failure time models:

Accelerated failure time **(**AFT) model is a failure time model which can be used for the analysis of time to event data. The model works to measure the effect of covariate to "accelerate" or to "decelerate" survival time. AFT model is one such model, and most commonly used are Exponential, Weibull, Log logistic, Lognormal and Generalized Gamma AFT models. Exponential and Weibull parametric models can work both in proportional hazards metric and in AFT metric. Log Logistic, Lognormal and Generalized Gamma models work only in AFT metric. AFT models are an alternative to the PH model for the analysis of survival data. Under AFT models, we measure the direct effect of explanatory variables on the survival time instead of hazards, as we do (Karimi, et al., 2016), (Subrat K. Acharya, 2014).

$(x_1, x_2 \ldots x_p)$ can be expressed as:

$$S(t) = S_0(\varphi t) \qquad (2.71)$$

where $S_0(\varphi)$ is the baseline survival function and φ is an acceleration factor defined to be:

$$\varphi = \exp\{(b_{1x1} + b_{2x2} + \cdots + b_{pxp})\}.$$

The ratio of two survival time is constant for any given survival probability. In order to explain this concept for the case of a single covariate (x_1) with two levels for example $x_1 = 0$, for a placebo group and $x_1 = 1$, for a new treatment group, The survival probabilities $S(t)$, for the placebo and new treatment groups are $S_0(t)$ and $S_0(\varphi t)$, respectively. The proportion of patients who are event-free in the placebo group at any time point t_1 is the same as the proportion of those who are event-free in the new treatment group at a time $t_2 = \varphi t_1$.The ratio of time $t_1/t_2 = \varphi$ or time ratio (TR) constant, where $\varphi > 1$ and $\varphi < 1$, which represent situations where the length of survival is increased and decreased in the new treatment group compared with the placebo, respectively.

Log –linear form of the accelerated failure time model

The AFT model is commonly rewritten as being log-linear with respect to time, giving

$$\log T_i = \mu + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p + \sigma \in \qquad (2.72)$$

T_i random variable associated with the lifetime of the individual in a survival study . In this model, $b_1, b_2, \ldots b_p$ are unknown confidents of the values of p explanatory

variables $, x_1, x_2 \ldots x_{p,}$ and μ, σ are two further parameters, known as intercept and scale parameter, respectively. The quantity $\in$ is a random variable used to model the deviation of the values of $\log T_i$ from the linear part of the model, and $\in$ is assumed to have a particular probability distribution. In this formulation of the model, the b-parameters reflect the effect that each explanatory variable has on survival times positive values suggest that the survival time increases and negative values survival time decrease (collett, 2003), assumed to have a particular probability distribution according to the probability distribution supposed to be followed by the survival time under study.

The AFT model is fitted by applying the maximum likelihood estimation method by using iterative Newton- Raphson procedure. For the sake of simplicity and ease of interpretation, the exponentiated regression coefficients ($\exp(\beta$) called time ratio (TR) is recommended to report like HR is reported in proportional hazards models. TR > 1 for a covariate implies that this slows down or prolongs the time to the event and TR < 1 for a covariate indicates the occurrence of earlier event is more likely to occur.

(William Aknaus, 1993; collett, 2003; Lemeshow, et al., 1999; Moechberger M.L, 1997).

2-14-3-1 .Lognormal Distribution

In its simplest form the lognormal distribution can be defined as the distribution of a variable whose logarithm follows the normal distribution, its origin may be traced as far back as 1879, when McAlister (1879) described explicitly a theory of the distribution. Most of its aspects have since been under study. Gaddum (1945a, b) gave a review of its application in biology, followed by Boag's (1949) applications in cancer research. Consider the survival time T such that log T is normally distributed with mean μ and variance σ^2. We then say that T is lognormally distributed and write T as $\Lambda(\mu, \sigma^2)$ that is mean lognormally distributed with parameters μ and σ^2.It should be noted that μ and σ^2 are not the mean and variance of the lognormal distribution

The probability density function and survivorship function are, respectively,

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log t - \mu)^2\right] \quad t > 0 \ , \sigma > 2 \qquad (2.73)$$

And

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_t^{\infty} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right] dx \qquad (2.74)$$

let $a = \exp(-\mu)$. Then $-\mu = \log a$, (2.60) and (2.61)can be written as

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log at)^2\right] \qquad (2.75)$$

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_t^{\infty} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2}(\log ax)^2\right] dx$$

$$= 1 - \Phi\left(\log\frac{at}{\sigma}\right) \qquad (2.76)$$

Where

$\Phi(y)$ is the cumulative distribution function of a standard normal variable.

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_0^y e^{\frac{-u^2}{2}} du \qquad (2.77)$$

The hazard function, from (2.62) and (2.64), has the form

$$h(t) = \frac{\frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log at)^2\right]}{1 - \Phi\left(\log\frac{at}{\sigma}\right)} \qquad (2.78)$$

The hazard function increases initially to a maximum and then decreases (almost as soon as the median is passed) to zero as time approaches infinity (Watson and Wells 1961). Therefore, the lognormal distribution is suitable for survival patterns with an initially increasing and then decreasing hazard rate.

The mean and variance of the two-parameter lognormal distribution are, respectively,

$$\exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

and

$$[\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$$

The coefficient of variation of the distribution is then

$$[\exp(\sigma^2) - 1]^{\frac{1}{2}}$$

The median survival time under this distribution is simply

$$t(50) = e^{\mu} \qquad (2.79)$$

And

The mode is $\exp(\mu + \sigma^2)$

Where

$\Phi(z)$ is the standard normal distribution function given by

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-u^2}{2}} du$$

The pth percentile of the distribution is then

$$t(p) = \exp[\sigma\Phi^{-1}(p/100) + \mu] \qquad (2.80)$$

Where $\Phi^{-1}(p/100)$, the pth percentile of the standard normal distribution is often called the probito f $p/100$.

The popularity of the lognormal distribution is due in part to the fact that the cumulative values of $y = \log t$ can be obtained from the tables of the standard normal distribution and the corresponding values of t are then found by taking antilog. Thus, the percentiles of the lognormal distribution are easy to find (Lee & Wang, 2003; collett, 2003).

2-14-3-2.Log –logistic distribution

One limitation of the Weibull hazard function is that it is monotonic function of time .however, situations in which the hazard function changes direction can rise. A particular form of unimodal is hazard the function

$$h(t) = \frac{\alpha\gamma t^{\gamma-1}}{1 + \alpha t^{\gamma}} \qquad (2.81)$$

For $0 \leq t < \infty, \gamma > 0$. This hazard function decreases monotonically if $\gamma \leq 1$, but if $\gamma > 1$, the hazard has a single mode. The survivor function corresponding to the hazard function in equation (2.81) is given by

$$S(t) = [1 + \alpha t^{\gamma}]^{-1} \qquad (2.82)$$

and the probability density function is

$$f(t) = \frac{\alpha\gamma t^{\gamma-1}}{(1+\alpha t^{\gamma})^2} \qquad (2.83)$$

This is the density of a random variable T that has a log-logistic distribution, with parameters α, γ . The distribution is so called because the variable log T has a logistic distribution, a symmetric distribution whose probability density function is very similar to that of the normal distribution.

The pth percentile of the log-logistic distribution is

$$t(p) = \left(\frac{pe^{-\alpha}}{100-p}\right)^{1/\gamma} \qquad (2.84)$$

and so the median of the distribution is

$$t(50) = e^{-\alpha/\gamma} \qquad (2.85).$$

(Lee & Wang, 2003; collett, 2003).

2-15.Assessment the suitability of parametric model:

Amore informative way of assessing whether a particular distribution for the survival times is plausible is to compare the survivor function for data of chosen model. This is greatly helped by transforming the survivor function to produce a plot that should give a straight line if assumed model is appropriate (collett, 2003). The basic idea of the three graphical methods is to see if the survival time itself, or a function of it, has a linear relationship with the distribution function and the cumulative hazard function of a given parametric distribution, or a function of the distribution function and the cumulative hazard function. If such a linear

relationship exists, it can be demonstrated graphically as a straight line. Thus, if one chooses the appropriate distribution and makes a probability, or hazard, plot, the result will be a straight line fit to the data. Parameters of the distribution chosen can be estimated from the probability or hazard plots without tedious numerical calculations (Lee & Wang, 2003).

2-15-1. Assessment the suitability of Weibull model

Since the survivor function for a Weibull distribution is.

$$S(t) = \exp(-\lambda t^{\gamma}) \qquad (2.86)$$

Taking logarithm of $S(t)$, multiplying by -1 , and taking logarithm a second time, gives.

$$\log\{-\log S(t)\} = \log\lambda + \gamma \log t. \quad (2.87)$$

We now substitute the Kaplan- Meier estimate of survivor function, $\hat{S}(t)$ for $S(t)$ in equation (2.87).

This property allows a graphical evaluation of the appropriateness of a Weibull model by plotting

$\log\{-\log\hat{S}(t)\}$ a against $\log t$ would then give an approximately straight line, the intercept and slope of the straight line will be $\log(\lambda)$and slope γ , respectively (collett, 2003).

2-15-2. Assessment the suitability of Exponential model

The exponential cumulative distribution function is

$$F(t) = 1 - \exp[-(\lambda t)] t > 0 \qquad (2.88)$$

The probability plot for the exponential distribution is based on the relationship between t and F(t), from (3.88),

$$t = \frac{1}{\lambda} \log \frac{1}{1 - F(t)} \qquad (2.89)$$

This relationship is linear between t and the function $\log\left[\frac{1}{(1-F(t))}\right]$

Thus, an exponential probability plot is made by plotting the ith ordered observed survival time $t_{(i)}$ versus $\log\left[\frac{1}{\left(1-\hat{F}(t_{(i)})\right)}\right]$

where $\hat{F}(t_{(i)})$ is an estimate of $F(t_{(i)})$ (collett, 2003),

2-15-3. Assessment the suitability of Gompertz model

From hazard function of Gompertz distribution, given equation

$\log h(t) = \lambda + \gamma t$,

If log of empirical hazard against. Time is linear, underlying distribution may be Gompertz

Empirical hazards:

$$\hat{h}(t_i) = \frac{-[\log(S(t_i)) - \log(S(t_{i-1}))]}{t_i - t_{i-1}} \qquad (2.90)$$

2-15-4. Assessment the suitability of lognormal model

From the survivor function of the lognormal distribution, given in equation (2.76)

$$\Phi^{-1}\{1 - S(t)\} = \frac{\log t - \mu}{\sigma} \qquad (2.91)$$

Where $\Phi^{-1}(.)$ is standard normal distribution function, and so a plot of $\Phi^{-1}\{1 - S(t)\}$ against $\log t$ should given a straight line, if the lognormal model is appropriate. The slope and intercept of this line provide estimate of $\sigma^{-1}, \frac{-\mu}{\sigma}$, respectively (collett, 2003).

2-15-5. Assessment the suitability of Log logistic model

The log logistic distribution function is

$$F(t) = \frac{\alpha t^{\gamma}}{1 + \alpha t^{\gamma}} \qquad t > 0, \quad \gamma > 0, \quad \alpha > 0 \qquad (2.92)$$

A probability plot for the log-logistic distribution is based on the following relationship obtained from (2.92):

$$\log t = \frac{1}{\gamma}\log\left[\frac{1}{1 - F(t)} - 1\right] - \frac{1}{\gamma}\log\alpha \qquad (2.93)$$

Thus a log logistic probability pot is graph of $\log(t_{(i)})$ versus $\log\left(\left\{\frac{1}{[1-\hat{F}(t_{(i)})]}\right\} - 1\right)$ where $\hat{F}(t_{(i)})$ is estimate of $F(t_{(i)})$ (Lee & Wang, 2003).

2-16.Fitting a parametric model to a single sample:

Parametric models can be fitted to an observed set of survival data using the method of maximum likelihood, first the situation where actual survival times have been observed for n individuals, so that there are no censored observations. If the probability density function of random variable associated with survival time is$f(t)$, the likelihood of the n observations $t_1, t_2, \ldots, t_n$ is simply the product

$$\prod_{i=1}^{n} f(t_i)$$

This likelihood will be a function of the unknown parameters in the probability density function, and the maximum likelihood estimates of these parameters are those values for which the likelihood function is maximum .in practice, it is generally more convenient to work with the logarithm of likelihood function

Suppose that r of the n individuals die at times $t_1, t_2, \ldots, t_r$, and that survival times of the remaining $n - r$ individuals, ${t_1}^*, {t_2}^*, \ldots, {t^*}_{n-r}$, are right –censored .therdeath times contribute a term of the form

$$\prod_{j=1}^{r} f(t_j)$$

To the overall likelihood function .Naturally, we cannot ignore information about the survival experience of $n - r$ individuals for whom a censored survival time has been recorded if a survival time is censored at time t^* , and the probability of this event is $p(T \geq t^*)$, which is $S(t^*)$ .thus each censored observation contributes a term of likelihood of n observations the total likelihood function is therefore

$$\prod_{j=1}^{r} f(t_j) \prod_{i=1}^{n-r} S(t^*) \qquad (2.94)$$

In which the first product is taken over the r death and the second over the $n-r$ censored survival times. Suppose that the data are regarded as n pairs of observation, where the pair for ith individual is

(t_i, δ_i) , $i = 1,2, \dots, n$ .In this notation,

δ_i is an indicator variable that take the value zero when the survival time t_i is censored and unity when t_i is an uncensored survival time. The likelihood function can then written as

$$\prod_{i=1}^{n} \{f(t_i)\}^{\delta_i} \{S(t_i)\}^{1-\delta_i} \qquad (2.95)$$

An alternative expression for the likelihood function can be obtained by writing expression (2.95)

$$\prod_{i=1}^{n} \left\{\frac{f(t_i)}{S(t_i)}\right\}^{\delta_i} ,$$

So that from equation (2.4) becomes

$$\prod_{i=1}^{n} [h(t_i)]^{\delta_i} S(t_i) \ . \qquad (2.96)$$

Estimates of the unknown parameters in this likelihood function are then found by maximizing the logarithm of likelihood function (collett, 2003).

2-16-1.Fitting the Exponential distribution

Suppose that the survival times of n individuals, $t_1, t_2, \ldots, t_n$, are assumed to have an exponential distribution .further suppose that the data give the actual death times of r individuals ,and that remaining n-r survival times are right –censored (collett, 2003). For the exponential distribution,

$$f(t) = \lambda e^{-\lambda t} , \quad S(t) = e^{-\lambda t}$$

The likelihood function for the n observations is given by

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} \left(\lambda e^{-\lambda t_i}\right)^{\delta_i} \left(e^{-\lambda t_i}\right)^{1-\delta_i} ,$$

Where

δ_i is zero if the survival time of the ith individual is censored and unity otherwise. After some simplification,

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} \lambda^{\delta_i} e^{-\lambda t_i},$$

And the corresponding log-likelihood function for uncensored individuals is

$$\log\mathcal{L}(\lambda) = \sum_{i=1}^{n} \delta_i \log\lambda - \lambda \sum_{i=1}^{n} t_i .$$

Since the data contain r deaths, $\sum_{i=1}^{n} \delta_i = r$ and the likelihood function becomes

$$\log\mathcal{L}(\lambda) = r \log\lambda - \lambda \sum_{i=1}^{n} t_i .$$

We now need to identify the value $\hat{\lambda}$, for which the log-likelihood function is a maximum .Differentiation with respect to λ gives

$$\frac{\mathrm{d}\log\mathcal{L}(\lambda)}{\mathrm{d}\,\lambda} = \frac{\mathrm{r}}{\lambda} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{t_i}\,,$$

And equating the derivative to zero and evaluating it at $\hat{\lambda}$ gives

$$\hat{\lambda} = \frac{\mathrm{r}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{t_i}} \qquad (2.97)$$

For the maximum likelihood estimator of λ .The mean of an exponential distribution is $\mu = \lambda^{-1}$, and so the maximum likelihood estimator of μ is

$$\hat{\mu} = \lambda^{-1} = \frac{1}{r}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{t_i}$$

where $\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{t_i}$ is the total time survived by the n individuals in the data set and r is number of deaths observed.

The standard error of either $\hat{\lambda}$or $\hat{\mu}$ can be obtained from the second derivative of the of the log-likelihood function. Differentiating $log\ \mathcal{L}(\lambda)$ a second time gives

$$\frac{d^2\log\mathcal{L}(\lambda)}{d\lambda^2} = -\frac{r}{\lambda^2}$$

and so asymptotic variance of $\hat{\lambda}$ is

$$Var(\hat{\lambda}) = \left\{-\mathrm{E}\left(\frac{d^2\log\mathcal{L}(\lambda)}{d\lambda^2}\right)\right\}^{-1} = \frac{\lambda^2}{r}$$

Consequently, the standard error of $\hat{\lambda}$ is given by

$$se(\hat{\lambda}) = {\hat{\lambda}}/{\sqrt{r}} \qquad (2.98)$$

This result could be used to obtain a confidence interval for mean survival time. The limits of A $100(1-\alpha)\%$ Confidence interval for λ are.

$$\hat{\lambda} \pm z_{\alpha/2} se(\hat{\lambda}) \ ,$$

where $z_{\alpha/2}$ is the upper $\alpha/2$-point of the standard normal distribution (collett, 2003).

Once an estimate of λ has been found, the estimated hazard function is $\hat{h}(t) = \hat{\lambda}$ and the estimated survivor function is $\hat{S}(t) = \exp(-\hat{\lambda}t)$. In addition, the estimated $p\ th$ percentile is given by

$$\hat{t}(p) = \frac{1}{\hat{\lambda}} log\left(\frac{100}{100-p}\right) \qquad (2.99)$$

and the estimated median survival time is that

$$\hat{t}(50) = {1}/{\hat{\lambda}} \log 2 \quad . \qquad (2.100)$$

The standard error of an estimate of the $p\ th$ percentile of distribution of survival times can be found using the result for the approximate variance of a function of a random variable given in equation (2.13).according to this result, an approximate to the variance of a function $g(\hat{\lambda})$ of $\hat{\lambda}$

Is such that

$$\text{var}\{g(\hat{\lambda})\} \approx \left\{\frac{dg(\hat{\lambda})}{d\hat{\lambda}}\right\}^2 \text{var}(\hat{\lambda}) \qquad (2.101)$$

Using this result, the approximate variance of the estimated pth percentile is given by

$$\text{var}\{\hat{\text{t}}(\text{p})\} \approx \left\{-\frac{1}{\hat{\lambda}^2}\log\left(\frac{100}{100-\text{p}}\right)\right\}^2 \text{var}(\hat{\lambda})$$

On simplifying this and taking the square root, we get

$$se\{\hat{\text{t}}(\text{p})\} = \frac{1}{\hat{\lambda}^2}\log\left(\frac{100}{100-\text{p}}\right) se(\hat{\lambda}),$$

and on further substituting for $se(\hat{\lambda})$ from equation (2.98) and $\hat{t}(p)$ from equation (2.100) we find

$$se\{\hat{\text{t}}(\text{p})\} = \frac{\hat{\text{t}}(\text{p})}{\sqrt{r}} \quad . \qquad (2.102)\ .$$

In particular, the standard error of the estimated median survival is

$$se\{\hat{\text{t}}(50)\} = \frac{\hat{\text{t}}(50)}{\sqrt{r}}. \qquad (2.103)$$

Confidence, intervals for a true percentile are best obtained from exponentiating the confidence limits for the logarithm of the percentile. This procedure ensures that confidence limits for the percentile will be non-negative. Again making use of the result in equation (2.101), the standard error of log $\hat{\text{t}}(\text{p})$ is given by

$$se\{log\,\hat{\text{t}}(\text{p})\} = \hat{t}(p)^{-1}\, se\{\hat{\text{t}}(\text{p})\},$$

and after substituting for $se\{\hat{\text{t}}(\text{p})\}$ from equation (2.102), this standard error becomes

$$se\{log\,\hat{\text{t}}(\text{p})\} = {}^{1}/_{\sqrt{r}}$$

Using this result, $100(1-\alpha)\%$ confidence limits for 100 pth percentile are

$$exp\left\{log\,\hat{t}(p) \pm z_{\alpha/2}\,{}^{1}/_{\sqrt{r}}\right\}, \quad \text{that is } \hat{t}(p)exp\left\{\pm z_{\alpha/2}\,{}^{1}/_{\sqrt{r}}\right\},$$

Where $z_{\alpha/2}$ is upper $\alpha/2$-point of standard normal distribution.

2-16-2.Fitting the Weibull distribution

The survival times of n individuals are now taken to be censored sample from a Weibull distribution with scale parameter λ and shape parameter γ . Suppose that there are γ deaths among the n individuals and $n-r$ right censored survival times. To obtain the likelihood of the sample data .the probability density, survivor and hazard function of a W (λ, γ) (collett, 2003) distribution are given by

$$f(t) = \lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma})$$

$$S(t) = \exp(-\lambda t^{\gamma})$$

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

and so, from expression (2.95),the likelihood of n survival times is

$$\prod_{i=1}^{n}\{\lambda\gamma t_i{}^{\gamma-1}\exp(-\lambda t_i{}^{\gamma}\}^{\delta_i}\{\exp(-\lambda t_i{}^{\gamma})\}^{1-\delta_i},$$

where δ_i is zero if the ith a survival time is censored and unity otherwise. Equivalently, from expression (2.96), the likelihood function is

$$\prod_{i=1}^{n}\{\lambda\gamma t_i^{\gamma-1}\}^{\delta_i} \exp(-\lambda t_i^{\gamma}).$$

This is regarded as a function of λ and γ , the unknown parameters in the weibull distribution, and so can be written $L(\lambda,\gamma)$. The corresponding log-likelihood function is given by

$$\log L(\lambda,\gamma) = \sum_{i=1}^{n}\delta_i \log(\lambda\gamma) + (\gamma-1)\sum_{i=1}^{n}\delta_i \log t_i - \lambda\sum_{i=1}^{n} t_i^{\gamma},$$

And noting that $\sum_{i=1}^{n}\delta = r_i$, the log-likelihood becomes

$$\text{Log } L(\lambda,\gamma) = r\text{Log}(\lambda\gamma) + (\gamma-1)\sum_{i=1}^{n}\delta_i \log t_i - \lambda\sum_{i=1}^{n} t_i^{\gamma}.$$

The maximum likelihood estimate of λ and γ are found by differentiating this function with respect to λ and γ , equating the derivatives to zero , and evaluating them at $\hat{\lambda}$ and $\hat{\gamma}$. The resulting equations are

$$\frac{r}{\hat{\lambda}} - \sum_{i=1}^{n} t_i^{\hat{\gamma}} = 0, \qquad (2.104)$$

And

$$\frac{r}{\hat{\gamma}} + \sum_{i}^{n}\delta_i \log t_i - \hat{\lambda}\sum_{i=1}^{n} t_i^{\hat{\gamma}} \log t_i = 0. \qquad . \quad (2.105)$$

From equation (2.104),

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} t_i^{\hat{\gamma}}}, \qquad (2.106)$$

and on substituting for $\hat{\lambda}$ in equation (3.105), we get equation

$$\frac{r}{\hat{\gamma}} + \sum_i^n \delta_i \log t_i - \frac{r}{\sum_{i=1}^n t_i^{\hat{\gamma}}} \sum_{i=1}^n t_i^{\hat{\gamma}} \log t_i = 0 \qquad (2.107)$$

This is a non – linear equation in $\hat{\gamma}$, which can only be solved using an iterative numerical procedure. Once the estimate, $\hat{\gamma}$, which satisfies equation (2.107), has been found, equation (2.106) can be used to obtain $\hat{\lambda}$. In practice, a numerical procedure, such as the Newton-Raphson algorithm, is used to find the values $\hat{\lambda}$ and $\hat{\gamma}$ which maximize the likelihood function simultaneously.

Once estimate of the parameters λ and γ have been fitting the Weibull distribution to the observed data, percentiles of the survival time distribution can be estimated using equation (2.65). The estimated pth percentiles of distribution is

$$\hat{t}(p) = \left\{\frac{1}{\hat{\lambda}} log\left(\frac{100}{100-p}\right)\right\}^{1/\hat{\gamma}} \qquad (2.108)$$

and so the estimate median survival time is given by

$$\hat{t}(50) = \left\{\frac{1}{\hat{\lambda}} \log 2\right\}^{1/\hat{\gamma}} \qquad (2.109)$$

The standard error of the estimated pth percentile can obtained using a generalization of the result in equation (2.101) to the case where the approximate variance of a function of two estimates is required.

$$se\{\hat{t}(p)\} = \frac{\hat{t}(p)}{\hat{\lambda}\hat{\gamma}^2}\left\{\hat{\gamma}^2 \, var(\hat{\lambda}) + \hat{\lambda}^2 (c_p - \log \hat{\lambda})^2 var(\hat{\gamma}) + 2\hat{\lambda}\hat{\gamma}(c_p - \log \hat{\lambda}) cov(\hat{\lambda}, \hat{\gamma})\right\}^{\frac{1}{2}}. \qquad (2.110)$$

where

$$C_p = \log\log\left(\frac{100}{100-p}\right)$$

The variances of $\hat{\lambda}$ and $\hat{\gamma}$, and their covariance, are found form the variance covariance matrix of estimate.

A confidence interval for the true value of the pth percentile, $t(P)$, is best obtained from the corresponding interval for $\log t(P)$ .The standard error of $\log \hat{t}(p)$ is

$$se\{\log \hat{t}(p)\} = \frac{1}{\hat{t}(p)} se\{\hat{t}(p)\} \qquad (2.111)$$

and $100(1-\alpha)\%$ confidence limits for $\log t(P)$ are

$$\log \hat{t}(p) \pm z_{\alpha/2} se\{\log \hat{t}(p)\},$$

2-16-3.Fitting Gompertz distribution

Estimation of λ and γ for data with or without censored observations

Assume that $t_1, t_2, \dots, t_n$ are the observed survival times from n individuals and the survival times follow the Gompertz distribution, without loss of generality, and assume that $t_1, t_2, \dots, t_r$ are uncensored and $t^+_{r+1}, t^+_{r+2}, \dots, t^+_n$ right-censored. The MLE of λ and γ can be obtained by solving the equations (Lee & Wang, 2003).Using the survival function and density function of the Gompertz distribution is given from equation (2.67) and (2.68).

$$r + \frac{e^{\lambda}}{\gamma}\left\{\sum_{i=1}^{r}[1-\exp(\gamma t_i)] + \sum_{i=r+1}^{n}[1+\exp(\gamma t_i^+)]\right\} = 0 \qquad (2.112)$$

$$\sum_{i=1}^{r} t_i - \frac{e^{\lambda}}{\gamma^2}\left\{\sum_{i=1}^{r}[1+(\gamma t_i - 1)\exp(\gamma t_i)] + \sum_{i=r+1}^{n}[1+(\gamma {t_i}^{+} - 1)]\exp(\gamma {t_i}^{+})\right\} = 0 \quad (2.113)$$

Using the Newton-Raphson iterative procedure

2-16-4.Fitting lognormal distribution

When the data are progressively censored, let $t_1, t_2, \dots, t_r$ be uncensored and ${t^+}_{r+1}, {t^+}_{r+2}, \dots, {t^+}_{n}$ be censored observations, the likelihood function (Lee & Wang, 2003), using (2.73), (2.80)

$$l(\mu, \sigma^2) = \frac{r\log(2\pi\sigma^2)}{2} - \sum_{i=1}^{r}\left(\log t_i + \frac{(\log t_i - \mu)^2}{2\sigma^2}\right) + \sum_{i=r+1}^{n}\log\left\{\int_{{t_i}^{+}}^{\infty}\frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}(\log x - \pi)^2\right]dx\right\}$$

And the MLE of μ and σ^2 can be obtained by solving the following two equations

$$\sum_{i=1}^{r}\frac{\log t_i - \mu}{\sigma^2} + \sum_{i=r+1}^{n}\frac{\int_{{t_i}^{+}}^{\infty}\frac{\log x - \mu}{x\sigma^2\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right]dx}{\int_{{t_i}^{+}}^{\infty}\frac{1}{x\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right]dx} = 0 \quad (2.114)$$

$$-\frac{n}{2\sigma^2} + \sum_{i=1}^{r}\frac{(log\, t_i - \mu)^2}{2\sigma^4} + \sum_{i=r+1}^{n}\frac{\int_{t_i^{+}}^{\infty}\frac{(log\, x - \mu)^2}{x2\sigma^4\sqrt{2\pi\sigma^2}}exp\left[-\frac{1}{2\sigma^2}(log\, x - \mu)^2\right]dx}{\int_{t_i^{+}}^{\infty}\frac{1}{x\sqrt{2\pi\sigma^2}}exp\left[-\frac{1}{2\sigma^2}(log\, x - \mu)^2\right]dx} = 0 \quad (2.115)$$

Again this can be done by applying the Newton-Raphson iterative procedure.

2-16-5.Fitting log logistic distribution

let $t_1, t_2, \dots, t_r$ be uncensored and $t^+_{r+1}, t^+_{r+2}, \dots, t^+_n$ be censored observations, the likelihood function, using the censored observations from n persons and the survival times follow the log-logistic distribution. Then the MLE of α and γ can be obtained from solving the following two simultaneous equations (Lee & Wang, 2003):

$$r - \alpha\left(2\sum_{i=1}^{r}\frac{t_i^{\gamma}}{1+\alpha t_i^{\gamma}} + \sum_{i=r+1}^{r}\frac{t_i^{+\gamma}}{1+\alpha\, t_i^{+\gamma}}\right) = 0 \qquad (2.116)$$

$$\frac{r}{\gamma} + \sum_{i=1}^{r}\log(t_i) - \alpha\left[2\sum_{i=1}^{r}\frac{t_i^{\gamma}\log(t_i)}{1+\alpha t_i^{\gamma}} + \sum_{i=r+1}^{n}\frac{t_i^{+\gamma}\log t_i^{+}}{1+\alpha\, t_i^{+\gamma}}\right] = 0 \quad (2.117)$$

Using the Newton-Raphson iterative procedure, if all the survival times observed are uncensored, the respective equations for all the MLE of α and γ can be obtained simply by replacing $r\ with\ n$ in (2.116) and (2.117).

2-17. Strategy for model selection:

One of these criteria is the information criterion of Akaike (AIC), the Baysian Information Criterion (BIC) and the Cox-Snell Information Criterion (CSIC), the latter of which is a graphic rather than a mathematical criterion, many of the criteria used to choose the best model from different models deal with the same data for prediction in the future.

2-17-1. Akaike's Information Criterion(AIC)

AIC: Comparisons may also be made on the basis of statistics between a varieties of potential models which do not necessarily need to be nested (Klein & Moeschberger, 1997; Pourhoseingholi, et al., 2007; Akaike, 1974; Collett, 2003).

$$\mathrm{AIC} = -2(\log \mathrm{likelihood}) + 2(\mathrm{P} + \mathrm{K}) \qquad (2.118) \quad .$$

Where P is the number of parameters, and K is the number of (excluding constant) coefficients in the model. For P=1, for P=2, for Weibull and Gompertz, for the exponential. The smaller the value of this statistic, the better the model, and the better this statistic is known as Akaike's knowledge criterion.

The statistic $-2\log \hat{\mathrm{L}} = -2(\log \mathrm{likelihood})$

To compare alternative models fitted to an observed set of survival data, a statistic that measures the extent to which the data are fitted by a particular model is required. For reasons given in sequel it is more convenient to use minus twice the logarithm of the maximized likelihood in comparing alternative models. If the maximized likelihood for a given model is denoted by $\hat{\mathrm{L}}$, the summary measure of agreement between the model and the data is $-2\log$.

$\hat{\mathrm{L}}$ is in fact the product of a series of conditional probabilities, and so this statistic will be less than unity .in consequence, $-2\log$ will always be positive, and for a given data set, the smaller the value of $-2\log$, the better the model.

The statistic $-2\log \hat{\mathrm{L}}$ cannot be used on its own as a measure of model adequacy. The reason for this is that the value of $\hat{\mathrm{L}}$, and hence of $-2\log \hat{\mathrm{L}}$ is dependent upon the number of observation in data set. The value $-2\log \hat{\mathrm{L}}$ is only useful when making comparisons between models fitted to the same data (collett, 2003).

2-17-2. Baysian Information Criteria (BIC)

$$\mathrm{BIC} = -2(\log \mathrm{likelihood}) + (\mathrm{P} + \mathrm{K}) * \log(\mathrm{n}) \qquad (2.119)$$

In the distribution, where P is the number of parameters, K is the number of coefficients and n is the number of observations. As the best-fit model, the distribution

that has the lowest BIC value is considered is given by (Saikia & Barman, 2017; Schwarz, (1978)) .

2-17-3. residuals for parametric models:

Suppose that T_i is the random variable associated with the survival time of the ith individual. $i = 1,2, \dots n,$ and that $x_{1i,}x_{2i}, \dots, x_{pi}$

Are values of p explanatory variables,$x_1, x_2, \dots x_p$ for this individual .assuming an accelerated failure time model for T_i, we have that

$\log T_i = \mu - \sigma_1 x_{1i} + \sigma_2 x_{2i} + \cdots + \sigma_p x_{pi} + \sigma\epsilon_i,$

2-17-3-1.Cox-Snell residuals

The Cox-Snell residuals are essentially the estimated values of cumulative hazard function for ith observation, at the corresponding event time, t_i. Residuals that have a similar form may also be used in assessing the adequacy of parametric models .the main difference is that now the survivor and hazard functions are parametric functions that depend on the distribution adopted for survival times. In particular, the estimate survivor function for ih individual, on fitting an accelerated failure time model given by

$$\hat{S}_i(t) = S_{\epsilon_i}\left(\frac{\log t - \hat{\mu} - \hat{\alpha}_1 x_{1i} - \hat{\alpha}_2 x_{2i} - \cdots - \hat{\alpha}_p x_{pi}}{\hat{\sigma}}\right), \qquad (2.120)$$

Where

S_{ϵ_i} is the survivor function of ϵ_i in accelerated failure time model, $\hat{\alpha}_j$ is the estimated coefficient of $x_{ji}, j = 1, 2, \dots, p,$ and $\hat{\mu}, \hat{\sigma}$ are estimated values of μ and σ.The form $S_{\epsilon_i}(t)$ for some commonly .

The Cox-Snell residuals for parametric model are defined is given by (Saikia & Barman, 2017; Cox & Snell, 1968) .

$$r_{Ci} = \hat{H}_i(t_i) = -\log\hat{S}_i(t_i), \qquad (2.121)$$

where $\hat{H}_i(t_i)$ is estimated cumulative hazard function, and $\hat{S}_i(t_i)$ is the estimated survivor function in equation (3.61), for ih individual, on fitting an accelerated failure time mode, evaluated at t_i .

2-17-3-2.Martingale residuals

The Martingale residuals provide a measure of difference between the observed number of deaths in the interval $(0, t_i)$ which is either 0 or 1 , and the number predicted by the model .observations with unusually large Martingale residuals are not well fitted by the model .the analogue of the Martingale residuals defined for the Cox regressions model . In question (3.43) is such that

$$r_{Mi} = \delta_i - r_{Ci}, \qquad (2.122)$$

Where δ_i is the event indicator for the ith observation, so that δ_i is unity if that observation is an event and zero if censored, and now r_{Ci} is the Cox -Snell given in equation (2.63) (collett, 2003).

2-17-3-3.Deviance residuals

The deviance residuals can regard as an attempt the martingale residuals symmetrically distributed about zero, and defined by

$$r_{D_i} = \text{sgn}(r_{Mi})[-2\{r_{Mi} + \delta_i \log(\delta_i - r_{Mi})\}]^{1/2}, \qquad (2.123)$$

It is important to note that these quantities are not components of the deviance for the fitted parametric model, but nonetheless it will be convenient to continue to refer to them as deviance residuals (collett, 2003).

Chapter Three

Basic Concept of Kidney Failure Disease

3-0 Preface:

The human kidneys are two relatively small but vital organs, located at the end of the rib cage, one on each side of the spine, behind the peritoneal cavity .Their main roles are removal of waste products from the blood, maintenance of homeostatic balance in the body, and secretion of important hormones. In order to perform their tasks, the kidneys have evolved over thousands of years into organs of an unusually sophisticated anatomy and physiology (Boulpaep, 2004), Kidney or renal failure means that the kidneys stop removing the waste from the blood. Stage 5 kidney failure is, also, known, as End- Stage Renal Disease (ESRD) and it is treated by dialysis or a kidney transplant (Prevention, 2014).

3-1. Kidney's function

The Kidney's function is filtering the blood. All the blood in our bodies passes through the kidneys several times a day. The kidneys remove wastes, control the body's fluid balance, regulate the balance of electrolytes, and the kidneys balance the salts and minerals such as calcium, phosphorus, sodium, and potassium that circulate in the blood. Your kidneys also make hormones that help control blood pressure, make red blood cells, and keep your bones strong. As the kidneys filter blood, they create urine, which collects in the kidneys' pelvis funnel shaped structures that drain down tubes called ureters to the bladder. Each kidney contains around a million units called nephrons, each of which is a microscopic filter for blood. It's possible to lose as much as 90% of kidney function without experiencing any symptoms or problems. (Matthew Hoffman, 2016; Cente, 2016).

3-2. Renal diseases

Considering their importance and complexity, it is not surprising that the impairment of the kidney functions has serious consequences for the organism. The most fundamental is dysfunction of other tissues and organs caused by the waste

products accumulating in the blood. Without appropriate treatment, that will very likely lead to a prompt death. Sadly, in Western countries renal diseases are becoming increasingly widespread. It has been estimated that 1 in 10 people has some kind of kidney disease, and 1% of them will in due course develop end stage renal disease and will need renal replacement therapy (either dialysis or transplantation) to stay alive. Although the final tragic outcome, the end stage renal disease, characterized by the loss of nephrons and deposition of extracellular matrix, is common for all sorts of non-malignant kidney diseases if they are not stopped at an earlier stage, the initial causes vary greatly. Renal disease may start with hypertension, glomerulonephritis (inflammation in the glomerulus, usually resulting from an infection or autoimmune reaction), formation of large cysts gradually replacing the normal tissue (like in the most frequent inherited kidney disease known as polycystic kidney disease), or overuse of medicines or "street" drugs]. The most common cause, however, is poorly controlled diabetes (Boulpaep, 2004)

3-3. Causes of kidney disease

There are many causes of kidney disease. In the United States, diabetes and high blood pressure are the two leading causes. Some conditions are inherited (run in families); people may be born with abnormal kidneys. The following are some of the most common causes of kidney damage.

Diabetes is a disease in which your body does not make enough insulin the hormone that processes sugar or cannot properly use normal amounts of insulin. The result is a high blood sugar level, which can cause problems in many parts of your body.

High blood pressure (also known as hypertension) is another common cause of kidney disease and other illnesses, such as heart attacks and strokes. When high blood pressure is controlled, the risk of kidney disease is decreased.

Glomerulonephritis (glo-mer-yoolow- nef-rite-iss) is a disease that causes inflammation of the kidney's tiny filtering units the glomeruli. Glomerulonephritis may happen suddenly, for example after a bout of strep throat, and the individual may get well again. However, the disease can also develop slowly over several years and it may cause loss of kidney function.

Polycystic kidney disease is the most common inherited kidney disease. It is characterized by the formation of cysts in the kidneys. These cysts enlarge over time and can seriously damage the kidneys or even cause kidney failure.

Kidney stones are a common problem. Having kidney stones may or may not lead to long-term kidney problems. Stones result from a build-up of extra wastes in the blood. The most common wastes are oxalate and uric acid. Sometimes extra fluid, diet, and medications can help prevent stones from forming. Kidney stones may cause severe pain in your back and side. Stones are sometimes too large to pass out of your body in urine. In these cases, the stones can be removed surgically or broken down into smaller pieces that can pass out of the body in urine.

Urinary tract infections (UTIs) happen when germs enter the urinary tract and multiply. Symptoms include feeling an increased need to urinate, pain and/ or burning during urination, cloudy or blood-stained urine, and a strong odor to the urine. These infections happen most often in the bladder, but they sometimes spread upwards to the kidneys. This causes fever and back pain. Kidney infections are serious and must be treated right away to avoid scarring kidney tissue.

Congenital diseases, ones that people are born with, may also affect the kidneys. These diseases usually begin with a problem that happens in a baby's urinary tract when it is growing in the womb. One of the most common congenital diseases happens when a valve in the bladder fails to work and allows urine to back up to the kidneys, causing infections and possible kidney damage over time.

Drugs and toxins can also cause kidney problems. Using large amounts of over-the-counter pain relievers (non-steroidal anti-inflammatory drugs (NSAIDs)) for a long time can be harmful to the kidneys. Certain other damage (Foundation, 2015).

3-4. Kidney Failure

In kidney failure, the kidneys lose their ability to filter enough waste products from the blood and to regulate the body's balance of salt and water. Eventually, the kidneys slow their production of urine, or stop producing it completely. Waste products and water accumulate in the body. This can lead to potentially life-threatening complications. Excess fluid can accumulate in the lungs and extreme changes in blood chemistry can affect the function of the heart and brain. General categories of kidney failure (also called renal failure) (Publications, 2017).

3-4-1.Acute renal failure

Kidney function stops or is abruptly reduced because of a sudden illness, a medication, a toxin or a medical condition that causes one of the following:

- which can occur during major surgery, severe burns with fluid loss through burned skin, massive bleeding (hemorrhage) or a heart attack that severely affects heart function
- Direct damage to kidney cells or to the kidneys' filtering units, which can be caused by inflammation in the kidneys, toxic chemicals, medications, contrast dye used for computed tomography (CT) scans and certain procedures (such as angiograms) that are guided by x-ray, and infections.
- Blocked urine flow from the kidney, which can occur because of obstructions outside the kidney, such as kidney stones, bladder tumors or an enlarged prostate (Publications, 2017).
-

3-4-2.Chronic kidney disease (chronic renal failure)

Chronic kidney disease (CKD) means your kidneys are damaged and can't filter blood the way they should. The disease is called "chronic" because the damage to your kidneys happens slowly over a long period of time. This damage can cause wastes to build up in your body. CKD can also cause other health problems (Cente, 2016).

Chronic kidney disease (CKD) is classified into five stages (stages 1–5) according to the Glomerular filtration rate (GFR).which can easily be estimated (eGFR), from measurement of the blood creatinine level, and taking into account, age, ethnicity and gender (Day, 2017). GFR is a test that measures the glomerular filtration rate. It compares the levels of waste products in the patient's blood and urine. GFR measures how many milliliters of waste the kidneys can filter per minute. The kidneys of healthy individuals can typically filter over 90 ml per minute. Glomerular filtration rate (GFR) is classified kidney disease five stages:

Stage 1 - GFR rate is normal. However, evidence of kidney disease has been detected.

Stage 2 - GFR rate is lower than 90 milliliters, and evidence of kidney disease has been detected.

Stage 3 - GFR rate is lower than 60 milliliters, regardless of whether evidence of kidney disease has been detected.

Stage 4 - GRF rate is lower than 30 milliliters, regardless of whether evidence of kidney disease has been detected.

Stage 5 - GFR rate is lower than 15 milliliters. Renal failure has occurred

3-4-3. End-stage renal disease

End-stage renal disease (ESRD) is the last stage of chronic kidney disease and is characterized by permanent irreversible kidney failure. Patients with ESRD include those who are treated with dialysis is a process that removes wastes and fluid from the body—and those who have a functioning kidney transplant. Because of the limited number of kidneys available for transplantation and variation in patients' suitability for transplantation, 70 percent of ESRD patients undergo maintenance dialysis (services, 2015)

End-stage renal disease (ESRD) which corresponds to GFR of 15 mL , initiation of maintenance dialysis or receipt of preemptive renal transplantation is classified as CKD stage 5.This also is called end stage renal failure .it's occurs when kidney function has deteriorated to the point that if dialysis treatments do not begin, the person will die.

3-4-3-1.Causes of end stage renal disease

The main causes of end stage renal disease are diabetes mellitus, renal vascular diseases, glomerulonephritis and hypertension (American, 2002). The risk of chronic kidney disease increases with ageing, but also lifestyle factors may play a role in the development of chronic kidney disease. It is known that obesity leads to chronic kidney disease through diabetes mellitus and hypertension, but emerging evidence indicates that obesity may also contribute directly to kidney damage through a cascade of additional hemodynamic, metabolic, and inflammatory mechanisms as well as by mechanical compression. (Iseki, et al., 2004).In addition, there is evidence that smoking may be a risk factor for chronic kidney disease. Furthermore, the prevalence of chronic kidney disease is 1.5 times increased in men compared with women, suggesting a sex difference in susceptibility (Vupputuri S, 2003).

3-4-3-2 .Symptoms chronic kidney disease and end-stage renal failure

Because the kidney damage in chronic renal failure occurs slowly over a long time, symptoms develop slowly, usually beginning when more than 80% of kidney function is lost. When this occurs, symptoms can include:

Headache, Fatigue, Weakness, Lethargy, Itching, Poor appetite, Vomiting, Increased thirst, Pale skin, High blood pressure, Slowing of growth in children, Bone damage in adults (publications, 2017)

3-5.Chronic kidney disease and end-stage renal failure diagnosis

The diagnosis of CKD starts with a medical history. A family history of kidney failure, high blood pressure, or diabetes may alert your doctor. However, other tests are necessary to confirm that you have CKD, such as:

Complete blood count

A complete blood count can show anemia. Your kidneys make erythropoietin, which is a hormone. This hormone stimulates your bone marrow to make red blood cells. When your kidneys are severely damaged, your ability to make erythropoietin decreases. This causes a decline in red blood cells, or anemia

Electrolyte level test

CKD can affect your electrolyte levels. Potassium may be high and bicarbonate levels may be low if you have CKD. There may also be an increase of acid in the blood (Elizabeth & Verneda Lights, 2017).

Blood test - a blood test may be ordered to determine whether waste substances are being adequately filtered out. If levels of urea and creatinine are persistently high, the doctor will most likely diagnose end stage kidney disease.

Urine test - a urine test helps find out whether there is either blood or protein in the urine.

Kidney scans - kidney scans may include a magnetic resonance imaging scan, computed tomography (CT) scan or an ultrasound scan. The aim is to determine whether there are any blockages in the urine flow. These scans can also reveal the size and shape of the kidneys - in advanced stages of kidney disease the kidneys are smaller and have an uneven shape.

Kidney biopsy - a small sample of kidney tissue is extracted and examined for cell damage. An analysis of kidney tissue makes it easier to make a precise diagnosis of kidney disease**.**

Chest X-ray - the aim here is to check for pulmonary edema (fluid retained in the lungs).

3-6. Risk factors

Factors that may increase your risk of chronic kidney disease include:

- Diabetes
- High blood pressure
- Heart and blood vessel (cardiovascular) disease
- Smoking
- Obesity
- Being African-American, Native American or Asian-American
- Family history of kidney disease
- Abnormal kidney structure
- Older age.(staff, 2017)

3-7.Treatment for end stage kidney disease

If your kidneys can't keep up with waste and fluid clearance on their own and you develop complete or near complete kidney failure, you have end stage kidney disease. At that point, you need dialysis or a kidney transplant.

3-7-1. Dialysis

Dialysis is a procedure that is performed routinely on persons who suffer from acute or chronic renal failure, or who have ESRD. The process involves removing waste substances and fluid from the blood that are normally eliminated by the kidneys. Dialysis may also be used for individuals who have been exposed to or ingested toxic substances to prevent renal failure from occurring. There are two types of dialysis that may be performed, including the following:

3-7-1-1.Peritoneal dialysis

Peritoneal dialysis is performed by surgically placing a special, soft, hollow tube into the lower abdomen near the navel. After the tube is placed, a special solution called dialysate is instilled into the peritoneal cavity. The peritoneal cavity is the space in the abdomen that houses the organs and is lined by two special membrane layers called the peritoneum. The dialysate is left in the abdomen for a designated period of time which will be determined by your doctor. The dialysate fluid absorbs the waste products and toxins through the peritoneum. The fluid is then drained from the abdomen, measured, and discarded. There are three different types of peritoneal dialysis: continuous ambulatory peritoneal dialysis (CAPD), continuous cyclic peritoneal dialysis (CCPD), and intermittent peritoneal dialysis (IPD). CAPD does not require a machine. Exchanges often referred to as passes, can be done three to five times a day during waking hours. CCPD requires the use of a special dialysis machine that can be used in the home. This type of dialysis is done automatically,

even while you are asleep. IPD uses the same type of machine as CCPD, but treatments take longer. IPD can be done at home, but usually is done in the hospital. Possible complications of peritoneal dialysis include an infection of the peritoneum, or peritonitis, where the catheter enters the body. Peritonitis causes fever and stomach pain. Your diet for peritoneal dialysis will be planned with a dietitian, who can help you choose meals according to your doctor's orders (Library, 2017).

3-7-1-2 Hemodialysis

Hemodialysis can be performed at home or in a dialysis center or hospital by trained health care professionals. A special type of access, called an arteriovenous (AV) fistula, is placed surgically, usually in your arm. This involves joining an artery and a vein together. An external, central, intravenous (IV) catheter may also be inserted, but is less common for long- term dialysis. After access has been established, you will be connected to a large hemodialysis machine that drains the blood, bathes it in a special dialysate solution which removes waste substances and fluid, then returns it to your bloodstream. Hemodialysis is usually performed several times a week and lasts for four to five hours. Because of the length of time hemodialysis takes, it may be helpful to bring reading material, in order to pass the time during this procedure. During treatment you can read, write, sleep, talk, or watch TV. At home, hemodialysis is done with the help of a partner, often a family member or friend. If you choose to do home hemodialysis, you and your partner will receive special training possible complications of hemodialysis include muscle cramps and hypotension (sudden drop in blood pressure). Hypotension may cause you to feel dizzy or weak, or sick to your stomach. Side effects are avoided by following the proper diet and taking medications, as prescribed by your doctor. A dietitian will work with you to plan your meals, according to your doctor's orders (Library, 2017).

3-7-2. Kidney transplant

A kidney transplant is a better option than dialysis for patients who have no other conditions apart from kidney failure. Even so, candidates for kidney transplant will have to undergo dialysis until they receive a new kidney.

The kidney donor and recipient should have the same blood type, cell-surface proteins and antibodies, in order to minimize the risk of rejection of the new kidney. Siblings or very close relatives are usually the best types of donors. If a living donor is not possible, the search will begin for a cadaver donor (dead person) (C. Poinier, et al., 2015).

3-8. Global summary of End-stage renal disease

End-stage renal disease (ESRD) is a major public health problem worldwide and is associated with considerable morbidity and mortality (Halle, et al., 2015).According to the 2010 Global Burden of Disease study chronic kidney disease was ranked 27th in the list of causes of total number of global deaths in 1990 (age- standardised annual death rate of 15・7 per 100 000), but rose to 18th in 2010 (annual death rate 16・3 per 100 000) (Jha, et al., 2013).Chronic kidney disease is at least 34 times more frequent in Africa than in developed countries.

Hypertension affects approximately 25% of the adult population and is the cause of chronic kidney failure in 21% of patients on renal replacement therapy in the South African Registry. The prevalence of diabetic nephropathy is estimated to be 14%16% in South Africa, 23.8% in Zambia, 12.4% in Egypt, 9% in Sudan, and 6.1% in Ethiopia. The current dialysis treatment rate ranges from 70 per million populations (pmp) in South Africa to < 20 pmp in the most of sub Saharan Africa. (Naicker, 2009).

End stage renal failure (ESRF) has become a major health problem in Sub Saharan Africa (SSA). There are limited data on the prevalence and incidence of ESRF in SSA due to lack of renal registries. Several studies pointed out to the magnitude of

the problem in SSA. In Nigeria a study reported an increase of hospital admissions because of ESRF from 6 to 16 % between the years 1989 and 2003 (FA, et al., 2011). In Senegal only 8.23 % of ESRF patients receive renal replacement therapy (RRT) (Diopt T, et al., 2003). In Ghana, a study pointed out that 5 % of total hospital admissions had renal disease of whom 27.1 % died, usually of ESRF (Plange-Rhule, et al., 1999). Hypertension is a leading cause of ESRF in Senegal and Ghana (Diouf B, et al., 2000; Matekole M, 1993) .

3-9. Kidney failure in Sudan

Very limited data are available about the causes of renal diseases leading to chronic renal diseases in all states of Sudan (Elsharif & Elsharif, 2011). However, hypertension and diabetes mellitus are the most commonly reported cause of kidney failure (Sarra Elamin ،2010 ؛Elsharif & Elsharif, 2011). Intermittent peritoneal dialysis (IPD) was introduced in Sudan in 1968. In the same year, a personal set-up for home hemodialysis (HD) marked the start of HD in Sudan. The first renal unit was opened in Khartoum Teaching Hospital in 1970, where the first kidney transplant in Sudan and the second kidney transplant in the Middle East took place in 1974 (Suliman SM, 1995). Although it's early dialysis in Sudan, but now there are some problems of renal failure in Sudan.

3-10. Problems of renal failure in Sudan

There are no accurate statistics on the causes of the outbreak of the disease clearly in Sudan .But the important reasons come at the forefront of pressure and diabetes, especially among the elderly. And glomerulonephritis in both young and middle age, as well as diseases common in Sudan such as kidney stones and bacterial infections of kidneys.

There is a complete absence of such a record, making it difficult to real, but through simple studies in some hospitals to find out the real reasons for the spread of such phenomenon in Sudan.

The absence of a regular follow-up system in the Sudan, many patients arrive in the late stages of kidney weakness and are waiting to dialysis. Simple medical interventions, such as urine testing, kidney functions, and regular sugar and pressure control, may enrich the human being from reaching this stage that needs dialysis.

The suffering of patients undergoing dialysis in the Sudan is mostly of many difficulties linked to financial support despite free Hemodialysis in Sudan, however, the patient needs a large number of persistent drugs, the cost of which has recently risen, forcing some patients to dispense with each other

In addition, some important drugs have been cut off from the market as a result of the non-importation of companies with regard to economic conditions and the instability of the exchange rate.

The state has recently opened some dialysis centers in the States of the Sudan, but the vast majority of dialysis centers are stationed in Khartoum, which creates considerable pressure on it as we note a large number of state patients who have to relocate their residency to Khartoum to continue dialysis. The opening of other state centers could greatly help to reduce pressure on the services of the centers in Khartoum. Hemodialysis treatment in Sudan is free and paid by the government (banaga, 2012).

There are limited data on the prevalence and causes of ESRF in Sudan apart from few studies. (Banaga, et al., 2015).

Chapter Four
Data Application and Analysis

4-0. Preface:

In this chapter we will use statistical methods mentioned in chapter three on the research data to obtain the required results of the study. The Microsoft Excel software 2013, SPSS 24 and STATA 2014 packages were used to analysis the study data.

4-1. Data research:

Data were collected from records in the governmental hospitals in Khartoum state for hemodialysis Patients, in the period from 2005 until December 2010 taking all available data for all patients in this period and patients were followed up until December 2015 .The study variables included age(date of diagnosis of the disease) , survival status even death or the date of last follow-up per months, Sex , Marital Status, Education status, occupation, Address, regular , Dialysis frequency per week, Hospitals , Diabetes Mellitus (yes or no), Hypertension (yes or no), Hypertension and Diabetes Mellitus (yes or no),polycystic kidney disease(yes or no), Renal obstructions (yes or no), Shrunken kidney(yes or no) Uncertain(yes or no), Other (yes or no).

4-2. Descriptive analysis of the variables of the study:

The descriptive statistical analysis, percentage, and frequency were measured using Microsoft Excel software and SPSS.

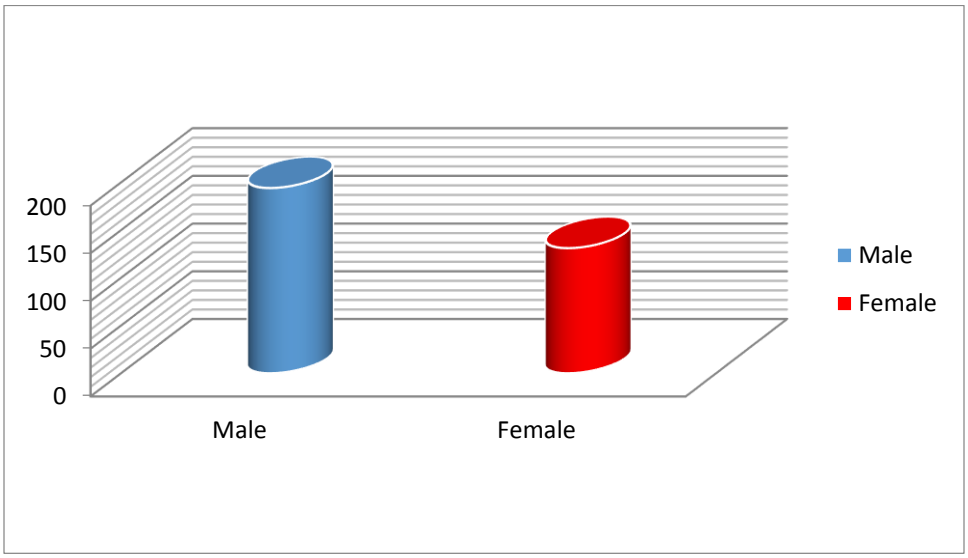
4-2-1.Qaltitative variables:

Table (4-1). Sex:

Sex	Frequency	Percentage%
Male	194	59.7%
Female	131	40.3%
Total	325	100.0%

Source: prepared by the researcher by using SPSS

Figure (4-1) .Bar chart of sex



Source: prepared by the researcher by using Excel

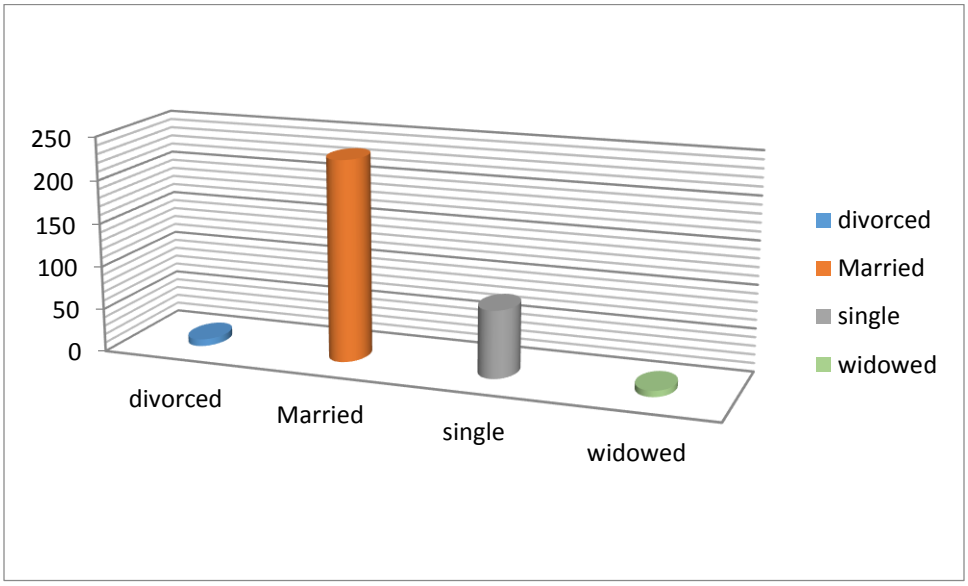
Seen from the table (4-1) and figure (4-1) that 194 patients by 59.7% were male, followed by female 131 patients by 40.3%. Note that the frequency of male hemodialysis patients more than female.

Table (4-2). Marital status:

Marital status	Frequency	Percentage%
Divorced	8	2.5%
Married	232	71.4%
Single	78	24%
Widowed	7	2.2%
Total	325	100%

Source: prepared by the researcher by using SPSS

Figure (4-2). Bar chart of marital status



Source: prepared by the researcher by using Excel

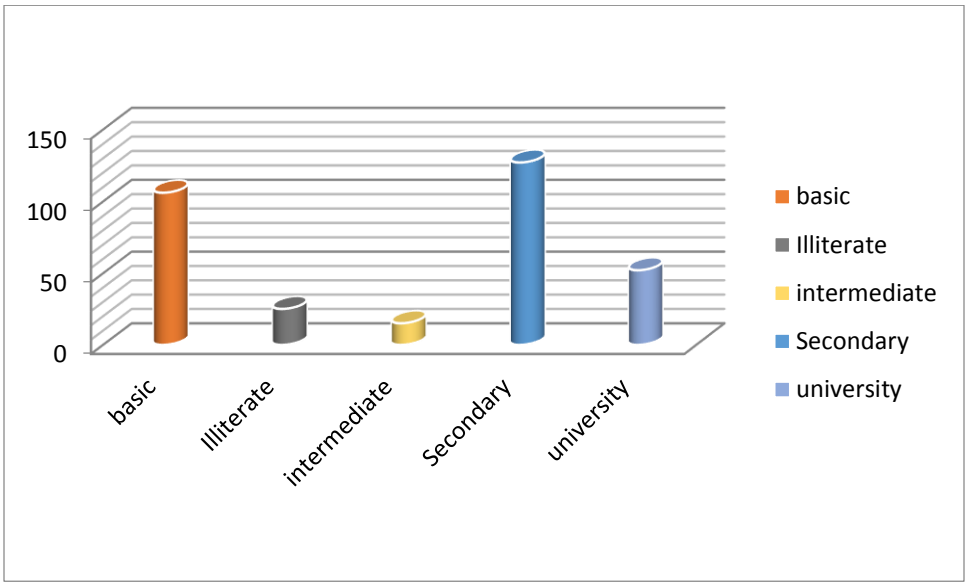
Seen from the table (4-2) and figure (4-2) that 8 divorced patients by 2.5%, followed by 232 patients by 71.4% were married, followed by single 78 patients by 24%, followed by widowed 7 patients by 2.2%. We note that the frequency of married patients is the largest among the other recurrences, in addition to the fact that most of them contracted the disease after their marriage

Table (4-3) .Education status:

Education status	Frequency	Percentage%
basic	106	32.6%
Illiterate	25	7.7%
intermediate	15	4.6%
Secondary	127	39.1%
university	52	16%
Total	325	100%

Source: prepared by the researcher by using SPSS

Figure (4-3) .Bar chart of education status



Source: prepared by the researcher by using Excel

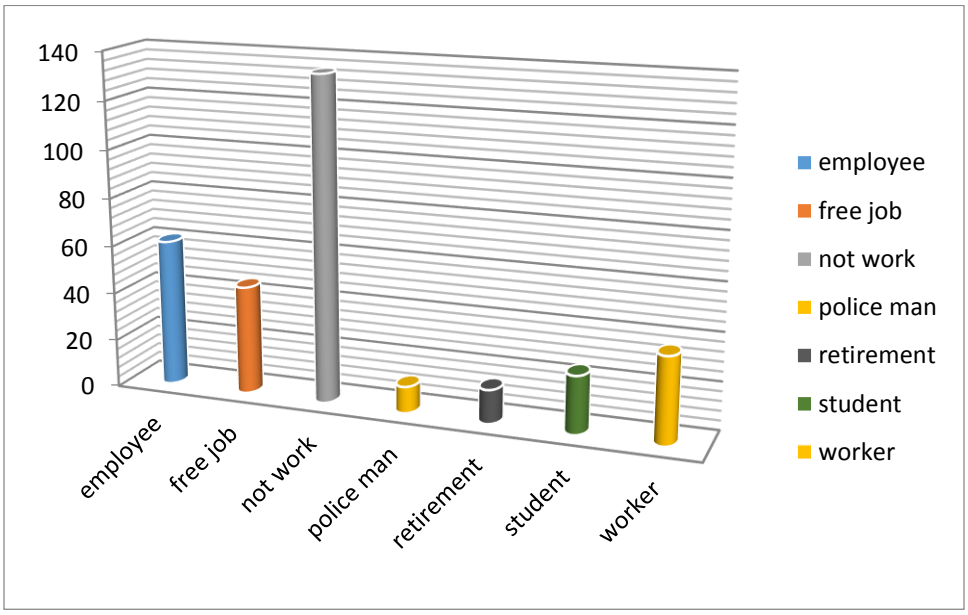
Seen from the table (4-3) and figure (4-3) that 106 patients by 32.6% were basic, followed by Illiterate 25 patients by 7.7%, followed by intermediate 15 patients by 4.6%, followed by Secondary 127 patients by 39.1%, followed by university 52 patients by 16%. We note that the frequency of Secondary patients is the largest among the other recurrences.

Table (4-4) .Occupation:

Occupation	Frequency	Percentage%
Employee	61	18.8%
Free job	45	13.8%
Not working	134	41.2%
Police man	11	3.7%
Retirement	14	4.3%
Student	24	7.4%
Worker	36	11.1.8%
Total	325	100%

Source: prepared by the researcher by using SPSS

Figure (4-4) .Bar chart of occupation



Source: prepared by the researcher by using Excel

Seen from the table (4-4) and figure (4-4) that 61 patients by. 18.8% were employee, followed by free job 45 patients by 13.8%, followed by not work 134 patients by 41.2%, followed by police man 11 patients by 3.7%, followed by retirement 14 patients by 4.3%, followed by student 24 patients by 7.4%, followed

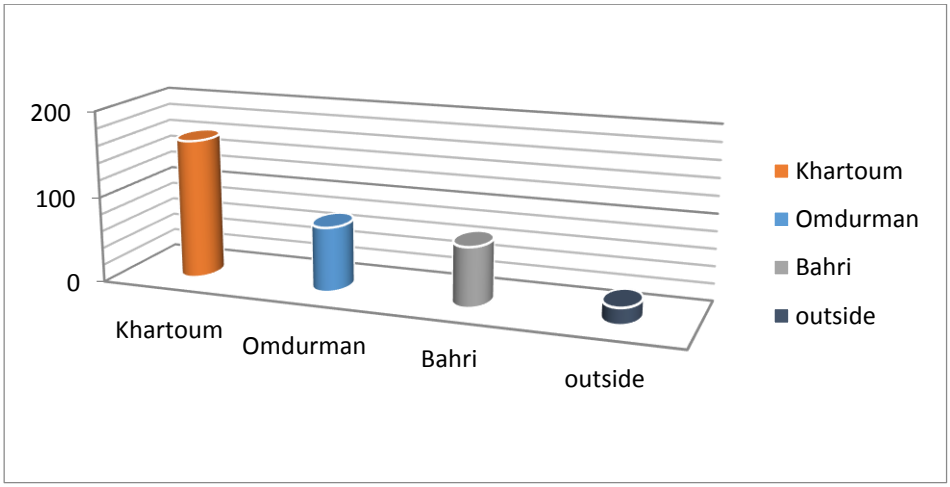
by worker 36 patients by 11.8.we noted that the number of the not work is large compared to others, and this indicates their inability to work and they suffer from kidney failure.

Table (4-5) .Address:

Address	Frequency	Percentage%
Khartoum	162	49.8%
Omdurman	75	23.1%
Bahri	69	21.2%
outside Khartoum state	19	5.8%
Total	325	100%

Source: prepared by the researcher by using SPSS

Figure (4-5) .Bar chart of address



Source: prepared by the researcher by using Excel.

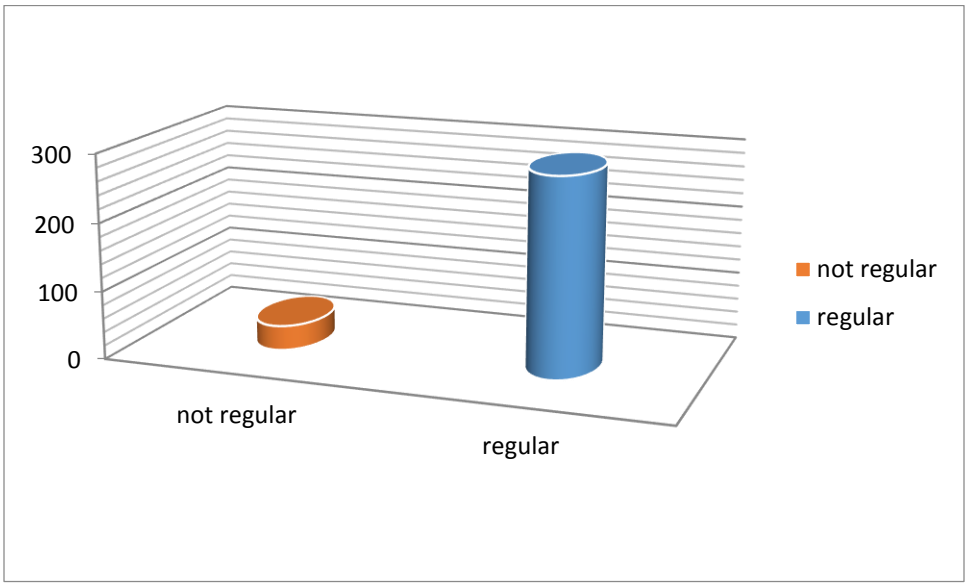
Seen from the table (4-5) and figure (4-5) that 162 patients by 49.8% were in Khartoum, followed by Omdurman 75 patients by 23.1%, followed by Bahri 69 patients by 21.2%, followed by outside 19 patients by 5.3.

Table (4-6). Regular

daily dialysis	Frequency	Percentage%
Not regular	36	11.1%
Regular	289	88.9%
Total	325	100%

Source: prepared by the researcher using by SPSS

Figure (4-6) .Bar chart of regular:



Source: prepared by the researcher by using Excel

Seen from the table (4-6) and figure (4-6) that 36 patients by 11.1% were not regular, followed by regular 289 patients by 88.9%,

Table (4-7). Dialysis frequency per week:

Source: prepared by the researcher by using SPSS

Table (4-8).Status * Dialysis frequency per week crosstabulation

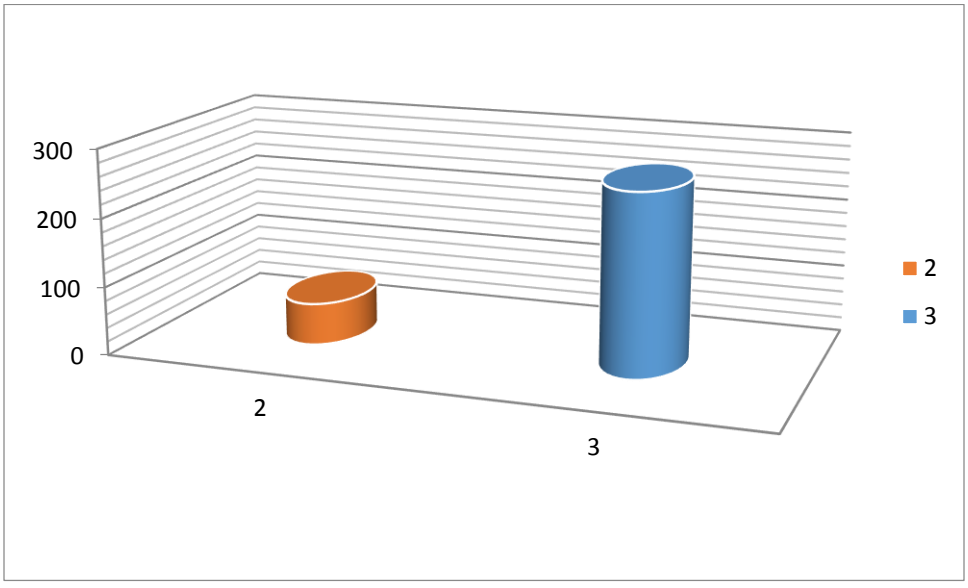
Dialysis frequency Per (wk)	Frequency	Percentage%
2 times	61	18.8%
3 times	264	81.2%
Total	325	100%

Source: prepared by the researcher by using SPSS

Count		Dialysis frequency		Total
		2 times	3 times	
status	Death	55	115	170
	Alive	6	149	155
Total		61	264	325

Source: prepared by the researcher by using SPSS

Figure (4-7) .Bar chart of Dialysis frequency per week



Source: prepared by the researcher by using Excel

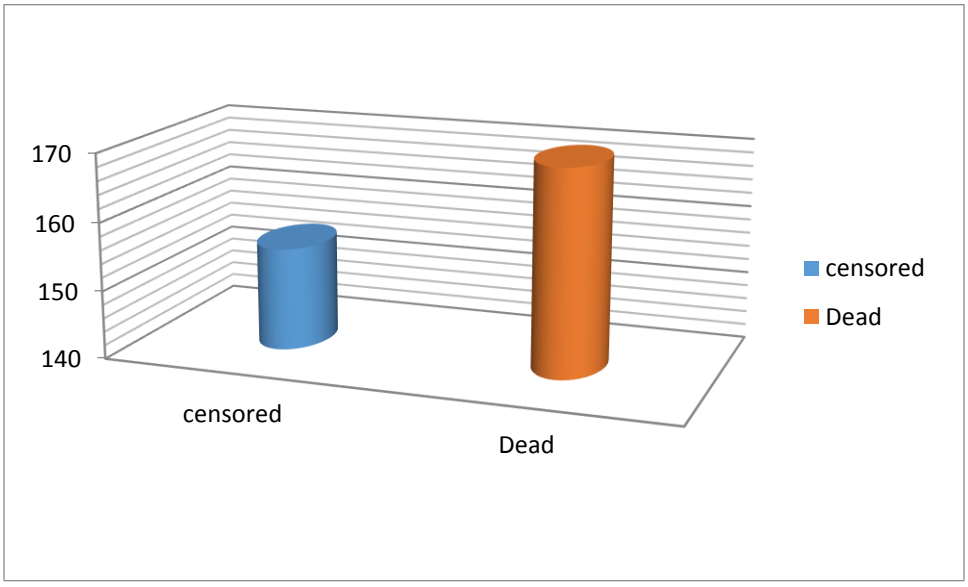
Seen from the table (4-7), (4-8) and figure (4-7) that 61 patients were (2) frequency dialysis in week their Percentage is 18.8% of whom 55 died, and the survivors are 6. They are followed by 264 patients with (3) frequency dialysis in week, their Percentage is 81.2%. 115 of them died and 149 survived.

Table (4-9).Survival status

survival status	Frequency	Percentage%
censored	155	47.7%
Dead	170	52.3%
Total	325	100%

Source: prepared by the researcher by using SPSS

Figure (4-8) .Bar chart of survival status:



Source: prepared by the researcher by using Excel

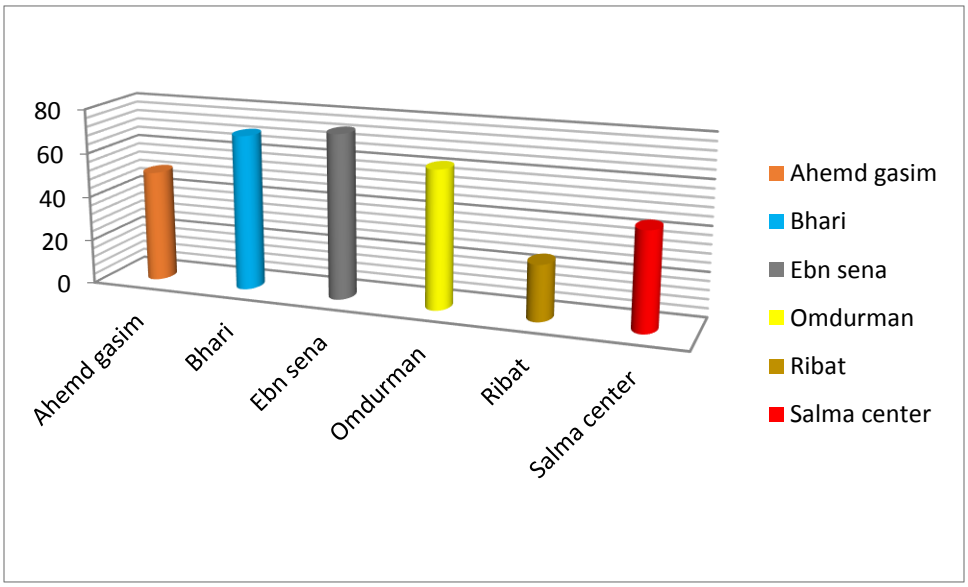
Seen from the table (4-8) and figure (4-8) that 155 patients by 47.7% were alive or censored, followed by dead 170 patients by 52.3%.

Table (4-10). Hospital

Hospital	Frequency	Percentage%
Ahemd Gasim	50	15.4%
Bhari	70	21.5%
Ebn Sena	74	22.8%
Omdurman	62	19.1%
Ribat	25	7.7%
Salma Center	44	13.5%
Total	325	100%

Source: prepared by the researcher by using SPSS

Figure (4-9). Bar chart of hospital:



Source: prepared by the researcher by using Excel

Seen from the table (4-9) and figure (4-9) that 50 patients by 15.4% were in Ahemd gasim, followed by Bahri 70 patients by 21.5%, followed by ebn sena 74 patients by 22.8%, followed by Omdurman 62 patients by 19.1%, followed by ribat 25 patients by 7.7%, followed by Selma center 44 patients by 13.5%.

Table (4-11). Diabetes mellitus:

diabetes mellitus	Frequency	Percentage%
No	236	72.6
Yes	89	27.4
Total	325	100.0

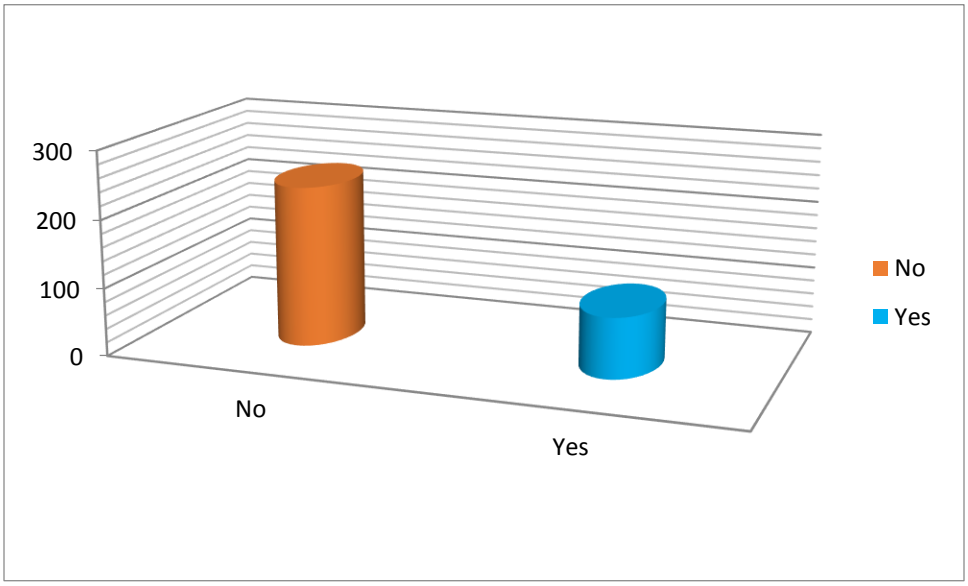
Source: prepared by the researcher by using SPSS

Table (4-12).Status * diabetes mellitus crosstabulation

		Diabetes mellitus		Total
		No	Yes	
status	Death	109	61	170
	Alive	127	28	155
Total		236	89	325

Source: prepared by the researcher by using SPSS

Figure (4-10).Bar chart of diabetes mellitus:



Source: prepared by the researcher by using Excel

Seen from the table (4-11), (4-12) and figure (4-10) that 236 patients were non-diabetes mellitus their Percentage is 72.6%, of whom 109 died, and the survivors are 127. They are followed by 89 patients with diabetes, their Percentage is 27.4%. 61 of them died and 28 survived.

Table (4-13). Hypertension:

Hypertension	Frequency	Percentage%
No	229	70.5%
Yes	96	29.5%
Total	325	100.0%

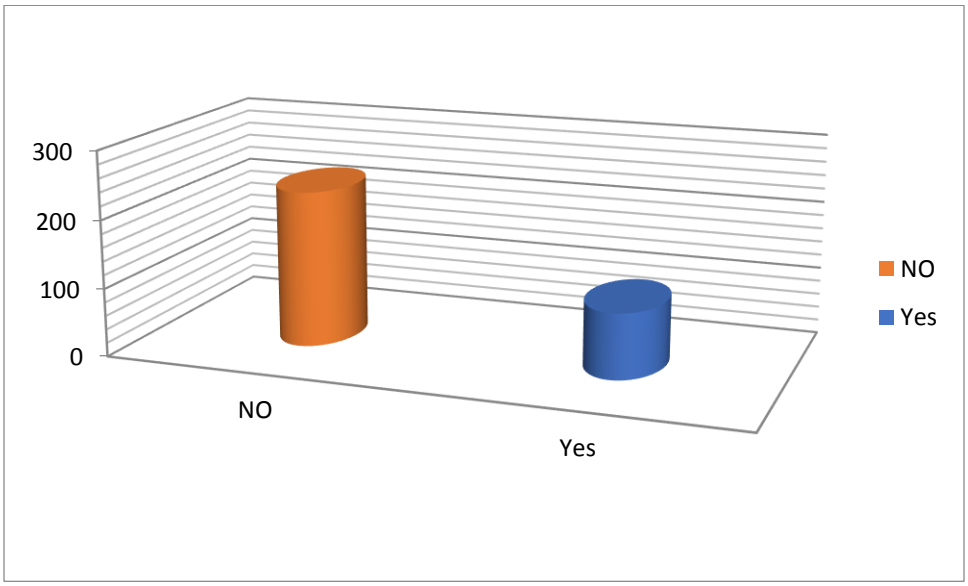
Source: prepared by the researcher by using SPSS

Table (4-14).Status * hypertension crosstabulation

		Hypertension		Total
		No	Yes	
status	Death	134	36	170
	Alive	95	60	155
Total		229	96	325

Source: prepared by the researcher by using SPSS

Figure (4-11). Bar chart of hypertension:



Source: prepared by the researcher by using Excel

Seen from the table (4-13), (4-14) and figure (4-11) that 229 patients were non hypertension their Percentage is 70.5%, of whom 134 died, and the survivors are 95 .They are followed by 96 patients with hypertension, their Percentage is 29.5%. 36 of them died and 60 survived.

Table (4-15) .Polycystic kidney disease:

Polycystic kidney disease	Frequency	Percentage%
No	308	94.8%
Yes	17	5.2%
Total	325	100.0%

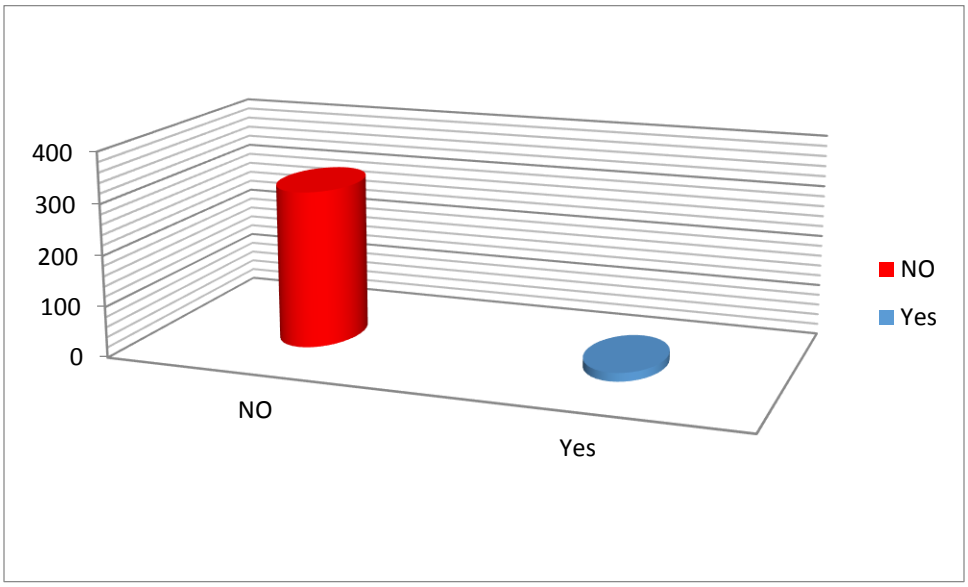
Source: prepared by the researcher by using SPSS

Table (4-16) .Status * polycystic kidney disease crosstabulation

		Polycystic kidney disease		Total
		No	Yes	
status	Death	158	12	170
	Alive	150	5	155
Total		308	17	325

Source: prepared by the researcher by using SPSS

Figure (4-12) .Bar chart of polycystic kidney disease:



Source: prepared by the researcher by using Excel

Seen from the table (4-15) , (4-16) and figure (4-12) that 308 patients were non - polycystic kidney disease their Percentage is 94.8%, of whom 158 died, and the survivors are 150.They are followed by 17 patients with polycystic kidney disease, their Percentage is 5.2%. 12 of them died and 5 survived.

Table (4-17). Diabetes mellitus and hypertension:

Diabetes mellitus and hypertension	Frequency	Percentage%
No	292	89.8%
Yes	33	10.2%
Total	325	100.0%

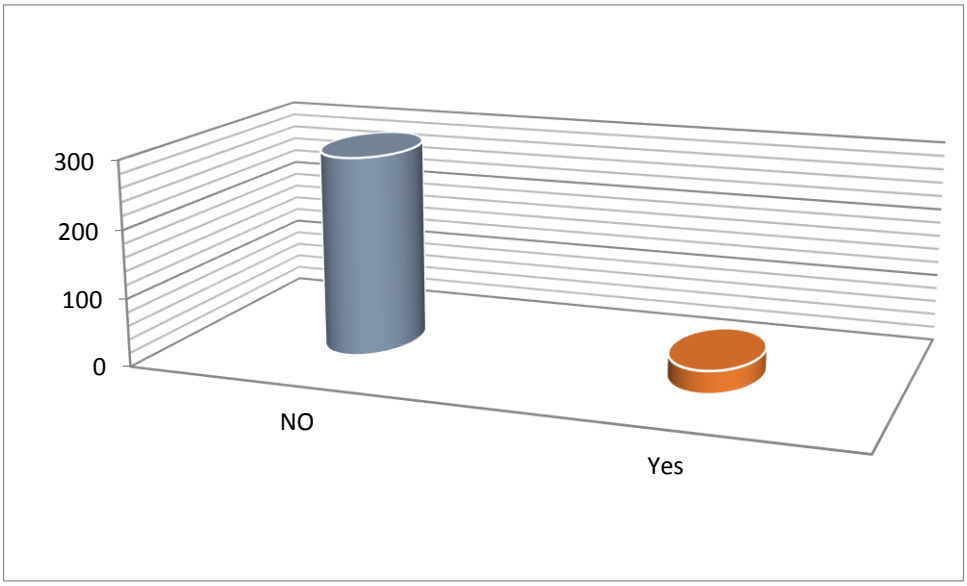
Source: prepared by the researcher by using SPSS

Table (4-18) .Status * diabetes mellitus and hypertension crosstabulation

		Diabetes mellitus and hypertension		Total
		No	Yes	
status	Death	143	27	170
	Alive	149	6	155
Total		292	33	325

Source: prepared by the researcher by using SPSS

Figure (4-13). Bar chart of diabetes mellitus and hypertension:



Source: prepared by the researcher by using Excel

Seen from the table (4-17), (4-18) and figure (4-13) that 292 patients were non - Diabetes mellitus and hypertension their Percentage is 89.8%, of whom 143 died, and the survivors are 149. They are followed by 33 patients with Diabetes mellitus and hypertension, their Percentage is 10.2%. 27 of them died and 6 survived.

Table (4-19). Renal obstructions:

Renal obstructions	Frequency	Percentage%
N0	298	91.7 %
Yes	27	8.3%
Total	325	100.0%

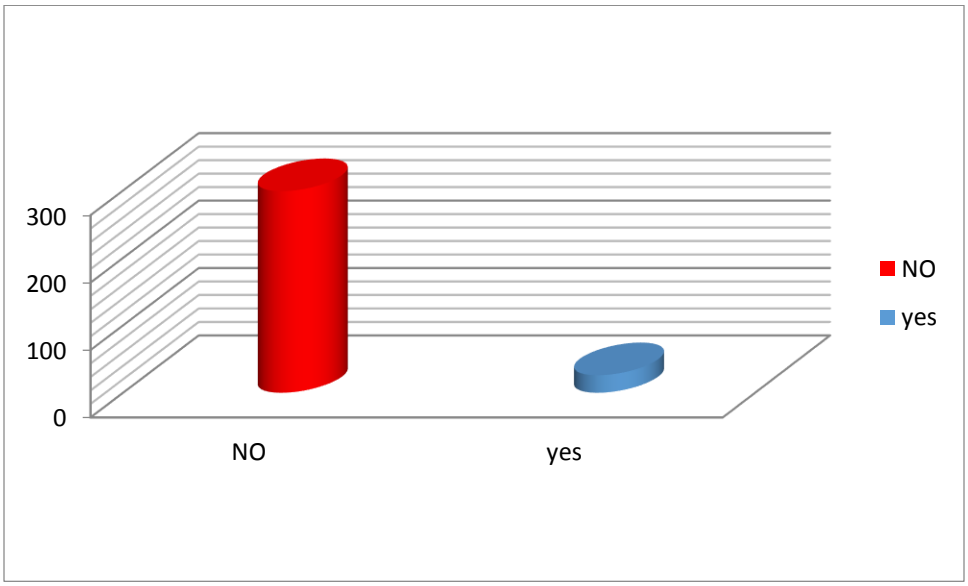
Source: prepared by the researcher by using SPSS

Table (4-20) .Status * renal obstructions crosstabulation

		Renal obstructions		
		No	Yes	Total
status	Death	156	14	170
	Alive	142	13	155
Total		298	27	325

Source: prepared by the researcher by using SPSS

Figure (4-14) bar chart of renal obstructions:



Source: prepared by the researcher by using Excel

Seen from the table (4-19) , (4-20) and figure (4-14) that 298 patients were non - renal obstructions their Percentage is 91.7%, of whom 156 died, and the survivors are 142. They are followed by 27 patients with renal obstructions, their Percentage is 8.3%. 14 of them died and 13 survived.

Table (4-21) .Shrunken kidneys:

shrunken kidney	Frequency	Percentage%
No	314	96.6%
Yes	11	3.4%
Total	325	100.0%

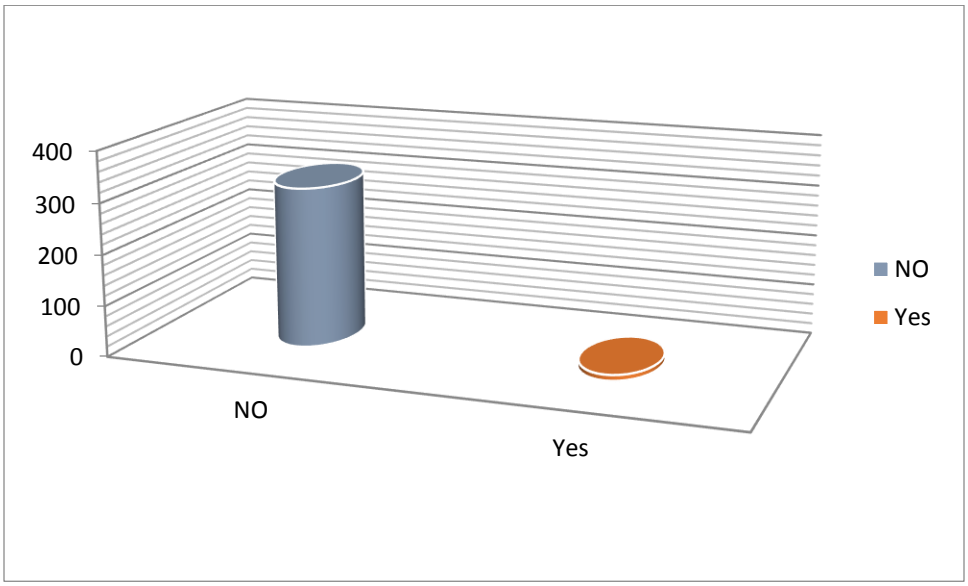
Source: prepared by the researcher by using SPSS

Table (4-22) .Status * shrunken kidneys crosstabulation

		Shrunken kidneys		
		No	Yes	Total
status	Death	169	1	170
	Alive	145	10	155
Total		314	11	325

Source: prepared by the researcher by using SPSS

Figure (4-15) bar chart of shrunken kidneys:



Source: prepared by the researcher by using Excel

Seen from the table (4-21), (4-22) and figure (4-15) that 314 patients were non - shrunken kidney their Percentage is 96.6%, of whom 169 died, and the survivors are 145. They are followed by 11 patients with shrunken kidney, their Percentage is 3.4%. 1 of them died and 11 survived.

Table (4-23) .Uncertain:

uncertain	Frequency	Percentage%
No	300	92.3%
Yes	25	7.7%
Total	325	100.0%

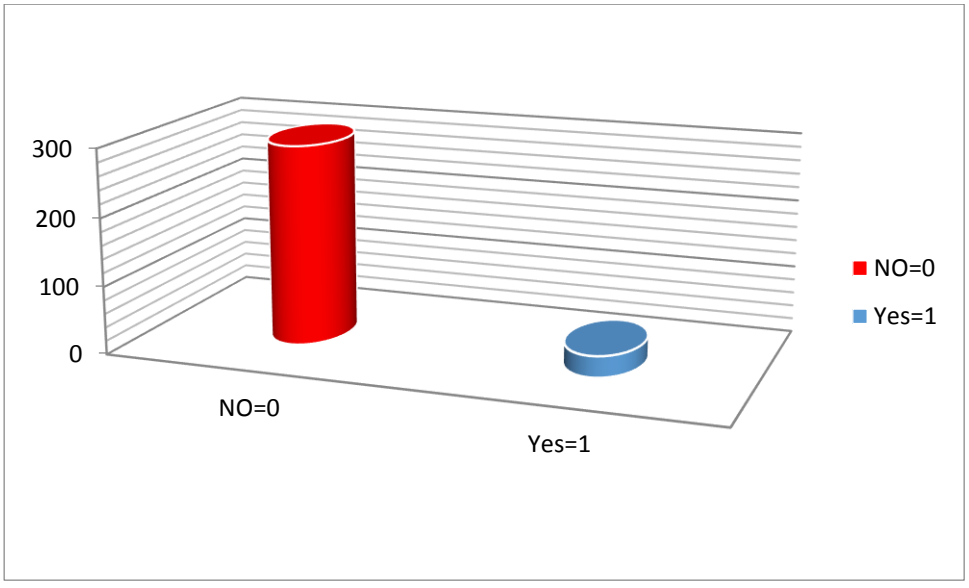
Source: prepared by the researcher by using SPSS

Table (4-24) .Status * uncertain crosstabulation

		Uncertain		Total
		No	Yes	
status	Death	160	10	170
	Alive	140	15	155
Total		300	25	325

Source: prepared by the researcher by using SPSS

Figure (4-16) bar chart Uncertain



Source: prepared by the researcher by using Excel

Seen from the table (4-23) , (4-24) and figure (4-16) that 300 patients were non - uncertain their Percentage is 92.3%, of whom 160 died, and the survivors are 140. They are followed by 25 patients with uncertain, their Percentage is 7.7%. 10 of them died and 15 survived. .

Table (4-25). Others:

other	Frequency	Percentage%
No	306	94.2%
Yes	19	5.8%
Total	325	100.0%

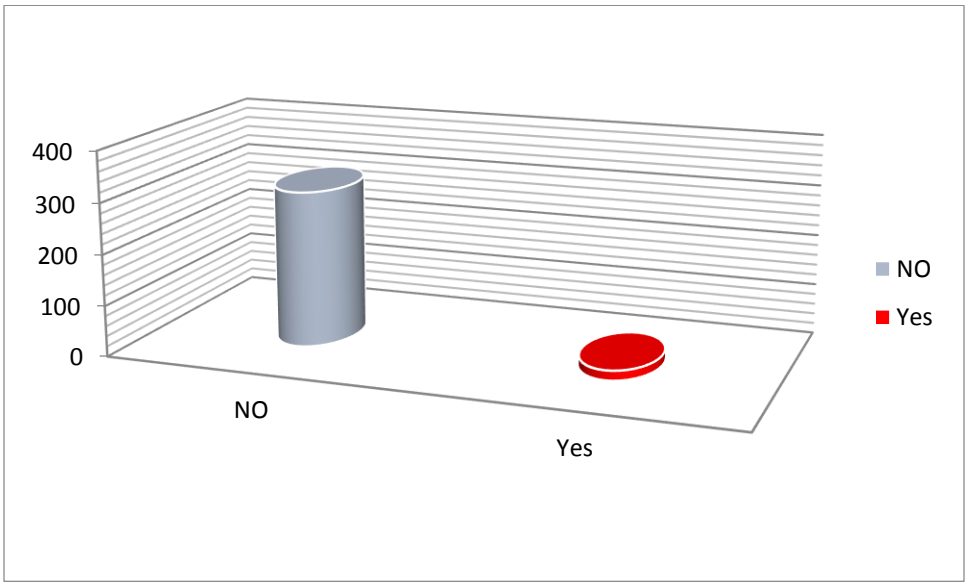
Source: prepared by the researcher by using SPSS

Table (4-26) .Status * other crosstabulation

		Others		Total
		No	Yes	
status	Death	195	5	170
	Alive	141	14	155
Total		306	19	325

Source: prepared by the researcher by using SPSS

Figure (4-17). Bar chart of other:



Source: prepared by the researcher by using Excel

Seen from the table (4-25), (4-26) and figure (4-17) that 306 patients were non – other their Percentage is 94.2%, of whom 195 died, and the survivors are 141. They are followed by 19 patients with other include (Systemic lupus erytherematosus, tropical disease (malaria), Gout, Food poisoning, cardiovascular disease, NSAID) their Percentage is 5.8%. 4 of them died and 14 survived. .

4-2-2.Quantitative variable

Table (4-27). Summarize of age

	N	Minimum	Maximum	First quartile	Median	Third quartile
age	325	6	88	46.03	45	75

Source: prepared by the researcher by using SPSS

Seen from the table (4-27) that the min age 6, the max age 88.the first quartile age 46.03, the median age 45, the third quartile 75.

Table (4-28). Summarize of urea:

	N	Minimum	Maximum	First quartile	Median	Third quartile
Urea	325	41	385	113	150	182.50

Source: prepared by the researcher by using SPSS

Seen from the table (4-28) that the min urea 41, the max 385, the first quartile urea 113, the median urea 150, the third quartile urea 182.50.

Table (4-29). Summarize of creatinine grouping:

Serum creatinine	N	Minimum	Maximum	First quartile	Median	Third quartile
	325	2	25	6	8	10.85

Source: prepared by the researcher by using SPSS

Seen from the table (4-29) that the min creatinine 2, the max creatinine 25, the first quartile creatinine 6, the median creatinine 8, the third quartile creatinine 10.85.

4-2.3.Normality distribution test

Table (4-30) Normality Test of Ahmed Gasim hospital

variable	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
urea	0. 236	50	0.200*
serum creatinine	0.199	50	0.200*
age	0.187	50	0.200*
*. This is a lower bound of the true significance.			

Table (4-31). Normality test of Bahri hospital

variable	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
urea	0.084	70	0.200*
serum creatinine	0.099	70	0.088
age	0.088	70	0.200*
*. This is a lower bound of the true significance.			

Table (4-32) .Normality test of Ebn Sena hospital

variable	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
urea	0.085	74	0.200*
serum creatinine	0.080	74	0.200*
age	0.082	74	0.200*
*. This is a lower bound of the true significance.			

Table (4-33). Normality test of Omdurman hospital

variable	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
urea	0.079	62	0.200*
serum creatinine	0.087	62	0.200*
age	0.087	62	0.200*
*. This is a lower bound of the true significance.			

Table (4-34). Normality test of Alribat hospital

variable	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
urea	0.122	25	0.200*
serum creatinine	0.128	25	0.200*
age	0.098	25	0.200*
*. This is a lower bound of the true significance.			

Table (4-35). Normality test of Selma Center

variable	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
urea	.098	44	0.200*
serum creatinine	.097	44	0.200*
ages	.101	44	0.200*
*. This is a lower bound of the true significance.			

Source: prepared by the researcher by using SPSS

The Kolmogorov-Smirnov test was used to determine whether the data follow the normal distribution or not .the results are as shown in the tables (4-30) to (4-35). Show that the probability value (sig) is greater than level of significance 0.05.thus, the distribution of data follows the normal distribution, so the parametric tests were uses.

4-3. Kaplan-Meier to estimate survival function, standard error and confidence intervals at the 5% level of significance:

Table (4-36): Kaplan-Meier to estimate survival function, standard error and confidence intervals at the 5% level of significance for patient's hemodialysis

Time	Total	event	sensors	Survivor Function	Std .Error	[95% Conf. Int.]	
						lower	upper
1	325	0	7	1.0000	.	. .	
2	318	2	2	0.9937	0.0044	0.9751	0.9984
3	314	3	1	0.9842	0.0070	0.9625	0.9934
4	310	2	0	0.9779	0.0083	0.9541	0.9894
5	308	2	0	0.9715	0.0094	0.9460	0.9851
6	306	7	1	0.9493	0.0124	0.9186	0.9686
7	298	2	0	0.9429	0.0131	0.9109	0.9637
8	296	1	0	0.9397	0.0134	0.9071	0.9611
10	295	3	0	0.9302	0.0144	0.8959	0.9535
11	292	2	1	0.9238	0.0149	0.8885	0.9483
12	289	20	10	0.8599	0.0196	0.8163	0.8938
13	259	1	1	0.8566	0.0198	0.8127	0.8909
14	257	3	1	0.8466	0.0204	0.8016	0.8821
15	253	1	0	0.8432	0.0206	0.7979	0.8791
16	252	2	2	0.8365	0.0210	0.7906	0.8732
17	248	1	2	0.8331	0.0211	0.7869	0.8702
18	245	1	0	0.8297	0.0213	0.7831	0.8672

20	244	1	2	0.8263	0.0215	0.7794	0.8642
21	241	1	0	0.8229	0.0217	0.7757	0.8611
22	240	4	3	0.8092	0.0224	0.7607	0.8489
23	233	2	0	0.8023	0.0227	0.7532	0.8426
24	231	16	13	0.7467	0.0250	0.6936	0.7920
25	202	1	3	0.7430	0.0252	0.6897	0.7886
26	198	4	2	0.7280	0.0258	0.6737	0.7748
27	192	0	2	0.7280	0.0258	0.6737	0.7748
30	190	1	1	0.7242	0.0259	0.6696	0.7713
31	188	1	0	0.7203	0.0261	0.6655	0.7677
33	187	0	2	0.7203	0.0261	0.6655	0.7677
34	185	1	0	0.7164	0.0262	0.6613	0.7642
36	184	6	2	0.6930	0.0270	0.6366	0.7426
37	176	1	0	0.6891	0.0272	0.6324	0.7389
40	175	1	0	0.6852	0.0273	0.6282	0.7353
44	174	1	1	0.6812	0.0274	0.6241	0.7316
46	172	1	1	0.6773	0.0276	0.6199	0.7279
48	170	5	5	0.6574	0.0281	0.5990	0.7093
50	160	1	0	0.6532	0.0283	0.5947	0.7055
55	159	1	0	0.6491	0.0284	0.5904	0.7016
58	158	0	1	0.6491	0.0284	0.5904	0.7016
59	157	0	1	0.6491	0.0284	0.5904	0.7016
60	156	20	28	0.5659	0.0302	0.5044	0.6227
61	108	1	0	0.5607	0.0304	0.4989	0.6178

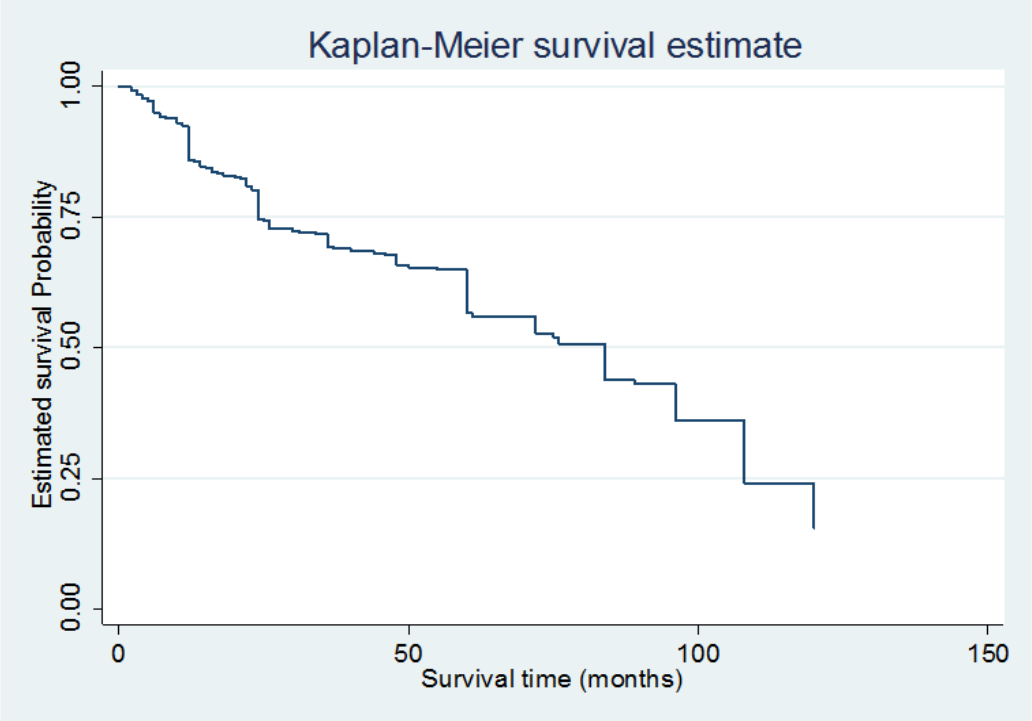
62	107	0	1	0.5607	0.0304	0.4989	0.6178
63	106	0	2	0.5607	0.0304	0.4989	0.6178
64	104	0	1	0.5607	0.0304	0.4989	0.6178
65	103	0	2	0.5607	0.0304	0.4989	0.6178
67	101	0	2	0.5607	0.0304	0.4989	0.6178
71	99	0	1	0.5607	0.0304	0.4989	0.6178
72	98	6	12	0.5263	0.0316	0.4626	0.5861
75	80	1	1	0.5198	0.0319	0.4555	0.5801
76	78	2	1	0.5064	0.0324	0.4412	0.5680
84	75	10	9	0.4389	0.0344	0.3707	0.5050
89	56	1	0	0.4311	0.0347	0.3624	0.4978
96	55	9	10	0.3605	0.0361	0.2903	0.4310
108	36	12	7	0.2404	0.0372	0.1715	0.3158
120	17	6	11	0.1555	0.0368	0.0918	0.2345

Source: prepared by the researcher by using STATA ,2014

Seen from table (4-36) that the first survival time observed is (1) month, and there are (325) individuals at risk, (0) individual is deaths, (7) individual is lost follow-up and the value of the estimated survival function is (1). The value of the function remains at this value until the time of death observed below. The second observed survival time is (2) months there are (318) individuals at risk, (2) individual is deaths, (2) individual is lost follow-up and the value of the estimated survival function is (0.9937), with standard error (0.0044) and confidence intervals (0. 0.9751 - 0.9984) at 5% significance level. The value of the function remains at this value until the time of death observed below. The third observed survival time is (3) months, and the number of individuals at risk is (314), there is (3) death, (1) individual is lost follow-up and the value of the estimated survival function is

(0.9842), with standard error (0.0070) and confidence intervals (0.9625– 0.9934) at 5% significance level. The value of the function remains at this value until the time of death observed below. This process continues until the last time of death (120) months, where there are (17) individual at risk and (6) died, (11) individual is lost follow-up and the value of the estimated survival function is (0.1555), with standard error (0.0368) and confidence intervals (0.0918– 0.2345) at 5% significance level.

Figure (4-18): Kaplan-Meier to estimate survival function, confidence intervals at the 5% level of significance for hemodialysis patients



Source: prepared by the researcher by using STATA, 2014.

The above figure is obtained from the survival function estimated in the table (4-35) and shows that there is a decreasing function defined by the estimated survival function, which decreases at the observed death times and Is fixed between these times because the estimated survival function curve is approximately zero at the last time of survival and 17 patients are at risk .The greatest time is a time of survival

4-4. Kaplan-Meier to estimate hazard function, standard error and confidence intervals at the 5% level of significance:

Table (4-37): Kaplan-Meier to estimate hazard function, standard error and confidence intervals at the 5% level of significance for patient's hemodialysis

Time	Total	event	sensors	hazard Function	Std. Error	[95% Conf. Int.]	
						Lower	upper
1	325	0	7	0.0000	.	..	. .
2	318	2	2	0.0063	0.0044	0.0016	0.0249
3	314	3	1	0.0158	0.0070	0.0066	0.0375
4	310	2	0	0.0221	0.0083	0.0106	0.0459
5	308	2	0	0.0285	0.0094	0.0149	0.0540
6	306	7	1	0.0507	0.0124	0.0314	0.0814
7	298	2	0	0.0571	0.0131	0.0363	0.0891
8	296	1	0	0.0603	0.0134	0.0389	0.0929
10	295	3	0	0.0698	0.0144	0.0465	0.1041
11	292	2	1	0.0762	0.0149	0.0517	0.1115
12	289	20	10	0.1401	0.0196	0.1062	0.1837
13	259	1	1	0.1434	0.0198	0.1091	0.1873
14	257	3	1	0.1534	0.0204	0.1179	0.1984
15	253	1	0	0.1568	0.0206	0.1209	0.2021
16	252	2	2	0.1635	0.0210	0.1268	0.2094
17	248	1	2	0.1669	0.0211	0.1298	0.2131
18	245	1	0	0.1703	0.0213	0.1328	0.2169

20	244	1	2	0.1737	0.0215	0.1358	0.2206
21	241	1	0	0.1771	0.0217	0.1389	0.2243
22	240	4	3	0.1908	0.0224	0.1511	0.2393
23	233	2	0	0.1977	0.0227	0.1574	0.2468
24	231	16	13	0.2533	0.0250	0.2080	0.3064
25	202	1	3	0.2570	0.0252	0.2114	0.3103
26	198	4	2	0.2720	0.0258	0.2252	0.3263
27	192	0	2	0.2720	0.0258	0.2252	0.3263
30	190	1	1	0.2758	0.0259	0.2287	0.3304
31	188	1	0	0.2797	0.0261	0.2323	0.3345
33	187	0	2	0.2797	0.0261	0.2323	0.3345
34	185	1	0	0.2836	0.0262	0.2358	0.3387
36	184	6	2	0.3070	0.0270	0.2574	0.3634
37	176	1	0	0.3109	0.0272	0.2611	0.3676
40	175	1	0	0.3148	0.0273	0.2647	0.3718
44	174	1	1	0.3188	0.0274	0.2684	0.3759
46	172	1	1	0.3227	0.0276	0.2721	0.3801
48	170	5	5	0.3426	0.0281	0.2907	0.4010
50	160	1	0	0.3468	0.0283	0.2945	0.4053
55	159	1	0	0.3509	0.0284	0.2984	0.4096
58	158	0	1	0.3509	0.0284	0.2984	0.4096
59	157	0	1	0.3509	0.0284	0.2984	0.4096
60	156	20	28	0.4341	0.0302	0.3773	0.4956
61	108	1	0	0.4393	0.0304	0.3822	0.5011

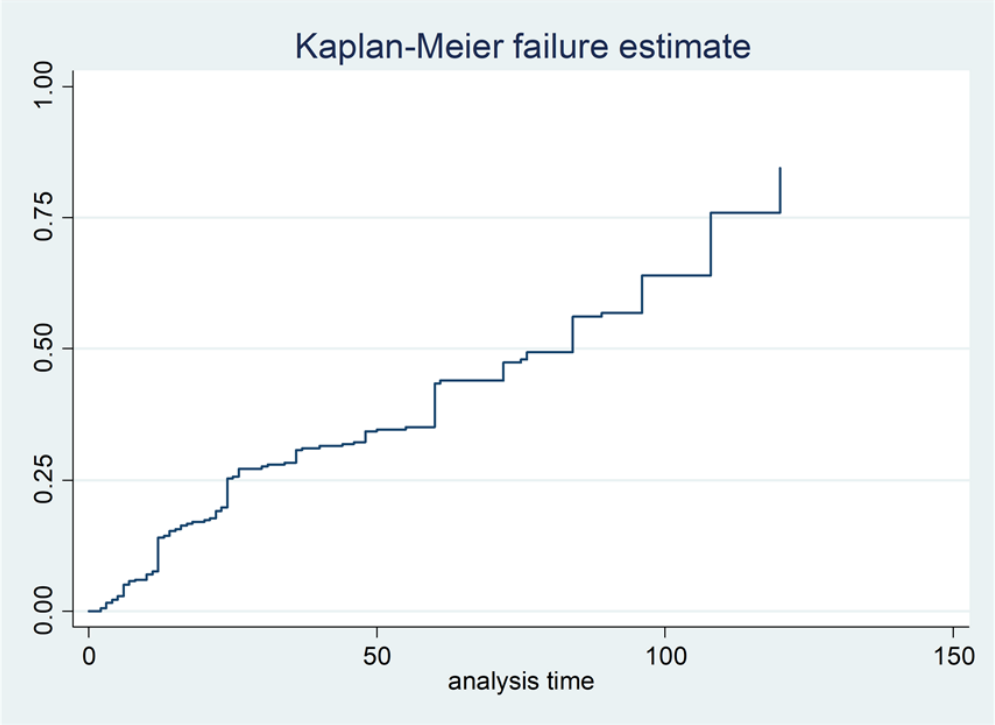
62	107	0	1	0.4393	0.0304	0.3822	0.5011
63	106	0	2	0.4393	0.0304	0.3822	0.5011
64	104	0	1	0.4393	0.0304	0.3822	0.5011
65	103	0	2	0.4393	0.0304	0.3822	0.5011
67	101	0	2	0.4393	0.0304	0.3822	0.5011
71	99	0	1	0.4393	0.0304	0.3822	0.5011
72	98	6	12	0.4737	0.0316	0.4139	0.5374
75	80	1	1	0.4802	0.0319	0.4199	0.5445
76	78	2	1	0.4936	0.0324	0.4320	0.5588
84	75	10	9	0.5611	0.0344	0.4950	0.6293
89	56	1	0	0.5689	0.0347	0.5022	0.6376
96	55	9	10	0.6395	0.0361	0.5690	0.7097
108	36	12	7	0.7596	0.0372	0.6842	0.8285
120	17	6	11	0.8445	0.0368	0.7655	0.9082

Source: prepared by the researcher by using STATA, 2014.

Seen from table (4-37) that the first deaths time observed is (1) month, and there are (325) individuals at risk, (0) individual is deaths, (7) individual is lost follow-and the value of the estimated hazard function is (0.0000). The second observed deaths time is (2) months there are (318) individuals at risk, (2) individual is deaths, (2) individual is lost follow-up and the value of the estimated hazard function is (0.0063), with standard error (0.0044) and confidence intervals (0.0016–0.0249) at 5% significance level. The value of the function remains at this value until the time of death observed below. The third observed deaths time is (3) months, and the number of individuals at risk is (314), there is (3) death, (1) individual is lost follow-up and the value of the estimated hazard function is (0.0158), with standard error (0.0070) and confidence intervals (0.0066– 0.0375) at 5% significance level. The

value of the function remains at this value until the time of death observed below. This process continues until the last time of death (120) months, where there are (17) individual at risk, (6) individual is deaths, (11) individual is lost follow-up and the value of the estimated hazard function is (0.8445), with standard error (0.0368) and confidence intervals (0.7655– 0.9082) at 5% significance level.

Figure (4-19): Kaplan-Meier to estimate hazard function, confidence intervals at the 5% level of significance for hemodialysis patients



Source: prepared by the researcher by using STATA, 2014

The above figure is obtained from the hazard function estimated in the table (4-36) and shows that there is an increasing function defined by the estimated hazard function, which increases at the observed death times and is fixed between these times because the estimated hazard function curve is approximately equal one at the last time of survival and 17 patients are at risk .The greatest time is a time of survival.

4-5. Estimation of Median and Quartiles:

4-5-1. Estimation of median and quartiles for survival time:

Table (4-38): Quartiles estimated for survival time:

Quartile	Estimate	Std. Error	[95% Conf. Int.]	
			Lower	Upper
25	24	2.905	24	36
50	84	3.659	61	89
75	108	4.527	108	120

Source: prepared by the researcher by using STATA, 2014

The first quartile of each individual is (24) months, this means that 25% of the individuals will live (24) months and it does not at least 24 months and not more than 36 months. The median of each individual is (84) months, this means that 50% of the individuals will live (84) months and it does not at least 61 months and not more than 89 months. The third quartile of each individual is (108) months, this means that 75% of the individuals will live (108) months and it does not at least 108months and does not more than 120 months.

4-5-2.Estimation median of survival time for sex:

Table (4-39): Estimation median of survival time sex

sex	Number of patients	Estimate 50%	Std. Error	[95% Conf. Int.]	
				lower	upper
male	194	84	6.257	72	108
female	131	75	4.301	60	84
total	325	84	3.659	61	89

Source: prepared by the researcher by using STATA, 2014

The median estimation for patients male is (84) months, this means that 50% of the patient's male will live (84) months and it does not at least 72 months and not more than 108 months. The median estimation for patients female is (75) months, this means that 50% of the patient's female will live (75) months and it does not at least 60 months and not more than 84 months.

4-5-3. Estimation median of survival time for age:

Table (4-40): Estimation median of survival time age

age	Number of Patients	Estimate	St. error	95% Confidence intervals	
				Lower	Upper
total	325	84	3.659	61	89

Source: prepared by the researcher by using STATA, 2014

The median estimation for patient's age is (84) months, this means that 50% of the patient's male will live (84) months and it does not at least 61 months and not more than 89 months.

4-6. Univariate analysis:

The Kaplan –Meier survival curves for different patient groups, and introduced the log-rank test to investigate differences between them. Both these methods are examples of univariate analysis; they describe the survival with respect to the factor under investigation, but necessarily ignore the impact of any others. It is more common, at least in clinical investigations, to have a situation where several (known) quantities or covariates, potentially affect patient prognosis. When investigating survival in relation to any one factor, it is often desirable to adjust for the impact of others (Bradburn, et al., 2003)

4-6-1. Log-rank test for equality of survival functions:

The log-rank test is a statistical test used to compare the survival distributions of two or more groups used to test the hypothesis where there is no difference between the categories for and variable. It does not provide any estimation of the actual size of the effect; in other words, it provides a statistical, but not a clinical, assessment of the effect of the factor. The use of a statistical model improves on these methods by allowing survival to be assessed with respect to several factors simultaneously, and in addition, offers estimates of the strength of effect for each constituent factor (Bradburn, et al., 2003).

4-6-1-1. Log-rank test for equality of survival functions for sex:

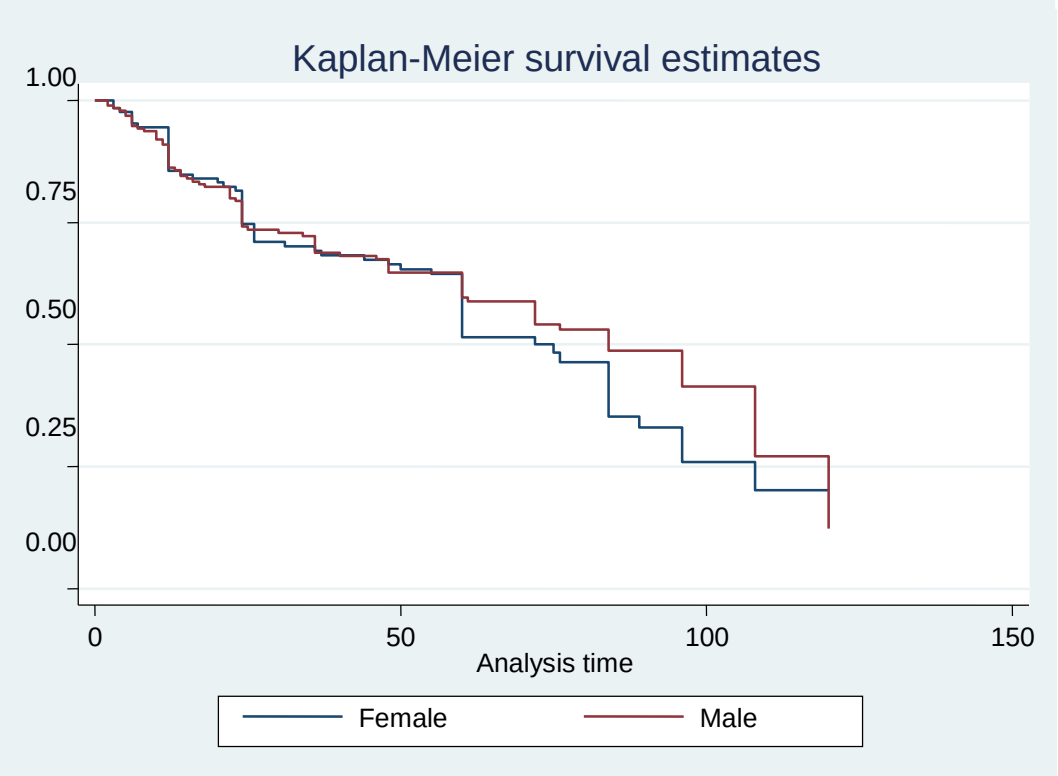
Table (4-41) Log-rank test for equality of estimated survival functions for sex

sex	Events Observed	Events Expected	Chi-square test	P-value
(male)	100	104.92	0.66	0.4164
(female)	70	65.08		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-41) that the value of chi-square test was (0.66), and the significant value to it (P-value = 0.42 > 0.05), there is no significant difference between the estimated survival functions for male and female in study.

Figure (4-20): survival curves of patient's sex



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-20) there is no difference between the curves of Kaplan - Meier survival functions for sex. Thus, the probability of survival does not vary according to sex.

4-6-1-2. Log-rank test for equality of survival functions for dialysis frequency per week:

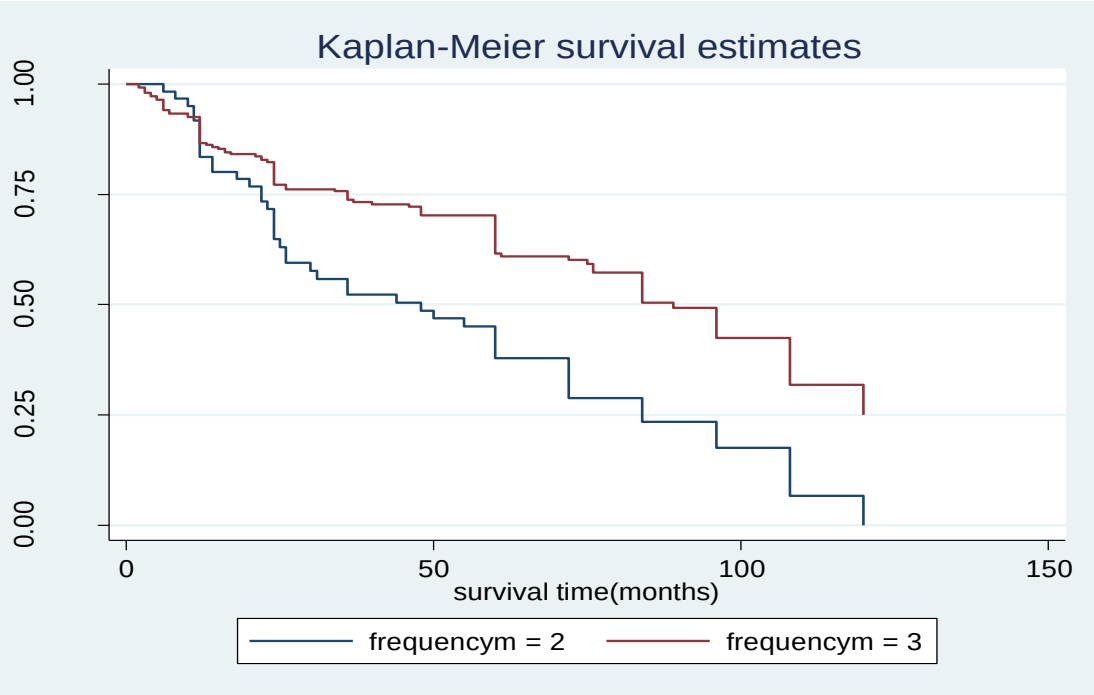
Table (4-42) Log-rank test for equality of estimated survival functions for dialysis frequency per week

dialysis frequency per week	Events Observed	Events Expected	Chi-square test	P-value
2	55	32.26	21.65	0.0000
3	115	137.74		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-42) that the value of chi-square test was (21.65), and the significant value to it (P-value = 0.000 < 0.05), there is significant difference between the estimated survival functions dialysis frequency per week

Figure (4-21): survival curves of patient's dialysis frequency per week



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-21) there is difference between the curves of Kaplan - Meier survival functions for dialysis frequency per week, since the estimated curve for those who dialysis three times a week is higher than the estimated curve, which is dialysis twice a week. Therefore ,at any point in time the survival rate those who dialysis three times is estimated to be longer than those who dialysis twice. Thus, the probability of survival does vary according dialysis frequency per week.

4-6-1-3. Log-rank test for equality of survival functions for diabetes mellitus:

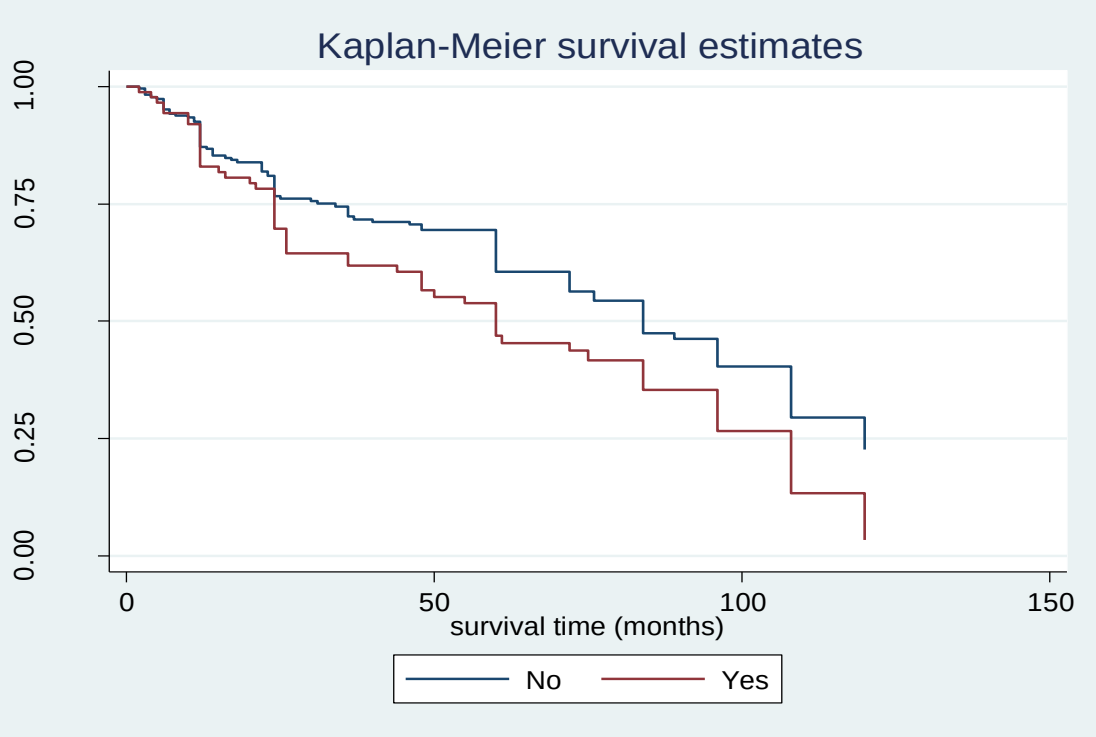
Table (4-43) Log-rank test for equality of estimated survival functions diabetes mellitus

diabetes mellitus	Events Observed	Events Expected	Chi-square test	P-value
No	109	123.91	7.21	0.0070
Yes	61	46.09		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-43) that the value of chi-square test was (7.21), and the significant value to it (P-value = 0.007 < 0.05), there is significant difference between the estimated survival functions diabetes mellitus.

Figure (4-22): survival curves of patients diabetes mellitus



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-22) there is difference between the curves of Kaplan - Meier survival functions for diabetes mellitus, since the estimated curve for those who no diabetes mellitus is higher than the estimated curve, which is diabetes mellitus. Therefore, at any point in time the survival rate those who no diabetes mellitus is estimated to be longer than those who have diabetes mellitus. Thus, the probability of survival does varies according to diabetes mellitus.

4-6-1-4. Log-rank test for equality of survival functions for hospitals:

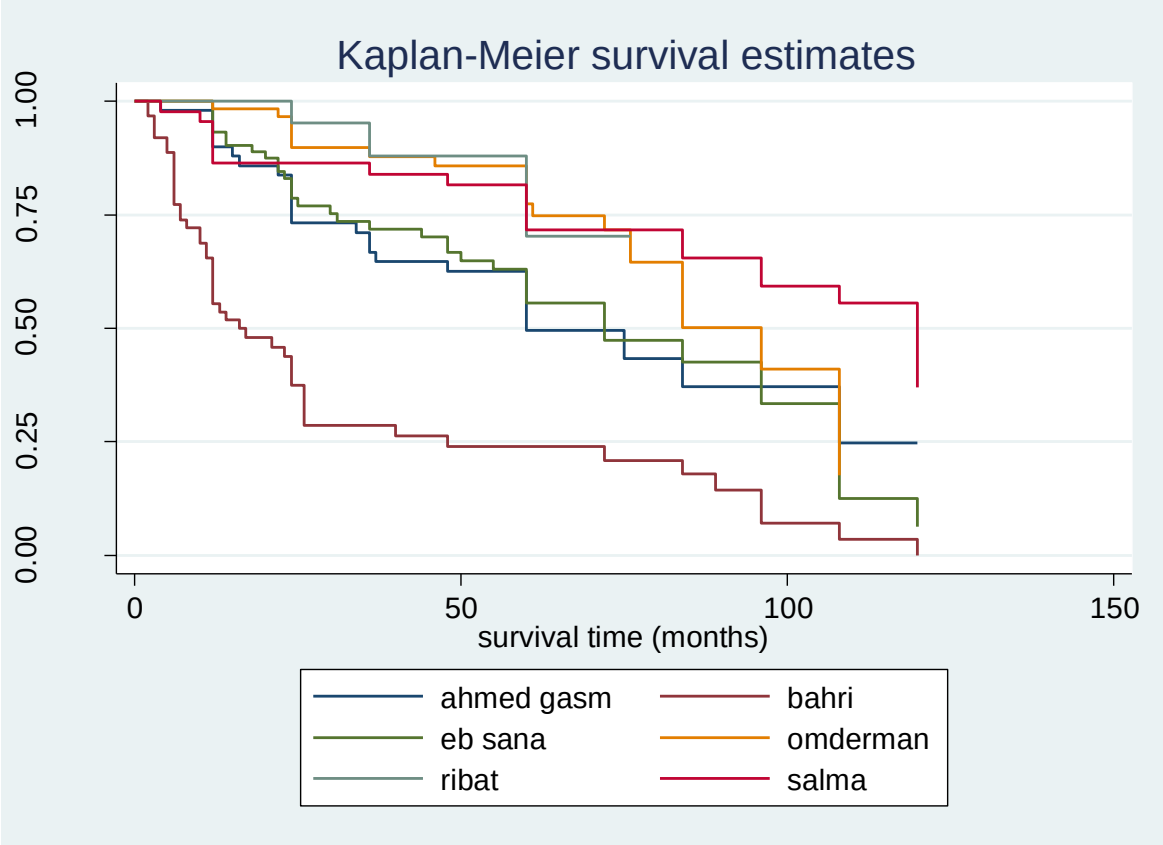
Table (4-44) Log-rank test for equality of estimated survival functions hospitals

marital status	Events Observed	Events Expected	Chi-square test	P-value
Ahmed gasm	27	25.73	88.57	0.0000
Bahri	49	16.47		
Ebn sena	43	40.52		
Omdurman	26	37.68		
Ribat	4	10.03		
Selma	21	39.57		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-44) that the value of chi-square test was (88.57), and the significant value to it (P-value = 0.000 < 0.05), there is significant difference between the estimated survival functions hospitals.

Figure (4-23): survival curves of hospitals



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-23) there is difference between the curves of Kaplan - Meier survival functions for hospitals, since the estimated curve for those who Ribat ,Salma center are higher than the estimated curve, which is Omdurman, Ebn Sana, and Bahri hospital under these curves estimated curve. Therefore, at any point in time the survival rate those whose higher estimated to be longer survival time than .Thus, the probability of survival does varies according to hospitals.

4-6-1-5. Log-rank test for equality of survival functions for hypertension:

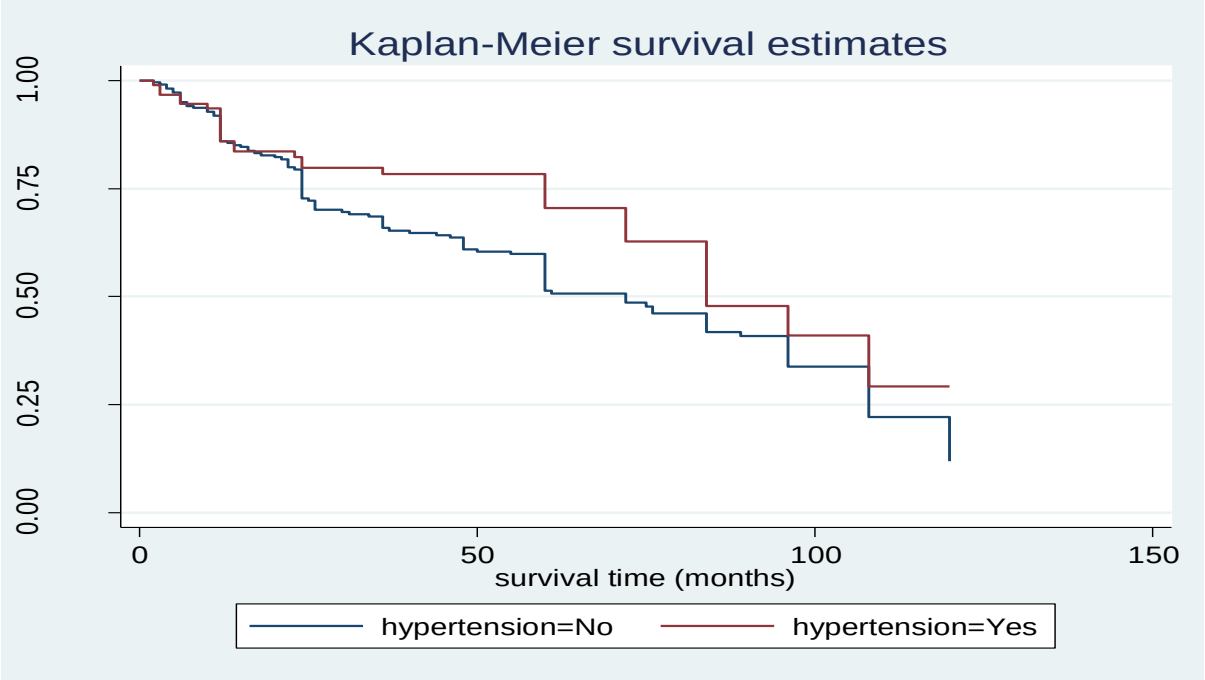
Table (4-45) Log-rank test for equality of estimated survival functions hypertension

hypertension	Events Observed	Events Expected	Chi-square test	P-value
No	134	121.80	4.70	0.0302
Yes	36	48.20		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-45) that the value of chi-square test was (4.70), and the significant value to it (P-value = 0.03 < 0.05), there is no significant difference between the estimated survival functions for no and yes hypertension in study.

Figure (4-24): survival curves of hypertension



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-24) there is no difference between the curves of Kaplan - Meier survival functions for hypertension. Thus, the probability of survival does vary according to hypertension.

4-6-1-6. Log-rank test for equality of survival functions for diabetes mellitus and hypertension:

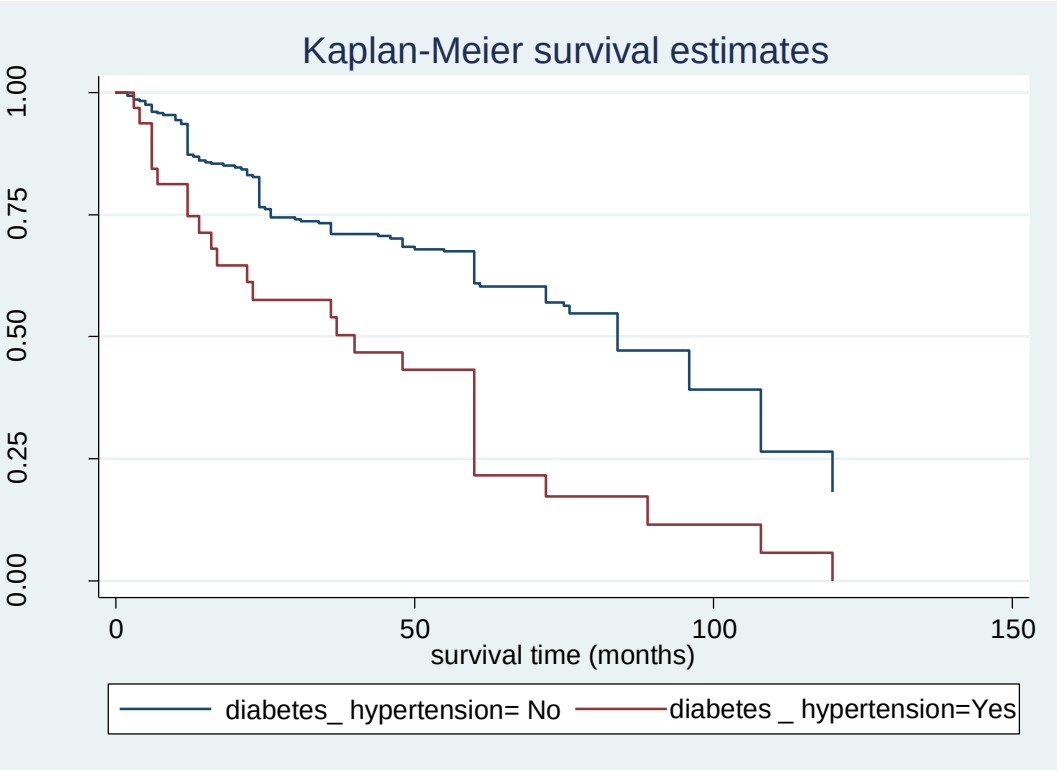
Table (4-46) Log-rank test for equality of estimated survival functions diabetes mellitus and hypertension

diabetes mellitus and hypertension	Events Observed	Events Expected	Chi-square test	P-value
No	143	157.66	20.31	0.0000
Yes	27	12.34		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-46) that the value of chi-square test was (20.31), and the significant value to it (P-value = 0.000 < 0.05), there is significant difference between the estimated survival functions diabetes mellitus and hypertension.

Figure (4-25): survival curves of functions diabetes mellitus and hypertension



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-25) there is difference between the curves of Kaplan - Meier survival functions for diabetes mellitus and hypertension, since the estimated curve for those who no diabetes mellitus and hypertension is higher than the estimated curve, which is diabetes mellitus and hypertension. Therefore, at any point in time the survival rate those who no diabetes mellitus and hypertension is estimated to be longer than those who have diabetes mellitus and hypertension. Thus, the probability of survival does varies according to diabetes mellitus and hypertension,

4-6-1-7. Log-rank test for equality of survival functions for regular:

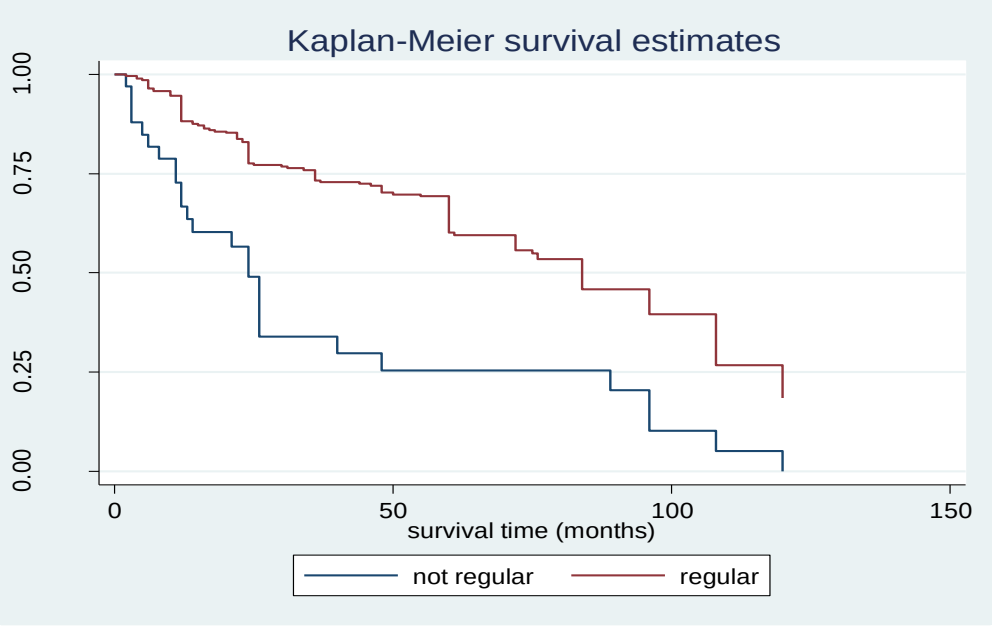
Table (4-47) Log-rank test for equality of estimated survival functions regular

regular	Events Observed	Events Expected	Chi-square test	P-value
not regular	27	10.94	27.38	0.0000
regular	143	159.06		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-47) that the value of chi-square test was (27.38), and the significant value to it (P-value = 0.000 < 0.05), there is significant difference between the estimated survival functions not regular and regular.

Figure (4-26): survival curves of regular



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-26) there is difference between the curves of Kaplan - Meier survival functions for regular, since the estimated curve for those who regular is higher than the estimated curve which is not regular. Therefore, at any point in time

the survival rate those who regular is estimated to be longer survival time than those who not regular. Thus, the probability of survival does vary according to regular.

4-6-1-8. Log-rank test for equality of survival functions for other:

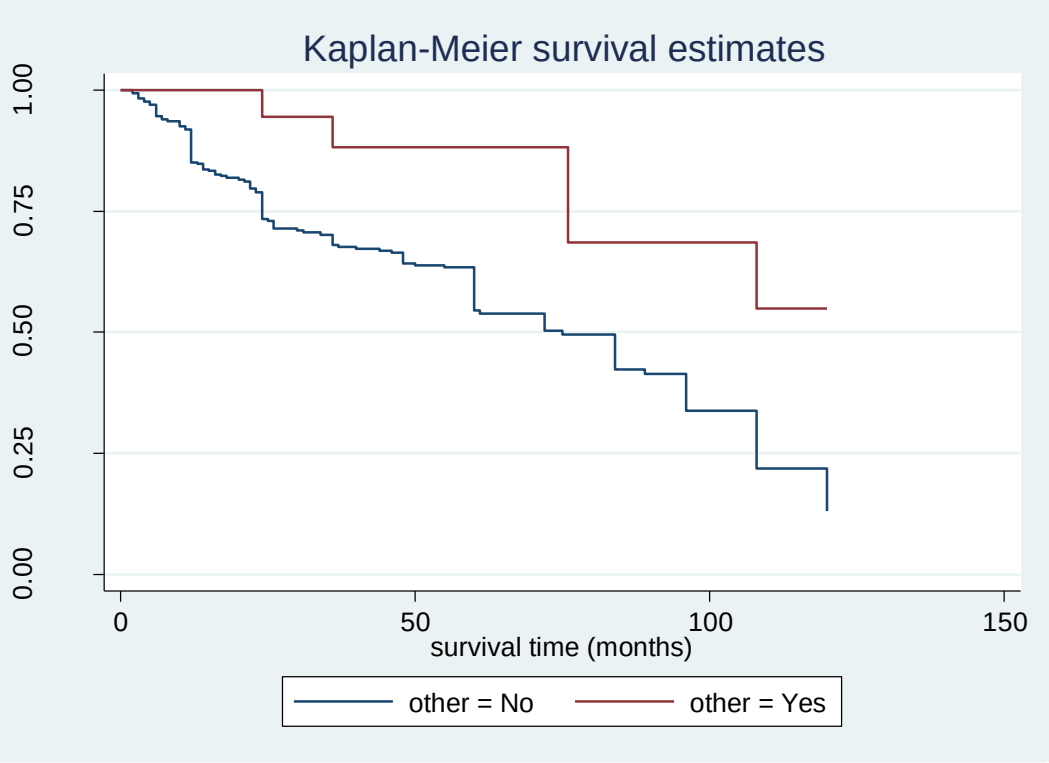
Table (4-48) Log-rank test for equality of estimated survival functions other

other	Events Observed	Events Expected	Chi-square test	P-value
No	165	155.47	7.63	0.006
Yes	5	14.53		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-48) that the value of chi-square test was (7.63), and the significant value to it (P-value = 0.006 < 0.05), there is significant difference between the estimated survival functions no, yes other.

Figure (4-27): survival curves of other



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-27) there is difference between the curves of Kaplan - Meier survival functions for regular, since the estimated curve for those who regular is higher than the estimated curve which is not regular. Therefore, at any point in time the survival rate those who regular is estimated to be longer survival time than those who not regular. Thus, the probability of survival does vary according to regular

4-6-1-9. Log-rank test for equality of survival functions for shrunken kidneys:

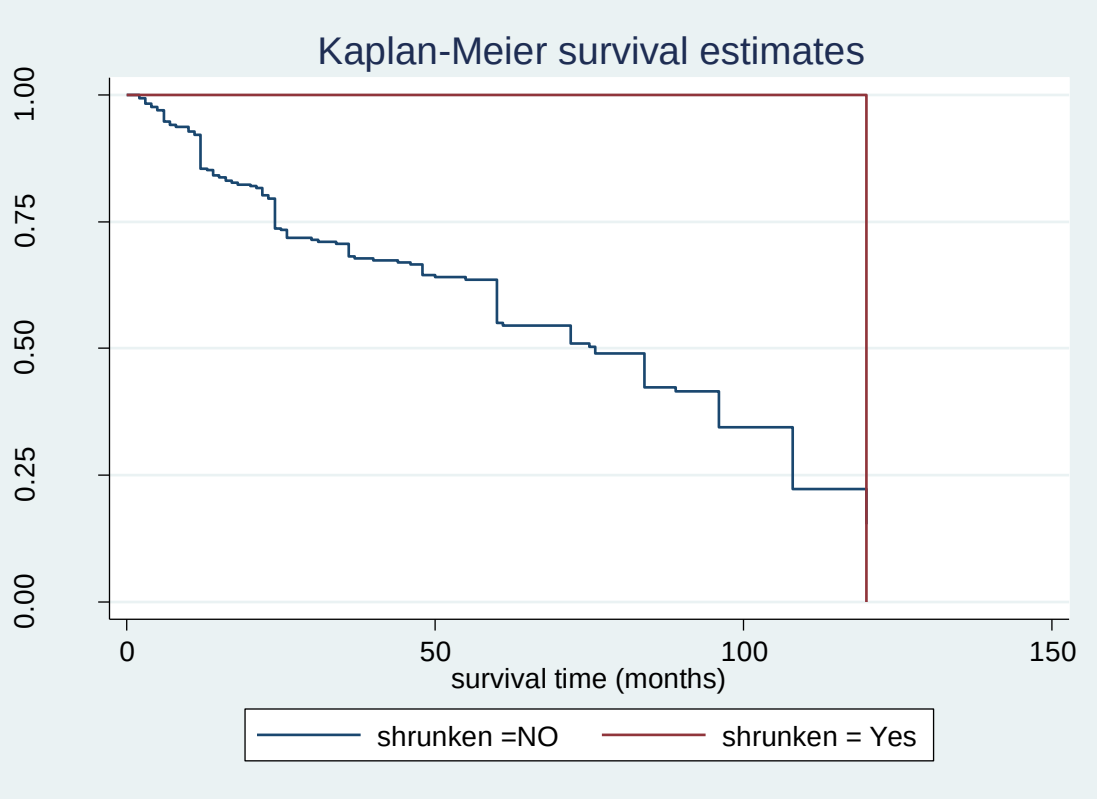
Table (4-49) Log-rank test for equality of estimated survival functions shrunken kidneys

Shrunken kidneys	Events Observed	Events Expected	Chi-square test	P-value
No	169	162.57	6.39	0.0115
Yes	1	7.43		
total	170	170.00		

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-49) that the value of chi-square test was (6.39), and the significant value to it (P-value = 0.012 < 0.05), there is significant difference between the estimated survival functions no, yes for shrunken kidneys

Figure (4-28): survival curves of shrunken kidneys



Source: prepared by the researcher by using STATA, 2014

We note from figure (4-28) there is no difference between the curves of Kaplan - Meier survival functions for shrunken kidneys. Thus, the probability of survival does vary according to shrunken kidneys

4-7.Estimate parametric models

4-7-1.Estimate parametric models for univariate analysis in hemodialysis patients:

Based on the log rank test, the variables considered to be important with p-value < 0.05 were entered in the parametric models, while other variables were not significantly excluded from the parametric models. The variables used in the parametric model were regular, dialysis frequency per week, hospitals, diabetes mellitus, hypertension, diabetes mellitus and hypertension, shrunken kidneys, other.in addition the urea and age and Serum creatinine

4-7-1-1. Estimate Exponential model for univariate analysis-- log relative-hazard form:

Table (4-50): Coefficient and hazard ratio

Variable	Coef.	95%Conf.Interval		Haz. Ratio	95% Conf. Interval		z	P-value	Prob <chi2
		upper	lower		upper	lower			
Age	0.021	0.011	0.030	1.020	1.011	1.030	4.36	0.000	0.000
Regular	-1.031	-1.442	-0.620	0.357	0.236	0.538	-4.91	0.000	0.000
Hospital	-0.225	-0.320	-0.131	0.798	0.727	0.877	-4.670	0.000	0.000
Diabetes mellitus	0.410	0.097	0.724	1.507	1.102	2.062	2.570	0.010	0.012
Diabetes mellitus and hypertension	0.833	0.422	1.244	2.300	1.525	3.471	3.970	0.000	0.003
Hypertension	-0.423	-0.791	-0.055	0.655	0.453	0.946	-2.250	0.024	0.019
Shrunken kidneys	-2.061	-4.027	-0.096	0.127	0.018	0.909	-2.060	0.040	0.002
Dialysis frequency per (wk)	-0.735	-1.056	-0.413	0.480	0.348	0.661	-4.480	0.000	0.000
Urea	0.006	0.004	0.009	1.006	1.004	1.009	4.510	0.000	0.000
Serum creatinine	0.077	0.044	0.110	1.080	1.045	1.116	4.550	0.000	0.000
Other	-1.076	-1.966	-0.186	0.341	0.140	0.830	-2.370	0.018	0.005
Coef; coefficient, HR; Hazard Ratio, p=value significant at < 0.05 level of significance, Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, cardiovascular disease, NSAID).									
Source: prepared by the researcher by using STATA, 2014									

We note from the table (4-50), the variable age is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between ages is 0.021 , and it does not at least 0.011 and not more than 0.030 with 95% confidence interval, the hazard ratio is 1.020, means that at any time during the study, the per-month rate of death among age is increase by 2% in the risk of age group. Prob >chi2 for univariate exponential model for age is significant (p-value = 0.000< 0.05).

The variable regular is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between regular is -1.031. , and it does not at least -1.442 and not more than -0.620 with 95% confidence interval, the hazard ratio is 0.357, means that at any time during the study, the per-month rate of death among regular patient is reduction by 64.2% in the risk of death of regular. Prob >chi2 for univariate exponential model for regular is significant (p-value = 0.000< 0.05).

The variable hospital is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between regular is -0.225. , and it does not at least -0.320 and not more than -0.131 with 95% confidence interval, the hazard ratio is 0 .798, means that at any time during the study, the per-month rate of death among hospital is reduction by 20.2% in the risk of death for hospital. Prob >chi2 for univariate exponential model for hospital is significant (p-value = 0.0000< 0.05).

The variable diabetes mellitus is significant value of Wald test (P-value=0.010 <0.05), so estimated coefficient is significant. That means the difference in the log hazard between diabetes mellitus is 0.410. , and it does not at least 0.097 and not more than 0.724 with 95% confidence interval, the hazard ratio is 1.507 , means that at any time during the study, the per-month rate of death among diabetes

mellitus is increase by 50.7% in the risk of death for diabetes mellitus. Prob >chi2 for univariate exponential model diabetes mellitus is significant (p-value = 0.012< 0.05).

The variable diabetes mellitus and hypertension is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between diabetes mellitus and hypertension is 0.833, and it does not at least 0.422 and not more than 1.244 with 95%confidence interval, the hazard ratio is 2.300, means that at any time during the study, the per-month rate of death is increase by 130% in the risk of death for of diabetes mellitus and hypertension. Prob >chi2 for univariate exponential model diabetes mellitus and hypertension is significant (p-value = 0.0003< 0.05).

The variable hypertension is significant value of Wald test (P-value=0.024<0.05), so estimated coefficient is significant. That means the difference in the log hazard hypertension is -0.423, and it does not at least -0.791 and not more than -0.055 with 95% confidence interval, the hazard ratio is 0.655 , means that at any time during the study, the per-month rate of death is reduction by 35.5% in the risk of death for hypertension . Prop >chi2 for univariate exponential model for hypertension is significant (p-value = 0.019< 0.05).

The variable shrunken kidneys is significant value of Wald test (P-value=0.040 <0.05), so estimated coefficient is significant. That means the difference in the log hazard shrunken is -2.061, and it does not at least -4.027 and not more than -0.096-with 95% confidence interval, the hazard ratio is 0.127, means that at any time during the study, the per-month rate of death is reduction by 87.3% in the risk of death for shrunken kidneys. Prob >chi2 for univariate exponential model for shrunken kidneys is significant (p-value = 0.002< 0.05).

The variable dialysis frequency per week is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard dialysis frequency per week is -0.735, and it does not at least -1.056 and not more than -0.413 with 95% confidence interval, the hazard ratio is 0.480, means that at any time during the study, the per-month rate of death is reduction by 52% % in the risk of death for dialysis frequency per week. Prob >chi2 for univariate exponential model for dialysis frequency per week significant (p-value = 0.0000< 0.05).

The variable urea is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means the difference in the log hazard urea is 0.006, and it does not at least 0.004 and not more than 0.009 with 95% confidence interval, the hazard ratio is 1.006, means that at any time during the study, the per-month rate of death is increase by 0.6% % in the risk of death for urea. Prob >chi2 for univariate exponential model for urea significant (p-value = 0.0000< 0.05).

The variable serum creatinine is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means the difference in the log hazard Serum creatinine is 0.077, and it does not at least 0.044 and not more than .110 with 95% confidence interval, the hazard ratio is 1.079, means that at any time during the study, the per-month rate of death is increase by 8% in the risk of death for Serum creatinine. Prob >chi2 for univariate exponential model for serum creatinine significant (p-value = 0.0000< 0.05).

The variable other is significant value of Wald test (P-value=0.018<0.05), so estimated coefficient is significant. That means the difference in the log hazard other is -1.076, and it does not at least -1.966 and not more than -0.186 with 95% confidence interval, the hazard ratio is 0.341, means that at any time during the

study, the per-month rate of death is reduction by 65.5% in the risk of death for other. Prop >chi2 for univariate exponential model for other is significant (p-value = 0.005< 0.05).

4-7-1-2. Estimate Weibull model for univariate analysis-- log relative-hazard form:

Table (4-51): Coefficient and hazard ratio

variable	Coef.	[95% Conf. Interval]		HR [exp(coef)]	[95% Conf. Interval]		z	P-value	Prob <chi2
Age	0.021	0.012	0.030	1.021	1.012	1.031	4.420	0.000	0.000
Regular	-1.091	-1.503	-0.678	0.336	0.222	0.508	-5.180	0.000	0.000
Hospital	-0.242	-0.336	-0.148	0.785	0.715	0.863	-5.040	0.000	0.000
Diabetes mellitus	0.415	0.101	0.728	1.514	1.107	2.071	2.590	0.009	0.011
Diabetes mellitus and hypertension	0.878	0.465	1.290	2.405	1.592	3.632	4.170	0.000	0.000
Hypertension	-0.418	-0.786	-0.050	0.658	0.456	0.951	-2.230	0.026	0.021
Shrunken kidneys	-2.077	-4.043	-0.111	0.125	0.018	0.895	-2.070	0.038	0.002
Dialysis frequency per (wk)	-0.736	-1.057	-0.415	0.479	0.347	0.661	-4.490	0.000	0.000
Urea	0.006	0.004	0.009	1.006	1.004	1.009	4.570	0.000	0.000
Serum creatinine	0.078	0.045	0.111	1.081	1.046	1.117	4.630	0.000	0.000
Other	-1.115	-2.005	-0.225	0.328	0.135	0.799	-2.450	0.014	0.003
Coef; coefficient, HR; Hazard Ratio, p=value significant at < 0.05 level of significance, Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, cardiovascular disease, NSAID).									
Source: prepared by the researcher by using STATA, 2014									

We note from the table (4-51) for the variable age is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between ages is 0 .021, and it does not at least 0.012 and not more than 0.030 with 95% confidence interval, the hazard ratio is 1.021, means that at any time during the study, the per-month rate of death among age is increase by 2.1% in the risk of death. Prob >chi2 for univariate Weibull model for age is significant (p-value = 0.000< 0.05).

The variable regular is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between regular is -1.091, and it does not at least -1.503 and not more than - 0.672 with 95% confidence interval, the hazard ratio is 0.336, means that at any time during the study, the per-month rate of death among regular is reduction by 66.4% in the risk of death for regular. Prob >chi2 for univariate Weibull model for regular is significant (p-value = 0.000< 0.05).

The variable hospital is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between hospital is -0.240 , and it does not at least -0.334 and not more than -0.146 with 95% confidence interval, the hazard ratio is 0.716, means that at any time during the study, the per-month rate of death among hospital is reduction by 28.4% in the risk of death for hospital . Prob >chi2 for univariate Weibull model for hospital is significant (p-value = 0.000< 0.05).

The variable diabetes mellitus is significant value of Wald test (P-value=0.009 <0.05), so estimated coefficient is significant. That means the difference in the log hazard between diabetes mellitus is 0.415. , and it does not at least 0.101 and not more than 0.728 with 95% confidence interval, the hazard ratio is 1.514, means that at any time during the study, the per-month rate of death among diabetes mellitus is increase by 51.4% in the risk of death for diabetes mellitus. Prob >chi2 for univariate Weibull model diabetes mellitus is significant (p-value = 0.011< 0.05).

The variable diabetes mellitus and hypertension is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between diabetes mellitus and hypertension is 0.878, and it does not at least 0.465 and not more than 1.290 with 95% confidence interval, the hazard ratio is 2.405, means that at any time during the study, the per-month

rate of death is increase by 140.2% in the risk of death for diabetes mellitus and hypertension. Prob >chi2 for univariate Weibull model diabetes mellitus and hypertension is significant (p-value = 0.026 < 0.05).

The variable hypertension is significant value of Wald test (P-value=0.026 <0.05), so estimated coefficient is significant. That means the difference in the log hazard hypertension is -0.418, and it does not at least -0.786 and not more than -0.050 with 95% confidence interval, the hazard ratio is 0.658, means that at any time during the study, the per-month rate of death is reduction by 34.2% in the risk of death for hypertension. Prob >chi2 for univariate Weibull model for hypertension is significant (p-value = 0.021< 0.05).

The variable shrunken kidneys is significant value of Wald test (P-value=0.038 <0.05), so estimated coefficient is significant. That means the difference in the log hazard shrunken is -2.077, and it does not at least -4.043 and not more than -0.111 with 95% confidence interval, the hazard ratio is 0.125 , means that at any time during the study, the per-month rate of death is reduction by 87.5% in the risk of death for shrunken Prob >chi2 for univariate Weibull model for shrunken is significant (p-value = 0.002<0.05)

The variable dialysis frequency per week is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard dialysis frequency per week is -0.736, and it does not at least -1.057 and not more than -0.415 with 95% confidence interval, the hazard ratio is 0.479, means that at any time during the study, the per-month rate of death is reduction by 52.1% in the risk of death for dialysis frequency per week. Prob >chi2 for univariate Weibull model for dialysis frequency per week significant (p-value = 0.000< 0.05).

The variable urea is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means the difference in the log hazard urea is 0.006, and it does not at least 0.004 and not more than 0.009 with 95% confidence interval, the hazard ratio is 1.006, means that at any time during the study, the per-month rate of death is increase by 0.6% in the risk of death for urea . Prob >chi2 for univariate Weibull l model for urea significant (p-value = 0.0000< 0.05).

The variable serum creatinine is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means the difference in the log hazard Serum creatinine is 0.078, and it does not at least 0.045 and not more than 0.111with 95% confidence interval, the hazard ratio is 1.081, means that at any time during the study, the per-month rate of death is increase by 8.1% in the risk of death for serum creatinine. Prob >chi2 for univariate Weibull model for Serum creatinine significant (p-value = 0.0000< 0.05).

The variable other is significant value of Wald test (P-value=0.014<0.05), so estimated coefficient is significant. That means the difference in the log hazard other is -1.115, and it does not at least -2.005 and not more than -0.225 with 95% confidence interval, the hazard ratio is 0.328, means that at any time during the study, the per-month rate of death is reduction by 67.2% in the risk of death for other. Prob >chi2 for univariate Weibull model for other is significant (p-value = 0.003< 0.05).

4-7-1-3. Estimate Gompertz model for univariate analysis-- log relative-hazard form:

Table (4-52): Coefficient and hazard ratio

variable	Coef.	95% Conf. Interval		Haz. Ratio	95% Conf. Interval		z	P>|z|	Prob <chi2
age	0.021	0.012	0.031	1.022	1.012	1.031	4.500	0.000	0.000
regular	-1.076	-1.488	-0.664	0.341	0.226	0.515	-5.120	0.000	0.000
hospital	-0.254	-0.348	-0.160	0.776	0.706	0.852	-5.290	0.000	0.000
diabetes mellitus	0.415	0.101	0.728	1.514	1.107	2.071	2.590	0.009	0.011
diabetes mellitus and hypertension	0.900	0.487	1.313	2.459	1.628	3.716	4.270	0.000	0.000
hypertension	-0.406	-0.774	-0.038	0.666	0.461	0.963	-2.160	0.031	0.025
shrunken kidneys	-2.072	-4.038	-0.106	0.126	0.018	0.899	-2.070	0.039	0.002
dialysis frequency per(wk)	-0.726	-1.048	-0.405	0.484	0.351	0.667	-4.430	0.000	0.000
urea	0.006	0.004	0.009	1.006	1.004	1.009	4.660	0.000	0.000
Serum creatinine	0.078	0.045	0.111	1.081	1.046	1.117	4.630	0.000	0.000
other	-1.137	-2.027	-0.246	0.321	0.132	0.782	-2.500	0.012	0.003

Coef; coefficient, HR; Hazard Ratio, p=value significant at < 0.05 level of significance, Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, cardiovascular disease, NSAID).

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-52), the variable age is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between ages is 0.021, and it does not at least 0.012 and not more than 0.031 with 95% confidence interval, the hazard ratio is 1.022, means that at any time during the study, the per-month rate of death among age is increase by 2.2% in the risk of death for age. Prob >chi2 for univariate Gompertz model for age is significant (p-value = 0.000< 0.05).

The variable regular is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between regular is -1.076 , and it does not at least -1.488 and not more than -0.664

with 95% confidence interval, the hazard ratio is 0.341 , means that at any time during the study, the per-month rate of death among regular is reduction by 65.9% in the risk of death for regular. Prob >chi2 for univariate Gompertz model for regular is significant (p-value = 0.000< 0.05).

The variable hospital is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard between regular is -0.254, and it does not at least -0.348 and not more than -0.160 with 95% confidence interval, the hazard ratio is 0 .776, means that at any time during the study, the per-month rate of death among hospital is reduction by 22.4% in the risk of death for hospital. Prob >chi2 for univariate Gompertz model for hospital is significant (p-value = 0.000< 0.05).

The variable diabetes mellitus is significant value of Wald test (P-value=0.009 <0.05), so estimated coefficient is significant. That means the difference in the log hazard between diabetes mellitus is 0.415. , and it does not at least 0.101 and not more than 0.728 with 95% confidence interval, the hazard ratio is 1.514 , means that at any time during the study, the per-month rate of death among diabetes mellitus is increase by 51.4% in the risk of death for diabetes mellitus. Prob >chi2 for univariate Gompertz model diabetes mellitus is significant (p-value = 0.011< 0.05).

The variable diabetes mellitus and hypertension is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means the difference in the log hazard between diabetes mellitus and hypertension is 0.900, and it does not at least 0.487 and not more than 1.313 with 95% confidence interval, the hazard ratio is 2.459, means that at any time during the study, the per-month rate of death is increase by 145.9% in the risk of death for diabetes mellitus and hypertension. Prob >chi2 for univariate Gompertz model diabetes mellitus and hypertension is significant (p-value = 0.000 < 0.05).

The variable hypertension is significant value of Wald test (P-value=0.031 <0.05), so estimated coefficient is significant. That means the difference in the log hazard hypertension is -0.406, and it does not at least -0.774 and not more than -0.038 with 95% confidence interval, the hazard ratio is 0 .666 means that at any time during the study, the per-month rate of death is reduction by 33.4% in the risk of death for hypertension. Prob >chi2 for univariate Gompertz model for hypertension is significant (p-value = 0.025 < 0.05).

The variable shrunken kidneys is significant value of Wald test (P-value=0.039 <0.05), so estimated coefficient is significant. That means the difference in the log hazard shrunken kidneys is -2.072, and it does not at least -4.038 and not more than -0.106 with 95% confidence interval, the hazard ratio is 0.126, means that at any time during the study, the per-month rate of death is reduction by 87.4% in the risk of death for shrunken. Prob >chi2 for univariate Gompertz model for shrunken kidneys is significant (p-value = 0.002<0.05)

The variable dialysis frequency per week is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in the log hazard dialysis frequency per week is -0.726, and it does not at least -1.048 and not more than -0.405 with 95% confidence interval, the hazard ratio is 0.484, means that at any time during the study, the per-month rate of death is reduction by 51.6% in the risk of death for dialysis frequency per week. Prob>chi2 for univariate Gompertz model for dialysis frequency per week significant (p-value = 0.0000< 0.05).

The variable urea is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means the difference in the log hazard urea is 0.006, and it does not at least 0.004 and not more than 0.009 with 95% confidence interval, the hazard ratio is 1.006 , means that at any time during the study, the per-month rate of death is increase by 0.6% in the risk of death for urea.

Prob >chi2 for univariate Gompertz model for urea significant (p-value = 0.000< 0.05).

The variable serum creatinine is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means the difference in the log hazard serum creatinine is 0.078, and it does not at least 0.045 and not more than 0.111 with 95% confidence interval, the hazard ratio is 1.081, means that at any time during the study, the per-month rate of death is increase by 8.1% in the risk of death for serum creatinine . Prob >chi2 for univariate Gompertz model for Serum creatinine significant (p-value = 0.0000< 0.05).

The variable other is significant value of Wald test (P-value=0.012 <0.05), so estimated coefficient is significant. That means the difference in the log hazard other is -1.137 , and it does not at least -2.027 and not more than -0.246 with 95% confidence interval, the hazard ratio is 0.321, means that at any time during the study, the per-month rate of death is reduction by 67.9% in the risk of death for other. Prob >chi2 for univariate Gompertz model for other is significant (p-value = 0.003 < 0.05).

4-7-1-4. Estimate Lognormal model for univariate analysis-- accelerated failure-time form:

Table (4-53): Coefficient and time ratio

variable	Coef.	[95% Conf. Interval]		time Ratio	[95% Conf. Interval]		z	P>|z|	Prob <chi2
age	-0.009	-0.019	0.000	0.990	0.981	1.000	-2.010	0.044	0.041
regular	1.181	0.731	1.631	3.258	2.078	5.108	5.150	0.000	0.000
hospital	0.269	0.175	0.362	1.308	1.192	1.436	5.650	0.000	0.000
diabetes mellitus	-0.356	-0.693	-0.019	0.700	0.500	0.981	-2.070	0.038	0.038
diabetes mellitus and hypertension	-0.855	-1.327	-0.384	0.425	0.265	0.681	-3.560	0.000	0.000
hypertension	0.332	-0.024	0.689	1.394	0.976	1.991	1.830	0.068	0.065
other	1.158	0.396	1.921	3.185	1.485	6.828	2.980	0.003	0.002
shrunken	2.028	0.572	3.483	7.597	1.770	32.576	2.730	0.006	0.001
dialysis frequency per(wk)	0.588	0.215	0.962	1.801	1.240	2.616	3.090	0.002	0.002
urea	-0.007	-0.010	-0.003	0.993	0.990	0.997	-3.850	0.000	0.000
Serum creatinine	-0.074	-0.111	-0.036	0.929	0.894	0.965	-3.810	0.000	0.000

Coef ; coefficient, HR; Hazard Ratio, p=value significant at < 0.05 level of significance
Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, cardiovascular disease, NSAID).

Source: prepared by the researcher by using STATA, 2014

We note from the table (4-53), for the variable age is significant value of Wald test (P-value=0.044 <0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between ages is -0.009, and it does not at least -0.019 and not more than 0.000 with 95% confidence interval, the time ratio is 0.990, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 1%. Prob >chi2 for univariate lognormal model for age is significant (p-value = 0.0000< 0.05).

The variable regular is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between regular is 1.181, and it does not at least 0.731 and not more than 1.631 with 95% confidence interval, the time ratio is 3.258, associated with prolonged survival time (longer survival) at any time during the study per-

month rate is 225.8%. Prob >chi2 for univariate lognormal model for regular is significant (p-value = 0.0000< 0.05).

The variable hospital is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means The difference in logarithm of ratios of survival times between hospital is 0.269 , and it does not at least .175 and not more than 0.362 with 95% confidence interval, the time ratio is 1.308, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 30.8%. Prob >chi2 for univariate lognormal model for hospital is significant (p-value = 0.0000< 0.05).

The variable diabetes mellitus is significant value of Wald test (P-value=0.038 <0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between diabetes mellitus is -0.356, and it does not at least -0.693 and not more than -0.019 with 95% confidence interval, the time ratio is 0.700, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 30%. Prob>chi2 for univariate lognormal model for diabetes mellitus is significant (p-value = 0.038< 0.05).

The variable diabetes mellitus and hypertension is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between diabetes mellitus and hypertension is -0.855, and it does not at least -1.327 and not more than -0.384 with 95% confidence interval, the time ratio is 0.425, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 57.5%. Prob >chi2 for univariate lognormal model for diabetes mellitus and hypertension is significant (p-value = 0.000< 0.05).

The variable hypertension is insignificant value of Wald test (P-value=0.068 <0.05), so estimated coefficient is insignificant. That means the difference in logarithm of ratios of survival times between hypertension is 0.332, and it does not

at least -0.024 and not more than 0.689 with 95% confidence interval, the time ratio is 1.394, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 39.4%. Prob >chi2 for univariate lognormal model for hypertension is insignificant (p-value = 0.065< 0.05).

The variable other is significant value of Wald test (P-value=0.003<0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between other is 1.158, and it does not at least 0.396 and not more than 1.921with 95% confidence interval, the time ratio is 3.185, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 218.5%. Prob >chi2 for univariate lognormal model for other is significant (p-value = 0.0018< 0.05).

The variable shrunken is significant value of Wald test (P-value=0.006 <0.05), so estimated coefficient is significant. That means The difference in logarithm of ratios of survival times between shrunken is 2.028 , and it does not at least 0.572 and not more than 3.483 with 95% confidence interval, the time ratio is 7.597, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 659.7%. Prob >chi2 for univariate lognormal model for shrunken is significant (p-value = 0.001< 0.05).

The variable dialysis frequency is significant value of Wald test (P-value=0.002 <0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between dialysis frequencies is 0.588, and it does not at least 0.215 and not more than 0.962 with 95% confidence interval, the time ratio is 1.801, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 80.1%. Prob >chi2 for univariate lognormal model for dialysis frequency is significant (p-value = 0.002< 0.05)

The variable urea is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios

of survival times between urea is -0.007, and it does not at least -0.010 and not more than -0.003 with 95% confidence interval, the time ratio is 0.993, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 0.7%. Prob>chi2 for univariate lognormal model for urea is significant (p-value = 0.000< 0.05).

The variable serum creatinine is significant value of Wald test (P-value=0.000 <0.05), so estimated coefficient is significant. That means The difference in logarithm of ratios of survival times between variable Serum creatinine is -0.074 , and it does not at least -0.111 and not more than -0.036 with 95% confidence interval, the time ratio is 0.929, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 7.1%. Prob >chi2 for univariate lognormal model for serum creatinine is significant (p-value = 0.000< 0.05).

4-7-1-5. Estimate log Logistic model for univariate analysis-- accelerated failure-time form:

Table (4-54): Coefficients and time ratio

Variable	Coef.	[95% Conf. Interval]		time Ratio	[95% Conf. Interval]		z	P-value	Prob >chi2
age	-0.012	-0.021	-0.003	0.988	0.979	0.997	-2.51	0.012	0.011
regular	1.196	0.749	1.643	3.307	2.115	5.172	5.24	0.000	0.000
hospital	0.265	0.173	0.356	1.303	1.189	1.428	5.68	0.000	0.000
diabetes mellitus	-0.367	-0.691	-0.043	0.693	0.501	0.958	-2.22	0.026	0.026
diabetes mellitus and hypertension	-0.834	-1.297	-0.371	0.434	0.273	0.690	-3.53	0.000	0.004
hypertension	0.394	0.044	0.744	1.482	1.045	2.103	2.21	0.027	0.026
Shrunken kidneys	1.819	0.335	3.304	6.167	1.398	27.210	2.4	0.016	0.001
dialysis frequency per(wk)	0.630	0.283	0.976	1.877	1.328	2.654	3.56	0.000	0.000
urea	-0.007	-0.010	-0.003	0.993	0.990	0.997	-3.81	0.000	0.000
Serum creatinine	-0.073	-0.110	-0.036	0.930	0.896	0.965	-3.87	0.000	0.000
other	1.026	0.286	1.767	2.791	1.331	5.851	2.72	0.007	0.003
Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, cardiovascular disease, NSAID).									
Source: prepared by the researcher by using STATA, 2014									

We note from the table (4-54) for the variable age is significant value of Wald test (P-value = 0.017 <0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between ages is -0.012, and it does not at least -0.012and not more than -0.021 with 95% confidence interval, the time ratio is 0.988, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 1.2. Prop >chi2 for univariate Log logistic model for age is significant (p-value = 0.011< 0.05).

The variable regular is significant value of Wald test (P-value=0.000<0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between regular is 1.196, and it does not at least 0.749 and not

more than 1.643 with 95% confidence interval, the time ratio is 3.307, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 230.7%. Prob>chi2 for univariate Log logistic model for regular is significant (p-value = 0.000< 0.05).

The variable hospital is significant value of Wald test (P-value = 0.000<0.05), so estimated coefficient is significant. That means The difference in logarithm of ratios of survival times between hospital is 0 .265 , and it does not at least 0.173 and not more than 0.356 with 95% confidence interval, the time ratio is 1.303, mean associated with prolonged survival time (longer survival) at any time during the study per-month rate is 30.3%. Prob >chi2 for univariate Log logistic model for hospital is significant (p-value = 0.000< 0.05).

The variable diabetes mellitus is significant value of Wald test (P-value=0.026 <0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between diabetes mellitus is -0.367 , and it does not at least -0.691 and not more than -0.043 with 95% confidence interval, the time ratio is 0.693, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 30.7%. Prob >chi2 for univariate Log logistic model for diabetes mellitus is significant (p-value = 0.026< 0.05).

The variable diabetes mellitus and hypertension is significant value of Wald test (P-value = 0.000<0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between diabetes mellitus and hypertension is -0.834 , and it does not at least -1.297 and not more than -0.371with 95% confidence interval, the time ratio is 0.434, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 56.6%.Prob>chi2 for univariate Log logistic model for diabetes mellitus and hypertension is significant (p-value = 0.0004< 0.05).

The variable hypertension is significant value of Wald test (P-value=0.027<0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between hypertension is 0.394, and it does not at least 0.044 and not more than 0.743with 95% confidence interval, the time ratio is 1.482, mean associated with prolonged survival time (longer survival) at any time during the study per-month rate is 48.2%. Prob >chi2 for univariate Log logistic model for hypertension is significant (p-value = 0.026< 0.05).

The variable shrunken kidneys is significant value of Wald test (P-value=0.016<0.05), so estimated coefficient is significant. That means The difference in logarithm of ratios of survival times between shrunken is 1.819, and it does not at least 0.335 and not more than3.304 with 95% confidence interval, the time ratio is 6.167, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 516.7%. Prob>chi2 for univariate Log logistic model for shrunken is significant (p-value = 0.001< 0.05).

The variable dialysis frequency is significant value of Wald test (P-value = 0.000<0.05), so estimated coefficient is significant. That means The difference in logarithm of ratios of survival times between dialysis frequency is 0.630, and it does not at least 0.283 and not more than 0.976with 95% confidence interval, the time ratio is 1.877, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 87.7% . Prob >chi2 for univariate Log logistic model for dialysis frequency is significant (p-value = 0.000< 0.05).

The variable urea is significant value of Wald test (P-value = 0.000<0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between urea is -0.007, and it does not at least -0.010 and not more than -0.003 with 95% confidence interval, the time ratio is 0.993, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 0.7%. Prob >chi2 for univariate Log logistic model for urea is significant (p-value = 0.0001< 0.05).

The variable Serum creatinine is significant value of Wald test (P-value = 0.000<0.05), so estimated coefficient is significant. That means The difference in logarithm of ratios of survival times between Serum creatinine is -0.073, and it does not at least -0.110 and not more than -0.036 with 95% confidence interval, the time ratio is 0.929, mean decreases the time of death (shorter survival) at any time during the study per-month rate is 7.4%. Prob >chi2 for univariate Log logistic model for Serum creatinine is significant (p-value = 0.000< 0.05).

The variable other is significant value of Wald test (P-value=0.007<0.05), so estimated coefficient is significant. That means the difference in logarithm of ratios of survival times between other is 1.026, and it does not at least 0.286 and not more than 1.767 with 95% confidence interval, the time ratio is 2.791, associated with prolonged survival time (longer survival) at any time during the study per-month rate is 179.1%. Prob >chi2 for univariate Log logistic model for other is significant (p-value = 0.003< 0.05).

4-7-2. Estimate parametric models for multivariate analysis in hemodialysis patients:

The estimate multivariate models include all statistically significant variables, according to the log-rank test used in the study.

4-7-2-1. Estimate multivariate Exponential mode- log relative-hazard form:

Table (4-55): Chi-square test:

Model	Chi-square test	P-value
	83.11	0.0000

From table (4-55) the multivariate exponential model is significant (p-value=0.000 <0.05).

Table (4-56). Coefficients

variable	**Coefficient**	**Std. Err.**	**[95% Conf. Interval]**		**z**	**P-value**
age	0.009	0.006	-0.002	0.020	1.660	0.098
regular	-0.756	0.242	-1.231	-0.282	-3.130	0.002
hospital	-0.141	0.058	-0.254	-0.028	-2.440	0.015
diabetes mellitus	0.017	0.212	-0.398	0.433	0.080	0.934
diabetes mellitus and hypertension	0.116	0.275	-0.424	0.656	0.420	0.674
hypertension	-0.488	0.237	-0.952	-0.024	-2.060	0.039
shrunken	-1.753	1.015	-3.742	0.237	-1.730	0.084
dialysis frequency per(wk)	-0.342	0.179	-0.692	0.008	-1.910	0.056
urea	0.004	0.002	0.000	0.008	2.110	0.035
Serum creatinine	0.025	0.021	-0.015	0.066	1.220	0.224
other	-0.820	0.477	-1.754	0.115	-1.720	0.086
Intercept	-2.863	0.735	-4.304	-1.422	-3.890	0.000
Source: prepared by the researcher by using STATA, 2014						

The table (4-56) .Based on multivariate analysis it was assessed that the risk Factors including regular (Coef=-0.756, p-value=0.002 <0.05), hospital (coef=-0.141, p-value=0.015 <0.05), hypertension (coef=-0.488, p-value=0.039 <0.05), Urea (coef = 0.004, p-value=0.001 <0.035), were significant results Wald test (P-value <0.05) comparing with other variables and it will be included in the multivariate model.

Table (4-57). Hazard ratios

variable	Haz. Ratio	Std. Err.	[95% Conf. Interval]	
age	1.009	0.006	0.998	1.021
regular	0.469	0.114	0.292	0.754
hospital	0.869	0.050	0.776	0.973
diabetes mellitus	1.018	0.216	0.672	1.541
diabetes mellitus and hypertension	1.123	0.309	0.655	1.926
hypertension	0.614	0.145	0.386	0.976
shrunken	0.173	0.176	0.024	1.267
dialysis frequency per(wk)	0.710	0.127	0.501	1.008
urea	1.004	0.002	1.000	1.008
Serum creatinine	1.025	0.021	0.985	1.068
other	0.441	0.210	0.173	1.122
Source: prepared by the researcher by using STATA, 2014				

According to hazard ratio (HR) variables including age (HR=1.009), that means the hazard of age at any time in study increase by 0.9%. The 95% CI indicates that the hazard ratio could be low as 0.998 or as large as 1.021.

Diabetes mellitus (HR=1.018), that means the hazard of diabetes mellitus at any time in study increase by 1.8%. The 95% CI indicates that the hazard ratio could be low as 0.672or as large as 1.541.

diabetes mellitus and hypertension (HR=1.123), that means the hazard of diabetes mellitus and hypertension at any time in study increase by 12.3%.The 95% CI indicates that the hazard ratio could be low as 0.655 or as large as 1.926.

Urea (HR=1.004), that means the hazard of urea at any time in study increase by 0.4%.The 95% CI indicates that the hazard ratio could be low 1.000 or as large as 1.008.

serum creatinine (HR=1.025) that means the hazard of serum creatinine at any time in study increase by 2.5%.the 95% CI indicates that the hazard ratio could be low as 0.985 or as large as 1.068.were higher significantly factor in hemodialysis patients.

On the other hand it was noted that other factors such as regular (HR=0.469), that means the hazard rate of regular at any time in study decreased by 53.1%.The 95% CI indicates that the hazard ratio could be low 0.292or as large as 0.754

hospital (HR=0.973), that means the hazard rate of hospital at any time in study decreased by 2.7%.The 95% CI indicates that the hazard ratio could be low 0.776or as large as 0.973.

hypertension (HR=0.614), that means the hazard rate of hypertension at any time in study decreased by 38.6%.The 95% CI indicates that the hazard ratio could be low 0.386 or as large as 0.976

shrunken kidneys (HR=0.173), that means the hazard of shrunken kidneys at any time in study decreased by 82.7%.The 95% CI indicates that the hazard ratio could be low 0.024 or as large as 1.267.

dialysis frequency per (wk) (HR=0.710), that means the hazard of dialysis frequency per (wk) at any time in study decreased by 29%.The 95% CI indicates that the hazard ratio could be low 0.501 or as large as 1.008.

other (HR=0.441), that means the hazard rate of other at any time in study decreased by 55.9%.The 95% CI indicates that the hazard ratio could be low 0.173 or as large as 1.122, had significant lower survival rate

4-7-2-2. Estimate multivariate Weibull model-log relative-hazard form:

Table (4-58). Chi-square test:

Model	**Chi-square test**	**P-value**
	87.46	0.0000

From table (4-58) the multivariate Weibull model is significant (p-value=0.000 <0.05).

Table (4-59): Coefficients

variable	**Coefficient**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
age	0.009	0.006	1.640	0.101	-0.002	0.021
daily dialysis	-0.775	0.244	-3.170	0.002	-1.254	-0.296
hospital	-0.159	0.058	-2.750	0.006	-0.272	-0.045
diabetes mellitus	0.014	0.212	0.070	0.946	-0.402	0.431
diabetes mellitus and hypertension	0.088	0.280	0.320	0.752	-0.459	0.636
hypertension	-0.506	0.238	-2.130	0.033	-0.972	-0.040
shrunken	-1.785	1.015	-1.760	0.079	-3.775	0.204
dialysis frequency per(wk)	-0.328	0.179	-1.830	0.067	-0.678	0.023
urea	0.004	0.002	2.000	0.045	0.000	0.007
Serum creatinine	0.026	0.021	1.240	0.213	-0.015	0.068
other	-0.846	0.477	-1.770	0.076	-1.780	0.089
Intercept	-3.843	0.804	-4.780	0.000	-5.419	-2.267
/ln_p	0.224	0.063	3.580	0.000	**0.101**	0.346
p	1.251	0.078			1.106	1.414
1/p	0.800	0.050			0.707	0.904
Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, Food poisoning, cardiovascular disease, NSAID).						
Source: prepared by the researcher by using STATA, 2014						

The table (4-59) .Based on multivariate analysis it was assessed that the risk

Factors including regular (Coef=-0.775, p-value=0.002 <0.05), hospital (coef=-0.159, p-value=0.006 <0.05), hypertension (coef=-0.506, p-value=0.033 <0.05), Urea (coef = 0.004, p-value=0.045 <0.05), were significant results Wald test (P-value <0.05) comparing with other variables were insignificant. It will be incorporated into the multivariate model

Table (4-60): hazard ratios

variable	**Hazard Ratio**	**Std. Err.**	**[95% Conf. Interval]**	
age	1.009	0.006	0.998	1.021
regular	0.461	0.113	0.285	0.744
hospital	0.853	0.049	0.762	0.956
diabetes mellitus	1.014	0.215	0.669	1.538
diabetes mellitus and hypertension	1.092	0.305	0.632	1.889
hypertension	0.603	0.143	0.378	0.960
shrunken kidneys	0.168	0.170	0.023	1.227
dialysis frequency per(wk)	0.721	0.129	0.508	1.023
urea	1.004	0.002	1.000	1.008
Serum creatinine	1.027	0.022	0.985	1.071
other	0.429	0.205	0.169	1.093
Source: prepared by the researcher by using STATA, 2014				

Source: prepared by the researcher by using STATA, 2014

According to hazard ratio (HR) variables including age (HR=1.009), that means the hazard of age at any time in study increase by 0.9%. The 95% CI indicates that the hazard ratio could be low as 0.998 or as large as 1.021.

Diabetes mellitus (HR=1.014), that means the hazard of diabetes mellitus at any time in study increase by 1.4%. The 95% CI indicates that the hazard ratio could be low as 0.669or as large as 1.538.

diabetes mellitus and hypertension (HR=1.092), that means the hazard of diabetes mellitus and hypertension at any time in study increase by 9.2%.The 95% CI indicates that the hazard ratio could be low as 0.632 or as large as 1.889.

Urea (HR=1.004), that means the hazard of urea at any time in study increase by 0.4%.The 95% CI indicates that the hazard ratio could be low 1.000 or as large as 1.008.

serum creatinine (HR=1.027) that means the hazard of serum creatinine at any time in study increase by 2.7%.the 95% CI indicates that the hazard ratio could be low as 0.985 or as large as 1.071.were higher significantly factor in hemodialysis patients.

On the other hand it was noted that other factors had significant lower hazard rate such as regular (HR=0.461), that means the hazard rate of regular at any time in study decreased by 53.9%.The 95% CI indicates that the hazard ratio could be low 0.258 or as large as 0.744

hospital (HR=0.853), that means the hazard rate of hospital at any time in study decreased by 14.7%.The 95% CI indicates that the hazard ratio could be low 0.762or as large as 0.956.

hypertension (HR=0.603), that means the hazard rate of hypertension at any time in study decreased by 39.7%.The 95% CI indicates that the hazard ratio could be low 0.378 or as large as 0.960

shrunken kidneys (HR=0.168), that means the hazard of shrunken kidneys at any time in study decreased by 83.2%.The 95% CI indicates that the hazard ratio could be low 0.023 or as large as 1.227.

dialysis frequency per (wk) (HR=0.721), that means the hazard of dialysis frequency per (wk) at any time in study decreased by 27.9%.The 95% CI indicates that the hazard ratio could be low 0.508 or as large as 1.023.

other (HR=0.429), that means the hazard rate of other at any time in study decreased by 57.1%.The 95% CI indicates that the hazard ratio could be low 0.169 or as large as 1.093, had significant lower survival rate

4-7-2-3.Estimate multivariate Gompertz model -log relative-hazard form:

Table (4-61): Chi-square test:

Model	**Chi-square test**	**P-value**
	88.59	0.0000

Source: prepared by the researcher by using STATA, 2014

From table (4-61) the multivariate Gompertz model is significant (p-value=0.000 <0.05).

Table (4-62): Coefficients:

variable	**Coefficient**	**Std. Err.**	**z**	P-value	**[95% Conf. Interval]**	
age	0.010	0.006	1.760	0.079	-0.001	0.021
regular	-0.723	0.246	-2.940	0.003	-1.205	-0.242
hospital	-0.172	0.058	-2.970	0.003	-0.286	-0.059
diabetes mellitus	0.018	0.213	0.090	0.931	-0.398	0.435
diabetes mellitus and hypertension	0.093	0.280	0.330	0.741	-0.457	0.642
hypertension	-0.491	0.238	-2.060	0.039	-0.957	-0.025
other	-0.848	0.477	-1.780	0.075	-1.783	0.087
shrunken	-1.783	1.015	-1.760	0.079	-3.773	0.207
dialysis frequency per(wk)	-0.301	0.179	-1.680	0.093	-0.652	0.050
urea	0.004	0.002	2.000	0.045	0.000	0.007
Serum creatinine	0.026	0.021	1.200	0.230	-0.016	0.068
_ Intercept	-3.418	0.754	-4.530	0.000	-4.895	-1.940
/gamma	0.011	0.003	4.440	0.000	0.006	0.016
Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, Food poisoning, cardiovascular disease, NSAID).						
Source: prepared by the researcher by using STATA, 2014						

The table (4-62) .Based on multivariate analysis it was assessed that the risk Factors including regular (Coef=-0.723, p-value=0.003 <0.05), hospital (Coef=-0.172, p-value=0.003 <0.05), hypertension (Coef=-0.491, p-value=0.039 <0.05), Urea (Coef = 0.004, p-value=0.045 <0.05), were significant results Wald test (P-value <0.05) comparing with other variables were insignificant. It will be incorporated into the multivariate model.

Table (4-63): Hazard Ratios

variable	**Hazard Ratio**	**Std. Err.**	**[95% Conf. Interval]**	
age	1.010	0.006	0.999	1.022
regular	0.485	0.119	0.300	0.785
hospital	0.842	0.049	0.752	0.943
diabetes mellitus	1.018	0.217	0.671	1.545
diabetes mellitus and hypertension	1.097	0.307	0.633	1.900
hypertension	0.612	0.146	0.384	0.975
shrunken kidneys	0.168	0.171	0.023	1.230
dialysis frequency per(wk)	0.740	0.133	0.521	1.051
urea	1.004	0.002	1.000	1.008
Serum creatinine	1.026	0.022	0.984	1.070
other	0.428	0.204	0.168	1.091
Source: prepared by the researcher by using STATA, 2014				

According to hazard ratio (HR) variables including age (HR=1.010), that means the hazard of age at any time in study increase by 1%. The 95% CI indicates that the hazard ratio could be low as 0.999 or as large as 1.022.

Diabetes mellitus (HR=1.018), that means the hazard of diabetes mellitus at any time in study increase by 1.8%. The 95% CI indicates that the hazard ratio could be low as 0.671or as large as 1.545.

diabetes mellitus and hypertension (HR=1.097), that means the hazard of diabetes mellitus and hypertension at any time in study increase by 9.7%.The 95% CI indicates that the hazard ratio could be low as 0.633 or as large as 1.900.

Urea (HR=1.004), that means the hazard of urea at any time in study increase by 0.4%.The 95% CI indicates that the hazard ratio could be low 1.000 or as large as 1.008.

serum creatinine (HR=1.026) that means the hazard of serum creatinine at any time in study increase by 2.6%.the 95% CI indicates that the hazard ratio could be low as 0.984 or as large as 1.070.were higher significantly factor in hemodialysis patients.

On the other hand it was noted that other factors had significant lower hazard rate such as regular (HR=0.485), that means the hazard rate of regular at any time in study decreased by 51.5%.The 95% CI indicates that the hazard ratio could be low 0.300 or as large as 0.785

hospital (HR=0.842), that means the hazard rate of hospital at any time in study decreased by 15.8%.The 95% CI indicates that the hazard ratio could be low 0.752 or as large as 0.943.

hypertension (HR=0.612), that means the hazard rate of hypertension at any time in study decreased by 38.7%.The 95% CI indicates that the hazard ratio could be low 0.384 or as large as 0.975

shrunken kidneys (HR=0.168), that means the hazard of shrunken kidneys at any time in study decreased by 83.2%.The 95% CI indicates that the hazard ratio could be low 0.023 or as large as 1.230.

dialysis frequency per (wk) (HR=0.740), that means the hazard of dialysis frequency per (wk) at any time in study decreased by 26%.The 95% CI indicates that the hazard ratio could be low 0.521 or as large as 1.051

other (HR=0.428), that means the hazard rate of other at any time in study decreased by 57.2%.The 95% CI indicates that the hazard ratio could be low 0.168 or as large as 1.091, had significant lower survival rate .

4-7-2-4. Estimate multivariate lognormal model - accelerated failure-time form:

Table (4-64): Chi-square test:

Mode	Chi-square test	P-value
	94.76	0.0000

From table (4-64) the multivariate Lognormal model is significant (p-value=0.000 <0.05).

Table (4-65): Coefficients:

variable	Coefficient	Std. Err.	z	P-value	[95% Conf. Interval]	
age	-0.002	0.005	-0.390	0.693	-0.011	0.007
regular	0.993	0.231	4.300	0.000	0.540	1.445
hospital	0.163	0.050	3.240	0.001	0.065	0.262
diabetes mellitus	-0.068	0.199	-0.340	0.732	-0.458	0.322
diabetes mellitus and hypertension	-0.470	0.258	-1.820	0.068	-0.976	0.035
hypertension	0.263	0.206	1.280	0.201	-0.140	0.667
shrunken kidneys	1.478	0.715	2.070	0.039	0.076	2.879
dialysis frequency per(wk)	0.194	0.179	1.080	0.278	-0.156	0.545
urea	-0.004	0.002	-2.490	0.013	-0.008	-0.001
Serum creatinine	-0.041	0.019	-2.140	0.032	-0.078	-0.003
other	0.772	0.383	2.010	0.044	0.021	1.523
Intercept	2.262	0.741	3.050	0.002	0.809	3.715
/ln_sig	0.088	0.056	1.570	0.117	-0.022	0.197
sigma	1.092	0.061			0.978	1.218
Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, Food poisoning, cardiovascular disease, NSAID).						
Source: prepared by the researcher by using STATA, 2014						

Shown in table (4-65), The Wald test is significant for regular (Coef =0.993,p-value=0.000<0.05) hospital (Coef=0.163,p-value=0.001<0.05), shrunken kidneys (Coef=1.478,p-value=0.039<0.05), and urea (Coef=-0.004,p-value=0.013<0.05), Serum creatinine (Coef=-0.041, p-value=0.032<0.05), other (Coef=0.772,=p-value=0.032<0.05). As a result, the calculated coefficients for these variables are significant. It will be incorporated into the multivariate model

Table (4-66): Time Ratio:

variable	**Time Ratio**	**Std. Err.**	**[95% Conf. Interval]**	
age	0.998	0.005	0.989	1.007
regular	2.699	0.623	1.717	4.243
hospital	1.178	0.059	1.067	1.300
diabetes mellitus	0.934	0.186	0.632	1.380
diabetes mellitus and hypertension	0.625	0.161	0.377	1.036
hypertension	1.301	0.268	0.869	1.949
shrunken	4.382	3.133	1.079	17.793
dialysis frequency per(wk)	1.214	0.217	0.855	1.724
urea	0.996	0.002	0.992	0.999
Serum creatinine	0.960	0.018	0.925	0.997
other	2.164	0.829	1.021	4.586
Source: prepared by the researcher by using STATA, 2014				

We note from the table (4-66), that time ratio of regular (2.699), hospital (1.178), hypertension (1.301), shrunken kidneys (4.382), dialysis frequency per week (1.214), and other (2.164) the factors improved the survival time to the event, implying that an investigator would wait longer for the event to occur. In the other hand, with certain factors such as age (0.998), diabetes mellitus (0.934), both diabetes mellitus, hypertension (0.625), urea (0.996) and Serum creatinine (0.960) a time ratio shorter than one leads to a patient's time to death being accelerated.

4-7-2-5. Estimate multivariate Log logistic model- accelerated failure-time form:

Table (4-67): Chi-square test:

Model	Chi-square test	P-value
	93.37	0.0000

Source: prepared by the researcher by using STATA, 2014

From table (4-67) the multivariate Log logistic model is significant (P value = 0.000 <0.05).

Table (4-68): Coefficients

variable	Coefficient	Std. Err.	z	P-value	[95% Conf.Interval]	
age	-0.001	0.005	-0.310	0.755	-0.010	0.008
regular	0.975	0.238	4.100	0.000	0.509	1.441
hospital	0.159	0.050	3.170	0.002	0.061	0.258
diabetes mellitus	-0.027	0.190	-0.140	0.887	-0.399	0.345
diabetes mellitus and hypertension	-0.434	0.268	-1.620	0.105	-0.958	0.091
hypertension	0.334	0.202	1.660	0.097	-0.061	0.729
shrunken	1.335	0.689	1.940	0.053	-0.016	2.686
dialysis frequency per(wk)	0.252	0.176	1.440	0.151	-0.092	0.597
urea	-0.004	0.002	-2.340	0.019	-0.008	-0.001
Serum creatinine	-0.036	0.019	-1.940	0.053	-0.073	0.000
other	0.670	0.361	1.860	0.063	-0.038	1.378
Intercept	2.082	0.760	2.740	0.006	0.593	3.570
/ln_gam	-0.481	0.063	-7.610	0.000	-0.605	-0.357
gamma	0.618	0.039			0.546	0.700
Other= (Systemic lupus erytherematosus, tropical disease (malaria), Gout, Food poisoning, cardiovascular disease, NSAID). Uncertain=unknown reason						
Source: prepared by the researcher by using STATA, 2014						

Shown in table (4-68), The Wald test is significant for regular (Coef =0.975, p-value=0.000<0.05) hospital (Coef=0.159, p-value=0.002<0.05), and urea (Coef=-0.004, p-value=0.019<0.05), As a result, the calculated coefficients for these variables are significant. It will be incorporated into the multivariate model

Table (4-69): Time Ratio:

_t	**Time Ratio**	**Std. Err.**	**[95% Conf. Interval]**	
age	0.999	0.005	0.990	1.008
regular	2.652	0.630	1.664	4.226
hospital	1.173	0.059	1.063	1.294
diabetes mellitus	0.973	0.185	0.671	1.412
diabetes mellitus and hypertension	0.648	0.173	0.384	1.095
hypertension	1.397	0.282	0.941	2.074
other	1.955	0.706	0.963	3.968
shrunken	3.801	2.620	0.984	14.676
dialysis frequency per(wk)	1.287	0.226	0.912	1.817
urea	0.996	0.002	0.992	0.999
Serum creatinine	0.964	0.018	0.930	1.000
Source: prepared by the researcher by using STATA, 2014				

We note from the table (4-69), that time ratio of regular (2.652), hospital (1.173), hypertension (1.397), shrunken kidneys (3.801), dialysis frequency per week (1.287), and other (1.955) the factors improved the survival time to the event, implying that an investigator would wait longer for the event to occur. In the other hand, with certain factors such as age (0.999), diabetes mellitus (0.973), both diabetes mellitus, hypertension (0.648), urea (0.996) and Serum creatinine (0.964) a time ratio shorter than one leads to a patient's time to death being accelerated.

4-8.Goodness of Fit of the multivariate parametric models

Cox-Snell Residuals and Akaike Information Criterion were used to assess these models

4-8-1. Akaike information Criterion (AIC) and The Baysian Information Criteria (BIC)

Table (4-70) Akaike information Criterion (AIC)

models	Exponential	Weibull	Gompertz	Lognormal	Log logistic
AIC	671.6734	662.0275	654.8547	667.4715	669.9081
BIC	717.0793	711.2172	704.0445	716.6612	719.0978

Source: prepared by the researcher by using STATA, 2014

Seen from the table (4-70), shows Akaike information Criterion (AIC) for the different considered methods. According to this Criterion, among the desired models, a model that has the lowest AIC, is the best and the most efficient one therefore, the Gompertz model was the best fitted model for hemodialysis patients data among other parametric models. The final multivariate Gompertz proportional hazard model is then given by:

$$h(t, x, \beta) = \hat{\lambda} e^{\hat{\gamma} t} e^{\beta x}$$

$$h(t, x, \beta) = 30.51\, e^{.90.91t} e^{(-0.723\ \text{regular} + -0.172\ \text{hospital} + -0.491\ \text{hypertension} + 0.004\text{urea})}$$

In the analysis of Gompertz find Stata provides estimate Shape parameter is the reciprocal of γ (gamma $= \frac{1}{\gamma}$) rather than for γ.the estimate of gamma is 0.011.there for, the estimate for γ is $1/(0.011) = 90.91$

(KALBFLEISCH & , 2002)

The shape parameter estimate is 90.91, so the hazard exponentially increases over time hazard $(shape\ parameter > 0)$.

$\hat{\lambda} = exp(-intercept) = exp(3.418) = 30.51$ .See table (4-54)

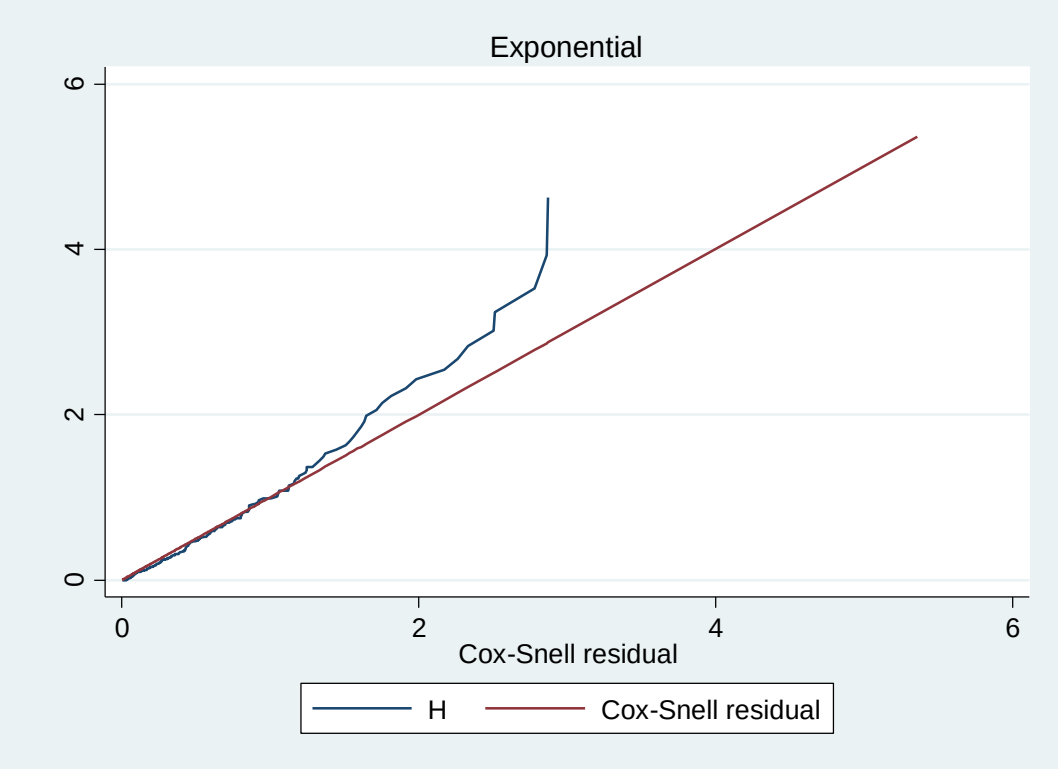
The final model is

$$h(t, x, \beta) = e^{(-0.723\ \text{regular} + -0.172\ \text{hospital} + -0.491\ \text{hypertension} + 0.004\text{urea})}\ 30.51\ e^{\cdot 90.91t}$$

4-8-2. Cox-Snell residual plot

The overall fit of the models are evaluated by using the diagnostic plot of Cox-Snell residuals. Cox-Snell residuals evaluated and compare models (Exponential, Weibull gompertz, log normal, and Log logistic). For each model, we calculated the Cox-Snell residuals, estimated their survival functions using Kaplan-Meier method and, then calculated the cumulative hazard functions for these estimations. According to Cox-Snell residuals considering that the closer the graph to the bisector the better fitted model to the data, there is some evidence of a systematic deviation from the straight line which gives us some concern about the adequacy of the fitted model.

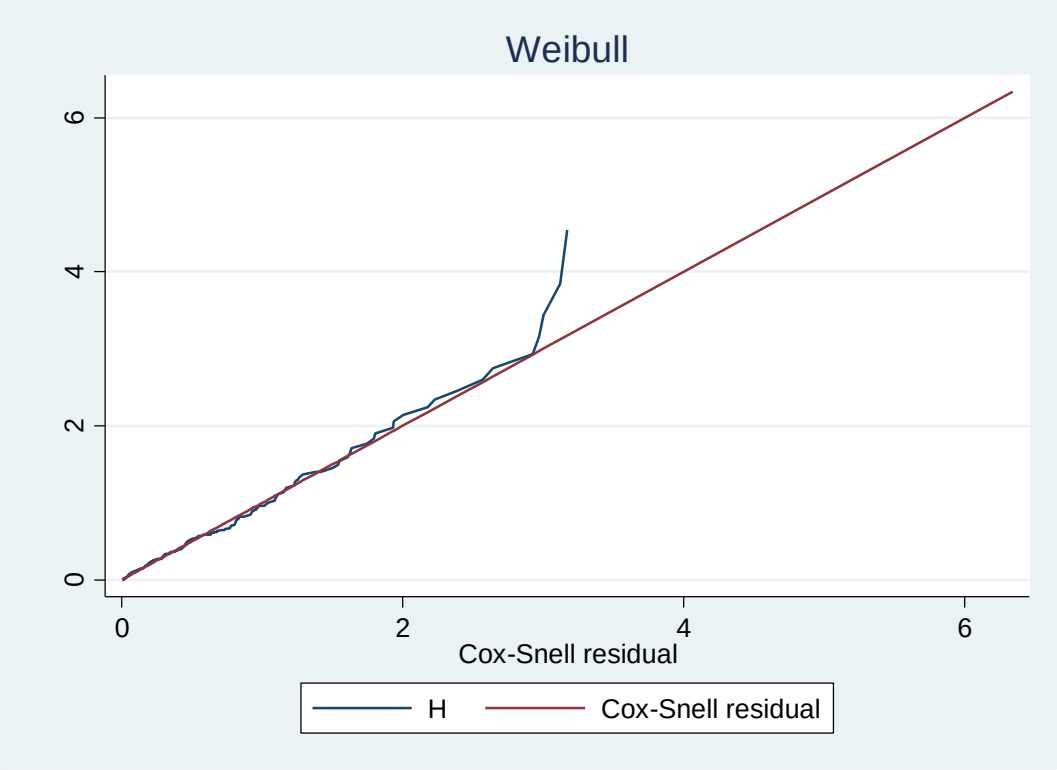
Figure (4-29) Cox -Snell residuals for Exponential



Source: prepared by the researcher by using STATA, 2014

We saw in figure (4-29) concluded that the Exponential model does not fit these data adequately.

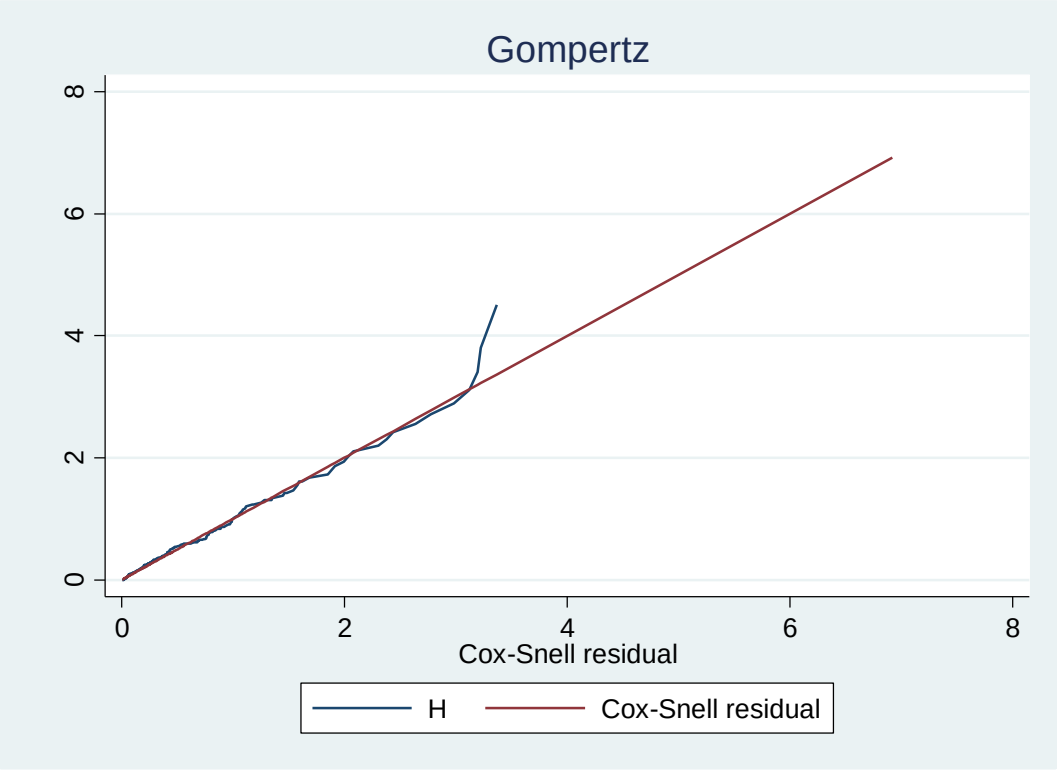
Figure (4-30). Cox -Snell residuals for Weibull



Source: prepared by the researcher by using STATA, 2014

We saw in Figure (4-30) concluded that the Weibull model does not fit these data adequately

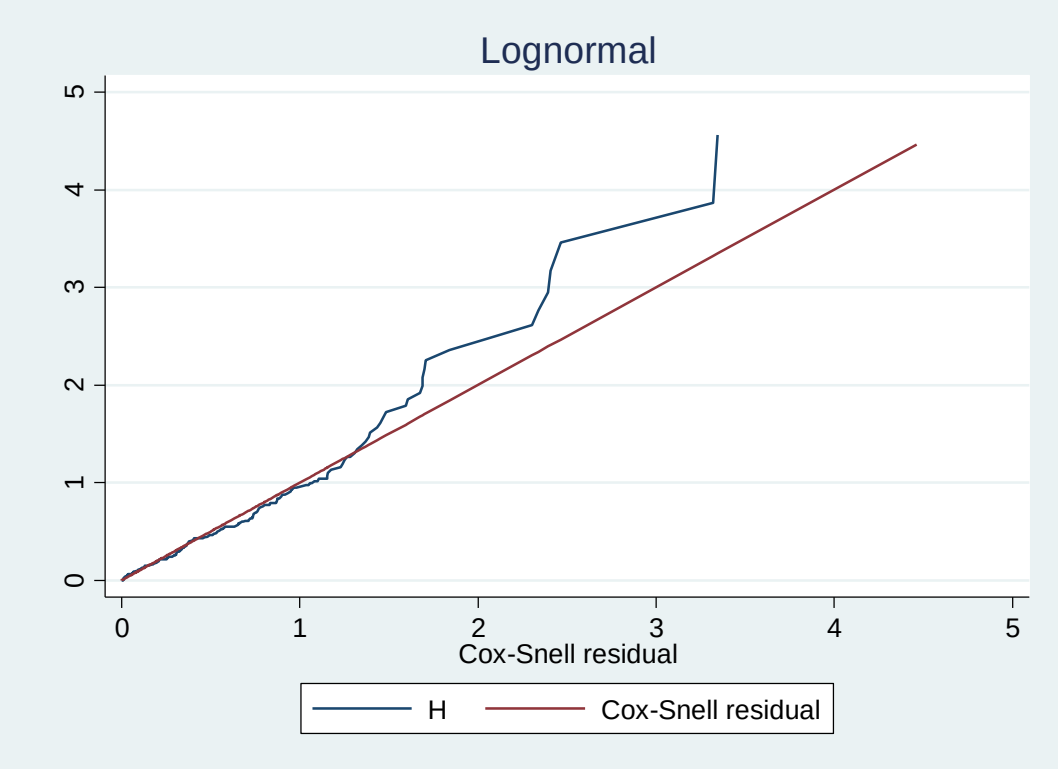
Figure (4-31) .Cox -Snell residuals for Gompertz



Source: prepared by the researcher by using STATA, 2014

We saw in Figures (4-31) the Gompertz model was the best fitted model for these data

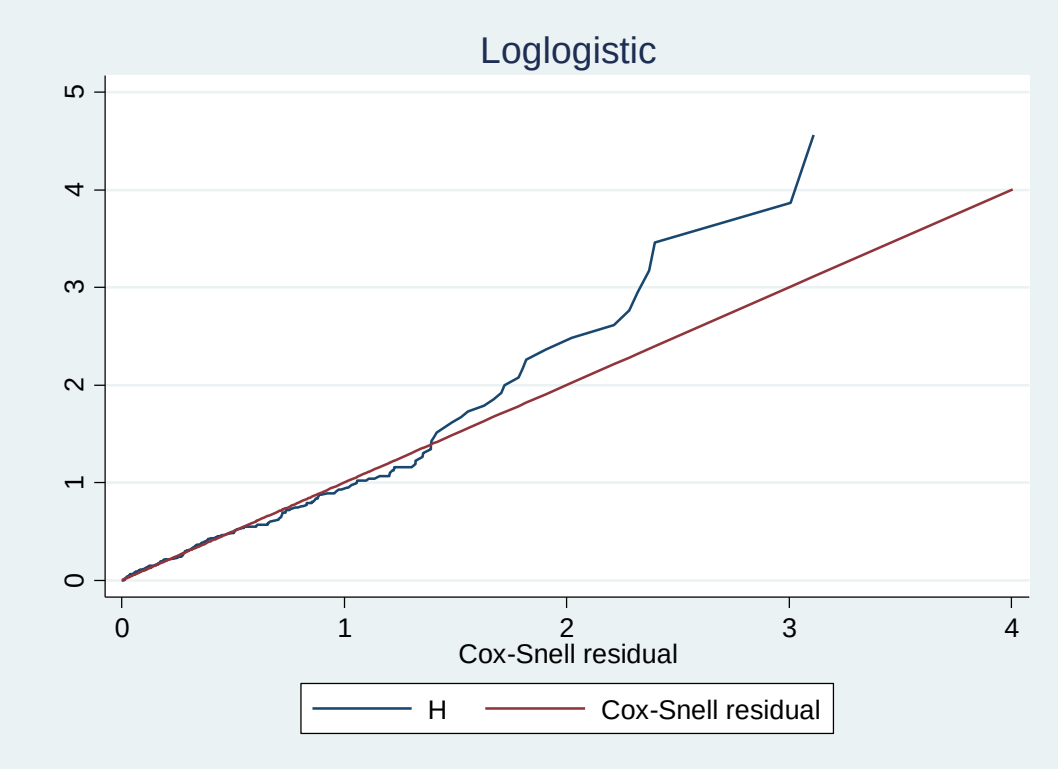
Figure (4-32) Cox -Snell residuals for Lognormal



Source: prepared by the researcher by using STATA, 2014

We saw in figure (4-32) concluded that the Lognormal AFT model does not fit these data adequately.

Figure (4-33) .Cox -Snell residuals for Log logistic



Source: prepared by the researcher by using STATA, 2014

We saw in figure (4-33) concluded that the lognormal AFT model does not fit these data adequately.

CHAPTER FIVE

Conclusions and Recommendations

5-0. Preface:

This chapter contains the results that have been reached through the practical side of the research, in addition to the proposed recommendations.

5-1. Conclusions:

1- The descriptive analysis showed that a total of 325 patients with hemodialysis were enrolled in this study. The demographic characteristics of the targeted patients showed that 59.7 % were male, 40.3 %, were female in terms of sex. By December 2015, 52.3 % of patients had died and 47.7 % were still alive, according to survival status. The marital status of the patients showed that 2.5% were divorced, 71.4% were married, 24% were single and 2.2% were widowed. Education revealed that 7.7 % of patients were illiterate, 32.6 % received basic education, 4.6 % were intermediate, 39.1 % completed secondary education and 16 % graduated. Patients' occupation wise shows that 18.8 % were employees, 13.8 % were freelancers, 41.2 % were unemployed, 3.7 % were police man, 4.3 % were retired 7.4 % were students, 11.8 % were workers .In regard to the qualitative variables such as age; the minimum age was 6years. The maximum age was 88years. The median age was 45years. Results of clinical characteristics showed that 88.9 % of patients with hemodialysis were regular and 11.1 % were irregular patients with hemodialysis, 27.4 % were diabetic mellitus and 72.6 % were not diabetic mellitus. 29.5 % had hypertension and 70.5 % had no hypertension. .89.8% had neither diabetes mellitus nor hypertension, and 10.2% had both diabetes mellitus and hypertension. 3.4 % had shrunken kidneys and 96.6 % had no shrunken kidneys. Dialysis frequency per week found that two times (8.8%) and three times (81.2%) had polycystic kidney disease and 94.8% had no polycystic kidney disease. 8.0 % had renal obstruction and 92.0 % had no renal obstruction. 9.5 % were uncertain and 90.5 % were uncertain. 5.8 % had each other, and 94.2 % had no other

2- The median overall survival time was estimated at 84 months and the trust level was found at 95% (61-89).

3- Based on the log rank test, the variables considered to be important with p-value < 0.05 were entered in the mean parametric model, while other variables were not significantly excluded from the parametric model. The variables used in the parametric model were regular, dialysis frequency per week, hospitals, diabetes mellitus , hypertension, diabetes mellitus and hypertension, shrunken kidneys, other.

4- The univariate analysis study, for five models (Exponential, Weibull, Gompertz, lognormal and log logistic) in term of hazard ratio and time ratio, all variables were significant effects but age wasn't in lognormal.

5- In univariate and multivariate analysis, According to HR and TR, the variables including age, diabetes mellitus, diabetes mellitus +hypertension, urea and serum creatinine were considered to be highly significant factors and increased the risk of death in patients (shorter survival) so that they could influence survival in hemodialysis patients in the five models used in this research. Whereas other factors, such as regular, hospital, hypertension, shrunk kidneys, dialysis frequency per week, other have decreased the risk of death (longer survival) and have a direct effect on the survival of the hemodialysis patient.

6- In Multivariate analysis for three models (Exponential, Weibull and Gompertz), that many variables were significance such as, regular, hospital, hypertension and urea but hypertension wasn't in lognormal and logistic. In addition to found the Serum creatinine and other were significant in lognormal but weren't in log logistic

7- The Gompertz model, which had the lowest AIC, BIC value, was selected as the most appropriate model. Although the AIC values of the parametric models (Exponential, Weibull, Gompertz, Lognormal and Log-logistic) were very close

to each other, so it is conclude that the Gompertz distribution is the best model for survival analysis of hemodialysis patients.

5-2. Recommendations:

The study recommended the following:

1. The possibility of using the root causes of kidney failure in Sudan should be conducted using the Cox Proportional Hazard model.
2. Estimation of Survival of kidney failure Patients, in the Presence of Prognostic Factors Using Accelerated Failure Time Model as an Alternative to Proportional Hazard Model.

3. The possibility of using the Accelerated failure time models, in the Presence of the economic, demographic and social consequences of kidney failure Patients and other chronic diseases.

4. Use analysis to keep in similar studies.

References

References

1. FA, A., AA, . S., MO, H. & A, A., 2011. The pattern clinical characteristics and outcome of ESRD in Ile-Ife Nigeria:is there a change in trend?. *Afr Health ScI,* Volume 11(4), pp. 594-601.

2. Mantel, . N., 1966. Evaluation of survival data and two new rank order statistcs arising in its consideration. *cancer Chemotherapy Report,* Volume 50, pp. 193-170.

3. Abdelaal, M. M. A. & Zakria, S. H. E. A., 2015. Modeling Survival Data by Using Cox Regression Model. *American Journal of Theoretical and Applied Statistics,* Volume Volmue 4(6), pp. 504-512.

4. Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans Automatic Control,* Volume 19, pp. 716-23.

5. Allison, P. D., 2010. Survival Analysis. In: R. O. Gregory R.Hancok, ed. *The Reviewer's Guide to Quantitative Methods in the Social Sciences.* New York: Taylor and Francis, p. 415.

6. American, N. K. f., 2002. K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification. *Am J Kidney 39(suppl 1):,* pp. S1-S000.

7. banaga, a., 2012. *Problems of renal failure in Sudan.* [Online] Available at: http://www.alrakoba.net/articles-action-show-id-17659.htm [Accessed 30 1 2015].

8. Banaga, A. . S. I. et al., 2015. Causes of end stage renal failur among haemodialysis patients in Khartoum State/Sudan. *BMC Res Notes,* 8(1), pp. 1-7.

9. Boulpaep, W. B. a. E., 2004. Medical Physiology: A CellularAr and Molecular Approach. *Saunders.*

10. C. Poinier, M., E. G. T., M. & A. H. M., 2015. *Web MD Medical Reference from Healthwise.* [Online] Available at: http://www.webmd.com/a-to-z-guides/hemodialysis-20667[Accessed 6 6 2017].

11. Cente, H. I., 2016. *The National Institute of Diabetes and Digestive and Kidney Diseases.* [Online] Available at: https://www.niddk.nih.gov/health-

information/kidney-disease/chronic-kidney-disease-ckd/what-is-chronic-kidney-disease[Accessed 28 6 2017].

12. Collett, D., 2003. *Modelling Survival Data in Medical Research.* Second Edition. London: Chapman & Hall.

13. Cox, D. & Snell, E., 1968. A General Definition of Residuals (with discussion). *Journal of the Royal Statistical Society,* Volume 30, pp. 248-275.

14. C. T., Bradburn, M. & Altman, L. S. a. D., 2003. Survival analysis.Part:1: basic concepts and first analyses. Br J Cancer. *Br J Cancer,* Volume 232-238, p. 89.

15. Day, W. K., 2017. *Chronic Kidney Disease.* [Online] [Accessed 21 6 2016].

16. DiouF, B. et al., 2003. Chronical renal failure in one Dakar Hospital Department. *Dakar Med,* Volume 48(3), p. 185–8.

17. Diouf, B. et al., 2000. Etiologies of chronic renal insufficiency in a adult internal medicine service in Dakar. *Dakar Med,* Volume 45(1), p. 62–5.

18. D. W. H. J., Lemeshow, S. & May, S., 1999. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data.* First Edition. New York: John Wiley & Sons.

19. D. W. H. J., Lemeshow, S. & May, S., 2008. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data. Seconed* Edition. New York: John Wiley & Sons.

20. Elamin, A. E., Mohammed, N. A. & Modawe, G., 2012. Aetiology of End-Stage Renal Disease among Adult Sudanese Patients. *Sudan JMS,* 7(4), pp. 255-258.

21. Elizabeth & Verneda Lights, B., 2017. *Chronic Kidney Disease: Overview, Causes, and Risk Factors.* [Online] Available at: http://www.healthline.com/health/chronickidneydisease [Accessed 6 26 2017].

22. Elsharif, M. . E. & Elsharif, . E. G., 2011. Causes of End-Stage Renal Disease in Sudan: A Single-Center Experience. *Saudi J Kidney Dis Transpl,* Volume 22(2), pp. 373-376.

23. Foundation, N. K., 2015. *YOUR KIDNEYS:Master Chemists of the Body.* [Online]
Available at: www.kidney.org

24. George, B., Seals, S. & Aban, I., 2014. Survival analysis and regression models. *J Nucl Cardiol,* Volume Volume 21(4), p. 686–694.

25. Gilbertson, D. T. et al., 2005. Projecting the Number of Patients with End-Stage Renal Disease in the United States to the Year 2015. *J Am Soc Nephrol,* Volume 16, p. 3736–3741.

26. Halle, M. P. et al., 2015. Epidemiological profile of patients with end stage renal disease in a referral hospital in Cameroon. *BMC Nephrology,* Volume 16(1), pp. 1-8.

27. Iseki, K. et al., 2004. Body mass index and the risk of development of end-stage renal disease in a screened cohort. *Kidney Int,* pp. 1870-1876.

28. Jha, V. et al., 2013. Chronic kidney disease: global dimension and perspectives. *Lancet,* p. 382: 260–72.

29. Kalbfleisch, . J. & L.Prentice, R. L., 2002. *The Statistical Analysis of Failure Time Data,*. Second Edition ed. Canada.: by John Wiley & Sons, Inc., Hoboken, New Jersey.

30. Karimi, A., Delpisheh, A. & K. S., 2016. Application of accelerated failure time models for breast cancer patients.survival in Kurdistan Province of Iran. *Cancer Research and Therapeutics,* Volume 12(3), pp. 1184-8.

31. Kleinbaum, D. G. & Klein, M., 2012. *Survival Analysis A Self- Learning Text Third Edition.* Springer New York:

32. Klein, J. P. & Moeschberger, M. L., 1997. *Survival Analysis :Techniques for Censored and Truncated Data.* New York: Springer-verlag.

33. Lawless & J. F., 2003. *Statistical Models and Methods for Lifetime Data, Second Edition.* Second Edition. New York: John Wiley & Sons.

34. Lee, E. T. & Wang, J. W., 2003. *Statistical Methods for Survival Data Analysis.* Third Edition. Canada: John Wiley & Sons, Inc., Hoboken, New Jersey.

35. Library, J. H. M. H., 2017. *End Stage Renal Disease (ESRD).* [Online] [Accessed 4 8 2017].

36. Matekole M, A. K. L. S. H. A. M. J. A. D., 1993. Hypertension and end-stage renal failure in tropical Africa. *Hum Hypertens,* Volume 7(5), p. 443–6.

37. Matthew Hoffman, M., 2016. *Picture of the Kidneys.* [Online] [Accessed 19 6 2017].

38. Moechberger M.L, K. J., 1997. *Survival Analysis :Techniques for Censored and Truncated Data Springer.* New York: s.n.

39. Naicker, S., 2009. End stage renal disease in sub- Saharan Africa. *Ethn Dis. Spring;,* Volume 19(1 Suppl 1), pp. S1-13-5.

40. Plange-Rhule, J. et al., 1999. Hypertension and renal failure in Kumasi,Ghana.. *Journal of human Hypertens,* Volume 13(1), p. 37–40.

41. Pourhoseingholi, M. A. et al., 2007. Comparing Cox Regression and Parametric Models fo rSurvival of Patients with Gastric Carcinoma. *Asian Pacific J Cancer Prev,* Volume 8, pp. 412-416.

42. Prevention, C. f. D. C. a., 2014. *National Chronic Kidney Disease Fact Sheet: General Information and National Estimates on Chronic Kidney Disease in the United States.* [Online] [Accessed 2017].

43. Publications, H. H., 2017. *Kidney Failure Guide: Causes, Symptoms and Treatment Options.* [Online] Available at:www.drugs.com/healthguide/kidneyfailure.html [Accessed 20 5 2017].

44. Ravanan, S. D. a. D. R., 2015. survival analysis of UIS patienst under parametric and non-parametric approach using R software. *International Journal of Scientific and Research Publications ,Volume 5(5),* pp. 2250-3153.

45. Saikia, R. & Barman, M. P., 2017. A Review on Accelerated Failure Time Models. *International Journal of Statistics and Systems,* Volume 12, pp. 311-322.

46. Sarra Elamin ،W .O .H .A.-A .2010 ،.Renal Replacement Therapy in Sudan,2009 .*Arab Journal of Nephrology and Transplantation* ،(2)3 ،pp. 31-36.

47. Schwarz, G. E., (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, Volume 6, pp. 461-469.

48. services, O. p. d., 2015. *Report to the Congress: Medicare Payment policy*, s.l.: s.n.

49. staff, M. C., 2017. *http://www.mayoclinic.org/*. [Online]
Available at: http://www.mayoclinic.org/diseases-conditions/chronic-kidney-disease/symptoms-causes/dxc-20207466
[Accessed 6 6 2017].

50. Subrat K. Acharya, S. P. K. V. S., 2014. Accelerated Failure Time Models:An Application in the Survival of Acute Liver Failure Patients in India. *International Journal of Science and Research (IJSR)*, 3(6), pp. 161-166.

51. Suliman SM, B. M. H. H., 1995. Dialysis and transplantation in Sudan. *Saudi J Kidney Dis Transplant*, pp. 6(3):312-4.

52. Vallinayagam, V., Parthasarathy, S. & Venkatesan, P., 2014. A Comparative Study of Life Time Models in the Analysis of Survival Data. *INDIAN JOURNAL OF APPLIED RESEARCH*, 4(1), pp. 344-374.

53. Vupputuri S, a. S. D. P., 2003. Lifestyle risk factors and chronic kidney disease. *Ann Epidemiol*, Volume 13, pp. 712-720.

54. William Aknaus, F. E. H. C. D. P. W. S. M., 1993. The clinical evaluation of new drugs for sepsis: aprospective study desigin based on survival analysis. *Jama*, Volume 1233-1241, p. 270(10).

Appendices

sex	age	Edcation_ status	occupation	martial_ status	hospital	DM	HTN	DM+ HTN	other	Plocycstic kidney	shrunken	uncertain	Renal obstructions	regular	Urea	Dialysis Frequency per(wk)	Serum eatinine	Survivl time	status
male	76	illiterate	free job	married	eb sana	yes	no	no	no	no	no	no	no	regular	155	2	11	120	dead
male	55	illiterate	free job	married	eb sana	yes	no	no	no	no	no	no	no	regular	190	2	9	60	dead
female	35	secondary	not work	single	eb sana	yes	no	no	no	no	no	no	no	regular	99	3	8	60	live
female	24	secondary	student	single	eb sana	no	no	no	no	yes	no	no	no	regular	99	2	7.5	96	live
male	65	university	employee	married	eb sana	no	no	yes	no	no	no	no	no	regular	150	2	11	72	dead
male	43	university	employee	married	eb sana	no	no	no	no	no	no	no	yes	regular	177	2	14	108	dead
male	51	secondary	worker	married	eb sana	yes	no	no	no	no	no	no	no	regular	206	2	15	12	dead
female	30	secondary	employee	single	eb sana	yes	no	no	no	no	no	no	no	regular	190	3	7	84	live
male	55	secondary	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	200	2	15	96	dead
male	32	secondary	free job	single	eb sana	no	no	no	no	no	no	no	yes	regular	99	3	4	24	live
male	40	secondary	free job	single	eb sana	no	no	yes	no	no	no	no	no	regular	130	3	6	24	live
female	36	secondary	not work	divorced	eb sana	no	no	yes	no	no	no	no	no	regular	140	3	9	72	live
female	44	secondary	employee	married	eb sana	no	no	yes	no	no	no	no	no	regular	168	3	8.5	96	live
female	64	basic	employee	married	eb sana	yes	no	no	no	no	no	no	no	regular	200	2	14	12	dead
female	25	secondary	not work	single	eb sana	no	no	no	no	no	yes	no	no	regular	200	3	5.6	48	live
female	27	university	employee	single	eb sana	no	no	yes	no	no	no	no	no	regular	156	3	6.6	60	live
female	50	basic	not work	married	eb sana	no	no	yes	no	no	no	no	no	regular	179	3	8	12	live
female	75	university	employee	divorced	eb sana	no	no	no	no	yes	no	no	no	regular	140	2	11	31	dead
male	21	secondary	worker	single	eb sana	no	no	yes	no	no	no	no	no	regular	180	2	9.8	12	dead
male	48	secondary	worker	married	eb sana	no	no	no	no	yes	no	no	no	regular	150	2	9	18	dead
male	44	basic	worker	married	eb sana	no	no	yes	no	no	no	no	no	regular	136	3	5	84	live
female	20	university	student	single	eb sana	no	no	no	no	no	yes	no	no	regular	145	2	4.3	24	live
male	34	university	free job	single	eb sana	no	yes	no	no	no	no	no	no	regular	165	2	6	12	dead
male	24	basic	free job	single	eb sana	no	no	yes	no	no	no	no	no	regular	160	2	8	14	dead
male	25	basic	free job	single	eb sana	no	no	yes	no	no	no	no	no	regular	200	3	9	13	live

male	55	university	employee	married	eb sana	no	no	no	no	yes	no	no	no	regular	220	2	15	25	dead
male	69	university	employee	married	eb sana	no	yes	no	no	no	no	no	no	regular	220	2	12	48	dead
male	44	secondary	free job	married	eb sana	yes	no	no	no	no	no	no	no	regular	189	2	9.9	72	dead
male	27	basic	employee	single	eb sana	yes	no	no	no	no	no	no	no	regular	166	3	7	108	live
male	52	secondary	employee	married	eb sana	no	no	no	no	no	no	yes	no	regular	183	3	8	24	live
female	31	secondary	not work	married	eb sana	no	no	no	no	no	no	no	yes	regular	185	2	8	24	live
female	46	illiterate	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	200	2	10	108	dead
female	27	basic	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	199	2	12	24	dead
female	69	secondary	not work	married	eb sana	no	no	no	yes	no	no	no	no	regular	179	2	10	24	dead
female	49	secondary	not work	married	eb sana	no	no	no	no	no	no	no	yes	regular	160	2	7	108	dead
male	59	secondary	not work	married	eb sana	no	no	no	no	no	no	yes	no	regular	145	2	6	36	dead
male	58	ntermediate	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	182	2	8.2	84	dead
male	36	ntermediate	free job	married	eb sana	no	no	no	no	no	no	yes	no	regular	178	2	10	22	dead
male	70	illiterate	employee	married	eb sana	no	no	no	no	no	no	no	yes	regular	193	2	9	24	dead
male	58	secondary	employee	married	eb sana	no	no	no	no	no	no	no	yes	regular	157	2	7	30	dead
female	37	basic	employee	divorced	eb sana	yes	no	no	no	no	no	no	no	regular	190	2	4	20	live
male	68	ntermediate	free job	married	eb sana	no	yes	no	no	no	no	no	no	regular	130	2	8	22	dead
male	61	university	employee	married	eb sana	no	yes	no	no	no	no	no	no	regular	145	2	6	23	dead
female	64	secondary	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	175	3	7	96	dead
female	67	ntermediate	employee	widowed	eb sana	no	yes	no	no	no	no	no	no	regular	188	2	8	11	live
male	59	basic	not work	married	eb sana	no	no	no	no	no	no	no	yes	regular	162	2	6	108	dead
female	36	basic	worker	married	eb sana	no	no	no	no	no	no	yes	no	regular	143	3	9	84	live
female	37	illiterate	worker	married	eb sana	no	no	yes	no	no	no	no	no	regular	158	2	7	84	live
male	58	illiterate	free job	married	eb sana	yes	no	no	no	no	no	no	no	regular	167	2	8	96	dead
male	43	university	employee	married	eb sana	no	no	no	no	no	no	no	yes	regular	190	3	9	12	live
female	57	ntermediate	not work	married	eb sana	no	yes	no	no	no	no	no	no	regular	168	2	15	14	dead
female	54	secondary	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	205	2	11	20	dead
male	50	basic	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	210	3	8	96	live
male	68	illiterate	worker	married	eb sana	no	yes	no	no	no	no	no	no	regular	163	3	6.4	72	live
male	64	basic	not work	married	eb sana	no	no	no	yes	no	no	no	no	regular	182	2	7.5	108	dead

male	80	university	employee	married	eb sana	no	no	yes	no	no	no	no	no	regular	166	3	7	36	live
male	60	basic	worker	married	eb sana	no	no	yes	no	no	no	no	no	regular	145	2	10	72	dead
male	45	secondary	employee	single	eb sana	no	no	no	no	yes	no	no	no	regular	163	3	6	120	live
female	40	secondary	free job	married	eb sana	no	no	no	no	no	no	no	yes	regular	130	3	5.5	24	live
male	37	secondary	free job	single	eb sana	no	no	yes	no	no	no	no	no	regular	140	3	6	22	live
male	68	secondary	employee	married	eb sana	no	no	yes	no	no	no	no	no	regular	172	2	12	72	dead
female	42	secondary	employee	married	eb sana	yes	no	no	no	no	no	no	no	regular	190	3	7	72	live
female	60	illiterate	not work	married	eb sana	no	no	no	no	no	no	no	yes	regular	200	2	12	84	dead
male	62	secondary	not work	married	eb sana	no	no	yes	no	no	no	no	no	regular	220	2	11	60	dead
male	38	university	free job	married	eb sana	no	no	no	no	no	no	no	yes	regular	163	3	6	48	live
female	47	university	employee	married	eb sana	no	no	yes	no	no	no	no	no	regular	96	3	7.9	60	live
female	37	secondary	employee	single	eb sana	yes	no	no	no	no	no	no	no	regular	159	2	14	44	dead
female	56	illiterate	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	180	2	3	55	dead
female	76	basic	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	200	2	11	50	dead
female	30	university	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	99	3	6.6	44	live
male	83	basic	not work	married	eb sana	yes	no	no	no	no	no	no	no	regular	119	3	5.8	48	dead
female	78	university	employee	widowed	eb sana	yes	no	no	no	no	no	no	no	regular	145	3	8	60	dead
male	60	illiterate	worker	married	eb sana	yes	no	no	no	no	no	no	no	regular	177	3	15	12	dead
female	65	illiterate	not work	married	eb sana	no	yes	no	no	no	no	no	no	regular	180	3	14	60	dead
male	55	illiterate	worker	married	salma	no	yes	no	no	no	no	no	no	regular	140	3	18	36	dead
female	65	basic	not work	married	salma	no	no	no	no	yes	no	no	no	regular	170	3	8	60	dead
male	65	illiterate	worker	married	salma	no	no	no	no	yes	no	no	no	regular	136	3	4	96	dead
female	36	secondary	not work	married	salma	no	no	no	yes	no	no	no	no	regular	140	3	7.5	120	live
male	40	university	not work	married	salma	no	no	no	no	yes	no	no	no	regular	99	3	7.5	120	live
male	33	university	not work	married	salma	no	no	no	yes	no	no	no	no	regular	95	3	5.8	108	live
female	52	secondary	not work	married	salma	yes	no	no	no	no	no	no	no	regular	142	3	9.8	60	dead
female	70	illiterate	not work	married	salma	no	no	no	no	yes	no	no	no	regular	111	3	10	60	dead
male	17	secondary	not work	married	salma	no	no	no	no	no	no	yes	no	regular	71	3	5.8	120	live
male	32	basic	not work	married	salma	no	no	yes	no	no	no	no	no	regular	118	3	6.6	96	live
male	14	basic	not work	single	salma	yes	no	no	no	no	no	no	no	regular	63	3	3.9	72	live

male	16	secondary	not work	single	salma	no	no	yes	no	no	no	no	no	regular	41	3	8	60	live
female	14	basic	student	single	salma	no	no	no	no	no	yes	no	no	regular	41	3	5.3	48	live
male	43	basic	employee	married	salma	no	no	yes	no	no	no	no	no	regular	97	3	6.6	12	live
male	23	secondary	student	single	salma	no	no	no	yes	no	no	no	no	regular	101	3	7	96	live
female	52	secondary	not work	married	salma	no	no	no	no	no	no	yes	no	regular	142	3	9.8	12	dead
male	14	basic	not work	single	salma	no	no	no	no	no	no	yes	no	regular	63	3	3.9	120	live
female	70	illiterate	not work	married	salma	yes	no	no	no	no	no	no	no	regular	111	3	10	60	dead
female	65	illiterate	not work	married	salma	no	yes	no	no	no	no	no	no	regular	156	3	15	4	dead
male	32	secondary	not work	single	salma	no	no	no	no	no	yes	no	no	regular	53	3	8	60	live
female	29	university	not work	single	salma	no	no	no	no	no	no	yes	no	regular	95	3	11	108	live
female	40	secondary	not work	married	salma	no	no	yes	no	no	no	no	no	regular	210	3	13	12	dead
male	60	ntermediate	not work	married	salma	yes	no	no	no	no	no	no	no	regular	199	3	9.9	120	dead
male	50	university	employee	married	salma	no	no	no	no	yes	no	no	no	regular	150	3	10	120	dead
male	72	basic	not work	married	salma	no	no	no	no	no	no	no	yes	regular	180	3	18	96	dead
female	30	secondary	not work	single	salma	yes	no	no	no	no	no	no	no	regular	177	3	10	12	dead
female	45	secondary	not work	married	salma	no	no	no	no	no	yes	no	no	regular	95	3	6.6	60	live
male	50	ntermediate	free job	married	salma	no	no	yes	no	no	no	no	no	regular	119	3	10	10	dead
female	32	secondary	not work	married	salma	no	no	no	yes	no	no	no	no	regular	147	3	12	60	live
male	34	secondary	not work	single	salma	yes	no	no	no	no	no	no	no	regular	155	3	12	120	dead
male	66	university	not work	married	salma	yes	no	no	no	no	no	no	no	regular	135	3	11	84	dead
male	50	ntermediate	free job	married	salma	yes	no	no	no	no	no	no	no	regular	141	3	11	108	dead
male	42	ntermediate	not work	married	salma	no	no	no	yes	no	no	no	no	regular	63	3	8	108	live
female	24	secondary	not work	single	salma	no	no	no	yes	no	no	no	no	regular	170	3	10	120	live
female	30	university	not work	married	salma	no	no	yes	no	no	no	no	no	regular	103	3	4.7	120	live
female	71	secondary	not work	widowed	salma	no	no	no	no	no	no	no	yes	regular	111	3	3	84	dead
female	77	basic	not work	married	salma	no	no	yes	no	no	no	no	no	regular	137	3	8.5	12	dead
male	55	university	employee	married	salma	no	no	no	no	no	no	no	yes	regular	155	3	12	48	dead
female	40	basic	not work	married	salma	yes	no	no	no	no	no	no	no	regular	189	3	6.5	120	live
female	60	basic	not work	married	salma	no	no	yes	no	no	no	no	no	regular	86	3	9.9	60	live
female	39	secondary	not work	married	salma	no	no	yes	no	no	no	no	no	regular	113	3	8.5	120	live

male	38	basic	free job	married	ahmed gasm	no	yes	no	no	no	no	no	no	regular	115	3	8	60	live
male	47	secondary	retirement	married	ahmed gasm	no	no	yes	no	no	no	no	no	regular	119	3	4.2	60	live
male	35	basic	worker	married	ahmed gasm	no	no	yes	no	no	no	no	no	regular	82	3	10	62	live
male	35	basic	worker	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	250	3	10	4	dead
male	53	basic	worker	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	250	3	3.5	16	dead
male	53	university	employee	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	190	3	8	15	dead
female	65	basic	not work	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	321	3	14	75	dead
female	60	basic	not work	married	ahmed gasm	no	no	no	yes	no	no	no	no	regular	84	3	11	60	live
male	24	basic	not work	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	91	3	6	60	live
female	80	basic	not work	married	ahmed gasm	no	yes	no	no	no	no	no	no	regular	186	3	10	37	dead
male	32	secondary	not work	married	ahmed gasm	no	no	no	no	no	no	yes	no	regular	150	3	8	60	live
female	35	basic	not work	married	ahmed gasm	no	no	no	no	no	no	no	yes	regular	53	3	7.5	72	live
female	57	basic	not work	widowed	ahmed gasm	no	yes	no	no	no	no	no	no	regular	184	3	5	60	dead
male	45	basic	worker	married	ahmed gasm	no	no	no	no	yes	no	no	no	regular	122	3	6.2	96	live
female	36	secondary	not work	divorced	ahmed gasm	no	no	no	no	no	no	no	yes	regular	108	3	5	60	live
male	20	secondary	student	single	ahmed gasm	yes	no	no	no	no	no	no	no	regular	111	3	6	36	dead
male	50	secondary	worker	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	91	3	9.5	24	live
male	64	university	free job	married	ahmed gasm	no	yes	no	no	no	no	no	no	regular	217	3	15	60	dead
female	45	secondary	not work	married	ahmed gasm	no	no	no	no	yes	no	no	no	regular	200	3	12	60	dead
female	42	illiterate	not work	divorced	ahmed gasm	yes	no	no	no	no	no	no	no	regular	250	3	9	24	dead
male	34	secondary	free job	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	88	3	3	12	live
female	28	university	student	widowed	ahmed gasm	yes	no	no	no	no	no	no	no	regular	89	3	8.7	12	live
male	17	secondary	student	single	ahmed gasm	no	no	no	no	yes	no	no	no	regular	105	3	6	24	dead
female	30	university	student	single	ahmed gasm	no	no	no	no	yes	no	no	no	regular	150	3	7	24	dead
male	29	secondary	free job	single	ahmed gasm	no	yes	no	no	no	no	no	no	regular	113	3	6	12	dead

female	35	basic	not work	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	210	3	9.2	12	dead
male	39	secondary	worker	single	ahmed gasm	no	no	no	no	yes	no	no	no	regular	202	3	12	34	dead
male	43	secondary	free job	married	ahmed gasm	no	no	no	no	no	no	no	yes	regular	150	3	2.9	22	dead
female	39	university	employee	married	ahmed gasm	no	no	no	no	no	no	no	yes	regular	130	3	5	60	dead
male	40	basic	not work	divorced	ahmed gasm	no	no	no	no	no	no	no	yes	regular	113	3	7	24	dead
male	37	secondary	free job	married	ahmed gasm	no	no	no	no	no	no	no	no	regular	102	3	8	84	live
male	58	university	employee	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	140	3	5	60	dead
female	48	basic	not work	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	161	3	7	12	dead
male	34	basic	free job	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	102	3	6	24	dead
female	48	basic	free job	married	ahmed gasm	no	no	no	yes	no	no	no	no	regular	156	3	3.2	36	dead
female	55	basic	worker	single	ahmed gasm	yes	no	no	no	no	no	no	no	regular	239	3	15	48	dead
female	42	secondary	not work	married	ahmed gasm	yes	no	no	no	no	no	no	no	regular	235	3	10	12	dead
female	18	secondary	student	single	ahmed gasm	no	no	no	no	no	no	yes	no	regular	200	3	8	72	live
female	44	university	employee	married	ahmed gasm	no	no	yes	no	no	no	no	no	regular	180	3	8	60	live
female	45	basic	not work	married	ahmed gasm	no	yes	no	no	no	no	no	no	regular	206	3	11	60	dead
male	61	secondary	free job	married	ahmed gasm	no	yes	no	no	no	no	no	no	regular	123	3	4	108	dead
female	48	basic	worker	married	ahmed gasm	no	no	yes	no	no	no	no	no	regular	80	3	3.5	120	live
male	19	secondary	free job	single	ahmed gasm	no	no	yes	no	no	no	no	no	regular	180	3	8	60	live
male	35	secondary	employee	single	ahmed gasm	yes	no	no	no	no	no	no	no	regular	215	3	7	60	live
female	17	secondary	student	single	ahmed gasm	no	no	no	yes	no	no	no	no	regular	113	3	6	60	live
male	31	secondary	free job	single	ahmed gasm	no	no	yes	no	no	no	no	no	regular	125	3	7	120	live
male	55	university	employee	married	ahmed gasm	no	no	yes	no	no	no	no	no	regular	385	3	25	84	dead
female	28	university	student	single	ahmed gasm	yes	no	no	no	no	no	no	no	regular	111	3	6.3	72	live
female	15	basic	student	single	ahmed gasm	no	no	no	no	no	yes	no	no	regular	111	3	6.3	72	live
male	28	secondary	free job	single	ahmed gasm	no	no	yes	no	no	no	no	no	regular	131	3	6	96	live
male	38	illiterate	free job	married	omderman	no	no	no	no	no	no	yes	no	regular	214	3	8	108	live

female	40	basic	free job	widowed	omderman	yes	no	no	no	no	no	no	no	regular	120	3	14	96	dead
male	36	basic	free job	married	omderman	no	no	no	no	no	no	no	no	regular	190	3	5	60	live
male	60	university	employee	married	omderman	no	no	no	no	no	yes	no	no	regular	76	3	6	108	live
female	28	secondary	student	single	omderman	yes	no	no	no	no	no	no	no	regular	200	3	7	84	dead
male	56	illiterate	not work	married	omderman	yes	no	no	no	no	no	no	no	regular	250	3	23	108	dead
male	80	basic	not work	married	omderman	yes	no	no	no	no	no	no	no	regular	144	3	17	12	dead
female	52	basic	not work	married	omderman	yes	no	no	no	no	no	no	no	regular	150	3	5	24	dead
male	53	basic	not work	married	omderman	no	no	no	no	no	no	yes	no	regular	130	3	5	22	dead
male	33	secondary	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	188	3	8	30	live
female	53	ntermediate	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	120	3	7	60	live
female	42	secondary	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	145	3	4	60	live
male	29	university	employee	single	omderman	yes	no	no	no	no	no	no	no	regular	150	3	6	72	live
male	71	secondary	not work	married	omderman	no	no	no	no	no	no	yes	no	regular	199	3	6	2	live
female	70	illiterate	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	230	3	7	84	dead
male	18	basic	not work	single	omderman	no	no	no	no	yes	no	no	no	regular	139	3	8	84	live
female	18	basic	not work	single	omderman	yes	no	no	no	no	no	no	no	regular	205	3	9	59	live
female	13	basic	not work	single	omderman	no	no	no	no	no	yes	no	no	regular	130	3	6	75	live
male	18	basic	not work	single	omderman	no	no	no	no	no	no	yes	no	regular	190	3	6	84	live
male	84	secondary	student	single	omderman	no	yes	no	no	no	no	no	no	regular	189	3	3	60	dead
male	60	secondary	student	single	omderman	yes	no	no	no	no	no	no	no	regular	156	3	6	61	dead
female	56	basic	student	single	omderman	yes	no	no	no	no	no	no	no	regular	166	3	6	64	live
male	38	secondary	student	single	omderman	no	no	no	no	no	no	yes	no	regular	136	3	3	71	live
male	27	secondary	free job	married	omderman	no	no	yes	no	no	no	no	no	regular	147	3	8	72	live
male	34	basic	worker	married	omderman	yes	no	no	no	no	no	no	no	regular	170	3	7	96	live
male	48	secondary	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	190	3	8	60	live
male	42	basic	free job	married	omderman	no	no	yes	no	no	no	no	no	regular	189	3	8	108	live
female	53	secondary	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	199	3	11	84	dead
male	66	secondary	retirement	married	omderman	no	no	no	no	no	no	yes	no	regular	185	3	7	96	dead
female	50	university	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	156	3	5	84	dead
male	64	basic	free job	married	omderman	no	no	yes	no	no	no	no	no	regular	168	3	6	108	dead

female	38	university	employee	married	omderman	no	no	no	yes	no	no	no	no	regular	119	3	5	84	live
male	39	secondary	free job	single	omderman	yes	no	no	no	no	no	no	no	regular	102	3	4	60	live
male	33	secondary	not work	single	omderman	no	no	yes	no	no	no	no	no	regular	164	3	9	96	live
male	67	basic	worker	divorced	omderman	yes	no	no	no	no	no	no	no	regular	164	3	12	108	dead
female	63	secondary	worker	married	omderman	no	no	yes	no	no	no	no	no	regular	176	3	11	60	dead
male	57	basic	worker	married	omderman	no	no	yes	no	no	no	no	no	regular	197	3	11	60	dead
male	36	basic	not work	married	omderman	no	no	no	yes	no	no	no	no	regular	111	3	4	60	live
male	57	secondary	employee	married	omderman	no	no	yes	no	no	no	no	no	regular	189	3	9	24	dead
female	31	ntermediate	not work	married	omderman	no	yes	no	no	no	no	no	no	regular	65	3	10	22	live
male	44	ntermediate	employee	married	omderman	no	no	no	no	no	no	yes	no	regular	175	3	7	24	live
male	65	basic	employee	single	omderman	no	no	no	no	no	no	no	yes	regular	189	3	10	60	live
male	80	secondary	retirement	divorced	omderman	yes	no	no	no	no	no	no	no	regular	250	3	10	24	live
male	50	basic	employee	married	omderman	no	no	yes	no	no	no	no	no	regular	150	3	11	27	live
male	6	university	employee	married	omderman	no	no	yes	no	no	no	no	no	regular	159	3	5	67	live
female	7	basic	employee	married	omderman	no	no	yes	no	no	no	no	no	regular	125	3	7	24	dead
male	14	basic	worker	married	omderman	no	no	no	no	yes	no	no	no	regular	125	3	9	72	dead
male	7	basic	worker	married	omderman	no	no	no	no	no	no	yes	no	regular	161	3	5	46	dead
female	31	basic	not work	married	omderman	no	no	no	yes	no	no	no	no	regular	150	3	8	65	live
male	40	university	not work	married	omderman	no	no	no	no	no	yes	no	no	regular	65	3	8	33	live
male	47	basic	worker	married	omderman	no	no	yes	no	no	no	no	no	regular	95	3	10	63	live
female	59	secondary	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	130	3	9	60	dead
male	58	secondary	employee	single	omderman	no	no	yes	no	no	no	no	no	regular	119	3	6	17	live
male	57	secondary	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	112	3	9	48	live
male	65	ntermediate	not work	married	omderman	no	no	no	yes	no	no	no	no	regular	215	3	8	76	dead
female	47	basic	not work	single	omderman	yes	no	no	no	no	no	no	no	regular	90	3	9	26	live
male	56	basic	not work	married	omderman	no	no	no	yes	no	no	no	no	regular	250	3	18	12	live
male	52	university	employee	single	omderman	yes	no	no	no	no	no	no	no	regular	90	3	10	25	live
female	29	secondary	not work	married	omderman	no	no	yes	no	no	no	no	no	regular	115	3	7	108	dead
male	49	secondary	free job	married	omderman	yes	no	no	no	no	no	no	no	regular	114	3	17	24	dead
male	53	secondary	not work	married	omderman	yes	no	no	no	no	no	no	no	regular	209	3	16	36	dead

female	60	basic	worker	married	omderman	no	no	no	yes	no	no	no	no	regular	161	3	7	76	dead
female	35	basic	not work	widowed	bahri	no	no	yes	no	no	no	no	no	regular	50	3	14	60	live
male	50	illiterate	worker	married	bahri	no	no	no	no	no	no	yes	no	regular	65	3	9.2	84	live
male	40	basic	worker	married	bahri	no	no	yes	no	no	no	no	no	regular	80	3	7	6	live
female	48	basic	free job	married	bahri	no	no	yes	no	no	no	no	no	regular	160	3	14	12	dead
male	53	secondary	employee	married	bahri	yes	no	no	no	no	no	no	no	regular	370	3	10	2	dead
male	46	secondary	free job	married	bahri	no	no	no	no	no	no	yes	no	regular	177	3	11	24	dead
male	55	secondary	retirement	married	bahri	yes	no	no	no	no	no	no	no	regular	111	3	5.4	6	dead
male	65	basic	worker	married	bahri	no	no	no	no	no	no	no	yes	regular	96	3	8.5	2	live
female	48	illiterate	not work	married	bahri	no	no	yes	no	no	no	no	no	regular	106	3	7	1	live
male	49	university	employee	married	bahri	no	yes	no	no	no	no	no	no	regular	153	3	9.3	17	dead
female	70	illiterate	not work	married	bahri	no	no	no	no	no	no	no	yes	not regular	90	3	14	1	live
male	16	secondary	student	single	bahri	no	no	no	no	no	no	yes	no	not regular	105	3	9.9	13	dead
male	47	illiterate	free job	single	bahri	yes	no	no	no	no	no	no	no	not regular	115	3	4	48	dead
female	70	illiterate	not work	married	bahri	no	no	no	no	no	no	yes	no	not regular	80	3	14	1	live
female	57	university	retirement	married	bahri	no	yes	no	no	no	no	no	no	not regular	150	3	10	89	dead
male	43	secondary	free job	married	bahri	no	no	yes	no	no	no	no	no	not regular	68	3	12	12	live
male	38	basic	free job	married	bahri	no	no	no	no	no	no	no	no	regular	100	3	15	5	dead
male	35	basic	free job	married	bahri	no	no	yes	no	no	no	no	no	single	147	3	9.3	6	dead
female	60	secondary	free job	single	bahri	no	no	yes	no	no	no	no	no	regular	197	3	13	23	dead
male	65	basic	retirement	married	bahri	no	no	yes	no	no	no	no	no	regular	130	3	5.6	12	dead
female	43	secondary	employee	married	bahri	no	yes	no	no	no	no	no	no	regular	190	3	8	16	dead
male	20	basic	worker	single	bahri	no	yes	no	no	no	no	no	no	regular	73	3	6	1	live
female	68	basic	not work	single	bahri	no	yes	no	no	no	no	no	no	regular	140	3	4	6	dead
female	23	university	student	married	bahri	no	no	yes	no	no	no	no	no	not regular	150	3	12	3	dead
male	58	secondary	retirement	single	bahri	no	no	yes	no	no	no	no	no	not regular	111	3	12	1	live
male	32	secondary	employee	married	bahri	no	yes	no	no	no	no	no	no	regular	170	3	15	6	dead
female	40	basic	not work	single	bahri	yes	no	no	no	no	no	no	no	regular	170	3	4.8	6	dead
male	25	university	employee	married	bahri	no	no	no	no	no	no	yes	no	regular	119	3	8.2	7	dead

male	25	basic	worker	married	bahri	no	yes	no	no	no	no	no	no	not regular	160	3	7.5	3	dead
male	68	secondary	employee	married	bahri	no	yes	no	no	no	no	no	no	not regular	118	3	2.7	40	dead
female	40	secondary	not work	married	bahri	yes	no	no	no	no	no	no	no	not regular	180	3	5.1	26	dead
female	57	basic	not work	married	bahri	yes	no	no	no	no	no	no	no	not regular	176	3	2.6	26	dead
male	14	basic	student	married	bahri	yes	no	no	no	no	no	no	no	not regular	160	3	7.3	5	dead
female	39	secondary	employee	married	bahri	yes	no	no	no	no	no	no	no	not regular	188	3	2	24	dead
male	88	basic	retirement	married	bahri	no	no	yes	no	no	no	no	no	not regular	99	3	8.8	14	live
female	50	secondary	not work	married	bahri	no	yes	no	no	no	no	no	no	regular	138	3	7.8	7	dead
male	22	basic	student	married	bahri	no	no	no	no	no	no	no	no	regular	91	3	8.8	3	live
male	67	university	retirement	single	bahri	yes	no	no	no	no	no	no	no	regular	100	3	13	10	dead
female	50	basic	not work	married	bahri	yes	no	no	no	no	no	no	no	regular	89	3	10	1	live
male	17	basic	student	married	bahri	no	no	no	no	no	no	no	no	regular	170	3	14	12	dead
male	27	university	student	single	bahri	no	no	no	no	no	no	no	no	regular	102	3	10	12	dead
male	57	university	retirement	single	bahri	no	no	yes	no	no	no	no	no	not regular	100	3	11	2	dead
male	57	basic	retirement	married	bahri	yes	no	no	no	no	no	no	no	not regular	112	3	7	108	dead
female	40	university	not work	married	bahri	no	no	yes	no	no	no	no	no	not regular	163	3	4.8	6	dead
female	50	secondary	not work	married	bahri	yes	no	no	no	no	no	no	no	regular	76	3	4	20	live
female	63	basic	not work	married	bahri	yes	no	no	no	no	no	no	no	not regular	170	3	3	21	dead
female	23	secondary	student	married	bahri	no	no	yes	no	no	no	no	no	not regular	144	3	3	3	dead
male	30	university	employee	single	bahri	no	no	yes	no	no	no	no	no	not regular	130	3	4	14	dead
male	32	secondary	employee	single	bahri	no	no	no	no	no	no	no	yes	not regular	122	3	5	24	dead
male	27	basic	worker	single	bahri	no	no	no	no	no	no	no	no	not regular	88	3	5	36	live
male	42	university	employee	single	bahri	no	no	yes	no	no	no	no	no	not regular	68	3	7.7	58	live
male	49	university	employee	married	bahri	no	no	yes	no	no	no	no	no	not regular	160	3	13	96	dead
male	43	secondary	free job	married	bahri	no	no	yes	no	no	no	no	no	not regular	98	3	8.9	16	live
female	75	secondary	not work	married	bahri	no	no	yes	no	no	no	no	no	not regular	170	3	12	12	dead

female	65	secondary	not work	married	bahri	yes	no	no	no	no	no	no	no	not regular	130	2	19	26	dead
male	24	basic	worker	married	bahri	no	no	no	no	no	no	no	no	not regular	130	2	25	11	dead
male	15	secondary	employee	single	bahri	yes	no	no	no	no	no	no	no	regular	192	2	8.8	10	dead
female	40	basic	not work	married	bahri	no	no	yes	no	no	no	no	no	regular	68	3	8.8	25	live
male	32	secondary	worker	married	bahri	no	no	yes	no	no	no	no	no	regular	87	3	9.9	12	live
female	28	secondary	not work	single	bahri	no	no	yes	no	no	no	no	no	regular	175	3	6	1	live
male	67	secondary	retirement	single	bahri	no	no	no	no	no	no	yes	no	not regular	130	2	14	8	dead
male	55	basic	retirement	married	bahri	no	yes	no	no	no	no	no	no	not regular	140	2	12	120	dead
male	53	secondary	employee	married	salma	no	no	no	no	no	no	no	yes	regular	71	3	5.5	24	live
female	65	secondary	not work	married	bahri	no	no	yes	no	no	no	no	no	not regular	120	2	11	96	dead
male	66	university	not work	married	salma	no	no	no	no	no	yes	no	no	regular	160	2	13	120	dead
female	68	basic	not work	single	bahri	no	yes	no	no	no	no	no	no	regular	112	2	13	72	dead
male	43	secondary	free job	married	bahri	no	no	yes	no	no	no	no	no	not regular	98	3	8.9	16	live
female	75	secondary	not work	married	bahri	no	no	yes	no	no	no	no	no	not regular	170	2	7.7	12	dead
female	65	secondary	not work	married	bahri	yes	no	no	no	no	no	no	no	not regular	130	2	19	26	dead
male	24	basic	worker	married	bahri	no	no	no	no	no	no	no	yes	not regular	130	2	25	11	dead
male	32	secondary	employee	married	salma	no	no	yes	no	no	no	no	no	regular	118	3	4	96	live
male	73	basic	retirement	married	bahri	no	no	yes	no	no	no	no	no	regular	180	2	15	84	dead
male	32	secondary	employee	married	bahri	no	yes	no	no	no	no	no	no	regular	170	2	15	6	dead
male	84	secondary	not work	married	ribat	no	yes	no	no	no	no	no	no	regular	189	2	20	60	dead
male	60	secondary	employee	married	ribat	no	no	no	yes	no	no	no	no	regular	156	3	6	24	live
female	56	basic	not work	married	ribat	yes	no	no	no	no	no	no	no	regular	166	3	13	22	live
male	38	secondary	police man	married	ribat	no	no	no	no	no	no	no	yes	regular	136	3	5	24	live
male	27	secondary	police man	single	ribat	no	no	yes	no	no	no	no	no	regular	147	3	6	60	live
male	45	basic	police man	married	ribat	no	no	yes	no	no	no	no	no	regular	190	3	13	72	live
male	34	secondary	police man	married	ribat	no	no	yes	no	no	no	no	no	regular	170	3	7	27	live
male	48	basic	police man	married	ribat	no	no	yes	no	no	no	no	no	regular	190	3	13	67	live
female	42	secondary	police man	married	ribat	no	no	yes	no	no	no	no	no	regular	189	3	12	24	live

male	53	secondary	police man	married	ribat	yes	no	no	no	no	no	no	no	regular	199	3	14	72	live
male	66	university	police man	married	ribat	no	no	yes	no	no	no	no	no	regular	185	3	10	46	live
female	50	basic	not work	married	ribat	no	no	yes	no	no	no	no	no	regular	156	3	5	65	live
male	64	basic	not work	married	ribat	no	no	no	yes	no	no	no	no	regular	168	3	14	33	live
male	38	secondary	police man	married	ribat	yes	no	no	no	no	no	no	no	regular	119	3	6	63	live
female	39	secondary	not work	married	ribat	no	no	yes	no	no	no	no	no	regular	102	3	3	60	live
male	33	secondary	employee	single	ribat	no	no	yes	no	no	no	no	no	regular	164	3	6	17	live
male	63	basic	not work	married	ribat	no	no	yes	no	no	no	no	no	regular	176	3	8	48	live
male	57	basic	not work	married	ribat	no	no	no	no	no	yes	no	no	regular	197	3	13	76	live
female	36	secondary	not work	single	ribat	no	no	yes	no	no	no	no	no	regular	111	3	10	26	live
male	57	ntermediate	not work	married	ribat	no	yes	no	no	no	no	no	no	regular	189	3	12	12	live
male	31	ntermediate	police man	single	ribat	no	no	no	no	no	no	yes	no	regular	215	3	8	25	live
female	44	basic	not work	married	ribat	no	no	no	no	no	no	no	yes	regular	175	3	7	12	live
male	65	secondary	not work	married	ribat	yes	no	no	no	no	no	no	no	regular	189	2	14	24	dead
male	80	basic	not work	married	ribat	no	no	yes	no	no	no	no	no	regular	250	2	17	36	dead
female	50	university	police man	married	ribat	no	no	yes	no	no	no	no	no	regular	150	2	10	60	Dead

DM= Diabetes Mellitus, HTN= Diabetes Mellitus