SURVIVAL ANALYSIS OF PATIENTS WITH END STAGE RENAL DISEASE: IN THE CASE OF ADAMA HOSPITAL MEDICAL COLLEGE.

BY:

MEKIYA HUSSEIN

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, COLLEGE OF NATURAL SCIENCES, JIMMA UNIVERSITY; PARTIAL FULFILMENT OF THE REQUIREMENTS OF MASTER OF SCIENCE (M.Sc.) IN BIOSTATISTICS.

JUNE, 2016

JIMMA, ETHIOPIA

SURVIVAL ANALYSIS OF PATIENTS WITH END STAGE RENAL DISEASE: IN THE CASE OF ADAMA HOSPITAL MEDICAL COLLEGE

BY**: MEKIYA HUSSEIN**

ADVISOR: **GEREMEW MULATA** (*ASS.PROF, M.Sc.*)

CO- ADVISOR: **DINBERU SEYOUM** (*PHD CANDIDATE*)

June, 2016

Jimma, Ethiopia

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the requirements of Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference.

I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate. Brief quotations from this thesis are allowed without requiring special permission if an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Mekiya Hussein

Date: _____

Signature: _____

Jun, 2016

Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis research advisors, we her by certify that we have read the thesis prepared by **Mekiya Hussein under our guidance, which is entitled "Survival Analysis of patients with End Stage Renal Disease : Application of Parametric and semi parametric Model**", in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

GEREMEW MULATA (*ASS.PROF, M.Sc.*)_____ _____

Advisor Signature Date

DINBERU SEYOUM (PhD Candidate) _____ _____

Co-AdvisorSignature Date

As the members of the board of examiners of M.Sc. thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____	_____	_____
Name of Chairman	Signature	Date
_____	_____	_____
Name of Advisor	Signature	Date
_____	_____	_____
Name of Co-Advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

ACKNOWLEDGEMENT

First, and foremost, I would like to thank my createrALLAH for his grace, wisdom, favor, faithfulness and wonderful help up to the present day.

My special thanks goes to Mr.Geremew Mulata, my major advisor and instructor, for his kind support, advice, valuable suggestions and comments beginning from the proposal development up to the final work of this thesis and will always be grateful for that.

Likewise, I am sincerely grateful to my co-advisor Mr. Dinberu Seyoum and instructor for his guidance and constructive comments beginning from the proposal development up to the final work of this thesis and my senior graduates particularly.

I would like to thank Adama Hospital Medical Collegestaff members for providing me the data used in this study more specifically: **Dr.Alemayew G. andDr. Boonsa K.**, as well as all member of record office.

I would like to thank Jimma University for the financial support for this study, instructors and for all the support I was given at any time I asked for.

Finally, my deepest and warmest gratitude goes to my family that has been a source of pride and courage throughout my work, and all my classmate encouragement and motivation during my studies.

DEDICATION

I dedicate this work to dear my family and to my beloved boyfriend Fuad for making me who I am today, for their support and for teaching me the value of education.

ABSTRACT

Background: Chronic kidney disease (CKD) with diagonesised end-stage renal disease (ESRD)is common public health problems worldwide. The aim of this study was to model and compare different parametric (Weibull, Log-logistic and Lognormal) and semi-parametric (Cox ph) regression survival models, using endstage renal disease (ESRD) data set.

Method: This study was conducted from 30, May 2012 to April 1st, 2016 and encompassed 500 ESRD patients at Adama Hospital Medical College. Retrospectives data were gathered by reviewing patients' medical and surgical wards history. The Cox ph regression and parametric Weibull, Log-logistic and log normal models were used for analyzing survival analysis of ESRD patient using R statistical package and STATA software. To compare these models Akaike Information criterion (AIC) and Cox-Snell residual were utilized.

Results: In this study, the totals of 500 ESRD patients were considered. 66.20% were female and 33.80% were male. Among those patients 72.40% and 27.60% were alive and died respectively. Concurrence to the both criteria (AIC and Cox-Snell residual), Lognormal survival model manifested the bestresults as compared with other models. Harmonyto this model, age at the time of admission (HR=0.94, p-value < 0.05), sex of patients (HR=0.54, p-value <0.05) and Family history (HR=0.45, p-value<0.05), had significant effect on survival of the ESRD patient

Conclusion: parametric survival model with baseline hazard lognormal distributionwas found appropriate to our dataset. Deal to the results of study, it conclude that having ESRD with complications increases the probability of death. The estimated survival and hazard rate (time ratio) of ESRD patients under age, sex and Family history had significant difference with p-values less than 0.05. Female patients have greater risk of death than males and based on the mean survival time age of patients greater than 53.34 years have a higher risk of death.

Keywords: Chronic Kidney Disease, ESRD, Risk Factors, parametric models, death

LIST OF ACRONYMS

AAU	Addis Ababa university
AHMC	Adama Hospitals Medical College
AIC	Akaike Information Criterion
AKI	Acute Kidney Injury
ARF	Acute Renal Failure
CKD	Chronic kidney disease
ESRD	End Stage Renal Disease
FDA	Food and Drug Administration
GFR	Glomerular Filtration Rate
HD	Hemodialysis
KM	Kaplan–Meier
NGOs	Non-Governmental Organizations
NKF	National Kidney Foundation
NKTI	National Kidney and Transplant Institute
PD	Peritoneal Dialysis
RRT	Renal Replacement Therapy
SSA	Sub-Saharan Africa
WHO	World Health Organization

Table of Contents

LIST OF TABLE

LIST OF FIGURES

1. INTRODUCTION

1.1. BACKGROUND OF THE STUDY

Chronic kidney disease (CKD) is a world-wide public health problem and it is associated with adverse outcomes of kidney failure, cardiovascular disease and premature death [**1**]. CKD is a condition in which your kidneys are damaged and cannot filter blood as well as healthy kidneys. Because of this, wastes from the blood remain in the body and may cause other health problems. Each of our kidneys has about a million tiny filters, called nephrons. If nephrons are damaged, they stop working. For a while, healthy nephrons can take on the extra work. But if the damage continues, more and more nephrons will be shut down. After a certain point, the nephrons that are left cannot filter your blood well enough to keep you healthy **[2**].

In 2010 it has been estimated that more than 500 million individuals globally have CKD, defined by either kidney damage or glomerular filtration rate (GFR) < 60 ml/min/1.73 m^2 for ≥ 3 months, regardless of the cause**[3**].This risk is due to high rates of diabetes and high blood pressure in these communities [**4**].The CKD stages were categorized based on the classification system established by the National Kidney Foundation - Kidney Disease Outcomes Quality Initiative classification. For the purposes of this study, CKD was defined as Kidney Disease Outcomes Quality Initiative CKD stages 1–5[**4, 5**]. End-stage renal disease (ESRD) which corresponds to an estimate GFR of <15 ml/min/1.73m^2, initiation of maintenance dialysis or receipt of preemptive renal transplantation is classified as CKD stage five. A person with ESRD needs treatment to replace the work of the failed kidneys [**6]**.

According to the latest WHO data published in April 2011, kidney disease deaths in Ethiopia reached 12,038 or 1.47% of total deaths. Globally, more than 100 countries (with combined population >1 billion) have no provisions for chronic maintenance dialysis or kidney transplantation and thus, more than 1 million people die annually from ESRD **[6-8]**. Kidney disease imposes disproportionate, incalculable human suffering and a catastrophic economic burden on the African continent in several respects: less than 2% of the patients with ESRD have access to RRT making ESRD a death sentence for most patients. The ESRD rate is

increasing at 6% to 8% per year on the African continent; Africa is experiencing an accelerated incidence of hypertension (60 million people) and type 2 diabetes mellitus (>12 million people), which are the underlying causes in >15% of CKD cases; and finally, at current estimates, none of the 54 countries in Sub-Saharan Africa (SSA) will be able to afford the cost of medical care associated with pre-dialysis CKD for their populations (estimated to be $2500 to $20,000 per patient annually) **[5, 11, 12**]. There is no cure for CKD, although treatment can slow or halt the progression of the disease and can prevent other serious conditions developing. [**9, 13, 14]**

The statistical analysis of lifetime data (time-to-event) plays an important role in medicine because we are observing something that develops dynamically over time. There are two points related to this development of survival data. First, Survival times are usually a mixture of discrete and continuous data that lend themselves to a different type of analysis. The Kaplan-Meier estimator of the survival function is a major step in the development of suitable models for such kind of data. Second, most evaluations are made conditionally on what is known at the time of the analysis, and changes over time [**16, 17]**.

The Cox proportional hazards (PH) model is now the most widely used for the analysis of survival data in the presence of covariates or prognostic factors. One of the reasons this model is so popular is because of the ease with which technical difficulties such as censoring and truncation are handled. This is due to the appealing interpretation of the hazard function as a risk that changes over time [**17**].We use parametric survival models when the assumptions of Cox- PH should not fulfilled. Parametric survival models are statistically more powerful than nonparametric or semi-parametric models [**19].**A survival analysis is conduct by using Weibull, Log-logistic and lognormal. This would be study by means of real dataset which is collected from CKD patients in Adama Hospital Medical College. Modelsare compared using AIC value and visual inspection method that is the residual plots.

1.2. Statement of the Problem

Kidney disease is an important public health issue worldwide. To address the CKD problem of Ethiopia sustained efforts from nongovernmental organizations (NGOs), governmental agencies, the pharmaceutical industry, and medical training programs are needed. The goal is to prevent renal failure and death from renal failure in the people. To achieve this goal, development of a high-quality chronic dialysis program is needed.

For many years the magnitude of ESRD in Ethiopia has not been studied except, by healthcare the prevalence of CKD and associated risk factor among diabetic patients. They are used logistic regression from statistical analysis and the health form of equation. [**1**]

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 provide reasonably accurate. The survival time for the ESRD patients who are follow up may depend on different factors, such as demographic factors, health conditions, the primary treatments given to the patients, the age, sex, hypertension Disease, diabetes mellitus and CKD diagnosis of ESRD and other risk factors patients that led them to death.; it is be attempt to find answers to the following questions:

1. What is the risk factor of CKDpatients' diagnosedwith ESRD that leads to death?
2. Parametric survival model comparison (weibull, log-logistic, lognormal) and Cox ph regression in order to find which one of them is the best model to describe the CKD patient?

1.3. Objectives of the Study

1.3.1 General Objective:

The main objective of this study is to identify the major predictive factors of survival analysis of patients with end stage renal disease under appropriate survival model for the death time in CKD patients: in the case of Adama Hospital Medical College.

1.3.2. Specific Objectives:

The specific objectives of this study are:

1. To identify predictive factor of hospital outcome (prolonged hospital stay and hospital mortality) in AHMC.
2. To determine the most significant of the risk factors ofsurvival time of CKD patients at ESRD.
3. To find out an appropriate survival model for the survival time in CKD patients.

1.4. Significance of the Study

Pieces of empirical information to be generated by this study would be paramount importance. Study of literature shows that much work has been carried out on statistical analysis on health related issues.The research gap isfor many years the magnitude of ESRD in Ethiopia has not been studied. Thus, this research could be a good stepping ground for other studies on kidney disease. The study would be used to identify risk factors of kidney associated mortality or death of patients.On model comparisonchoosing various parametric survival models, the research would reduce uncertainties on survival models specifically in fitting data such as kidney disease.Researcher's that would be willing to carryout relevant study in the future and also the analysis would be great significance. And also this research would be useful to researchers, and government and nongovernmental organization which are working in the health sector especially concerning on kidney disease (the factors affecting survival/death status of CKD are captivating RRT) for policy formulation.

2. LITERATURE REVIEW

2.1 Definition and General Overview of Chronic Kidney Disease

The kidneys regulate the composition and volume of blood, remove metabolic wastes in the urine, and help control the acid/ base balance in the body. They activate vitamin D needed for calcium absorption and produce erythropoietin needed for red-blood-cell synthesis. One of the most useful tools is the nutritional care plan **[5**].

CKD has been defined as decreased kidney function and/or kidney damage persistent for at least 3 months. Kidney dysfunction is indicated by a glomerular filtration rate (GFR) of less than 60 mL/min/1.73m^2, while kidney damage most frequently is manifested as increased urinary albumin excretion.2 within this framework, CKD has been categorized into five stages: Stage 1: Kidney damage with GFR ≥90 mL/min/1.73 m^2(normal), Stage 2: Kidney damage with GFR 60–89 mL/min/1.73 m^2(increased risk damage), Stage 3: GFR 30–59 mL/min/1.73 m^2(decrease GFR), Stage 4: GFR 15–29 mL/min/1.73m^2(regardless of kidney damage) and Stage 5: GFR <15 mL/min/1.73m^2(kidney failure) treated by dialysis or transplantation [**11**].

Multiple well-designed randomized controlled human trials have evaluated both the efficacy and safety of protein restriction in patients with progressive CKD [**12**]. Moderate protein restriction (0.6 to 0.8 g/kg per day) is associated with significant benefit of protein restriction in delaying renal dialysis [**12, 20, 22**]. It is generally well tolerated and does not lead to malnutrition in patients with CKD providing caloric goals are met, dietary protein is of high biologic value, and metabolic acidosis is avoided. For CKD stage 3, low protein, low phosphorus diets may retard dialysis **[21**]. A diet providing about 0.60–0.75 g protein/kg/day, of which at least 0.35 g/kg/day is high biologic value protein, is needed to ensure a sufficient intake of the essential amino acids. For CKD stage 4 and 5, the potential advantages of using a low-protein, low- phosphorus diet are more compelling. A low protein diet will generate less nitrogenous compounds that are potentially toxic both systemically and to the kidney itself. In addition, it generally contains less phosphorus and potassium; reductions which are usually imperative at this advanced stage of renal failure. The same energy intakes are recommended for people with stage 3 or 4 CKD (i.e.GFR 60 mL/

minute/1.73 m^2). Not all individuals with chronic renal disease are willing and able to adhere to diets providing 0.60 g protein/kg/day or less. For this reason it is necessary to have a close follow-up and nutrition counseling by a registered dietician at least 3-4 times a year. [**21-24**]

A diagnosis of End Stage Renal Disease (ESRD) means that you are in the final stage of kidney disease and your kidneys are not functioning well enough to meet the needs of daily life. It also used to be an automatic death sentence; however, advancements in treatments now allow patients to live much longer than ever before [**15**]. Without dialysis or a kidney transplant, death will occur. The outcome of the treatment depends on the patient. Having kidney disease increases chances of having cardiovascular disease, heart attacks, and strokes. Also, the body can hold in too much fluid which could lead to swelling in your arm and legs, high blood pressure, or fluid in your lungs called pulmonary edema [**20**]. To analysis the random cause of the morbidity among Ethiopia people, the researchers generated a mathematical survival model that would predict the absolute risk of death with End Stage Renal Disease patients [**9**].

2.2 Empirical Literature Review

Chronic kidney disease (CKD) is a worldwide public health problem that affects 26 million Americans in the U.S., with 600,000 requiring dialysis or kidney transplantation. Despite its prevalence, there are fewer clinical trials for kidney disease than any other common disease. In December 2012, the National Kidney Foundation (NKF) and the Food and Drug Administration (FDA), challenged the research community to evaluate the current definition of kidney disease progression and examine whether improvements were possible. At that time, NKF and FDA officials viewed favorably emerging evidence that a decline in estimated kidney function was promising as a reliable indicator of kidney disease progression. This research grew directly out of the NKF-FDA challenge**[5].**For the study, a global Chronic Kidney Disease Prognosis Consortium led by Josef Coresh, MD, PhD, MHS, a professor at the Bloomberg School, analyzed data from 1.7 million participants recruited into 35 cohorts in dozens of countries from 1975-2011 and followed for an average of 5 years. The researchers' points of comparison were the FDA's current definition of CKD

disease onset for clinical trials a doubling of serum creatinine, a blood marker that assesses kidney function and the emerging evidence on a decline in estimated kidney function. [**40**]

Researchers first analyzed kidney disease progression among all participants during a baseline period of two years. Thenthey examined how this progression predicted subsequent disease progression to the observed 12,344 cases of end-stage renal diseases (ESRD) and 223,944 deaths. The study found that the current serum-creatinine standard used in clinical trials which is associated with a 57% reduction in kidney function carried very high risk a 32-fold increased risk of ESRD and 3.7-fold increased risk of mortality but only occurred in less than 1% of participants in the two-year baseline period**[17-21**]. In contrast, a 30% decline in kidney function, also measured by serum creatinine levels, occurred in 7% of participants in the two-year baseline period. This was associated with a 5-fold higher risk of end-stage renal disease (ESRD) and 1.8-fold higher risk of mortality. This level of risk is high and yet common and early enough to facilitate testing if new therapies are working.

Although mortality rates have decreased over the last decade, they still remain on a very high level. According to the United States Renal Data System, 147 per 1,000 Medicare CKD patients age 66 and older died in 2009. Only 50% of dialysis patients and 82% of those who receive a preemptive transplant are still alive three years after the start of ESRD therapy. The mortality rate in 2009 for ESRD and dialysis patients 65 and older was 274 and 313 per 1,000 respectively. The Renal Association states that patients who present with uncomplicated AKI have a mortality rate of up to 10%, in patients presenting with AKI and multiorgan failure the rate increases to over 50%, and rises further to as high as 80% if renal replacement therapy is required. In mainland China, the number of CKD patients was estimated to be around 119.5 million based on a national survey in 2010 [**12**]. Data from the Chinese Renal Data System revealed that there were 270,000 patients undergoing HD, while only 30,000 were received PD treatment [**21-23**], suggesting that HD is the major treatment modality in China, accounting for approximately 90% of the total ESRD patients.

In Italy, chronic kidney disease (CKD), defined as Glomerular Filtration Rate (GFR) < 60 ml/min, represent about 6% of population **[7].** The end-stage renal disease (ESRD) for which life can be sustained only with renal replacement therapy such as dialysis or kidney

transplantation is growing and in Italy the last report from RIDT (Italian registry of dialysis and transplant) show an incidence of about 160 parts-per-million (p-pm) and prevalence of 788 p-pm. There are currently approximately 40,000 patients undergoing maintenance dialysis.In 2012, the National Kidney and Transplant Institute (NKTI) cited kidney failure as the 9th leading cause of death among Filipinos [**20**].

According to the latest WHO data published in April 2011, kidney disease deaths in Ethiopia reached 12,038 or 1.47% of total deaths.Among adult patients with Acute Kidney Injury (AKI), septic abortion and falciparum malaria were the leading causes in studies in the 1980's and 90's. Like all other chronic non-communicable diseases, data on the prevalence of Chronic Kidney Disease (CKD) and the incidence and prevalence of kidney disease in Ethiopia are not available. However, there are some hospital based studies and observations indicating the causes of CKD and several studies both in hospital and in the community looking at the major risk factors for CKD, namely hypertension and diabetes. Based on some unpublished hospital based studies and estimates: more recent data from TikurAnbessa Hospital, the main teaching hospital of AAU Medical School, indicate that chronic glomerulonephritis, diabetes and hypertension are the leading causes of CKD. Renal diseases accounted for 1.2-6 % of adult hospital medical admissions in reports from various parts of the country**[9**].

2.3Socio-Economic Factors theDeath of End Stage Renal Disease

Chronic renal failure, the ultimate stage of chronic kidney disease, represents a major issue for the National Health Service due to the high increase of its incidence and prevalence as well as for the high social costs associated with the management of the disease. More specifically, and with a particular reference to medical and non-medical direct costs, an average annual cost per patient under dialysis was estimated to be about €34,071.70 (i.e. €653.43 per week). The low protein diet, as reported in scientific literature, has demonstrated to be the best alternative in order to delay dialysis [**12, 23**]. However, in terms of healthcare politics and planning, it is essential to demonstrate also an economic advantage (sustainability) deriving from such care. A mixed model of care may be a way to subsidize health care services for the poor. Malaysia provides its dialysis services via a blend of model

of care of public hospitals, for-profit private centers, and not-for-profit organizations, such as religious groups, the National Kidney Foundation and the Rotary club [**4, 24**].

Africa is experiencing an accelerated incidence of hypertension (60 million people) and type 2 diabetes mellitus (>12 million people), which are the underlying causes in >15% of CKD cases; and finally, at current projections, none of the 54 countries in Sub-Saharan Africa (SSA) will be able to afford the cost of medical care associated with pre-dialysis CKD for their populations (estimated to be $2500 to $20,000 per patient annually) **[2, 7, 9].** Even more out of reach is the annual cost of dialysis treatment which amounts to $20,000 to $30,000 per person per year in SSA -Medicare covers this cost expenditure for the 500,000 US citizens with ESRD. In contrast, fewer than 5% of the 500,000 new cases of ESRD in SSA gain access to even a limited period of dialysis [**38, 39**]. These distressing facts elevate the urgent need for research to mitigate the incidence and severity of kidney disease globally and particularly in SSA.

For many years the magnitude of ESRD in Ethiopia has not been studied. The use of dialysis in the country as a treatment strategy for ESRD dates less than a decade. In addition, access for dialysis is limited and is a highly unaffordable for the general public. Each dialysis session costs about $100 (1700 Birr) excluding the costs for other supportive cares. Because of the low socio-economic status, dialysis is thus considered as luxury care in the country. There is currently no dialysis center in Public hospitals in Ethiopia with a population surpassing 85 million. In addition, there is no national strategy for prevention and care of patients with CKD[**13, 20**].

2.4 Overview of Model Families

2.4.1 Identification of Survival Indicators

Survival models have been extensively used in medical research during last several years. Especially the Cox proportional hazards model. In fact, Cox proportional hazards model become workhorse of regression analysis for censored data. In Cox proportional hazards model, they included explanatory variables or covariates to study the effect of covariates on distribution of survival times [**26**]. As the use of survival analysis grew, parametric models

gave way to non-parametric and semi parametric approaches for their appeal in dealing with the ever growing field of clinical trials in medical research. Survival models have the capability of handling censored data. Cox and Oakes used survival analysis in modeling human lifetimes [**16**].

The Cox Proportional Hazard model is the most popular technique to analyze the effects of covariates on survival time but under certain circumstances parametric models may offer advantages over Cox's model. The results of data analysis using parametric models are, similar to the Cox regression. Although the Hazard Ratio in Cox and parametric models are approximately similar but Weibull and Exponential are the most favorable for survival analysis of the data [**18, 19, 23**].

Moreover, characterizing the different survival distributions that correspond to different subgroups within a heterogeneous population is the objective of many studies. In a review of survival analyses in CKD study, it was found that only 5 % of all studies used the Cox model with respect to checking the underlying assumptions. If this assumption does not hold, the Cox model can lead to unreliable conclusions. Therefore, the parametric models such as Lognormal, Weibull, Exponential and log logistic are the common options. These models provide the interpretation based on a specific distribution for duration times without need to proportional hazard assumptions [**26, 28**].

The study conducted in Iran (2013) reported that the variables survival included the following variables: age (p=0.0001), heart failure (p = 0.001), Hypertension rat (p = 0.003), relative family (p=0.037), serum albumin (p = 0.0001) were significantly influences. The results indicated the general preference of parametric models over Cox semi-parametric model. The lognormal model had the highest efficiency among parametric models. Andsuggested that the lognormal models has the best fit among parametric models and may be used as a substitute for Cox model in analyzing kidney transplant survival, they included this variable in the final model in order to shape the effect of the nephrological monitoring on survival of incident dialysis patients [**35, 41, 47**].They apply their models to bivariate survival data set of related to Kidney disease and compare these above models using AIC comparison techniques [**48]**

3. DATA AND METHODOLOGY

This section describes the data and methods used in this study to come up with the development of a survival model and estimate the probability of surviving all causes of death for a specified time interval calculated from the cohort of ESRD cases.

3.1 Source of Data

The data used for this study were obtained from Adama Hospital Medical College, Adama town, Oromia regional state, Middle East Ethiopia, 99 km away from Addis Ababa. The hospital serves as a referral hospital to the patients from different parts of Oromia region and other nearby regions such as Amhara, Somali and Afar regional states. The survival data were extracted from the patient's chart which contains epidemiological, laboratory and clinical information of all CKD patients.

3.2 Study Design and Period

Retrospectives data were gathered for patients in the medical and surgical wards. Data were collected on all patients from their medical records with laboratory result.Theclinicians follow result between 30, May 2012 to 1st, April 2016 in AHMC.
The study was a retrospective study i.e. all the events-exposure had already occurred in the past, which reviews the patient cards and patent's information sheet.

3.3 Study Population

The total number of Chronic Kidney Disease patients diagnosed with ESRD from 30, May 2012 to 1st, April2016 in AHMC. These consist of the age, gender, hypertension Disease, diabetes mellitusandCKD diagnosis of ESRD patients that led them to death.

3.3.1 Data Gathering Procedure

In this study, all of the dataneeded by the researcher were obtained from the records in Adama Hospital Medical College. To compute the risk factors of death on CKD patients diagnosed with ESRD, the status of each patient are needed to be observed.

Inclusion and exclusion criteria

Inclusion criteria: - The CKD patients' end-stage renal/ESR (**15-29 ml/min/1.73m^{2}GFR)** with diagnoses risk factors.

Exclusion criteria:-the patient's chronic renal insufficiency, late renal insufficiency, pre-ESRD (>=**30 ml/min/1.73m^{2}GFR**) and other related CKD patients.

3.4. Variables of the Study

3.4.1 The Response Variable

The response /dependent variable in this study is the survival time (time to death) measured in Days from the date ofthe treatment start until the date of the patient's death or censor (Patients who lost to follow; transferred to another Hospital for referral).The status of the patients is **1** if the person died and **0** if the person is censored over the given time.

3.4.2 Explanatory Variables

Some predictors are considered in this study to investigate the determinant factors for the survival time of CKD patients and are given below:-

Table1. Description of independent variables used in the analysis.

Covariate	Values/codes
Sex	**0= Male 1=Female**
Age	**Years**
Residence	**0=Rural 1= Urban**
Diabetes mellitus	**0=No 1= Type I2= Type II**
Hipertensión	**0=No1= Yes**
Vascular access	**1= Fistula 2= Catheter 3=Graft,**
Family history	**0=Absent1= Present**
Obesity	**0= Absent 1= Present**

3.4.2.1 OPERATIONAL DEFINITION

1. **Age** = age of the ESRD patients in year.
2. **Sex** = sex of the ESRD patients.
3. **Residence** = the place of the patients lived.
4. **Diabetes mellitus** = a chronic condition associated with abnormally high levels of sugar (glucose) in blood
5. **Hypertension (high blood pressure)** =as a systolic blood at or above 140mmHg and/or a diastolic blood pressure at or above 90 mmHg.
6. **Vascular access=**the most common treatment for kidney failure.
7. **Obesity=** use a measurement called BMI (body mass index) whichis the in dividual's weight multiplied by 703 and then divided by twice the height i n inches.BMI of 25.9-29 is considered overweight; BMI over 30 is considered. Obesityis an abnormal accumulation of body fat.
8. **Family history =** the relative or family history of kidney patients.

3.5. METHODS OF DATA ANALYSIS

3.5.1 The Survival Model

Survival analysis is survival time which is also called failure time. Survival time is a length of time that is measured from time origin to the time that event of interest occurred. To determine survival time precisely, there are three requirements: A time origin must be unambiguously defined, a scale for measuring the passage of time must be agreed upon and finally the definition of event (often called failure) must be entirely clear. The specific difficulties in survival analysis arise largely from the fact that only some individuals have experienced the event and other individuals have not had the event in the end of study and thus their actual survival times are unknown. This leads to the concept of censoring. Censoring occurred when we have some information about individual survival time, but we do not know the survival time exactly. There are three types of censoring: right censoring, left censoring, and interval censoring.

Right censoring: is said to occur if the event occurs after the observed survival time. Let C denote the censoring time, that is, the time beyond which the study subject cannot be observed. The observed survival time is also referred to as follow up time. It starts at time 0 and continues until the event Y or a censoring time C, whichever comes first. Let C_1, $C_{2...}C_n$be a sample of censoring times. And T_1, T_2….T_n, be event times. We observe a sample of couples, (y_1, δ_1), (y_2, δ_2), …, (y_n, δ_n), where for i=1,2,….n.

$$Y_i = \min (T_i, C_i) = \begin{cases} T_i, if T_i \leq C_i \\ C_i\ , if T_i \geq C_i \end{cases}$$

$$\delta_i = I\ (T_i \leq C_i) = \begin{cases} 1\ , if T_i \leq C_i \\ 0\ , if T_i \geq C_i \end{cases}$$

The actual occurrence time of event is known within an interval of time. An important assumption for methods presented in survival analysis studies for the analysis of censored survival data is that the individuals who are censored are at the same risk of subsequent failure as those who are still alive and uncensored i.e. a subject whose survival time is censored at time *C* must be representative of all other individuals who have survived to that time. If this is the case, the censoring process is called non-informative. Statistically, if thecensoring process independent of the survival time.In this study, we assumed that the censoring is randomrightcensoring. The response variable in survival analysis is survival time and is no longer limited to only time to death. It is a non-negative random variable used loosely for the time period from a starting time point to the occurrence of any event. In this study context useda random right censored; survival time is the length of survival time of ESRD patients which is measured in days in Adama Hospital Medical College follow up from May 30, 2012 to 1st, April 2016.

3.5.2Survival Functions

The survivor function is defined to be the probability that the survival time of a randomly selected subject is greater than or equal to some specified time. Thus, it gives the probability that an individual surviving beyond a specified time. Moreover, the distribution of survival time is characterized by three functions: The survivorship function, the probability density function, andthe hazard function.

Let T be a continuous random variable associated with the survival times, t be the specified value of the random variable T and f (t) be the underlying probability density function of the survival time T. The cumulative distribution function F(t), which represents the probability that a subject selected at random will have a survival time less than some stated value t, is given by:

$$F(t) = P(T < t) = \int_0^t f(u)du \ , t \geq 0$$

The survivor function $S(t)$is given by;

$$S(t) \ = \ \mathrm{P}(\mathrm{T} \geq t) = 1 - \ \mathrm{F}(\mathrm{t}), \mathrm{t} \geq 0 \qquad (1)$$

From equations (1) and (2) the relationship between $f(t)$and $S(t)$can be derived as

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}\big(1 - S(t)\big) = \frac{-d}{d}S(t) \geq 0 \qquad (2)$$

Theoretically, as t ranges from 0 to infinity, the survivor function can be graphed as a smooth curve. Survivor functions have the characteristics that:

1. They are non-increasing
2. At time t = 0, S(0) = 1; that is, at the start of the study, since no one has experienced the event yet, the probability of surviving past time 0 is one and
3. At time $t \rightarrow \infty$, $S(\infty) \rightarrow 0$; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually converge to zero.

3.5.3 Hazard Function

The hazard function h(t) gives the instantaneous potential for failing at time t, given that the individual has survived up to time t. The hazard function is a measure of the probability of failure during a very small interval, assuming that the individual has survived at the beginning of the interval. The hazard function is not a probability as it does not lie between 0 and 1.

The hazard function, $h(t) \geq 0$ is given as;

$$h(t) = \lim_{\Delta t \to 0} \frac{P\{an individual fails in the time interval (t, t+\Delta t) given survives until time t\}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P(t \le T \le t+\Delta t \mid T \ge t)}{\Delta t} \quad (3)$$

By applying the theory of conditional probability and the relationship in equation (3), the hazard function can be expressed in terms of the underlying probability density function and the survivor function becomes:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \ln S(t). \quad (4)$$

The corresponding cumulative hazard function, H(t), is defined as:

$$H(t) = \int_0^t h(u)du = -\ln S(t), \quad (5)$$

Then; $S(t) = \exp(-H(t)$ And

$$f(t) = h(t)S(t) \quad (6)$$

The survival function is most useful for comparing the survival progress of two or more groups. The hazard function gives a more useful description of the risk of failure at any time point.

3.5.4 Non-Parametric Methods

Nonparametric methods are often very easy and simple to understand as compared to parametric methods. Furthermore, nonparametric analyses are more widely used in situations where there is doubt about the exact form of distribution. Survival data are conveniently summarized through estimates of the survival function and hazard function. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time. These methods are said to be non-parametric methods since they require no assumptions about the distribution of survival time. Preliminary analysis of the data using non-parametric methods provides insight into the shape of the survival function for each group and get an idea of whether or not the groups are proportional, i.e., if the

estimated survival functions for two groups are approximately parallel (do not cross). In order to compare the survival distribution of two or more groups, log-rank tests can be used.

3.5.4.1 The Kaplan-Meier Estimator of Survival Function

The Kaplan–Meier (KM) method is a nonparametric survival analysis that is used to estimate the two survival probabilities curves dealing with differing survival times especially when not all of the subjects continue in the study. KM estimator incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t. The technique becomes a little more complicated but still manageable when censored times are included. Let ordered survival times are given by $0 \leq t_1 \leq t_2 \leq t_j \leq \infty$, then:

$$\hat{S}(t)=\begin{cases}1, & if\, t < t_1 \\ \prod_{j:t_j \leq t}[1-\frac{d_j}{r_j}], & \text{if } t \geq t_1\end{cases} \qquad (7)$$

Where; d_j is the observed number of events at time t_j and r_j is the number of individuals at risk at time t_j.

The Kaplan-Meier estimator, $\widehat{S}(t)$ is a step function with jumps at the observed event times. The size of the jump at a certain event time t_j depends on the number of events observed at t_j, as well as on the pattern of the censored event times before t_j. Kaplan-Meier also known as the Product-Limit estimator. The variance ofthe Product-Limit estimator is estimated by Greenwood's formula **[25**], and is given by:

$$\text{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j:t_j \leq t} \frac{d_j}{r_j(r_j-d_j)};\ j= 1, 2, \ldots, r \qquad (8)$$

Since the distribution of survival time tends to be positively skewed, the median is preferred for a summary measure. The median survival time is the time beyond which 50% of the individuals under study are expected to survive, i.e., the value of t_{50} at $\hat{S}(t_{50}) = 0.5$. The

estimated median survival time is given by t_{50} = min $\{t_i/\hat{S}(t)< 0.5\}$, where t_i is the observed survival time for the i^{th} individual, i= 1,2,…, n[**28**].

3.5.4.2.Comparison of Survival Distributions

The log–rank test: is a non-parametric test that provides an overall comparison of the KM curves being compared. The main objective is to compare two survival curves by treatment group. The log rank test statistic for comparing two groups is given by:

$$Q=\frac{[\sum_i^m w_i(d_{1i}-\hat{e}_{1i})]^2}{\sum_i^m w_i^2 \hat{V}_{1i}} \sim \chi^2_{k-1}, \qquad (9)$$

Where: $\hat{e}_{1i} = \frac{n_{1i}d_i}{n_i}$ And $\hat{V}_{1i} = \frac{n_{1i}n_{0i}d_i(n_{1i}-d_i)}{n_i^2(n_i-1)}$

n_{0i}= the number at risk at observed survival time $t_{(i)}$ in group 0

n_{1i}= the number at risk at observed survival time $t_{(i)}$ in group 1

n_i = the total number of individuals or risk before time $t_{(i)}$

d_{1i} = is the number of observed event in group 1

d_i = the total number of event at $t_{(i)}$

k = is number of groups in each category

3.5.5Semi-Parametric Cox Ph Regression Model

The non-parametric method does not control for covariates and it requires categorical predictors. When we have several prognostic variables, we must use multivariate approaches. But we cannot use multiple linear regression or logistic regression because they cannot deal with censored observations. We need another method to model survival data with the presence of censoring. One very popular model in survival data is the Cox proportional hazards model. The Cox proportional hazards (PH) regression model (introduced in a seminal paper by Cox, [**26**]; a broadly applicable and the most widely used method of survival analysis. Survival models are used to quantify the effect of one or more

explanatory variables on failure time. This involves specification of a linear -like model for the log hazard. A parametric model based on the exponential distribution may be parameterized as follows:

$$\log h_i(t \mid x) = \alpha + \beta_1 x_{i1} + \beta_{21} x_{i2} + \cdots + \beta_k x_{ik}$$

Equivalently;

$$h_i(t \mid x) = exp(\alpha + \beta_1 x_{i1} + \beta_{21} x_{i2} + \cdots + \beta_k x_{ik}) = \exp(\alpha)\exp(\beta' X)$$

In this case the constant α represents the log-baseline hazard since $\log h_i(t) = \alpha$ when the entries are zero. The Cox PH model is a semi-parametric model where the baseline hazard $\alpha(t)$ is allowed to vary with time.

$$\log h_i(t \mid x) = \alpha(t) + \beta_1 x_{i1} + \beta_{21} x_{i2} + \cdots + \beta_k x_{ik}$$

$$h_i(t \mid x) = h_o(t) exp(\beta_1 x_{i1} + \beta_{21} x_{i2} + \cdots + \beta_k x_{ik})$$

$$h_i(t \mid x) = h_o(t) exp(X_i^T \beta) \qquad (10)$$

Where , $h_o(t)$ is the baseline hazard function; $\mathbf{X}_i$ is a vector of covariates and $\boldsymbol{\beta}$ is a vector of parameters for fixed effects.

The corresponding survival function for Cox-PH model is given by:

$$S(t,X) = [S_o(t)]^{\exp\{\sum_{i=1}^{p} \beta_i X_i\}} \qquad (11)$$

where, $S_o(t)$ is the baseline survival function.

In this model, no distributional assumption is made for the survival time; the only assumption is that the hazards ratio does not change over time (i.e., proportional hazards) that is why this model is also known as semi -parametric model. Even though the baseline hazard is not specified, we can still get a good estimate for regression coefficients β, hazard ratio, and adjusted hazard curves. If all of the x's are zero the second part of equation(11) equals 1 so, $h_i(t) = h_o(t)$. For this reason the term $h_o(t)$ is called the baseline hazard function. With the Cox proportional hazards model the outcome is described in terms of the hazard ratio. The measure of effect is called hazard ratio. The hazard ratio of two individuals with different covariates X and X^* is given by:

$$\widehat{HR} = \frac{h_o(t)\exp(\hat{\beta}'X)}{h_o(t)\exp(\hat{\beta}'X^*)} = \exp\{\sum \hat{\beta}'(X - X^*)\}$$

This hazard ratio is time-independent, which is why this is called the proportional hazards model. The parameter of the Cox proportional hazard model refers to the hazard ratio of one group in comparison to the other groups for categorical covariates and change in hazard ratio with a unit change of the covariate for the continuous variables when other covariates are fixed.

The change in hazard ratio for the continuous covariate is given by:

$$\frac{h_i(t,x_k+1)}{h_k(t,x_k)} = \exp(\beta_k). \qquad (12)$$

Which represent change in the hazard when there is a unit change in the covarite while other covariates keeps constant.

For catagorical explanatory variable X with a levels, the model containes (α-1) dummy variables defind as D_i = 1, if x =i, and 0 otherwise for i= 1,2,…, α-1. Let $\beta_1, \beta_2, \ldots, \beta_{a-1}$ denote the coefficient of the levels of dummy variables. The ratio of the hazard of two subjects, one with X at level j and other with k (j,k = 1,2,…, α-1), provided that the value of all other explanatory variables for this subject are the same, the hazard ratio between these two categories is given by:

$$\frac{h(t \mid D_j)}{h(t \mid D_k)} = \frac{\exp(\beta_j)}{\exp(\beta_k)} = \exp(\beta_j - \beta_k).$$

The quantity $\exp(\beta_j - \beta_k)100\%$ signifies the ratio of hazard function for subject at level j and k of covariates, given that the effect of other covariate keeps fixed.

3.5.5.1 Partial Likelihood Estimation for Cox Ph Model

Fitting the Cox proportional hazards model, we estimate $h_o(t)$ and β. A more popular approach is proposed by Cox [**27**] in which a partial likelihood function that does not depend on $h_o(t)$ is obtained for β. Partial likelihood is a technique developed to make inference about the regression parameters in the presence of nuisance parameters ($h_o(t)$) in the Cox PH model. In this part, we construct the partial likelihood function based on the proportional hazards model.

The data in survival analysis based on the sample size n are denoted by the triplet (T_i,δ_i, X_i), i= 1, 2,…, n where T_i is the time at which the i^{th} individual experience the event (in this research context; death), δ_i = 1 if the event has occurred, $\delta_i = 0$ if censored, X_i is the vector of covariate or risk factors for the i^{th} individual.

We assume:

Given X_i the life time and the censoring times are independent (non-informative censoring).

$\tau_1 < \tau_1 < \ldots < \tau_D$ be the D ordered distinct event times

We assume that there are no tied event times.

Let us define by;

I_j is the identity of the individual who give death at time τ_j

V_j the time of the j^{th} failure at time τ_j and all information about censoring in $[\tau_{j-1}, \tau_j]$

The observable data (T_i,δ_i, X_i) is represented by $\{I_j\}$ and $\{V_j\}$. Hence;

$P(Data) = P(\{I_1,V_1,\ldots,I_D,V_D\})$

$$= P(\{I_1,V_1\}) \times P(\{I_2,V_2\}/\{I_1,V_1\}) \times \ldots \times P(\{I_D,V_D\}/\{I_1,V_1,\ldots,I_{D-1},V_{D-1}\})$$

$$= \prod_{j=1}^{D} P(I_j \mid I_1, V_1, \ldots, I_{j-1}, V_{j-1}, V_j) \times P(V_j \mid \{I_1, V_1, \ldots, I_{j-1}, V_{j-1}\})$$

Due to the non-informative censoring, the second term does not add much information about the parameters β.

Hence, we define the partial likelihood as;

$$L^{partial}(\beta) = \prod_{j=1}^{D} P(I_j \mid \{I_1, V_1, \ldots, I_{j-1}, V_{j-1}, V_j\}) = \prod_{j=1}^{D} P(I_j \mid H_j)$$

Where, H_j is the "history" of the data, up to j th failure and including the failure time, but not the identity of the failing.

At each failure, we note that the quantity $P(I_j \mid H_j)$ is the conditional probability that a specific individual fails at time τ_j given all the individuals that had not fail before τ_j.

We denote by R(t) the set of all the individuals under study just prior to time t.

$P(I_j \mid H_j)$= P (individuals I_j fails | one individual fails in R (τ_j))

$$= \frac{P(\text{individuals Ij fails} \mid \text{at risk at } \tau_j)}{\sum_{l \in R(\tau_j)} P(individuallfails \mid atriskat\tau_j)}$$

$$= \frac{\lambda(\tau_j \mid X_j)d\tau_j}{\sum_{l\in \mathrm{R}(\tau_j)} \lambda(\tau_j \mid X_j)d\tau_j} = \frac{\lambda_o(\tau_j)\exp(\beta^T X_j)}{\sum_{l\in \mathrm{R}(\tau_j)} \lambda_o(\tau_j)\exp(B^T X_l)} = \frac{\exp(\beta^T X_j)}{\sum_{l\in \mathrm{R}(\tau_j)} \exp(B^T X_l)}$$

We get the partial likelihood;

$$L^{partial}(\beta) = \prod_{j=1}^{D} \frac{\exp(\beta^T X_j)}{\sum_{l\in \mathrm{R}(\tau_j)} \exp(B^T X_l)}. \quad (13)$$

This is the partial likelihood defined by Cox. Note that, it does not depend on the underlying hazard function $h_o(.)$. Cox recommends treating this is as an ordinary likelihood for making inferences about β in the presence of the nuisance parameter $h_o(.)$.

The likelihood function in equation (13) can be expressed by;

$$L^{partial}(\beta) = \prod_{j=1}^{D} \left[\frac{\exp(\beta^T X_j)}{\sum_{l\in \mathrm{R}(\tau_j)} \exp(B^T X_l)}\right]^{\delta_i} \quad (14)$$

The partial likelihood given by equation (14), although it describes only part of the data, could be regarded as a likelihood function allowing the estimation of β with standard procedures.

In general, large sample properties like normality and consistency of maximum likelihood estimators of β based on partial likelihood have been shown to be the same as those of any estimator from complete likelihood[**28**].

Checking of Cox PH assumption

The main assumption of the Cox proportional hazards model is proportional hazards. Proportional hazards means that the hazard function of one individual is proportional to the hazard function of the other individual, i.e., the hazard ratio is constant over time. There are several methods for verifying that a model satisfies the assumption of proportional hazard.

1. **Graphical Method:** There are two types of graphical techniques to check the proportional hazardassumptions. The most common technique is by comparing the estimatedln$(-lnS(t))$. If the two survival curves do not intersect and are parallel, it evidently provides evidence against the assumptions.Another graphical approach is by comparing observed with predicted survival curves. If the twosurvival curves are close, then the proportional hazard assumption is plausible
2. **Test of GoodnessFit**: This approach provides a Chi–Square statistics where the computed values for each variable rely on p–values. This p–value is used for

evaluating the PH assumption for each variable. If the p–values are large then PH is satisfied, whereas small p–values provide violation of PH assumption.

3.5.6 PARAMETRIC REGRESSION MODELS

In the previous topics it was focused entirely on the use of semi-parametric model and proportional hazards Cox regression model, but a parametric survival model assumes that the survival time follows a known distribution. Many models using different distribution have been developed. The commonly applied models are lognormal, weibull and log-logistic models.

3.5.6.1 Weibull RegressionModel

The Weibull distribution is worldwide used to model life data. The distribution can handle increasing, decreasing or constant failure-rates and can be created for data with and without suspensions (no failures). The Weibull distribution is flexible and fits to a wide range of data, including Normal distributed data. Only log–normal data does not fit in the Weibull distribution and needs separate analyses. For creating the plot you need to record the time to failure that can be expressed in mileage, cycles, minutes, strength, stiffness or similar continuous parameters. Suppose that survival times are assumed to have a Weibull distribution with scale parameter and shape parameter, the Weibull density function can be expressed as:

$$f(t,\mu,\alpha) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \exp\left(\left(-\frac{t}{\mu}\right)^{\alpha}\right), \text{ where } \mu > 0 \text{ and } \alpha > 0 \qquad (15)$$

And the baseline hazard of this model for the i^{th} subject is:

$$h_0(t_i;\mathbf{X}) = \frac{\alpha}{\mu}\left(\frac{t}{\mu}\right)^{\alpha-1} \qquad (16)$$

Independent observation (t_i,δ_i), $i=1,2,\ldots,n$ with survival time t_i, and censoring indicator δ_i which has value of one if i^{th} observation is not censored and zero when the i^{th} observation is censored and let β be the unknown parameter. The likelihood function is:

$$L(\beta) = \prod_{i=1}^{n} \left\{ f(t_i)^{\delta_i} \left(s(t_i)\right)^{1-\delta_i} \right\} = \prod_{i=1}^{n} \left\{ \left(\frac{f(t_i)}{s(t_i)} \right)^{\delta_i} s(t_i) \right\} \quad (17)$$

Parameter zings the Weibull distribution using $\lambda = \mu - \alpha$then$h_o = \lambda\alpha t\alpha - 1$ would be the baseline hazard function. Now incorporate covariates matrix X in the hazard function the Weibull regression model becomes:

$$h(t; \mathbf{X}) = \lambda\alpha t\alpha - 1 \exp(\mathbf{X\beta}) \quad (18)$$

The model assumes that individuals' i and j with covariates $X_i and X_j$ have proportional hazard functions of the form:

$$\frac{h(t; X_i)}{h(t; X_j)} = \frac{\exp(X_i\beta)}{\exp(X_j\beta)} = \exp((X_i - X_j)\beta) \quad (19)$$

The quantities $\exp(\beta)$can be interpreted as hazard ratios.

A different parameterization is used with intercept v and scaled parameter σ and covariate effects γ_j having relationship with original parameterization as $\beta = -\frac{\gamma j}{\sigma}, \alpha = \sigma - 1$ and $\mu = \exp(v)$.

3.5.6.2 Log-Logistic RegressionModel

An alternative model to the Weibull distribution is the Log logistic distribution. The log-logistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, and increasing, as well as hump-shaped that is, initially increases and then decreases **[29]**. The distribution imposes the following functional forms on the density, survival and hazard.

3.5.6.3Log Normal Regression Model

Skew distributions, low values mean high variance and non-negative values like species variety, distribution of minerals in earth's crust and generally based on lognormal distribution are commonly used for responding to stimulant biologic substances, most types of survival data, time distribution of hardware repair, financial researchers and studying the load price. As the application of lognormal distribution was tested, there were some examples of geology, metallurgy, health, environment, ecology, linguistics, social sciences

and economics [42].The log-normal model assumes that $\epsilon \sim N(0,1)$. Let h(t) be the hazard function of T for

$$Qs = U'_{H_0} I^{-1}{}_{pxp}(\hat{\beta} = 0) U_{H_0} \sim \chi^2 \quad (20)$$

When $\beta = 0$ i.e. $\beta_0 = \beta_1 = \cdots \beta_p = 0$. then, it can be shown that h(t) has the following functional form:

$$h(t) = \frac{\phi(\log(t)/\delta)}{[1-\Phi(\log(t)/\delta)]} \quad (21)$$

Where, $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp({}^{-t^2}/_2)$ is the probability density function, and $\Phi(t) = \int_{-\infty!}^{t} \frac{1}{\sqrt{2\pi}} \exp\left({}^{-u^2}/_2\right) du$ is the cumulative distribution function of the standard normal distribution. Then, the log-hazard function of T at any covariate value X can be expressed as:

$$\log h(t|X) = \log h_0\left(te^{-\beta' x}\right) - \beta' x \quad (22)$$

Obviously we no longer have a proportional hazards model. If the baseline hazard function is desired, it can be obtained from equation **(**22**)** by setting x = 0. The survival function S (t|X) at any covariate X can be expressed as:

$$S(t|x) = \phi[\beta_0^* + \beta_1^* + \cdots + \beta_p^* x_p - \alpha \log(t)] \quad (23)$$

Where, $\alpha = {}^{1}/_{\delta}, \beta_j^* = {}^{\beta_j}/_{\delta} for j = 0,1, \dots, p$. This is the final survival model with intercept depending, with t.

3.5.6 Model Selection

Akaike Information Criterion (AIC)

To select the model that can predict the survival time CKD, we would use Akaikie information criterion (AIC). Akaikie [**35**] proposed an informative criterion (AIC) statistic to compare different non-nested models. For survival model the valueof AIC is computed as:

AIC=−2LogL+2(k+c+1), (24)

Where k is the number of covariates and c the number of model specific distributional parameters. This thesis used the AIC to compare various candidates of non- nested parametric models.

3.5.7 Model Diagnostics

A. The Cox Snell Residuals

For the parametric regression problem, analogs of the semi parametric residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates **[26,37**]. The first such residual is the Cox–Snell residual that provides a check of the overall fit of the model. The Cox–Snell residual, r_j, is defined by:

$$r_j=\widehat{H}(T_j \mid X_j) \tag{25}$$

where $\widehat{H}$ is the cumulative hazard function of the fitted model. If the model fits the data, then the r_j's should have a standard (λ =1) exponential distribution, so that a hazard plot of r_j versus the Nelson–Aalen estimator of the cumulative hazard of the r_j's should be a straight line with slope 1. For the three baseline hazard functions considered in this thesis, the Cox–Snell residuals are:

$$\text{Forweibull, } r_j= \hat{\lambda}\exp(\widehat{\beta}'X_j)t_j^{\rho}, \tag{26}$$

$$\text{For log-logistic, } r_j = \ln\left(\frac{1}{1+\hat{\lambda}\exp(\widehat{\beta}'X_j)t_j^{\rho}}\right) \tag{27}$$

If the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be negative.

B. Deviance Residual

The i^{th}*deviance residual*, denoted by *Di*, is the square root of the i^{th} term of the deviance, augmented by the sign of the$\widehat{mi}$:

$$D_i = sign(\widehat{m}_i)x\sqrt{-2(\log\left(\hat{L}_i(\tilde{\mu}_i,\hat{\delta})\right) - \log(\widehat{L_{si}}))} \tag{28}$$

These residuals are expected to be symmetrically distributed about zero. Hence, their plot is easier to interpret. But we caution these do not necessarily sum to zero.

The model deviance is:-

$$D = \sum_{i=1}^{n} D_i^{\,2} \;=\; \text{the sum of the squared deviance residuals}$$

C. Coefficient of Determination (R^2)

In proportional hazards regression model as in all regression analyses there is no single, simple method of calculating and interpreting R^2, because in this model, R^2 depends on the proportion of the censored observations in the data. A perfectly adequate model may have what, at face value, seems like a terribly low R^2 due to high percent of censored data [**28**]. Cox and Snell [**37**]proposed model assessment using R^2 similar to the one used in linear regression which is given by:

$$R^2 = 1 - exp\left[\frac{2}{n}(LL_0 - LL_\beta)\right] \qquad (29)$$

Where LL_o is the log likelihood for zero models or without covariates, LL_β is the log likelihood including covariates, n is the number of subjects included in the study.

3.5.8 ETHICAL CONSIDERATION

The Ethical clearance was taken from Jimma University, department of Statistics. The official ethical clearance was obtained also from Adama Hospital medical college. Careful recruitment and training for data collectors was carry out. To maintain the confidentiality, the data collector and the supervisor extracted the necessary data from the patient baseline and follow up card. The data obtained was code carefully for the analysis.

4 RESULTS AND DISCUSSION

4.1 Demographic Details and Characteristics of Patients

In this study, the totals of 500 ESRD patients were considered.Among those patients 72.40% and 27.60% were alive and died respectively. The medical cards of those patients were reviewed, 169 (33.80%) were males from them 32 (18.93%) were dead and 331 (66.20%) were females while106 (32.02%) were dead. It also shows that there are more female ESRD patients than male patients.

Table 4.1 Summary of covariates characteristics and associated with ESRD Patients dataset from May 2012 to April 2016 at AHMC.

			Status of patient	
Variable	Categories	Frequency (present %)	Death (%)	Censored (%)
Sex	Male	169(33.80%)	32(18.93%)	137(81.07%)
	Female	331(66.20%)	106(32.02%)	225(67.98%)
Residence	Rural	203(40.60%)	46(22.66%)	157(77.34%)
	Urban	297(59.40%)	92 (30.98%)	205(69.02%)
Hypertension	no	120(24%)	1(0.83%)	119(99.17%)
disease	yes	380(76%)	137(36.05%)	243(63.95%)
Diabetic	No	208(41.60%)	1(0.48%)	207(99.52%)
mellitus(type	type I	110(22.00%)	44(40.00%)	66(60.00%)
of diabetic)	type II	182(36.40%)	93(51.10%)	89(48.90%)
Family history	Absent	141(28.20%)	27(19.15%)	114(80.85%)
	Present	359(71.80%)	111(30.92%)	248(69.08%)
Vascular	fistula	172(34.40%)	52(30.23%)	120(69.77%)
access	catheter	132(26.40%)	33(25.00%)	99(75.00%)
	graft	196(39.20%)	53(27.04%)	143(72.96%)
Obesity	Absent	157(31.40%)	1(0.64%)	156(99.36%)
	Present	343(68.60%)	137(39.94%)	206(60.06%)

Different covariates characteristics are displayed in the above Table 4.1.Vascular -access 82.27%, Residents of patients 53.64%, Family history 50.07 %, Obesity 40.58 % and Hypertension disease 36.88%, were caused for the death of the patients.

The total patients those lived in urban were 297 (59.40%) among those 92 (30.98%) were dead. the patients who had complications with Hypertension disease were 380(76%), among those 137(36.05%) were dead. On another hand, the ESRD patients their BMI (obesity) had > 30Kg/m were 343(68.60%) among those 137(39.94%) were dead. Diabetic mellitus (type of diabetic) was categorized into No diabetic, type I and type II. Among those three categories who had no diabetic patients were 208(41.60%), whereas 110(22.00%) and 182(36.40%) were those who included under the categories of type I and type II respectively. Further, from the patients those had no diabetic 0.48% of them were died at the same time as the category of type I and II 40.00% and 51.10% of the patients were died respectively.

Among 500 ESRD patients, 359(71.80%) of them had family history of kidney disease (FH-KD)and141(28.20%) of them had no family history of kidney disease (FH-KD).Thus, the death percentile was higher for those patients who had family history of kidney disease111 (30.92%), while smaller for patients with no family history of kidney disease (FH-KD) 27(19.15%).

Moreover, vascular access (the method used to treat patients) was categorized into fistula, catheter and graft. As it has been observed that 172(34.40%) patients were those who got treatment through fistula, likewise 132(26.40%) and 196(39.20%) of them were those who treated through catheter and graft respectively. On another hand, among the patients those who treated through the above three methods 52(30.23%) of them the patients who got treatment through fistula and died. similarly, among the treated patients through catheter 33(25.00%) and graft 53(27.04%) of them were also died respectively.

Table 4.2 Descriptive output for the survival time

		Censored		Mean		Median	
Total (n)	#of events	Total	Percent	Estimate	St. Err	Estimate	St.Err
500	138	362	72.40%	1295.6	35.5	1680	58.4

St. Err: standard error, # = number

According to the above table 4.2 showed the mean survival time of the patients to death is 1295.6 days with standard error of 35.5. Likewise the median of follow up time for dead patient was 1680 days with standard deviation of 58.4.Moreover the mean age of the patients was also 53.34 years with standard deviation of 0.71 and median age was 54.50. On the other hand, as table 4.3 in the appendix revealed that maximum age of the patients was 91 years old while the minimum age of the patients was 18 years old.

As shown in figure 4.1, about the overall result of the study illustrated that 138 (28.00%) patients had diedand 362 (72.00%) of them censored until the end of the study.

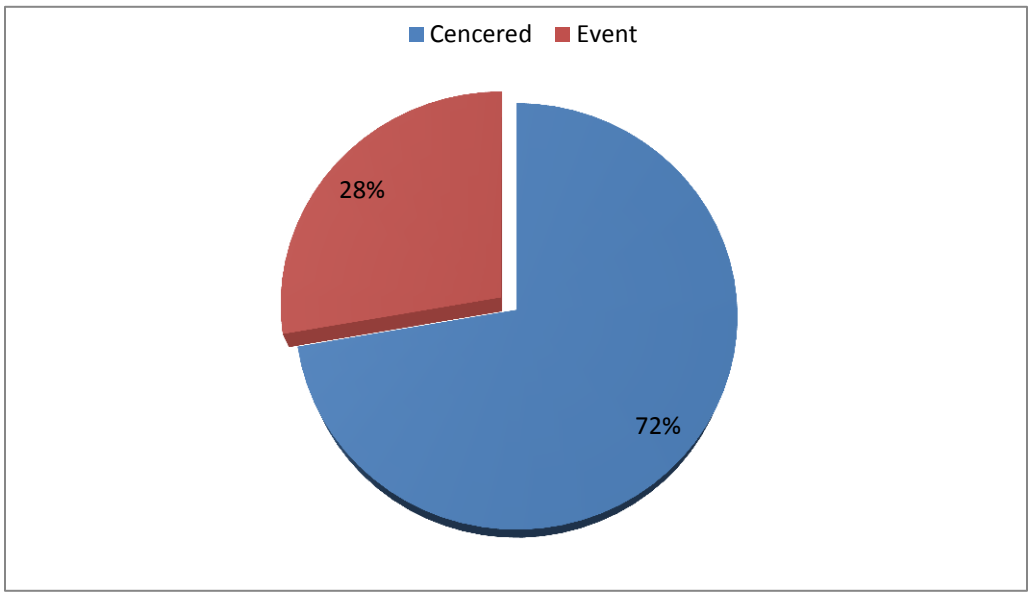

Figure 1 pie chart of the censored and event

4.2. Non-parametric Survival Analysis

4.2.1. The Kaplan- Meier Estimate of ESRD patients

Non-parametric survival analysis is very important to visualize the survival of time -to-death of the ESRD patient's different levels of the covariate. Moreover, it gives information on the shape of the survival and hazard functions of ESRD data set. Survival time distributions for time-to-death was estimated for each covariates group using the K-M method and in order to compare the survival curves of two or more groups, log-rank test has been in use.

Table 4.4 Mean survival time(CI) and long rank test of the ESRD dataset

			Log rank test		
Variable	**Categories**	**Mean [95%CI]**	**Chi.sq**	**Df.**	**P-value**
Age		53.34 [51.9,54.7]	149	69	0.000
Sex	Male	1445.78 [1336,1555]			
	Female	1223.73 [1137,1310]	8.2	1	0.004
	Rural	1381.95 [1275,1488]			
Residence	Urban	1250.38 [1161,1339]	2.3	1	0.128
Hypertension	No	1782.31 [1747,1816]			
disease	Yes	1167.00 [1085,1248]	45.1	1	0.000
Diabetic	No	1790.86 [1773,1808]			
mellitus	type one	1040.88 [891, 1190]	129	2	0.000
	type two	923.49 [804, 1042]			
Family	Absent	1452.62 [1337,1567]			
history	Present	1235.00 [1150,1319]	7.2	1	0.007
Vascular	fistula	1249.45 [1128,1370]			
access	catheter	1339.79 [1207,1472]	0.9	2	0.624
	graft	1308.83 [1198,1419]			
Obesity	Absent	1788.17 [1765,1811]			
	Present	1107.46 [1021,1193]	68	1	0.000
Total	Overall mean	1295.59 [1226,1365]			

The estimated mean time (std.err) long rank test and 95% confidence interval for time-to-death patients with different covariates characteristics are summarized in the above Table 4.4. Accordingly, the mean survival time of male patients' to death had been (1445.78 days) greater than females 1223.73 days. On the other hand the mean survival time to death of

ESRD patients lived in the rural 1381.95days, (95% CI: [1275, 1488]) greater than from those of lived in the 1250.38days, with its (95% CI: [1275, 1488]). The mean survival time to death of ESRD patients, those who had no hypertension disease (1782.31 days) were greater than from those ESRD patients with hypertension disease (1167days) respectively. Regarding the mean survival time to death (1235 days)of the patients who had family history of kidney disease [FH-KD] was less than those patients those had no family history of kidney disease [FH-KD].

The mean survival time of the patients who had no diabetic mellitus were 1790.86which is greater than those of categorized under type I diabetic (1040.88) and type II diabetic (923.49)days respectively. The mean survival time of patients' for vascular access user those treated through catheter (1339.79) were greater than those treated through graft (1308.83) and fistula (1249.45)days respectively. Generally, the overall mean of survival time of ESRD patients which indicated in the above table 4.4 was 1295.59days, with (95% CI: [1226, 1365]).

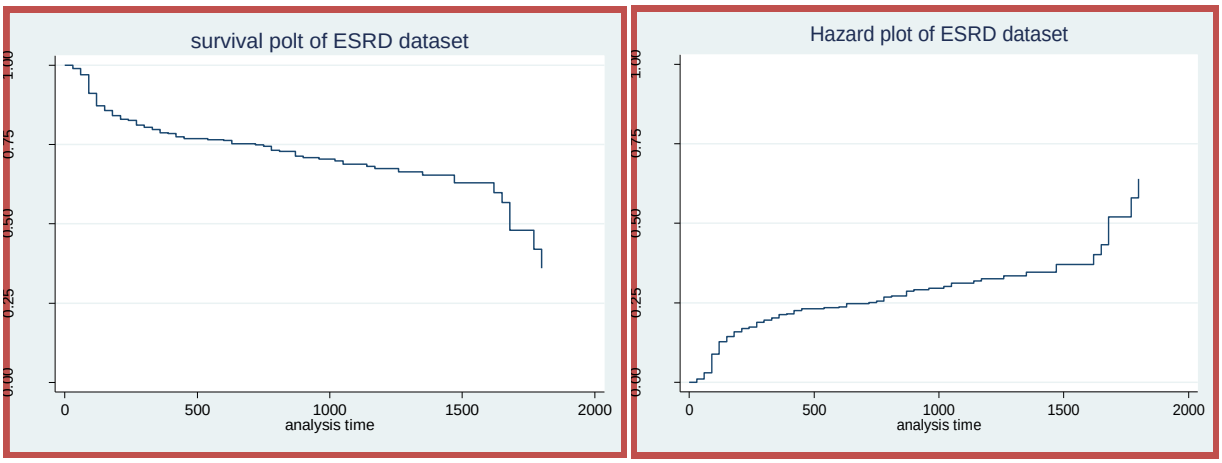

Figure 4.2: K-M plots of Survival and hazard functions of ESRD patients

Plots of the KM curves to the survival and hazard experience of time- to- death is shown in figure 4.2. The survival plot decreases at increasing rate at the beginning and decreases at decreasing rate latter. Thus, it implies that to investigate the significance differences between the survival probabilities of patients by ESRD dataset.

4.2.2. Kaplan Meier survival curve of ESRD patients by different covariates

4.2.2.1. **Kaplan Meier curve forSex of patients**

Figure 4.3 shows the survival functions according to sex of the patients ESRD Kaplan-Meier survivor estimates for the two sex groups are plotted. The plots stand for the probability of survival of patients with respect to sex it showed that males had higher survival time than female patients. The result of the log rank test Table 4.4 also revealed the some idea ($p<0.05$) means this difference is statistically significant.

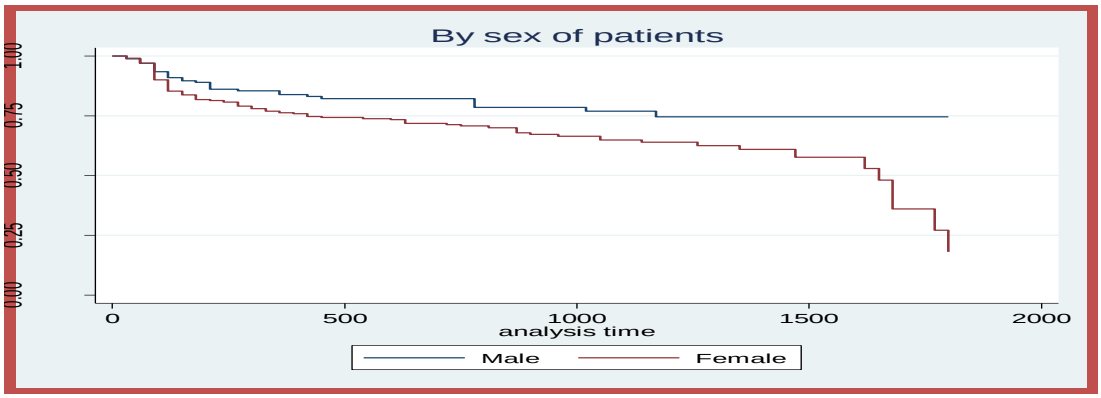

Figure 4.3K- M plot survivals Time-to death by sex of ESRD patients

4.2.2.2. Kaplan Meier curve for their Residence of patients

The survival plot for time-to-death of patients by place of residence is shown in figure 4.4. The plot indicates that the risk of interval is similar for both patients lived in rural and urban at the beginning of the curve. But, the difference becomes started at the middle of the curve. The survival plot ESRD of patients those lived in rural is above that of urban lived. This implied that the survival of ESRD patients those lived rural is higher than urban However, the difference in survival were not supported by Statistical tests, since log-rank test in Table 4.4also exposed the some idea ($p>0.05$) showed that there is insignificant difference between rural and urban lived patients with respect to survival time.

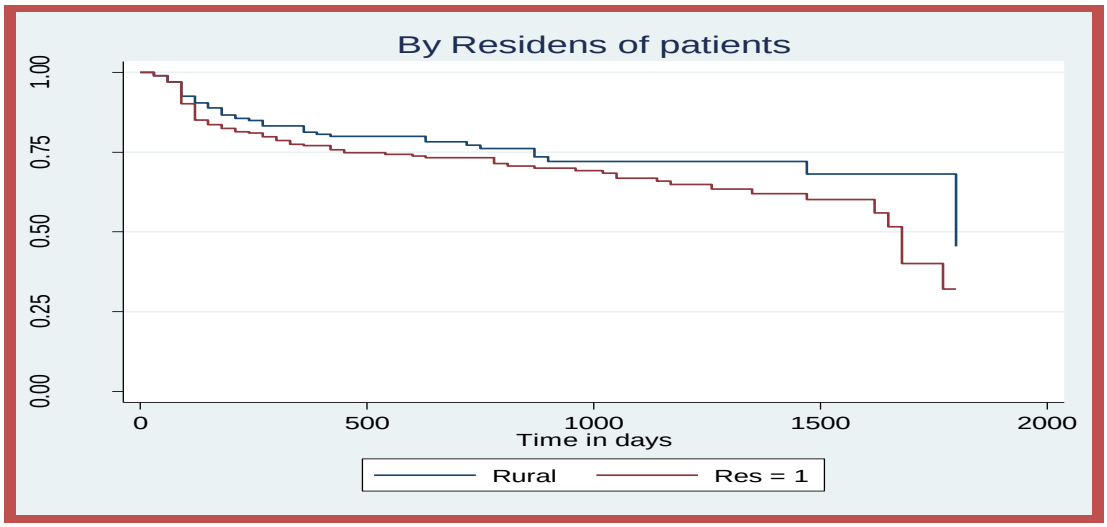

Figure 4.4K- M plot survivals Time -to -death by Resides of patients

4.2.2.3. Kaplan Meier curve for Hypertension diseases patients

Figure 4.5 showed the survival functions with respect to ESRD with hypertension diseases of the patients. Kaplan-Meier Estimates is represented by the survival curves for without hypertension diseases are above those the patients' complications with hypertension. This implied that the patients without hypertension more survival than with hypertension. The log rank test in Table 4.4 also revealed that hypertension diseases had significant association to time- to- death of ESRD patients ($p<0.05$).

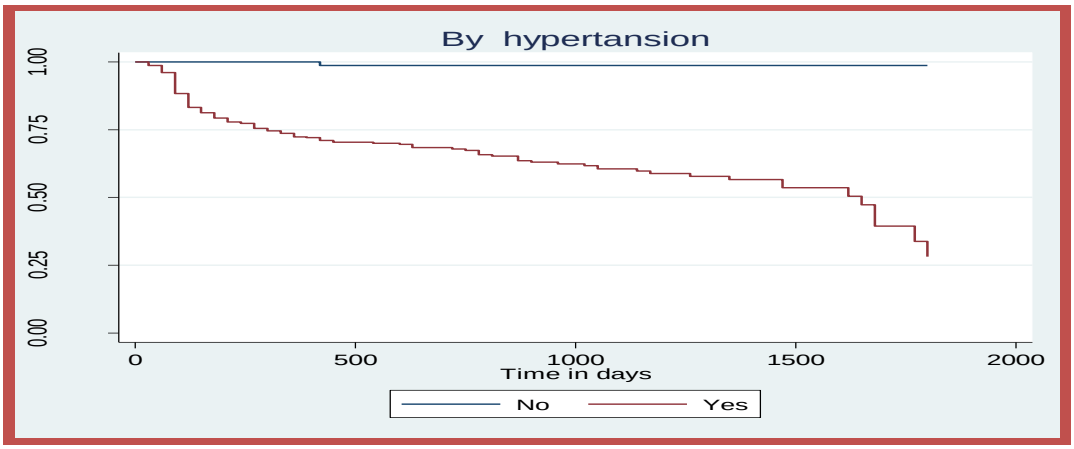

Figure 4.5K- M survival curve by Hypertension diseases

4.2.2.4. Kaplan Meier curve for diabetic mellitus (type of diabetic) patients

The Figure 4.6 of survival functions of time-to-death of patients in ESRD with barrier diabetic mellitus (type of diabetic)showed below. It indicates that the Kaplan-Meier Estimates which represented by the survival curves for patients' without diabetic mellitus is above the rest of patients' with diabeticsmellitus.Type I and type II diabetic becomes cross each other at the end of the plot. The differences that are displayed in survival curve emphasize that the survival time of patients' without diabetic the longest than with diabetic patients. The result of the log rank test in Table 4.4 revealed the difference is significant ($p<0.05$).

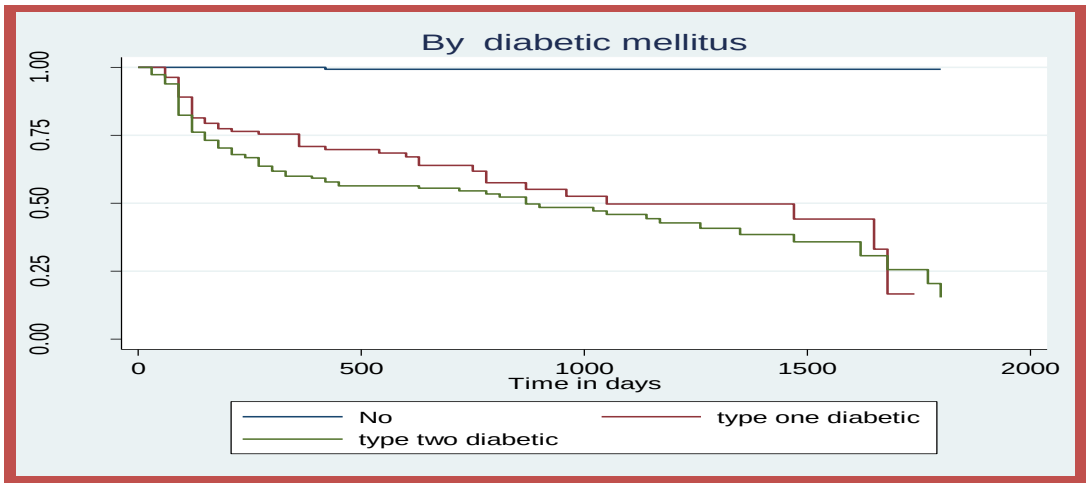

Figure 4.6 K- M survival plot time to death by diabetic mellitus

4.2.2.5. Kaplan Meier curve for Family history of kidney disease

Kaplan-Meier survival estimates are to comparing the survivor function between who had family history of kidney disease and who had not are family history of kidney disease were plotted in (Figure 4.7 in the appendix). The Figure showed the patients who had not family history of kidney disease were higher survival probability than those who had FH-KD. Statistical test is made by using log-rank test in Table 4.4 seen the difference is significant ($P<0.05$) it shown there is significant difference between patients who had family history of kidney disease and do not have FH-KD.

4.2.2.6. Kaplan Meier Curve For Types of Treatment Taken

The other categorical variable included in this study was Vascular Access method; Kaplan-Meier survivor estimate for this covariate was plotted in Figure 4.8 in the appendix. This plot suggested that the risk of death is similar for the three types (Fistula, Catheter, and Graft) at the beginning and little beat difference is observed at the middle of the curve and it becomes separate at end of the plot. But, the fistula is survival plot is below that of the Catheter, and Graft used of treatment. This implied that the risk of death fistula higher than catheter, and graft used of treatment and the result of the log rank test in Table 4.4 insignificant at ($p > 0.05$); it showed the vascular access taken treatment of patient had no significant difference for the survival of patients with ESRD.

4.2.2.7. Kaplan Meier survival curve for ESRD of the patients withObesity

Figure 4.9 in the appendix; Kaplan-Meier survivals estimates are to comparing the survivor functions between the ESRD patients who had obesity and had not are plotted. The plot showed as the survival curve for the ESRD patients who had notObesity is above those the patient had Obesity. This implied that the patients who had notObesity more survival than who had Obesity. And also the log rank test in Table 4.4; Obesity had significant difference for the survival of patients with ESRD patients at ($p < 0.05$).

4.3 Semi-Parametric Cox Ph Regression Model

4.3.1 Univariate Analysis of Cox Ph RegressionModel

The relationship between each predictor and survival probability of ESRD patients after doing individual covariate Cox proportional hazards model analysis is given in Table 4.5 in the appendix. Univariable analysis is an appropriate approach that is used to identify out potentially important variables before directly included in the multivariate model. As can be seen from this Table, Univariate Cox proportional hazards regression analysis revealed that Age, sex of patients, and Family history of kidney disease were significant risk factors for all-cause mortalityrelated to ESRD while Residence and Vascular access were not significant

at level of significance. To deal with multiple covariates it is necessary to use the modest level (25%)significance level.

4.3.2. Multivariate Analysis ofCoxPhRegression Model

From the p values of the output of individual covariate analysis at 10% significance level some important predictor variables were ignored and excluded from the model. So, multiple covariates analysis must be done to check whether the excluded variables in single covariate analysis are significant in the inclusion of collection of variables.When include all the covariate those are significant and insignificant in the Univariate analysis andcheck the below Table showed as the summary of multivariate analysis of Cox ph regression model.

Table 4.6 Summary of Multivariate Cox PH Analysis to the ESRD dataset from AHMC

Covariate	Category	coef (β)	HR	se(coef)	Wald	Pr(>|z|)	[95%Cl HR]
Age		0.05	1.04	0.006	58.5	0.000	[1.0 1.1]
Sex	(0)male	Ref					
	(1)female	0.55	1.7	0.2	7.4	0.006	[1.0 1.1]
Residence	(0)Rural	Ref					
	(1)urban	0.21	1.23	0.18	1.3	0.257	[0.9,1.8]
Family history	(0)Absent	Ref					
	(1)Present	0.63	1.88	0.22	8.5	0.004	[1.2 2.9]
Vascular access	(1) Fistula	Ref					
	(2)Catheter	-0.18	0.83	0.22	0.68	0.409	[0.5 1.3]
	(3) Graft	-0.21	0.81	0.2	1.1	0.293	[0.6 1.2]

Ref=reference, Coef= coefficient, se(coef) = standard error of coefficient; HR= hazard ratio; p=value significant at ≤ 0.01level of significance,

Covariates which become insignificant in the multivariate analysis were removed from the model by using stepwise elimination technique. Accordingly, residence and vascular Access were excluded. To check the assumption of ph and finally, the effect of interactions terms were also tested and found to be statistically insignificant in multivariable Cox ph model at 5% level of significance see in (Table 4.8 in the appendix). The final model kept the main effect of the covariate age, sex of patients, family history kidney disease patients.

4.3.2.1Assessment of Model Adequacy

Test of Goodness fit: Rho is a relation between time and residuals. The test of correlation (rho) is insignificant that indicates proportional hazards assumption is fulfilled. Variables Age, sex of patients, and family history of kidney disease are fulfilled the assumption because all the p values are greater than 0.01. In Schoenfeld if the p value is greater than 0.01or 0.05 it indicates that the Cox proportional hazards assumptions are fulfilled. Moreover it is also possible to see its global test and if it is greater than 0.01or 0.05 the assumption have satisfied by the covariates in the model.

Graphically test ph Assumption: From figure 4.9 in appendixes. The variables included in the final model that means sex of patients, Age of the patients, and family history of patients are fulfilled the proportional hazards assumption because$\ln(-lnS(t))$ vs. $\ln(time)$ are no cross each other and slightly parallel. In (Table 4.8 in the appendix) it showed thatthe time-dependent covariates (interaction of covariates with logarithm of time) were not significant which justifies the proportional hazard assumption holds at 5% level of significance.

4.3.2.2 Interpretation and Presentation Final Model of Cox Ph Analysis

Model adequacies are presented in section 4.3.2.1. It suggested that the model is in good fit. Thus the Cox regression coefficients in the final model are interpreted as follows.

After adjusting other covariates, the hazard rate for being ESRD of a patient with Age, sex and family history in multivariate Cox models, age of the patients a significant [HR = 1.04, P = <0.01or p<0.05, (95% CI;[1.03, 1.05])], when the age of the patients increases at ten years the hazard ratio is $[HR = e^{(\beta age)} = e^{0.05age*10} = 1.649age \approx 1.65]$ increase by 65%. Thus, in this study, age was one of the most important contributors to patient hazard in ESRD patients. The hazard of ESRD female patients was 70.00% greater than male patients. The hazard of Family history is 88% greater than no Family [HR=1.88, (90% CI;[1.2, 2.9]), p<0.01].

4.4 Parametric Regression Analysis

For the data on ESRD patients the parametric models were fitted. The common applicable criterion to select the model is the Akaikie information criterion (AIC) statistic proposed by Akaikie (1972)**.** From Table 4.9 the lognormal regression model has the least AIC value which shows that the lognormal regression model well fitted to data ESRD patients and The Cox snell plot in Figure 4.4.1 & Figure 4.4.2 (in section 4.4.2 detail) also indicates the lognormal model fit the data is better than the rest parametric models.

Table 4.9 The AIC value for different parametric regression models

Models type	Weibull	Log-logistic	Lognormal
AIC values	721.4	710	703.9
Log-Likelihood	-350.7	-345	-341.9

AIC=Akaike's information criteria

As lognormal regression is selected, according to the lognormal analysis of single covariate, the selected risk factors for further analysis and interpretation are made here below in table 4.11 and also see Table 4.10in the appendix presents the results in detail. From the Table 4.10 the Log normal regression model has the least AIC value which showed that the Log normal regression model well fitted to data of ESRD. Hazard ratio for Cox regression and Weibull models, and related risk for lognormal and log-logistic were estimated.

Based on section 4.4.2; both criteria (AIC and Cox-Snell residual), lognormal survival model had the best fit compared to other parametric models, following with Weibull, log-logistic and Cox ph regression (Table 4.10 and Table 4.12 in the appendix). Started the lognormal model, age of patients (HR=0.94, p-value < 0.05), Female (HR=0.5, p-value <0.05), family history(HR=0.5, p-value < 0.05), Residence (HR=0.7 p-value>0.05).The AIC value results indicated the general preference of parametric models over Cox semi-parametric model because Cox had the heights' AIC value. Among parametric models, log-logistic, lognormal and Weibull models close values, even though the lognormal model had the lowest AIC value. Thus, lognormal model was the best fit model over the data of ESRD survival time in

this study; age, sex of patients and family history of kidney disease were significant in all models. These variables were introduced into the model as covariates in multivariate analysis. This implies they are factors on survival time of patients.

Table 4.11ANOVA table for Univariate covariates in lognormal regression model

Variables	Df	Deviance	Resid.Df	-2*LL	Pr(>chi)
Null	NA	NA	498	2383.796	NA
Age	1	65.87	497	2317.9	$4.81e\text{-}^{16}$
as.factor(Sex)1	1	5.805	496	2312.1	$1.60e\text{-}^{02}$
as.factor(Family)1	1	9.42	495	2302.7	$2.14e^{-03}$

*NA=Not applicable, DF=*Degrees *of freedom LL=Log likelihood*

As we can see in Table 4.11 showed the risk factors which are associated with the ESRD were age, sex of patients, and family history of kidney disease. The LRT per the ANOVA function, with a p -value of 1.6×10^{-02}, provides strong evidence the lognormal model with the predictor variable femaleis adequate while the p -value of 2.14×10^{-03} for the estimated coefficient of family historyof kidney disease provides the evidence.Consistently, the p -value of 4.81×10^{-16} for the estimated coefficient of age of the patients provides this strongly evidence.

4.4.1 LogNormal Regression Analysis

Multivariate lognormal regression analysis, including variables with p-values less than 0.05 on Univariate analysis, established that age, sex of patients and Family history were significant independent determinants at significance level (Table 4.12 in the appendix). Results presented in the following table, it indicate the parameter estimates of coefficients for the covariates in the final lognormal regression model along with the associated standard error, significance level, Hazard ratio (related risk)and 95% confidence interval of the related risk. Survival time of ESRD patients were significantly related with age, sex of patients, and family history of kidney disease can be seen in the below Table 4.13. The Wald statistics for the parameter estimates indicate that at least one of the parameters in each covariate was significantly different from zero at 0.05 levels of significance. From,the lognormal

regression model, after fixing other coefficients, theHazard ratioof a patient with the related riskof ESRDage of patients is0.94. The related riskof ESRD female patients is decreases0.54times than ESRD male patients. The related riskof Family history of kidney disease is decreases0.45times than no Family historyof kidney disease.

Table 4.13 Multivariate analysis in the Log normalregression model of ESRD dataset

Covariate	Category	$\hat{\beta}$	Std. Err.	Wald	RR	P>|z|	[95% CI RR]
Age		-0.06	0.01	53	0.94	0.000	[0.93 0.96]
Sex	Male	Ref					
	Female	-0.63	0.13	6.2	0.53	0.013	[0.33 0.88]
Family history	Absent	Ref					
	Present	-0.80	0.12	9.1	0.45	0.003	[0.27 0.76]

β= coefficient, se(coef) = standard error of coefficient; RR=relative risk*; p=value significantat $\leq$ 0.01level of significance*

4.4.2Model Diagnostics

The Cox Snell Residuals Plots

The Cox snell plot indicates the data is better fitted by lognormal distribution from the parametric models and also Cox ph model because the 45^0 lines are more expressed by lognormal distribution. Another common applicable method to select a model that fit the data is the AIC statistic given above table 4.9.In this studying through Cox snell plotmethod over time, stating that the hazard ratio is constant.

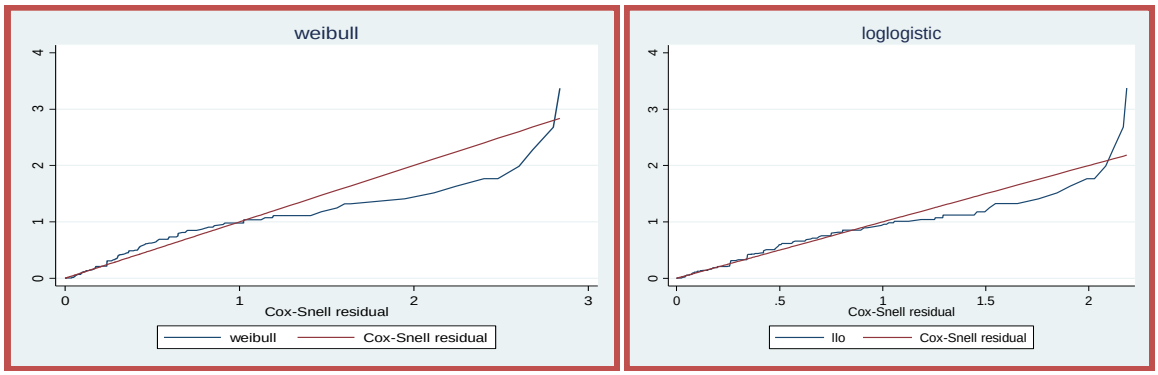

Figure 4.4.1 Weibull and log-logistic Cox snell residual plot for the ESRD data.

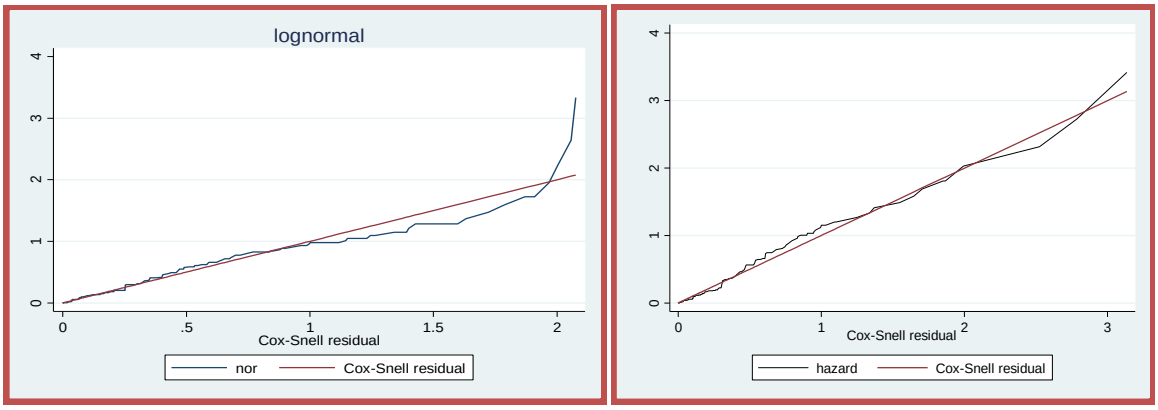

Figure 4.4.2 Lognormal and Cox regression Cox snell residual plot for the ESRD data.

Deviance Residual Plots

As Figure 4.4.3 below The Plots of Cox ph regression, Weibull, Log-logistic, and Log-normal baseline distributions for survival time-to ESRD data set is deviance residual against the linear predication shows that the deviance residuals seem to be approximately symmetric distributed about zero and there exists no clearly outlying observation. Therefore, I have almost some concern about the adequacy of the fitted Cox PH model.

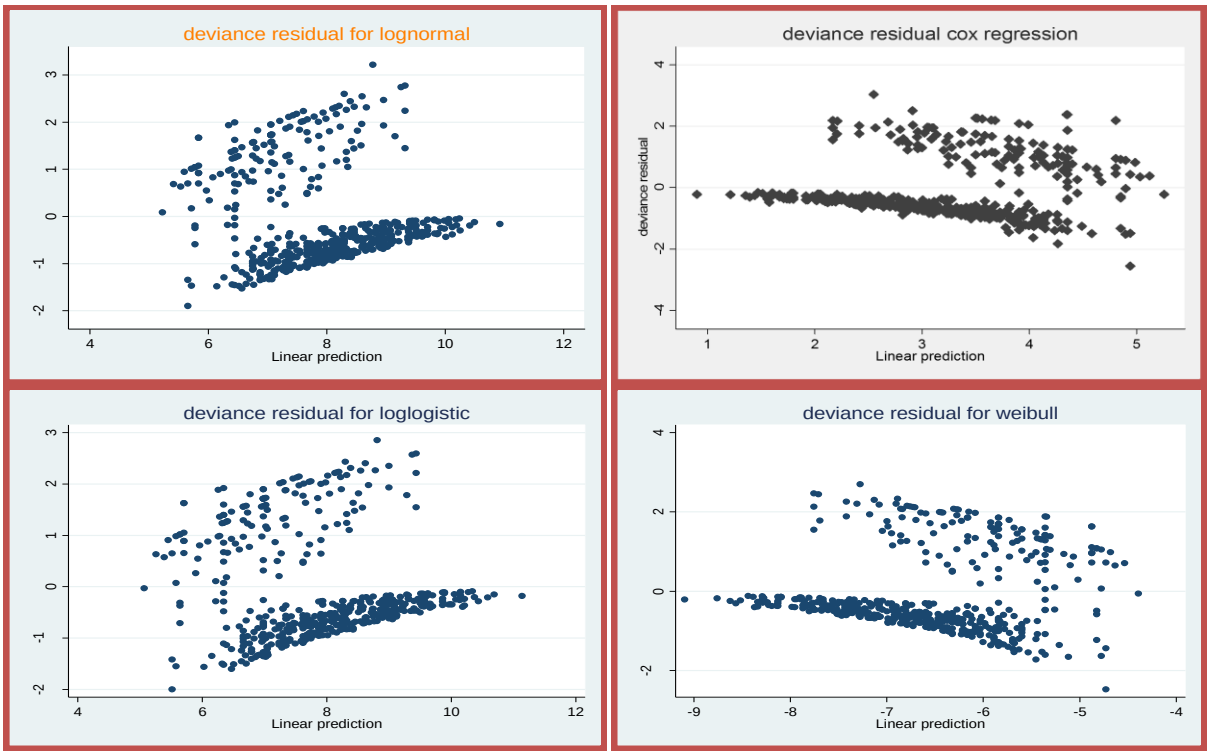

Figure 4.4.3 Plots of Cox ph regression, Weibull, Log-logistic, and Log-normal baseline distributions for survival time-to ESRD data set.

Table 4.14 The Likelihood ratio and significance of the lognormal regression model

Log-lik (LL_0)	Log-lik (LLβ)	χ^2	Df	P value	Scale	Intercept
-1151.3	-1058.5	81.1	3	0.000	1.88	12

LL_0;log-likelihood intercept, LLβ; log-likelihood model

From the likelihood ratio test Table 4.14, it can be seen that the model is significant and in using the log likelihood values of the null model and the full model it can be seen that the model has a significant.There is an improvement in this model because the log-likelihood has increased from -1151.3 to -1058.5 when covariates are included.

Where, LL_0= -1151.3, LLβ= -1058.5, n=500, then the equation (29) is :

$$R^2=1-exp\left[\frac{2}{n}(LL_0 - LL_\beta)\right] = 1 - \exp\left[\frac{1}{500}(-1151.3 + 1058.5)\right] = 0.16906192$$

$0.16906192 * 100 = 16.906192 \sim 16.91$, therefore, R^2=16.91% this indicates the model is fitted well because R^2 is small. In survival analysis the coefficient of determination is expected to be small because of there are a lot of censors in such type of data. In this paper also coefficient of determination is small and the study concludes it to be good fitted model.

As it is observed above the lognormal distribution is selected for this data because of its AIC is small compared to the other three models and this is also supported by Cox snell on the graph above. Using the regression model of equation (22) and with the parameters found, the survival time of ESRD patients with lognormal distribution can expressed; $\alpha = {}^{1}/_{\delta} = \frac{1}{1.88} = 0.531914893 \sim 0.53$,$\beta_j^* = {}^{\beta_j}/_{\delta} = \frac{\beta_0}{\delta} = \frac{12}{1.88} = 6.38297872 \sim 0.64$, j=0….p

4.5 DISCUSSION

In this study to estimate and compare the Cox ph regression and parametric survival models by considering three baseline distributions: Weibull, log-logistic, and log-normal distributions.For this study the source of data were in a single center study with 500 common and occurrence end stage renal disease patients. It presented the results of 5 years patient survival and risk factors of mortality during a mean of follow-up of 1295.6 days. Covariate which were included in the study were age, sex, residence, hypertension ,diabetic mellitus, family history, vascular access and obesity of patients and the outcome variable of interest were the survival of time-to-death of the ESRD patients in days.

Most of the patients with ESRD at the hospital had prior diagnosis of diabetes mellitus and followed hypertension with complexity obesity.It is even more significant for developing countries which now face the double burden of infectious diseases and growing problems of non-communicable diseases such as obesity, diabetes and hypertension **[40].** Other studies in Ethiopia like at St Gabriel General Hospital and at Butajira hospital, and in Africa were found to be the commonest cause of CKD [**1, 38,44**] as there was no routine renal biopsy and other necessary investigations for diagnosis of renal diseases, the real causes of ESRD in this setting cannot be easily stated **[38].** In addition, findings from this study can never be used as a representative data for the country as most of the participants were from Adama town and its rural where the living standard is better and hygiene is good.

The Cox's proportional hazard model fitted using complete case analysis found five variables that can serve as predictive factors on the survival analysis of with ESRD. These are age, sex of the patients, residence, family history of kidney disease and vascular access. With regard to the parametric regression models also included in this study which do not assume constant baseline hazard except for exponential regression model. This is consistent with findings from other studies. **[35, 39-41]**

Univariate Cox proportional hazards regression analysis revealed that age, sex of patients, and Family history of kidney disease were significant risk factors for all-cause mortality with ESRD while Residence and Vascular access were not significant at 1% significancelevel. In multivariate Cox models, age of the patients a significant [HR = 1.04, P = <0.01or p<0.05, 95% CI: (1.03, 1.05)]. As for risk factors, this study found a significant association between older age and CKD whatever then equations were used. This is consistent with findings from other studies**.** It was reported that the estimated Glomerular Filtration Rate (eGFR) diminishes with age, and when age increases the hazard of the patients will have increases **[1, 39]**. Thus, to screen kidney in this age is an important strategy for the detection of CKD and to improve the outcomes.

The hazard of ESRD female patients was 70.00% greater than male patients. Sex differences have a significant association with ESRD in this study when renal function was assessed by the in Univariate and Multivariate Cox models. But, not with the Japanese's study in which the male gender was reported to be a non-modifiable risk factor for CKD [**47]**. Furthermore there was a strong association between female sex and CKD in the UK and Sweden studies [**45, 46**]. The possible explanation for this might be due to the higher proportion of obese females than males.

The hazard of Family history kidney disease was 88% greater than no Family history [HR=1.88, 90% CI:(1.2, 2.9), p<0.01],There was a significant association between ESRD and the presence of longer duration a family history of kidney disease in this study in log rank test and Cox ph model. This corresponds with the findings of several studies, which

reported that the like hood of developing reduced eGFR was greater among patients with longer duration of diabetes and among those with or whose parents has kidney disease [**1,36-41].**

In other way, various parametric models like Weibull, lognormal and log-logistic are widely used for analyzing the survival data. These models can interpret the survival time based on a specific distribution irrespective of proportional hazard hypothesis. If the survival times use to analysis with parametric models will be stronger. This means that, under special conditions, parametric models such as Weibull, log-logistic and lognormal may have more accurate results than Cox model **[43].** Therefore: parametric, non-parametric or semi-parametric models. Moreover, it has more flexibility in adding covariates to the model. In this study, some variables that were significant through the log-rank test and had the proportional hazard assumption were introduced into the model.

By using AIC value, the models developed from Weibull, lognormal, log-logistic and Cox model were evaluated **[35].** Results indicated the general preference of parametric models over Cox semi-parametric model. Among parametric models, than others did from log-logistic, lognormal and Weibull models with close values, the lognormal model had the lowest AIC quantity. Thus, lognormal model was the best fit model over the data of ESRD survival time in this research. This result is in line with the study in **[43].**

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 CONCLUSIONS

Chronic kidney disease (CKD) with diagonesised end-stage renal disease (ESRD)is common public health problems worldwide. In this study, the totals of 500 chronic kidney diseases with diagnosed ESRD patients are considered. Among those patients 72.40% were alive andthe rest27.60%was diedat the observation of end of the study.Thus, the result having ESRD with complications increases the probability of death.

Based on the hospital outcome the most common causes of death were diabetic mellitus and hypertension; In addition, female patients had greater risk of death than male patients and based on the mean survival time age of patients greater than 53.34 years hadthe higher risk of death than thosehad mean survival time of age less than 53.34years. The estimated survival and hazard rates of ESRD patients under age, sex and Family history by Kaplan-Meier Method and log-rank test has no significant difference with p-values greater than 0.05.

The Cox ph regression analysis showed that the major factors that affect the ESRD are age, sex of the patients and Family history of kidney disease.The developed the hazard ratio of death in ESRD patients generated by Cox ph regression andrelative risk of patients by lognormal Distribution; at certain age of the patientsdetermination of medical experts to improve the quality of medications and technologies that resolve lessen the risk of death in ESRD and to patients with or without ESRD as a base of healthy lifestyle living.

5.2 RECOMMENDATIONS

Based on the findings of our study different factors were identified for the survival time of ESRD patients. In summary the key recommendations emerging from this study for policy makers, clinicians and the public at large are presented as follows:

For ESRD patients who have progressed to the need for dialysis, morbidity and mortality can be reduced and quality of life enhanced through faithfulness to an appropriate dietary and medical regimen, along with regular physical activity.

The concerned bodies should work on RRT clinics and dialysis center are not widespread in all corners the country. And it would be useful to initiate programs that emphasize coping strategies for improving the performance over and above testing and educational campaigns.

This study shows that main predictive factors for the survival time of ESRD patients are more health variables, so health workers should be cautious when a patient are with the proper control of blood pressure (hypertension), Body Mass Index (obesity) and blood glucose (diabetic mellitus), in combination with appropriate medications.

Future researchers may are warranted by including other factors such as, blood lipids, blood counts or tests, blood pressure and urine test as independent variables for the risk of death. This stud's suggest the use lognormal model and more samples from different hospitals that would make the model a general representation for estimating the risk of death.

5.3 LIMITATION OF THE STUDY

Although important findings were obtained from the present study, there are some limitations worth mentioning:-The study was based on chart revision where incomplete documentation, inappropriate chart labeling and lost records made it difficult to include all patients registered for dialysis.

Scope of the Study

This study was mainly focus on to apply survival techniques to model time-to-event in case of kidney diseases based on the data were obtained from the follow up record of the patients in AHMC.

REFERENCE

1. Temesgen F., Mehidi K, Tilahun Y. (2014). Prevalence of chronic kidney disease and associated risk factors among diabetic patients in southern Ethiopia, American Journal of Health Research, 2(4): 216-221(http://www.sciencepublishinggroup.com/j/ajhr)
2. James MT, Quan H, Tonelli M, et al. (2009 Jul); CKD and risk of hospitalization and death with pneumonia. American Journal of Kidney Diseases. 54(1):24-32. PMID19447535
3. Atkins RC, Zimmet P. World Kidney Day 2010 Feb: Diabetic Kidney Disease—Act Now or Pay Later. Am J Kidney Dis. 55(2):205–8?
4. Zilisteanu, D.S.(2013)Late nephrology referral and impact on morbidity and mortality of patients with chronic renal disease, Ph.D. Thesis, Carol Davila University of Medicine and Pharmacy, Bucharest.
5. National Kidney Foundation. (2002 Feb); K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis.39 (2 Suppl 1):S1-266. PMID 11904577
6. USRDS. United States Renal Data System 2008 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; www.usrds.org/adr_2008.htm. Accessed on Jan. 2010
7. Gilbertson DT, Liu J, Xue JL, et al. (2015); projecting the number of patients with end stage renal disease in the United States to the year. Journal of the American Society of Nephrology; 16(12):3736-41. PMID 16267160
8. USRDS. United States Renal Data System 2010 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.
9. Zewdu W. (1994) Acute renal failure in Addis Ababa, Ethiopia: a prospective study of 136 patients. Ethiop Med J; 32(2):79-87.
10. Wali RK. (2010) Aspirin and the prevention of cardiovascular disease in chronic kidney disease: time to move forward? J Am Coll Cardiol; 56:966–8.
11. Schieppati A, Remuzzi G. (2003) The future of renoprotection: frustration and promises. *Kidney Int* ; 64: 1947–1955

12. Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, Chen M, He Q, Liao Y, Yu X, Chen N, Zhang JE, Hu Z, Liu F, Hong D, Ma L, Liu H, Zhou X, Chen J, Pan L, Chen W, Wang W, Li X, Wang H; (2012)Prevalence of chronic kidney disease in China: a cross-sectional survey. *Lancet*, 379:815-822.
13. Randeree IG, Czarnocki A, Moodley J, Seedat YK, Naiker IP. (1995)Acute renal failure in pregnancy in South Africa. Ren Fail; 17(2):147-153
14. Cheng, C. L., Kao, Y. H., Lin, S. J., Lee, C. H. & Lai, M. L. (2011) Validation of the National Health Insurance Research Database with ischemic stroke cases in Taiwan. Pharmacoepidemiology and drug safety 20, 236–242
15. Naicker S. End-stage renal disease in sub-Saharan Africa. Ethn Dis 2009; 19(S1):13-15.
16. Kaplan, E. L., and P. Meier. (1958). Nonparametric estimation from incomplete observations.Journal of the American Statistical Association 53: 457–481.
17. Oakes D. (1977),The asymptotic information in censored survival data. Biometrika. 64,441-8
18. Lawless, J. F., (1982). Statistical Methods and Model for Lifetime Data; Wiley, New York
19. Kleinbaum D, Klein M. Survival analysis: A self- learning text, New York, Springer-Verlag. 2005
20. Anand S, Bitton A, Gaziano T: (2013) The gap between estimated incidence of end-stage renal disease and use of therapy. *PloS One*, 8:e72860
21. US Renal Data System. USRDS 2002 Annual Data Report. Bethesda: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.
22. BMC Cardiovascular Disorders 2009, 9:39 Diabetes Metab Syndr.2012 Jan-Mar; 6(1):36-41. Eth. Med J 2004; 42:17-22;
23. Weiner DE, Tabatabai S, Tighiouart H, et al. (2006 Sep); cardiovascular outcomes and all-cause mortality: exploring the interaction between CKD and cardiovascular disease. American Journal of Kidney Diseases; 48(3):392-401. PMID 16931212
24. Khatri M, Nickolas T, Moon YP, et al. (2009 Nov); CKD associates with cognitive decline. Journal of the American Society of Nephrology. 20(11):2427-32. PMID 19729443
25. Greenwood, M. The natural duration of cancer. Reports on Public Health and Medical Subjects 33 (1926), 1.26.75

26. Kalbfleisch, J. D., and Prentice, R. L. Marginal likelihoods based on Cox.s regression and life model. Biometrika 60 (1973), 267.278.
27. Cox, (1972). Regression models and life tables with discussion Journal of the Royal Statistical Society.
28. Hosmer DW, Lemeshow S. (1999) Regression modeling of time-to-event data. Wiley, NewYork.
29. David W. Hosmer, Stanley Lemshow, Susanne May, (2012). Survival analysis: Regression modeling of time to event data. 2nd ed. Wiley series
30. Kalbfleisch, J. D., and Prentice, R. L. (2002) The statistical Analysis of Failure Time Data, 2nd ed. Wiley, New York.
31. Kay R. and Kinnersley N., (2002). "On the use of the accelerated failure time model as an alternative to the proportional hazard model in the treatment of time to event data: A case study in Influenza", Drug information Journal, 36, 571-79.
32. Sedgwick, P. Survival (time to event) data: censored observations, British Medical Journal; Vol. 343, 2011, d4816
33. Boneg, Y. T. (2001). Weibull frailty for modeling heterogeneity; Unpublished Ph.D. Dissertation, University of Guelph,
34. Dätwyler, C. and Stucki, T. (2011). Parametric Survival Model. Handout
35. IBRAHIM, J. G., CHEN, M.-H., and SINHA, D. Bayesian Survival Analysis. Springer-Verlag, New York, 2001
36. Klein, J. P., and Moeschberger, M. L. (1997) Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York.
37. Therneau T, Grambsch P. (2000) Modeling survival data: Extending the Cox Model, New York, Springer- Verlag.
38. Fatiu A, Rashad S. (2008) CKD Prevention in Sub-Saharan Africa: A Call for Governmental, Nongovernmental, and Community Support. Am J Kidney Dis. 51:515–523.doi: 10.1053/j.ajkd.2007.12.006? [PubMed]
39. Go AS, Chertow GM, Fan D, et al. (2004(13) Sep 23) Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. New England Journal of Medicine; 351(13):1296-305. PMID 15385656

40. US Renal Data System. USRDS 2002 Annual Data Report. Bethesda: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.
41. Laird, N., and Olivier, D.(2012) Covariance analysis of censored survival data using log-linear analysis
42. McGilchrist, C. A. and Aisbett, C. W. (1991), Regression with frailty in survival analysis. Biometrics 47, 461-466.
43. Limpert, E., W.A. Stahel and M. Abbt, 2001. 24. Nardi, A. and M. Schemper, 2003. Log-normal distributions across the sciences: keys and clues. Biosciences, 51(5): 341-52.
44. Tamiru Sh., Esayas K., Belete H., Amare D. and Tewodros A.,**(**2013**)** Survival patterns of patients on maintenance Hemodialysis for end stage renal disease in Ethiopia: summary of 91 cases, BMC Nephrology. 14: 127, PMCID: PMC3693969
45. Retnakaran R, Cull CA, Thorne KI, Adler AI, Holman RR. (2006) Risk Factors for Renal Dysfunction in Type 2 Diabetes U.K. Prospective Diabetes Study 74. Diabetes. 55(6):1832–9.
46. Garg AX, Kiberd BA, Clark WF, Haynes RB, Clase CM. (2002) Albuminuria and renal insufficiency prevalence guides population screening: results from the NHANES III.Kidney Int. 61:2165–75.
47. Ohta M, Babazono T, Uchigata Y, Iwamoto Y. (2010) Comparison of the prevalence of chronic kidney disease in Japanese patients with Type 1 and Type 2 diabetes. Diabet Med. 27(9):1017–23.
48. Amir H. H., Behrouz B., Mansour R.andDariush R. (2013);A Comparison Between Cox Regression and Parametric Methods in Analyzing Kidney Transplant Survival**,** World Applied Sciences Journal 26 (4): 502-507,

APPENDIX

Table 4.3 Summary of Age of the patients						
	Media	min	max	mean	Std. Err	[95% C I]
Age	54.50	18.00	91.00	53.34	0.7107257	[51.94 54.74]

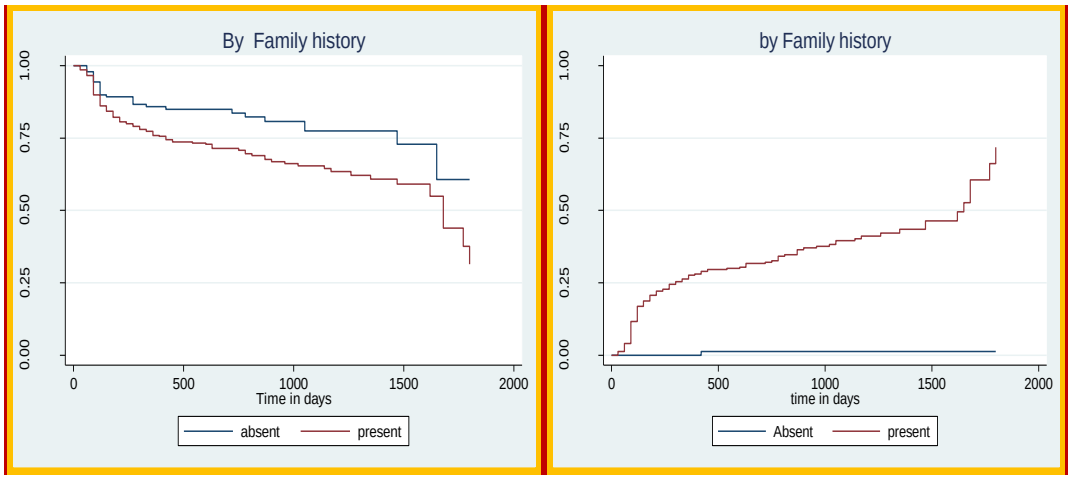

Figure 4.7 The Kaplan- Meier survival and Hazard curve by Family history

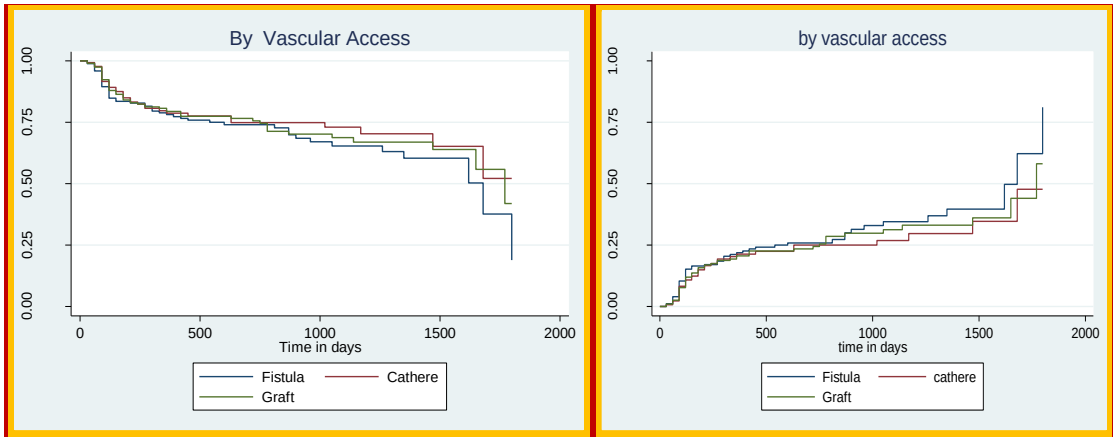

Figure 4.8 The Kaplan- Meier survival and Hazard curve of Vascular Access

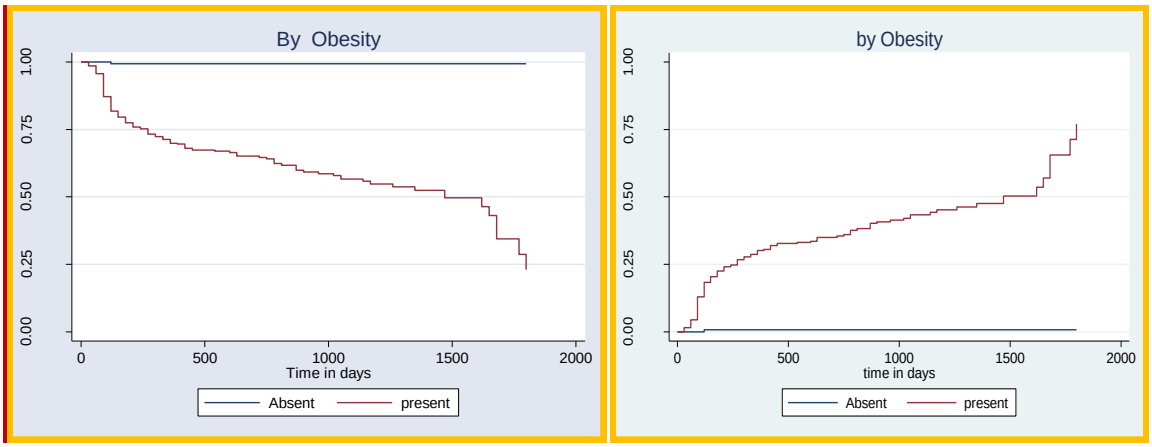

Figure 4.9 The Kaplan- Meier survival and Hazard curve of Obesity

Table 4.5 summary Univariate Cox PH Analysis to the ESRD dataset from the AHMC

Covariate	Category	coef(β)	HR	se(β)	Wald	Pr(>|z|)	[95%ClHR]
Age		0.05	1.0	0.01	56	0.00*	[1.03, 1.06]
Sex	(0)male	Ref					
	(1)female	0.57	1.8	0.2	8	0.005*	[1.2 , 2.6]
Residence	(0)Rural	Ref					
	(1)urban	0.28	1.3	0.18	2.3	0.126	[0.93, 1.88]
Family history	(0)Absent	Ref					
	(1)Present	0.56	1.8	0.21	6.97	0.008*	[1.16 , 2.68]
Vascular access	(1) Fistula	Ref					
	(2)Catheter	-0.21	0.8	0.22	0.87	0.35	[0.52 , 1.26]
	(3) Graft	-0.13	0.9	0.2	0.42	0.52	[0.60 , 1.29]

Ref=reference, Coef= coefficient, se(β) = standard error of coefficient; HR= hazard ratio ;p=value significant at ≤ 1% level of significance

Table 4.7 Schoenfeld residual for each covariate

Covariate	Category	Rho	χ^2	DF.	Prob>chi2
Age		-0.16	3.61	1	0.0574
Sex	Male	Ref			
	Female	0.11	1.81	1	0.18
Family history	Absent	Ref			
	Present	-0.002	0.00	1	0.9816
Global test		NA	5.29	3	0.152

NA=Not applicable, Df=Degrees of freedom

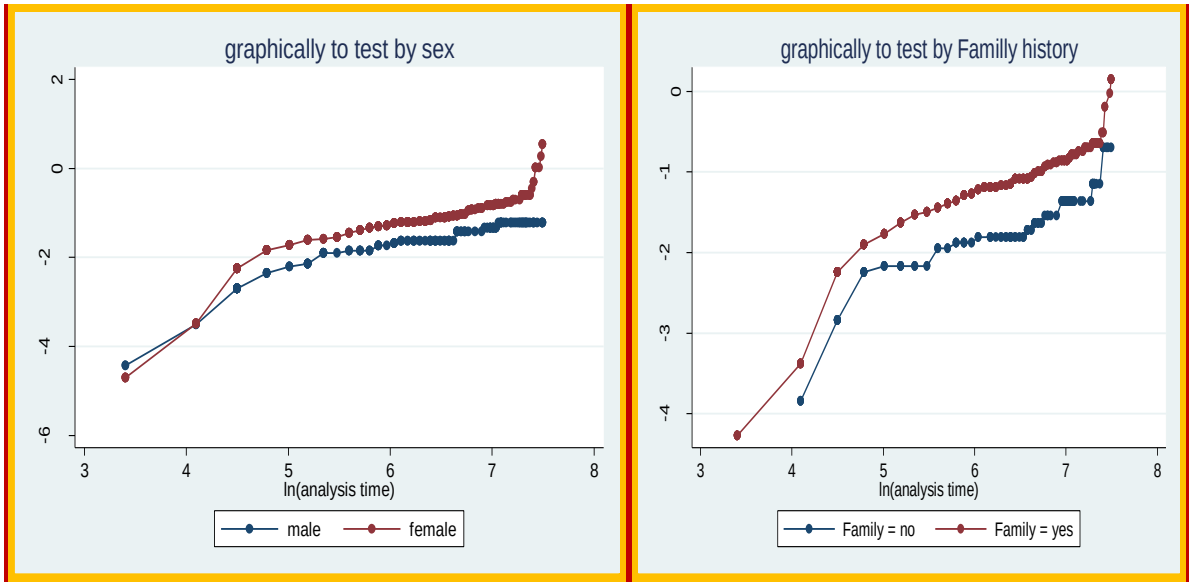

Figure 4.10graphically ph test assumption by sex and Family history kidney disease

Table 4.8 Statistical Test for Ph Assumption of the Covariates and Their Interaction with Log(Time)

Variables	coef (β)	**HR**	**se(coef)**	**Wald**	**Pr(>|z|)**
Age of the patients	4.58e-^{02}	1.047	3.63E-02	1.6	0.207
Sex (0)male	Ref	1	1	1	1
as. factor(Sex)1	-5.33e-^{01}	5.87e-^{01}	1.16	0.21	0.646
FH-KD (0)Absent	Ref	1	1	1	1
as.factor(Family)1	-1.322	2.67e^{-01}	1.279	1.07	0.301
Age: log(Time)	-5.76e-^{04}	9.99e-^{01}	6.79e-03	0.01	0.932
log(Time):as.factor(Sex)1	2.16e-^{01}	1.241	2.20e-01	0.96	0.326
log(Time):as.factor(Family)1	2.97e-^{01}	1.346	2.35e-01	1.6	0.206

FHKD: family history kidney disease, HR: hazard ratio

Table 4.10 Summary of Cox-**ph** and **Parametric models** for Univariate analysis in ESRD dataset from May 2012 to April 2016.

Variables	Cox-ph regression		Weibull		log-logistic		Lognormal	
	HR	P>|z|	HR	**P>|z|**	TR	P>|z|	TR	P>|z|
Age	1.04	0.00	1.05	0.00	0.9	0.00	0.94	0.00
Sex								
male	Ref	1	1	1	1	1	1	1
female	1.75	0.01	1.75	0.01	0.5	0.01	0.5	0.01
Residence								
Rural	Ref	1	1				1	1
urban	1.31	0.13	1.3	0.15	0.7	0.15	0.7	0.15
FH								
Absent	Ref	1	1	1	1	1	1	1
Present	1.75	0.01	1.8	0.01	0.4	0.00	0.5	0.01
VA								
Fistula	Ref	1	1	1	1	1	1	1
Catheter	0.81	0.36	0.83	0.41	1.3	0.46	1.3	0.46
Graft	0.88	0.52	0.89	0.55	1.2	0.57	1.2	0.55

HR: hazard ratio, RR: relative ration, VA: vascular access, FH: family history

Table 4.12 Summary of parametric estimatesUnivariate analysis in ESRD dataset from the AHMC

Variables	Cox-ph regression $\hat{\beta}$(95%CI $\hat{\beta}$)	weibull $\hat{\beta}$(95%CI $\widehat{\beta}$)	log-logistic $\hat{\beta}$(95%CI $\hat{\beta}$)	Lognormal $\hat{\beta}$(95%CI$\hat{\beta}$)
Age	0.044(0.03,0.06)	0.046(0.03,0.058)	-0.07(-0.08,-0.0.05)	-.061(-0.08,-0.04)
Sex				
male	Ref	1	1	1
female	0.560(0.17,0.96)	0.561(0.17,0.96)	-0.736 (-1.27,-0.2)	-0.66 (-1.2,-0.15)
Residence				
Rural	Ref	1	1	1
urban	0.27(-2.4,0.63)	0.26 (-0.09,0.62)	-0.37(-0.87,0.13)	-0.35(-0.84,0.13)
Family history				
Absent	Ref	1	1	1
Present	0.56 (0.14,0.98)	0.607 (0.18,1.03)	-0.83 (-1.39,-0.27)	-0.78(-1.35, -0.24)
Vascular access				
Fistula	Ref	1	1	1
Catheter	-0.21(-0.65,0.23)	-0.19 (-0.62,0.25)	0.23 (-0.39,0.85)	0.23(-0.37,0.83)
Graft	-0.12(-0.5,0.25)	-0.12 (-0.49,0.26)	0.16(-0.39,0.69)	0.16 (-0.37,0.69)

$\hat{\beta}$ Estimate (coefficient); 95%CI ; 95% of coefficientconfidence interval; Ref is reference