2018-07-18

JOINT MODELING ON LONGITUDINAL GLOMERULAR FILTRATION RATE MEASUREMENT AND TIME-TO-DEATH OF RENAL FAILURE PATIENTS TREATED UNDER HEMODIALYSIS: SEPARATE AND JOINT MODEL

HIWOT, ABEL

BAHIR DAR UNIVERSITY

COLLEGE OF SCIENCE

DEPARTMENT OF STATISTICS

JOINT MODELING ON LONGITUDINAL GLOMERULAR FILTRATION RATE MEASUREMENT AND TIME-TO-DEATH OF RENAL FAILURE PATIENTS TREATED UNDER HEMODIALYSIS: SEPARATE AND JOINT MODEL

By: HIWOT ABEL

ADVISOR: DEMEKE LAKEW (Ass. Prof.)

A THESIS TO BE SUBMITTED TO THE DEPARTMENT OF STATISTICS IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS (BIOSTATISTICS)

JUNE, 2018

BAHIR DAR, ETHIOPIA

Joint Modeling on Longitudinal Glomerular Filtration Rate Measurement and Time-to-Death of Renal Failure Patients Treated Under Hemodialysis: A Comparison of Separate and Joint Model

By:

Hiwot Abel

A Thesis to be Submitted to the Department of Statistics in Partial Fulfilment of the Requirements for the Degree of

Master of Science

In

Statistics (Biostatistics)

Department of Statistics

College of Science

Bahir Dar University

June 2018

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 Bahir Dar University in the partial fulfilment for the Degree of Master of Science in statistics (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.

Declared by:

Hiwot Abel _____ _____

Name signature date

Place of submission: Department of Statistics, College of Sciences, Bahir Dar University

Date of submission: June, 2018

This thesis has been submitted for examination with my approval as a University advisor.

Demeke Lakew (Asst. prof.) _____ _____

Advisor signature date

APPROVAL SHEET

This is to certify that, the thesis prepared by Hiwot Abel, entitled: Joint Modeling on Longitudinal Glomerular Filtration Rate Measurement and Time-to-Death of Renal Failure Patients Treated Under Hemodialysis: A Comparison of Separate and Joint Model , and submitted in partial fulfillment of the requirements for the Degree of Masters of Science in Statistics (Biostatistics) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Approved by the Board of Examiners:

_____	_____	_____
chairperson	signature	date
_____	_____	_____
Internal Examiner	Signature	Date
_____	_____	_____
External Examiner	Signature	Date

ACKNOWLEDGEMENT

This thesis could not have been accomplished without assistance of a number of persons and I would like to take these few lines to thank them warmly for their help.

First of all, I thank God for being with me in all aspects during my life and giving the opportunity to pursue my graduate study

My special gratitude goes to my advisor Demeke Lakew (Ass. prof.), for his invaluable suggestions and comments that contributed to the successful realization of this study. I also extend my sincere gratitude to my instructors, staffs of Statistics Department and friends for their unreserved knowledge sharing and cooperation.

Finally, my deepest and warm gratitude goes to my lovely mother Tejitu Kenna that has been a source of pride and encouragement throughout my work.

TABLE OF CONTENTS

LIST OF TABLE

List of figure

ABSTRACT

Background*: Renal failure is one among the slowly progressive diseases of kidney function characterized generally by low glomerular filtration (GRF). The replacement therapy of renal failure by hemodialysis involves the removal of excessive toxic fluids and toxic metabolic end products from the body. One continuous and five categorical predictors were included in the analysis. The mean baseline age of renal failure patients was 36.64 years. Out of 269 renal failure patients 118(43.87%) were female and 150(55.76%) were hypertensive. Joint models typically combine linear mixed effects models for repeated measurements and Cox models for censored survival outcomes. Thus, the aim of this study was to present joint modelling on longitudinal glomerular filtration rate measurement and time-to-death of renal failure patients treated under hemodialysis.*

Methods: *Hospital based retrospective study was conducted among renal failure patients attending hemodialysis between 2016 and 2018 at Saint Paulo's Hospital Millennium Medical College, Addis Ababa, Ethiopia. The longitudinal eGFR and the time to event (i.e. death) data with the separate modeling approach and the joint modeling approach was fitted. A total of 269 renal failure patients screened who were under hemodialysis follow-up at Saint Paulo's Hospital Millennium Medical College.*

Results: *The results for separate and joint models were quite similar to each other but not identical. However, the estimates of the association parameters in the joint analysis were significantly different from zero, providing evidence of association between the two sub-models. The relationship between kidney function as measured by eGFR and the hazard for death was negatively significant. Thus, death is less likely to occur in patients with higher eGFR.*

Conclusions: *When evaluating the overall performance of both the separate and joint models in terms of model parsimony, goodness of fit, smaller total AIC, and the statistical significance of both the association parameters, the joint model performs better. Thus, authors concluded that the joint model was preferred for simultaneous analyses of repeated measurement and survival data.*

Key words: *Hemodialysis, Chronic Kidney Disease, Joint Model, Longitudinal Data, Survival Data Cox PH model.*

ABBREVIATIONS AND ACRONYMS

AIC: Akakian Information Criteria

BUN: Blood Urea Nitrogen

CKD: Chronic Kidney Disease

CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration

eGFR: Estimated Glomerular Filtration Rate

ESRD: End-Stage Renal Disease

GEE: Generalized Estimating Equation

GLMs: Generalized Linear Models

HD: Hemodialysis

MDRD: Modification of Diet in Renal Disease

mGLMM: multivariate Generalized Linear Mixed Model

RRT: Renal Replacement Therapy

SAS: Statistical Analysis system

SCr: Serum Creatinine

SPSS: Statistical Package for Social Science

WHO: World Health Organization

CHAPTER ONE

1. INTRODUCTION

1.1.Background of the Study

Kidneys are bean-shaped organs that constitute the upper part of the urinary system. Their roles include filtering blood, i.e. removing waste products, and regulating blood volume and pressure, amongst others (Field et al., 2011). The glomerulus is the kidney's filtration unit. A single kidney includes about one million glomeruli. It is accepted that glomerular filtration rate (GFR) is the best overall measure of kidney function/ health (Stevens et al., 2006). Normal GFR values are expected to be approximately 130 ml/min/1.73m^2 of body-surface area for a young man, and 120 ml/min/1.73m^2 for a young woman (depending on age and body size). GFR less than 60 ml/min/1.73m^2 indicates chronic kidney disease (CKD), and GFR less than 15 ml/min/1.73m^2 indicates end-stage renal disease and preparation for renal replacement therapy (RRT), i.e. dialysis or transplantation. Direct measurement of GFR is expensive and difficult in routine clinical practice. Alternatively, estimated versions, called eGFR, are widely used. There are many formulae to obtain eGFR which combine kidney function biomarkers, e.g. serum creatinine (SCr), and demographic factors, e.g. gender, age and ethnicity, in a deterministic manner. Two popular ones are the Modification of Diet in Renal Disease (MDRD) (Levey et al., 1999)and Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI), (Levey et al., 2009). Here, kidney function biomarkers are easy to measure in a routine blood test, but are known to be subject to substantial measurement error, which is also inherited into eGFR. These noisy variables are biochemical that are expected to be removed from the body by the kidneys, but their levels in blood might also be associated with other factors. For example, SCr is a muscle breakdown product, and heavy drinkers are known to lose muscle. Such a person might have a higher level of SCr than expected, yet have a pair of well-functioning kidneys. Another example is that some drugs are known to inhibit creatinine clearance, and might lead to measuring higher levels of SCr in the tests (Stevens et al., 2006).

Hemodialysis is one of the renal replacement therapy (Nisha et al., 2017). The technique plays a vital role in the process for the extracorporeal removal of waste products such as creatinine, urea and free water from the blood, when the kidneys are impaired. The principle behind hemodialysis is the diffusion of solutes through a semi permeable membrane. Hemodialysis is usually performed

with uremic patients for two to three times a week and the required times for dialysis vary from two to four hours (Bloembergen et al., 1996). The difference in the time of dialysis depends on various factors, including kidney function, amount of waste in body, level of salts and body weight. Dialysis improves many symptoms of kidney failure, but some problems including hypertension, anemia and itch often require additional drug treatments as well (Tomás et al., 2008). While HD does not cure renal disease, its use does allow patients with ESRD to survive (Weisbord et al., 2007). Nevertheless, HD is a lifelong treatment that significantly and sometimes adversely affects patients both physically and mentally (Kimmel, 2001).

The progression of kidney damage is marked by the rise in two important chemical substances in the blood, Creatinine and Urea whose evaluation in Serum helps to assess Glomerular Filtration Rate (GFR) followed by renal function. However, neither Creatinine nor Urea is directly toxic and they are only a measure of kidney function (Devi et al., 2009). Creatinine is produced from muscles and is excreted through the kidneys along with other waste products. Creatinine concentration in serum is maintained by the balance between its generation and excretion by the kidneys. It has been estimated that 2% of the body's Creatine is converted into creatinine every day, resulting in the daily generation of creatinine at a fairly constant rate of 20 to 25 and 15 to 20 mg/kg/day for male and female respectively (Iseki et al., 1997). The quantity of creatinine in serum depends on their generation, glomerular filtration and tubular secretion of serum creatinine. Calculations based on serum creatinine and the age groups of the patient are used to estimate more precisely the degree of kidney function (Clark et al., 1998). These calculated values are called the estimated glomerular filtration rate or eGFR. Sometimes a 24 hour urine collection and blood test together are used to measure the efficiency of kidney's removal (or clearance) of Creatinine from the body. These results are known as Creatinine clearance (Yassin et al., 2014). Urea is an organic compound and plays a vital role in the metabolism of nitrogen-containing compounds. It is a waste product from dietary protein and is also filtered into urine by the kidneys (Dorgalaleh et al., 2013). Urea nitrogen is a normal waste nitrogen product found in blood that comes from the breakdown of protein from foods. Healthy kidneys remove urea nitrogen from blood, but the level of urea in blood rises with kidney failure occurs(Rusul Arif AA and S, 2014).

Renal failure is one among the slowly progressive diseases of kidney function characterized generally by low glomerular filtration (GRF). The replacement therapy of renal failure by hemodialysis involves the removal of excessive toxic fluids and toxic metabolic end products from

the body(Nisha et al., 2017). In renal failure dialysis, the patient must be connected to a machine that mechanically filters the blood. Dialysis does not treat renal failure, but instead keeps a person alive by performing the crucial functions of the kidneys. A person may have to undergo dialysis as often as several times a day or as little as weekly, depending on the severity of the renal failure. A person with acute, reversible renal failure may need dialysis while the kidneys recover (Tuso, 2009).

Chronic kidney disease (CKD) and end-stage renal disease (ESRD) are two common worldwide public health problems in recent years ((Modi and Jha, 2006). Hemodialysis (HD) is one of the main modalities of renal replacement therapy in ESRD patients (Mousavi et al., 2015).

CKD is increasingly recognized as a global public health problem (Eckardt et al., 2013). Kidney failure is the most severe form of CKD, and is fatal if not treated by renal replacement therapy (RRT), which can be dialysis or kidney transplantation. The prevalence and associated burden of CKD is rising worldwide; with the fastest growth occurring in low-income and middle-income countries(Vos et al., 2016).

In end stage renal disease (ESRD), the kidneys are not able to do the metabolic actions and create liquids and metabolites equilibrium in the body(Smeltzer et al., 2010). In such cases, the survival of the patient is only possible with a kidney transplant. Although there are geographical differences in the treatment of ESRD, hemodialysis remains the most common therapy for this disease (Clark et al., 1998). Due to increased access to dialysis, many ESRD patients now live longer than ever before. Worldwide, the ESRD-related mortality rate in Europe and Japan is relatively low, whereas it is very high in developing countries due to limited access to hemodialysis. In the U.S., the mortality rate of patients receiving dialysis is almost 18% - 20%, and their 5-year survival rate is about 30% - 35 % (Fauci A et al., 2008).

Chronic kidney disease (CKD) is an important challenge for health systems around the world (Foundation, 2002). Consuming a huge proportion of health care finances (Arogundade and Barsoum, 2008, Grassmann et al., 2005). It is even more significant for developing countries (Foundation, 2002) which now face the double burden of infectious diseases and growing problems of non-communicable diseases such as obesity, diabetes and hypertension (Krzesinski et al., 2006). About 85% of third world population where CKD prevention programs are either rudimentary or virtually nonexistent (Grassmann et al., 2005). Morbidities and mortalities

emanating from CKD in these countries are immense and related to limited access for treatment options (Arogundade and Barsoum, 2008). Renal replacement therapy (RRT) is the mainstay of care for patients with end stage renal disease (ESRD). Dialysis as an option of RRT prolongs survival, reduces morbidities and improves quality of life. However, despite many technical advances, morbidities and mortalities of patients on dialysis remain unacceptably high and their quality of life is often poor (Bethesda, 2003). Common independent predictors of survival are age, race, serum albumin at the start of dialysis, activity level at the start of dialysis, and presence of certain comorbidities such as heart failure and cancer (Bleyer et al., 1996).

Renal disease is common throughout the world. In the United States alone, almost 100,000 people began renal replacement therapy (RRT) for end-stage renal disease (ESRD) in 2001(Kimmel and Peterson, 2005); by 2008, this number had increased to 485,000 patients (Collins et al., 2010).More than 90% of these patients were started on hemodialysis (HD), while only 8.5% began RRT with peritoneal dialysis (PD) (Kimmel and Peterson, 2005). In Korea, the number of dialysis centers and machines has continuously increased, and 62.1% of patients receiving RRT were being treated with HD (Son et al., 2009). An international comparison showed that Taiwan has the greatest incidence and second-greatest prevalence of ESRD (Kuo et al., 2007). Furthermore, renal disease is one of the top 10 causes of death in Taiwan, and roughly 95% of ESRD patients are on HD (Hsieh et al., 2007).

Treatment options for CKD are not readily available for most countries in sub-Saharan Africa. The region contributes to less than 5% of patients on RRT worldwide (Jafar et al., 2006). Dialysis and transplant programs in this part of the world are dependent on the availability of external funding and donors. As a result, only less than 5% of patients with diagnosed ESRD are able to get treatment for longer than 3 months (Arogundade and Barsoum, 2008).

The beginning of dialysis activity in Africa was in 1957,(Barsoum, 2012) only 12 years following Willem Kolff's breakthrough in the Netherlands.(Barsoum et al., 2015)The sole general physician in Krugersdorp, a small town in South Africa, built the first dialysis machine in the continent, which was a cross between a Kolff coil and a rotating drum. He used it to treat 2 patients with acute renal failure and although both died shortly after, the event was a historic landmark. The next attempt was made a year later in Egypt. Professor Nagy El-Mahallawy of Ein-Shams University in Cairo imported a primitive Alwall dialyzer, which he used to treat a woman with

acute renal failure who died after a few sessions. ,(Barsoum, 2012)Efforts were resumed in both countries in 1962 to 1963, when both peritoneal dialysis (PD) and hemodialysis were used routinely for the management of acute renal failure and poisoning in Cairo and Johannesburg university hospitals. Two North African universities in Tunisia and Algeria and one in Kenya joined the club during the same period. The first patient in the continent to receive hemodialysis by a Scribner shunt was treated at Kasr-El-Aini medical school of Cairo University in February 1964. In the following years, dialysis services were started in Nigeria (1965), Sudan (1968), Libya (1972), Zimbabwe (1972), and Morocco (1977). Dialysis for the management of acute renal failure subsequently was adopted in other leading teaching institutions in the rest of Africa.(Barsoum, 2012).

Outcomes of dialysis in Africa generally are suboptimal, with annual survival ranging from 20% to 70% and relatively poor quality of life.(Arogundade et al., 2005) Both the unavailability and the inadequacy of dialysis services have been attributed to insufficient financial and human resources and illiteracy, in addition to malnutrition and concomitant infections such as hepatitis C virus in the Sahara(Barsoum, 2013) HIV (human immunodeficiency virus) in the Sub Sahara(Swanepoel et al., 2012)and tuberculosis(Shigidi et al., 2012) and parasitic infections(Barsoum, 1991) throughout the continent. In most countries, suboptimal primary care often fails to prevent and treat chronic non communicable conditions, such as diabetes, hypertension, and CKD complications.

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 socioeconomic 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. (Shibiru et al., 2013)

1.2. Statement of the Problem

Renal failure is one among the slowly progressive diseases of kidney function characterized generally by low glomerular filtration (GRF). The replacement therapy of renal failure by

hemodialysis involves the removal of excessive toxic fluids and toxic metabolic end products from the body (Nisha et al., 2017). If kidneys fail, it needs treatment to replace the work they normally do. The treatment options are dialysis or a kidney transplantation.

Patients who are on dialysis also face different problems that may end up with losing life. As far as research knowledge concerned, there is no study that has been conducted in Ethiopia that reports the potential factors for the death of renal failure HD patients. Knowing the causes of the death of renal failure patients will help in taking appropriate care for that problem. This enables to give attention due to these problems so that we can prolong the life of HD patients.

Many well established methods exist for analyzing longitudinal and survival data separately; including linear mixed effects models for longitudinal data, and parametric or semi-parametric models for survival data. But their separate use may be inappropriate when the longitudinal variable is correlated with patient health status, (Guo and Carlin, 2004) and unbiased statistical inferences are more likely to be obtained via a joint model (Tsiatis et al., 1995, Wulfsohn and Tsiatis, 1997). Joint models of longitudinal and survival data, on the other hand, incorporate all information simultaneously and provide valid and efficient inferences. But, by separate modeling, the interrelationships of the two responses cannot be investigated. For example, the GFR are measured at different times for each patient, hence, the GFR level changes from time to time for each patient and also there will be a considerable difference in the GFR fluctuations (variability) among subjects. Therefore, separate modeling would not able to examine the effect of these differences of the longitudinal response on the survival outcome but joint modeling does.

Asar et al., (2015) develop the joint model of longitudinal estimated GFR measurement and time to initiation of renal replacement therapy (RRT). And compare the results with those obtained by the more widely used approaches of conducting separate analyses of the repeated measurements and survival times based on a linear mixed effects model and a Cox model, respectively. They used Linear Mixed Model for the longitudinal part by analyzed log-transform eGFR(Y=log (eGFR)) data. In this study Linear Mixed Effect Model were used for the longitudinal part and Cox proportional hazard model for survival part.

The main aim of this thesis was studying the joint model of longitudinal estimated GFR measurement and time to death of renal failure patients treated under hemodialysis at Saint Paulo's

Hospital Millennium Medical College. Specifically determine factors influencing the survival time of renal failure patients and identifying the typical patterns of progression in estimated GFR.

1.3. Objective of the Study

1.3.1. General Objective

The general objective of this study was to investigate the survival status of renal failure patients treated under hemodialysis with joint model of estimated glomerular filtration rate at Saint Paulo's Hospital Addis Ababa.

1.3.2. Specific Objective

- To investigate factors influencing the survival time of renal failure patients in Saint Paulo's Hospital.
- Fit appropriate joint models of the repeated measurements and time-to-death from dialysis.
- To determine the prognostic factors affect the Progression rate of GFR.

1.4. Significance of the Study

The importance of this study was identifying the significant factors affecting the progression rate of estimated GFR and time to death of renal failure patients treated under hemodialysis. On top of this, it may help to provide information for health professionals, policy makers and other governmental and non-governmental organizations to give attention for this silent killer disease and minimize the risk of death in the country as well as in the study area. In addition, it helps as a reference for other researchers.

1.5. Limitation of the Study

The study is conducted based on secondary data which might have incomplete biased information and poor data recording on the different patient charts. The modeling problem is failing to convergence due to in inability of computer computing capacity. Thus, here in the separate and joint model, authors did not see the interaction effect of the predictors over time.

CHAPTER TWO

2. LITERATURE REVIEW

Renal failure refers to temporary or permanent damage of the kidneys that result in loss of normal kidney function. There are two different types of renal failure that are termed as acute and chronic. Acute renal failure has an abrupt onset and is potentially reversible. Chronic renal failure progresses slowly for at least three months and can lead to permanent renal failure. The causes, symptoms, treatments, and outcomes of acute and chronic are different (Krzesinski et al., 2006).

Dialysis is a procedure that is performed routinely on persons who suffer from acute or chronic renal failure, or who have end stage renal disease (ESRD). The process involves removing waste substances and fluid from the blood that are normally eliminated by the kidneys. Dialysis may also be used for individuals who have been exposed to or ingested toxic substances to prevent renal failure from occurring. There are two types of Dialysis treatments which are hemodialysis and peritoneal dialysis. People with ESRD are living longer than ever. Dialysis treatments (both hemodialysis and peritoneal dialysis) are not cures for ESRD, but will help you feel better and live longer. Over the years, ESRD can cause other problems such as bone disease, high blood pressure, nerve damage, and anemia (having too few red blood cells) (WHO, 2006).

Diseases of the kidneys are amongst the most important causes of death and disability in many countries throughout the world (Verma et al., 2006). Renal failure is a systemic disease and usually turns into a route cause for several kidney and urinary tract diseases. Renal failure induces a slow and progressive decline of kidney function enhanced by various factors including infections, auto immune diseases, diabetes and other endocrine disorders, cancer, and toxic chemicals (Chielle et al., 2015). It is usually a result of complications arising from other serious medical conditions. Unlike acute renal failure, which happens quickly and suddenly, chronic renal failure occurs gradually - over a period of weeks, months, or years - as the kidneys slowly stop working, leading to an end-stage renal disease (ESRD) (Rusul Arif AA and S, 2014). High blood pressure is one of the leading causes of kidney failure. It may also damage the blood vessels in the kidney affecting the secretion of waste products. Waste may secrete extra cellular fluids and further raise the blood pressure eventually leading to ESRD (Nisha et al., 2017). Anemia parallels the degree of renal impairment and it is a most important cause is the failure of renal erythropoietin secretion.

Ueda et al., (2003) investigated factors affecting progression of renal failure by measuring the doubling of serum creatinine (sCr) as an end point in cohort of 85 type 2 diabetic patients with chronic renal failure. According to their Cox proportional hazard analyses, lower serum albumin, lower hemoglobin, sex (male), and lack of insulin therapy was significant factors for the progression of renal failure. Kaplan-Meier analysis showed that patients without insulin therapy had significantly faster progression of renal failure. Due to multivariate Cox proportional hazards analysis, lower serum albumin concentration, lower hemoglobin value, higher mean blood pressure, and lack of use of insulin was significant factors favoring progression of renal failure (Ueda et al., 2003).

Younespour et al., (2016) was modeling joint of the longitudinal measurements of serum creatinine level and time-to-event data of time to graft failure. The data involved 413 renal transplantation patients and investigated the etiological role of recipient characteristics in serum creatinine changes within the follow-up period and in relation to the graft failure risk, as well as evaluated whether or not the serum creatinine level represents an indicator of graft failure following renal transplantation. According to the result of survival sub model, patients who received a living donor kidney had a higher risk of graft failure than patients who received a deceased donor kidney transplant. Also based on the results of the longitudinal sub-model, the serum creatinine values significantly decreased over time. The significant model association parameter revealed a positive association between the serum creatinine levels and graft failure, which means that graft failure is more likely to occur in patients with higher serum creatinine levels (Younespour et al., 2016).

According to Silins et al., (1989), 8432 patients' data in Canada was used to assess the mortality rates among patients with end stage renal disease (ESRD). By using life table method they estimated the probability of dying during the first 5 years after registration from the registered data. Univariate analyses was performed to determine the probability of dying by risk factor. The risk factors studied was age, sex, race, blood type, year of registration, primary diagnosis, treatment, and transplantation history. Significant differences in the probability of dying was found between those with and without diabetes mellitus, between those who had received a renal transplant and those who had not, between white and nonwhite patients and between various age groups. Primary diagnosis and treatment was significantly associated with the risk of dying among the ESRD patients. For those who had received a transplant, the length of time spent waiting for a transplant was positively associated with the risk of death from ESRD. Patients who had received

peritoneal dialysis before transplantation had a higher risk of death than those who had received either hemodialysis or transplantation as the first treatment. Three survival models namely Cox's, exponential, and weibull, was compared in assessing the effect of each risk factor on mortality. Results from the exponential model are presented, as it agreed excellently with but was simpler than the Cox proportional hazards and weibull models (Silins et al., 1989).

A study was carried out by Montaseri et al., (2016) on the application of Parametric Model to investigate Survival in Hemodialysis Patients and used 270 hemodialysis patients' data who were referred to Imam Khomeini and Fatima Zahra hospitals between November 2007 and November 2012. Also used the Akaike information criterion (AIC) and residuals to compare the performance of the parametric models. According to the results of a multivariate analysis of the variables in the parametric models the mean serum albumin and the clinic attended was the most important predictors in the survival of the hemodialysis patients. Among the parametric models tested, the results indicated that the performance of the Weibull model was the highest. The Weibull model seemed to show the best fit among the parametric models of the survival of hemodialysis patients (Montaseri et al., 2016).

Shibiru et al., (2013) used Kaplan-Meier survival analysis to assess survival patterns of patients on maintenance hemodialysis for end stage renal disease. The study was conducted at Saint Gabriel General Hospital, Ethiopia. They used descriptive statistics and chi-square tests to test the association among different variables. According to the result Kaplan-Meier survival analysis showed that type of vascular access used and erythropoietin treatment are significantly affected both short term and long term survival patterns (Shibiru et al., 2013).

Sá Carvalho et al., (2003) used Cox proportional hazards model to analyze survival of 11579 patients on hemodialysis in 67 health facilities in Rio de Janeiro state. They applied a frailty random effect models to investigate difference in mortality between health centers not explained by measured characteristics. According to the result the variables that are significantly associated with outcomes is age and diabetes. And there is significant frailty effects among centers (Sá Carvalho et al., 2003).

Vahedi et al., (2016) used data from kidney center of Hasheminejad in Terhan, Iran. The study was aimed to provide the survival analysis of maintenance hemodialysis patients using different parametric survival model. The different parametric survival models (exponential weibull,

gompertz, lognormal and log-logistic) was compared by using AIC and cox-snell residual. According to AIC and cox-snell residual weibull survival model manifested better results as compared to others. And they reported age at the time of admission, walking ability, diabetes mellitus, hemoglobin level, serum creatinine, and serum protein as a significant effect on survival of the hemodialysis (Vahedi et al., 2016).

Jaffa et al., (2015) also applied multivariate Generalized Linear Mixed Models (mGLMM) to investigate multiple longitudinal kidney function outcomes collected over 3 years on a cohort of 110 renal transplantation patients. They determined the correlated outcomes blood urea nitrogen (BUN), serum creatinine (Cr), and estimated glomerular filtration rate (eGFR) and the effects of various covariates such as gender, age, and race by using different mGLMMs. And assessed different mGLMMs such as shared random intercept (SHRI), shred random intercept and slope (SHRIS), separate random intercept (SPRI), separate random intercept and slope (SPRIS) to identify the one that has the best fit and most accurate estimates. Also conducted a bootstrap pseudo-simulation to gauge the tradeoff between the complexity and accuracy of the models. According to the result SPRI provided most accurate estimates (Jaffa et al., 2015).

Asar et al., (2015) develop the joint modelling framework and compare the results with those obtained by the more widely used approaches of conducting separate analyses of the repeated measurements and survival times based on a linear mixed effects model and a Cox model, respectively. They used a data set from the Chronic Renal Insufficiency Standards Implementation Study (CRISIS) run by Salford Royal NHS Foundation Trust (SRFT). According to the result, the conventional linear mixed effects model and the longitudinal component of the joint model was found to be similar. However, there was considerable differences between the results for the Cox model with time-varying covariate and the time-to-event component of the joint model. In Cox model analysis the relationship between kidney function as measured by eGFR and the hazard for initiation of RRT was significantly underestimated that treats eGFR as a time-varying covariate, because the Cox model does not take measurement error in eGFR into account. Regarding to these result Joint models should be preferred for simultaneous analyses of repeated measurement and survival data, especially when the former is measured with error and the association between the underlying error-free measurement process and the hazard for survival is of scientific interest (Asar et al., 2015).

2.1.Joint Model

Joint models are a class of models to describe the joint behavior of a biomarker process (longitudinal response) and an associated survival process, where the biomarker process is observed at a series of times and the survival process gives rise to censored event times. When association between the two processes exists, less biased and more efficient inferences will be obtained by using joint model (Guo and Carlin, 2004) and unbiased statistical inferences are more likely to be obtained via a joint model (Tsiatis et al., 1995, Wulfsohn and Tsiatis, 1997). Also the use of separate models, in such cases, can cause biased estimates (Ratcliffe et al., 2004), with joint models resulting in a reduction in the standard error of estimates. Thus, with more accurate parameter estimates, valid inferences concerning the effect of covariates on the longitudinal and survival processes can be obtained through joint models.

In general, a joint model consists of two sub models, which (Henderson et al., 2000)referred to as the measurement model for the longitudinal process, and the intensity model for the survival process, and a *latent association* function of the random effects in which the two sub models are linked. And these two processes are assumed to be conditionally independent given unobserved random effects (Wulfsohn and Tsiatis, 1997,)

CHAPTER THREE

3. DATA AND METHODS

3.1. Study Design and Area

A retrospective study was carried out from January 2016 to March 2018 among renal failure patients attending dialysis at Saint Paulo's Hospital. The survival analysis were used to investigate the survival time of renal failure patients that are treated under hemodialysis with the joint model of longitudinal estimated glomerular filtration rate. The study was conducted in Saint Paulo's Hospital Millennium Medical College which is located in the capital city of Ethiopia, Addis Ababa.

3.2. Study Population

All renal failure patients who were treated under hemodialysis from January 2016 to March 2018 at Saint Paulo's Hospital Millennium Medical College and fulfill the inclusion criterion was considered in this study.

3.3. Measurements and Study Variables

3.3.1. Response Variables

The continuous longitudinal eGFR and the survival outcome of patients were the two response variables considered in this study. The continuous longitudinal outcome variable was the number of estimated glomerular filtration rate (GFR) per $ml/\min^{-1}/1.73(m^2)^{-1}$. GFR is the total of the filtration rates of the functioning nephrons in the kidney. GFR is measured using the Modification of Diet in Renal Disease (MDRD) Study equation (Levey et al., 2007) as:-

$$eGFR = 175 \times \{SCr\}^{-1.154} \times \{Age\}^{-0.203} \times 0.742(Sex) \times 1.21(Race)$$

Where eGFR denotes estimated glomerular filtration rates measured by $ml/\min^{-1}/1.73(m^2)^{-1}$,SCr denotes serum creatinine measured by mg/dl , Sex coded as male=0 and female=1, age in year and Race coded as white=0, black=1.

The survival outcome variable was time to an event of clinical interest (death) occurs from a defined origin. Typically, survival times T, can be either observed or censored, the latter meaning

that observation of the subject in question is terminated before the event of clinical interest occurs; hence the data tell us that T is at least T_0, but we do not know the exact value of T. Censoring in this study was considered when patient transferring to another hospital or alive or loss to follow before March 2018.

3.3.2. Independent Variables

Six independent variables was used for either the separate or joint analyses. One among the six predictors was continuous while the remaining five were categorical predictors.

Table 1 Independent variables

Variables	Categories
Sex	1 =Male, 0 =female
Age	in years
Hypertension	1 =Yes, 0 =No
Diabetes incidence	1 = type I or type II diabetes, 0 =No diabetes
Anemia	1 =anemia, 0 =no anemia
Cardiac complication	1 = Cardiac complication, 0 =No Cardiac complication,

3.4. Method of Data Collection

Secondary data was used for this study. The data was collected from patient charts based on those variables to be considered in this study. Laboratory tests were performed in each month, both the longitudinal and survival data are extracted from the patient's chart which contains epidemiological, laboratory and clinical information of all renal failure patients under hemodialysis follow up.

3.5. Method of Data Analysis

Data was checked to ensure that all the information were properly collected and recorded. Before and during data processing the information would be checked for completeness and internal consistency. The analyses of Linear Mixed effect Model for the longitudinal data, Cox model for

time to event data, and joint model was coded and analyzed using the statistical packages SPSS 21 and R 3.4.3

In this study the longitudinal eGFR data and the time to event (i.e. death) data with the separate modeling approach and the joint modeling approach were fitted. Suppose longitudinal response data and time-to-event data are available from m subjects. Denote $Y_i = \{y_i(t_{ij}), j = 1,...,n_i$ to be a set of longitudinal response measurements for the i^{th} subject collected at times $\{t_{ij}, j = 1,...ni\}$. In addition, each subject provides a (possibly right censored) failure time $Ti = \min(Ti^*, Ci)$ and an event indicator $\delta i = \mathrm{I}(Ti^* \leq Ci)$ which indicates whether the observed failure time is a true failure time, $Ti*$ or a censoring time Ci.

3.5.1. Separate Analysis of Longitudinal Data

3.5.1.1. Exploratory Data Analysis

Exploratory data analysis were conducted in order to investigate various associations, structures and patterns exhibited in the data set. This consists of obtaining the summary statistics such as frequencies and percentages in a particular group. In addition, the individual profile plots and mean structure plots were obtained in order to gain some insights of the data (Verbeke and Molenberghs, 2000). Smoothing techniques that highlight the typical response as a function of an explanatory variable without reliance on specific parametric models.

3.5.1.2. Linear Mixed Model (LMM)

Three classes of models are commonly used for analysis of longitudinal data; mixed effects model (or random effects model), marginal models (generalized estimating equations (GEE) models) and transition models. (Linear) Mixed effects models (LMM) are widely used in which random effects are introduced to incorporate the between-subjects variation and within subject correlation in the data. In marginal models, the mean structure and the correlation (covariance) structure are modeled separately without distribution assumptions for the data while in the transitional models, the within subject correlation is modeled via Markov structures.

Mixed models extend classical linear regression models by including random or subject-specific effects next to the (traditional) fixed effects in the structure for the mean. The random effects not

only determine the correlation structure between observations on the same subject, they also take account of heterogeneity among subjects, due to unobserved characteristics.

In linear mixed effects model, the sequence of the longitudinal measurements $y_{i1,}y_{i2,,,,,,,,}y_{in_i}$ for the *i*th subject at times $t_{i1,}t_{i2,,,,,}t_{in_i}$ is modeled as:(Rizopoulos, 2012b).

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \varepsilon_i \\ b_i \sim \mathrm{N}(0, D) \\ \varepsilon_i \sim \mathrm{N}(0, \sigma^2 I_{ni}) \end{cases}$$

Where X_i and Z_i are known design matrices, for the fixed-effects regression coefficients β, and the random-effects regression coefficients b_i, respectively, and I_{ni} denotes the *ni*-dimensional identity matrix. The random effects are assumed to be normally distributed with mean zero and variance-covariance matrix *D*, and are assumed independent of the error terms ε_i, i.e., cov (bi, ε_i)=0.

In general, in mixed effects models, random effects b_i are introduced for each subject to incorporate the correlation between the repeated measurements within subject. Since each subject shares the same random effects, the measurements within subject are correlated. Moreover the random effects facilitate subject specific inference. A mixed effects model specifically incorporates both sources of variations: it uses random effects or subject effects to represent deviations of subject longitudinal trajectories from the population average. Thus, a mixed effects model allows subject specific inference, in addition to standard population average inference.

3.5.2. Separate Analysis of Survival Data

3.5.2.1. Kaplan Meier and Log-rank Test

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).

The standard estimator of the survival function, proposed by (Kaplan and Meier, 1958)is called the Product-Limit estimator.

This estimator incorporates information from all the observations available, both uncensored (event times) and censored, by considering survival to any point in time as a series of steps defined at the observed survival and censored times. We use the observed data to estimate the conditional probability of confirmed survival at each observed survival time and then multiply them to obtain an estimate of the overall survival function.

We assume that we have a sample of independent observations denoted $(t_i, \delta i), i = 1,2,...,n$ of the underlying survival time variable T and censoring indicator variable (if censored) $\delta(\delta_i = 0)$

Assume also that $r \leq n$ of the observations are recorded death times. The rank-ordered survival times are denoted by $t_{(1)} < t_{(2)} < ... < t_{(r)}$

Let n_i =the number at risk of dying at $t_{(i)}$ and d_i =the observed number of deaths at $t_{(i)}$. Then the KM estimator of the survivor function at time t is

$$\hat{s}(t) = \begin{cases} 1, & if\ t < t_{(1)} \\ \prod_{t(i) \leq t} (\frac{n_i - d_i}{n_i}), & if\ t \geq t_{(1)} \end{cases}$$

Log-rank test is used also for statistically testing whether the survival experiences of groups of a variable is the same or not. It is obtained by constructing a 2×*k* table at each distinct event (default) time, and comparing the default rates among groups, conditional on the number at risk in the groups.

3.5.2.2. Cox Proportional Hazard Models

In the analysis of survival data interest centers on the risk or hazard of death at any time after the time origin of the study. As a consequence the hazard function is directly modeled in survival analysis. Objective of the modeling process is to determine which combination of the potential explanatory variables affects the form of the hazard of death can be studied. And also to obtain an estimate of the hazard function itself for an individual. From this an estimate of the survivor functions and hence the median survival time can be obtained.

The basic model for survival data is the proportional hazards model, proposed by Cox (1972), and is called the Cox regression model. Although the model is based on the assumption of proportional hazards, no particular form of probability distribution is assumed for the survival times. The model

is therefore referred to as a semi-parametric model. The widely used semi-parametric survival regression model is the *Cox Proportional Hazards (PH)* model in which the *hazard* at time *t* can be expressed as:

$$h(t, x, \beta) = h_0(t) r(x, \beta)$$

This hazard function is the product of two functions. The function $h_0(t)$, characterizes how the hazard function changes as a function of survival time. The other function $r(x, \beta)$, characterizes how the hazard function changes as a function of subject covariates. The functions must be chosen such that $r(x, \beta) > 0$. We note that $h_0(t)$ is the hazard function when $r(x, \beta) = 1$.

When the function $r(x, \beta)$ is parameterized such that $r(x = 0, \beta) = 1$, $h_0(t)$ is frequently referred to as the *baseline hazard function.*

The ratio of the hazard functions for two subjects with covariate values denoted x_1 *and* x_0 is

$$HR(t, x_1, x_0) = \frac{h(t, x_1, \beta)}{h(t, x_0, \beta)} = \frac{h_0(t) r(x_1, \beta)}{h_0(t) r(x_0, \beta)} = \frac{r(x_1, \beta)}{r(x_0, \beta)}$$

The hazard ratio (HR) depends only on the function $r(x, \beta)$. Cox (1972) suggested $r(x, \beta) = \exp(x\beta)$. With this parameterization the hazard function is $h(t, x, \beta) = h_0 e^{x\beta}$ and the hazard ratio is;

$$HR(t, x_1, x_0) = \frac{h(t, x_1, \beta)}{h(t, x_0, \beta)} = e^{\beta(x_1 - x_0)}$$

This model is referred to as the Cox model, the Cox proportional hazards model or simply the proportional hazards model. The term proportional hazards refers to the fact that, the hazard functions are multiplicatively related (i.e., their ratio is constant over survival time). If the proportional hazard assumption is not true, the parametric survival models might be applied.

3.5.3. The Joint Models Structure

The extended Cox model is only appropriate for exogenous time-dependent covariates and therefore cannot handle longitudinal biomarkers. When primary interest is in the association between such endogenous time-dependent covariates and survival, an alternative modeling framework has been introduced in the literature, known as the joint modeling framework for

longitudinal and time-to-event data. The motivating idea behind these joint models is to couple the survival model, which is of primary interest, with a suitable model for the repeated measurements of the endogenous covariate that will account for its special features. To introduce this modeling framework, we use following notation.

We denote by T_i^* the true event time for the i^{th} subject, T_i the observed event time, defined as the minimum of the potential censoring time C_i and T_i^*, and by $\delta_i = I(T_i^* \le C_i)$ the event indicator. For the endogenous time-dependent covariate (eGFR) we let $y_i(t)$ denote its observed value at time point *t* for the *i*th subject. We should note that we do not actually observe $y_i(t)$ for any time *t*, but rather only at the very specific occasions t_{ij} at which measurements were taken. Thus, the observed longitudinal data consist of the measurements $y_{ij} = \{y_i(t_{ij}),\ j = 1,...,n_i\}$ and $m_i(t)$ denote the true and unobserved value of the respective longitudinal outcome at time *t*, uncontaminated with the measurement error value of the longitudinal outcome so it is different from $y_i(t)$.

3.5.3.1. The Longitudinal Sub Model

In the survival sub model we used the true unobserved value of the longitudinal covariate $m_i(t)$. Taking into account that the longitudinal information $y_i(t)$ is collected with possible measurement errors, the first step towards measuring the effect of the longitudinal covariate to the risk for an event is to estimate $m_i(t)$, in order to reconstruct the complete true history $M_i(t)$ to each subject. The main goal, in this study, is to jointly model the longitudinal eGFR measurements and time to death, with a special attention to the effect of GFR variability on the risk of death. In most joint models studied in the past decade, longitudinal data are delineated by a conventional linear mixed model assuming homogeneous within subject variance(Breslow and Clayton, 1993).Then, the linear mixed model can be rewritten as,

$$\begin{cases} y_i(t) = m_i(t) + u_i(t) + \varepsilon_i(t) \\ m_i(t) = x_i^T(t)\beta + z_i^T(t)b_i \\ b_i \sim N(0, \mathbf{D}) \\ \varepsilon_i(t) \sim N(0, \sigma^2) \end{cases}$$

Where we explicitly note that the design vectors $x_i(t)$ for the fixed effects β, and $z_i(t)$ for the random effects b_i, as well as the error terms $\varepsilon_i(t)$, are time dependent. We assume that error terms are mutually independent, independent of the random effects, and normally distributed with mean zero and variance σ^2.

This mixed model accounts for the measurement error problem by postulating that the observed level of the longitudinal outcome $y_i(t)$ equals the true level $m_i(t)$ plus a random error term. We could incorporate an additional stochastic term $u_i(t)$. This last term is used to capture the remaining serial correlation in the observed measurements, random effects are unable to capture. $u_i(t)$ is considered as a mean-zero stochastic process, independent of and Moreover, the time structure in the definitions of $x_i(t)$ and $z_i(t)$, and the use of subject-specific random effects allows to reconstruct the complete path of the time-dependent process $M_i(t)$ for each subject.

3.5.3.2. The Survival Sub Model Specification

Our aim is to measure the association between the longitudinal eGFR level and the risk for death, while accounting for the special features of the former. To achieve this we introduce the term $m_i(t)$ that denotes the true and unobserved value of the longitudinal outcome at time *t*. Note that $m_i(t)$ is different from $y_i(t)$, with the latter being the contaminated with measurement error value of the longitudinal outcome at time *t*. To quantify the strength of the association between $m_i(t)$ and the risk for death, a straightforward approach is to postulate a relative risk model of the form:

$$h_i(t/M_i(t),w_i) = \lim_{dt\to 0} p\frac{\{t \le T_i^* < t+dt / T_i^* \ge t, M_i(t), w_i\}}{dt} \quad \text{.....(1)}$$
$$= h_0(t)\exp\{\gamma^T w_i + \alpha m_i(t)\} \qquad ,t>0$$

where $M_i(t) = \{m_i(s), 0 \le s < t\}$ denotes the history of the true unobserved longitudinal process up to time point t, $h_0(.)$ denotes the baseline risk function, and w_i is a vector of baseline covariates with a corresponding vector of regression coefficients γ. Similarly, parameter α quantifies the effect of the underlying longitudinal outcome to the risk for an event. The interpretation of γ and α is $\exp(\gamma_j)$ denotes the ratio of hazards for one unit change in w_{ij} at any time *t*, whereas $\exp(\alpha)$ denotes the relative increase in the risk for an event at time *t* that results from one unit increase in

$m_i(t)$ at the same time point. Moreover, note that the relative risk expression (1) postulates that the risk for an event at time *t* depends only on the current value of the time-dependent marker $m_i(t)$. However, this does not hold for the survival function. In particular, using the known relation between the survival function and the cumulative hazard function, we obtain that:

$$S_i(t/M_i(t),w_i) = p(T_i^* > t/M_i(t),w_i) \quad \text{.....(2)}$$
$$= \exp(-\int_0^t h_0(s)\exp\{\gamma^T w_i + \alpha m_i(s)\}ds)$$

Which implies that the corresponding survival function depends on the whole covariate history $M_i(t)$ .Reminding, again that both are written as a function of a baseline hazard ($h_0(S)$). Regardless of the fact that the literature recommends to leave $h_0(S)$ completely unspecified, in order to avoid the impact of misspecifying the distribution of survival times, in the joint modeling framework it can lead to an underestimation of the standard error of the parameter estimates(Hsieh et al., 2006) when the data satisfies Cox assumption will use it.

In order to incorporate a time dependent covariate within this framework, we let S_0 denote an absolutely continuous baseline survival function, and we follow the formulation of Cox and Oakes (Cox and Oakes, 1984)that postulates,

$$(\int_0^{T^*}\{\gamma w + \alpha m(s)\}ds) \sim S_0$$

This can be re-expressed in terms of the risk rate function for subject as:

$$h_i(t/M_i(t),w_i) = h_0(v_i(t))\exp\{\gamma^T w_i + \alpha m_i(t)\} \text{.....(3)}$$
$$\textit{where} \quad v_i(t) = \int_0^t \exp\{\gamma^T w_i + \alpha m_i(s)\}ds$$

Similarly, to (1), the baseline risk function $h_0(t)$ can be assumed of a specific parametric form or modeled flexibly. An important difference of (3) compared to (1) is that in the former the entire covariate history $M_i(t)$ is assumed to influence the subject-specific risk (due to the fact that $h_0(t)$ is evaluated at $v_i(t)$, whereas in the latter the subject-specific risk depends only on the current value

of the time-dependent covariate $M_i(t)$ .Also, the survival function for a subject with covariate history $M_i(t)$ equals $S_i(t/M_i(t)) = S_0(v_i(t))$.

3.5.4. Parameter Estimation

In joint modeling maximum likelihood estimation are methods to estimate the model parameters maximum likelihood approach will be used to estimates the model parameters for both processes of this study. (Rizopoulos, 2012a) has also used the likelihood method for joint models, as perhaps the most commonly used approach in the joint literature.

3.5.4.1. Joint Model Maximum likelihood

The maximum likelihood estimates are derived as the modes of the log-likelihood function corresponding to the joint distribution of the observed outcomes $\{\mathrm{T_i}, \delta_i, \mathrm{y_i}\}$. To define this joint distribution we will assume that the vector of time-independent random effects b_i underlies both the longitudinal and survival processes. This means that these random effects account for both the association between the longitudinal and event outcomes, and the correlation between the repeated measurements in the longitudinal process (conditional independence). Formally, we have that

$$p(T_i, \delta_i, y_i / b_i; \theta) = p(T_i, \delta_i / b_i; \theta)\, p(y_i / b_i; \theta)$$
$$p(y_i / b_i; \theta) = \prod_j p\{y_i(t_i) / b_i; \theta\}$$

where $\theta = (\theta_t^T, \theta_y^T, \theta_b^T)^T$ denotes the full parameter vector, with θ_t denoting the parameters for the event time outcome, θ_y the parameters for the longitudinal outcomes and θ_b the unique parameters of the random-effects covariance matrix, and y_i is the $n_i \times 1$ vector of longitudinal responses of the ith subject.

Under the assumptions the log-likelihood contribution for the ith subject can be formulated as follows

$$\log p(T_i, \delta_i, y_i; \theta) = \log \int p(T_i, \delta_i, y_i, b_i; \theta) db_i \qquad \text{.....(4)}$$
$$= \log \int p(T_i, \delta_i / b_i; \theta_t, \beta)[\prod p\{y_i(t_{ij}) / b_i; \theta_y\}] p(b_i; \theta_b) db_i$$

With the conditional density for the survival part $p(T_i, \delta_i \mid b_i; \theta_t, \beta)$ taking the form

$$p(T_i, \delta_i \mid b_i; \theta_t, \beta) = h_i(T_i \mid M_i(T_i); \theta_t, \beta)^{\delta_i} S_i(T_i \mid M_i(T_i); \theta_t, \beta$$
$$= [h_0(T_i)\exp\{\gamma^T w_i + \alpha m_i(T_i)\}]^{\delta_i} * \exp(-\int_o^{T_i} h_0(s)\exp\{\gamma^T w_i + \alpha m_i(s)\}ds)$$

Where $h_0(.)$ can be any positive function of time. The joint density for the longitudinal responses together with the random effects is given by

$$p(y_i \mid b_i; \theta)p(b_i; \theta) = \prod_j p\{y_i(t_{ij}) \mid b_i; \theta\}p(b_i; \theta_b)$$
$$= (2\pi\sigma^2)^{-n_i/2}\exp\{-\frac{\|y_i - x_i/\beta - z_i b_i\|^2}{2\sigma^2}\} * (2\pi)^{-qb/2}|\mathbf{D}|^{-1/2}\exp(-b_i^T \mathbf{D^{-1}} b_i/2)$$

Where qb denotes the dimensionality of the random-effects vector, and $\|x\| = \{\sum_i x_i^2\}^{1/2}$ denotes the Euclidean vector norm.

Maximization of the log-likelihood function $\ell(\theta) = \sum_i \log p(T_i, \delta_i, y_i; \theta)$ with respect to θ requires a combination of numerical integration and optimization algorithms, because both the integral with respect to the random effects in (4) and in the survival function given by (2) do not have an analytical solution. Despite some authors have employed standard numerical integration techniques, such as Monte Carlo or Gaussian quadrature, the Expectation-Maximization (EM) algorithm described by (Wulfsohn and Tsiatis, 1997) has been traditionally preferred. The intuitive idea behind the EM algorithm is to maximize the log-likelihood in two steps: the Expectation step, where missing data are filled, so we replace the log likelihood of the observed data with a surrogate function, and the Maximization step, where this surrogate function is then maximized. Furthermore (Rizopoulos et al., 2009) has introduced a direct maximization of the observed data log-likelihood which is a quasi-Newton algorithm. Therefore hybrid optimization approaches start with EM and then continue with direct maximization.

3.5.4.2. Optimization Control

To control the optimization process, the EM algorithm starts with a fixed number of iterations, and if convergence is not achieved, it switches to a quasi-Newton algorithm until convergence is obtained. The following two criteria are used to declare convergence,

$$\max\{\ \left|\theta^{(it)}-\theta^{(it-1)}\right|/(\left|\theta^{(it-1)}\right|+\varepsilon_1)\ \}<\varepsilon_2,$$
$$\ell(\theta^{(it)})-\ell(\theta^{(it-1)})<\varepsilon_3\{\left|\ell(\theta^{(it-1)})\right|+\varepsilon_3\}$$

Where $\theta^{(it)}$ denotes the parameters values at the i^{th} iteration. In addition, the values for $\varepsilon_1,\varepsilon_2$ that are frequently used are about 10^{-3} or 10^{-4}, and for ε_3 it is about 10^{-8}

3.5.4.3. Numerical Integration

As mentioned before, a numerical approach is necessary to approximate the integrals of the survival function (2), as well as the integral with respect to the random effects (4), the latter becoming more computationally demanding as its dimensionality increases.

In addition to the possibility of using the Gauss-Hermite (GH) quadrature to approximate these integrals' solutions, (Rizopoulos, 2012a) proposed an alternative approach, called the adaptive Gauss-Hermite (AGH) rule that decreases the computational burden to some degree. Now, we used GH quadrature to approximate the integrals' solutions.

3.5.5. Estimation and Inference for Joint Model

3.5.5.1. Estimation of the Random Effects

The estimation of the random effects presented in (Rizopoulos, 2012a) is based on Bayes theory. Assuming that $p(b_i,\theta)$ is the prior distribution, and that $p(b_i\mid T_i,\delta_i,y_i;\theta)p(y_i\mid b_i;\theta)$ is the conditional likelihood part, the corresponding posterior distribution is,

$$p(b_i\mid T_i,\delta_i,y_i;\theta)=\frac{p(T_i,\delta_i,\mid b_i;\theta)p(y_i\mid b_i;\theta)p(b_i;\theta)}{p(T_i,\delta_i,y_i;\theta)}\propto p(T_i,\delta_i,\mid b_i;\theta)p(y_i\mid b_i;)p(b_i\theta)$$

It does not have a closed form solution so it has to be numerically computed. It means that we use Bayesian Sampling method using MCMC technique. Two types of estimators typically used are,

$$\begin{cases}\overline{b_i}=\int b_i p(b\mid T_i,\delta_i,y_i;\theta)\\ \hat{b}=\arg\max_b\{\log p(b\mid T_i,\delta_i,y_i;\theta\}\end{cases}$$

3.5.5.2. Asymptotic Inference for Joint Models

After getting the regression model, we go for assessing the significance of the coefficients and the construction of the confidence interval as well. (Rizopoulos, 2010)explained that the three different tests used to test the null hypothesis of the regression parameters.

$$H_0 : \theta = \hat{\theta}_0 \qquad v_s \qquad H_a : \theta \neq \hat{\theta}_0$$

a **Likelihood Ratio Test**, with the test statistic defined as:

$$LRT = -2\{\ell(\hat{\theta}_0) - \ell(\hat{\theta})\}$$

Where $\hat{\theta}_0$ and $\hat{\theta}$ denote the maximum likelihood estimates under the null and alternative hypotheses, respectively.

a **Score Test**, with the test statistic defined as:

$$U = S^T(\hat{\theta}_0)\{I(\hat{\theta}_0)\}^{-1} S(\hat{\theta}_0)$$

Where $S(.)$ and $I(.)$ denotes the score function and the observed information matrix of the model under the alternative hypothesis;

a **Wald Test**, with the test statistic defined as:

$$W = (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0)$$

Under the null hypothesis, they are asymptotically $\chi_p{}^2$ distributed, with p denoting the number of parameters being tested. In particular, the Wald test for a single parameter θ_j is equivalent to $\frac{(\hat{\theta}_j - \theta_{0j})}{\hat{se}(\hat{\theta}_j)}$ which under the null hypothesis follows an asymptotic standard normal distribution.

Despite of being asymptotically equivalent, the behavior of the tests is different in finite samples. The selection of any of these procedures depends on the limitations of each one. Specifically, regarding the computational cost of fitting, the Wald test only requires to fit the model under the null hypothesis, and the score test under the alternative. However, the likelihood ratio test requires fitting the model under both hypotheses, being more computationally expensive. But other issues must be considered, such as that the Wald test does not take into account the variability introduced

by estimating the variance components, apart from ignoring the fact that we need to estimate the survival process. Also, the implementation of the score test needs extra steps to calculate the required components. Therefore, if there is a disagreement among the three tests of the significance of the coefficient, the partial likelihood ratio test will prevail.

A general drawback of these tests is that they are only appropriate for the comparison of two nested models. In order to carry out the comparison of non-nested models, information criteria could be used, such as the Akaike's Information Criterion (AIC) defined as,

$$AIC = -2\ell(\hat{\theta}) + 2n_{par}$$

Where n_{par} denotes the number of parameters in the model and the model is the best model if it has smallest AIC.

3.5.6. Missing Data Treatment

Missing values are a common issue in a lot of practical data situations. There are different imputing missing values in longitudinal study. The most popular imputation method to handling missing value is multiple imputations (Singer and Willett, 2003). Participants of studies drop out, devices fail to measure values, or questions of a survey are not answered; there are many reasons that lead to incomplete data. It is one way to address this problem by imputing multiple values for a single unobserved one. Consequently the uncertainty of the imputation can be taken into account, by including the discrepancy between the imputed values in the final estimation (Singer and Willett 2003). The multiple imputations replace each missing item with two or more acceptable values, representing a distribution of possibilities. The advantage of the method is that once the imputed data set have been generated.

3.5.7. Joint Model Diagnostics

The standard tools to assess model assumptions are residual plots. Properties and features of residuals, when longitudinal and survival outcomes are separately modeled, have been extensively studied in the literature.

3.5.7.1. Residuals for Joint Models

A) Residuals for the Longitudinal model

In the mixed-effects model, two types of residuals are often used, namely the subject-specific (conditional) residuals, and the marginal (population averaged) residuals. The subject-specific residuals aim to validate the assumptions of the hieriarchical version of the model, i.e,

$$\begin{cases} y_i = x_i\beta + z_i b_i + \varepsilon_i \\ b_i \sim N(0, \mathbf{D}) \end{cases}$$

And are defined as

$$r_i^{ys}(t) = \{y_i(t) - x_i^T(t)\hat{\beta} - z_i^T(t)\hat{b}_i\}$$

With corresponding standardized version

$$r_i^{yss}(t) = \{y_i(t) - x_i^T(t)\hat{\beta} - z_i^T(t)\hat{b}_i\}/\hat{\sigma}$$

Where, $\hat{\beta}$ and $\hat{\sigma}$ denote the MLEs, and $\hat{b}_i$ the empirical Bayes estimates for the random effects. These residuals predict the conditional errors $\varepsilon_i(t)$, and can be used for checking the homoscedasticity and normality assumptions. On the other hand, the marginal residuals focus on the marginal model for y_i implied by the hierarchical representation, that is,

$$\begin{cases} y_i = x_i\beta + \varepsilon_i^* \\ \varepsilon_i^* \sim N(0, Z_i\mathbf{D}Z_i^T + \sigma^2 I_{ni}) \end{cases}$$

Are defined as

$$r_i^{ym} = y_i - x_i\hat{\beta}$$

With corresponding standardized version

$$r_i^{ysm} = \hat{V}_i^{-1/2}(y_i - x_i\hat{\beta})$$

where $\hat{V}_i = Z_i \mathbf{D} Z_i^T + \sigma^2 I_{ni}$ Denotes the estimated marginal covariance matrix of y_i. These marginal residuals r_i^{ym} predict the marginal errors $\mathrm{y_i} - x_i\beta = z_i b_i + \varepsilon_i$, and can be used to investigate misspecification of the mean structure $x_i\beta$ as well as to validate the assumptions for the within-subjects covariance structure V_i.Both types of residuals can be used to check the assumptions of the longitudinal part of a joint model as well.

B) Residuals for the Survival model

A standard type of residuals for the relative risk sub model of the joint model is the martingale residuals. These are based on the counting process notation of time-to-event data, and in particular on the subject-specific counting process martingale, which is defined for the *i*th subject as

$$r_i^{tm}(t) = N_i(t) - \int_0^t R_i(s) h_i(s \mid \hat{M}_i(s); \hat{\theta}) ds$$

$$= N_i(t) - \int_0^t R_i(s) \hat{h}_0(s) \exp\{\hat{\gamma}^T w_i + \hat{\alpha} \hat{m}_i(s)\} ds$$

where $N_i(t)$ is the counting process denoting the number of events for subject *i* by time *t*, $R_i(t)$ is the left continuous at risk process with $R_i(t) = 1$ if subject *i* is at risk at time *t*, and $R_i(t) = 0$ otherwise, $\hat{m}_i(t) = x_i^T(t)\beta + z_i^T(t)\hat{b}_i$ and $\hat{h}_0$ denotes the estimated baseline risk function. The main use of these residuals is for a direct identification of excess events (i.e., to reveal subjects that are poorly fit by the model) and for evaluating whether the appropriate functional form for a covariate interest has been used in the model.

CHAPTER FOUR

4. ANALYSIS AND RESULTS

4.1. Data Description

The data consists of 269 patients who were renal failure and who were treated under hemodialysis between January 2016 and March 2018 in Addis Ababa Saint Paulo's Hospital Millennium Medical College. All patients who started dialysis before March 2018 are excluded.

The continuous longitudinal eGFR and the survival outcome of patients were the two response variables considered in this study. The continuous longitudinal outcome variable was the number of estimated glomerular filtration rate (GFR) per $ml/\min^{-1}/1.73(m^2)^{-1}$. GFR is the total of the filtration rates of the functioning nephrons in the kidney. GFR is measured using the Modification of Diet in Renal Disease (MDRD) Study equation (Levey et al., 2007). Scr were measured approximately every one month; at study entry and every one month. There is a sharply increasing degree of missing data over time due to missed clinic visits.

The survival end point of interest is death. Censored patients are those patients who missed contact due to lost to follow up or transferring to other hospital or alive. Hence, the time-to-death or death time in months was created by subtracting the date of dialysis entry from the date of the last visit. Thus, 72 (26.77%) patients were died while the rest 197 (73.23%) were censored patients.

One continuous and five categorical predictors were included in the analysis. The mean baseline age of renal failure patients was 36.64 years. Out of 269 renal failure patients 118(43.87%) were female and 150(55.76%) were hypertensive. Among the sampled renal failure patients 116(43.12%) were with cardiac complication. Table 2 also revealed that the percentage of event (death) was higher in male (17.10%) than that of female renal failure patients. The percentage of renal failure patients who has hypertension and diabetic complication was higher than that of patients who did not such complication. Whereas, the percentage of patients who was anemic and cardiac smaller event (death) than that of who did not (Table 2).

Table 2 Frequencies and Percentages for Baseline Categorical Variables together with the Observed Number of death in each category

No	Variable	Categories	Total (%)	Observed events (%)
1	Sex	Female	151 (56.13%)	26 (9.67%)
		Male	118 (43.87%)	46 (17.10%)
2	Hypertension	No	119 (44.24%)	17 (6.32%)
		Yes	150 (55.76%)	55 (20.45%)
3	Diabetes	No	104 (38.66%)	14 (5.20%)
		Yes	165 (61.34%)	58 (21.57%)
4	Anemia	No	139 (51.67%)	43 (15.99%)
		Yes	130 (48.32%)	29 (10.78%)
5	Cardiac	No	153 (56.88%)	38 (14.13%)
		Yes	116 (43.12%)	34 (12.64%)
		Continuous variable		
6	Age	Mean	Sd	
		36. 64	15.25	

4.2. Separate Analysis of Longitudinal Data

Before any data analysis, the assumptions of the data must be checked and hence the Q-Q plots was used to check the normality of the longitudinal measures of eGFR.

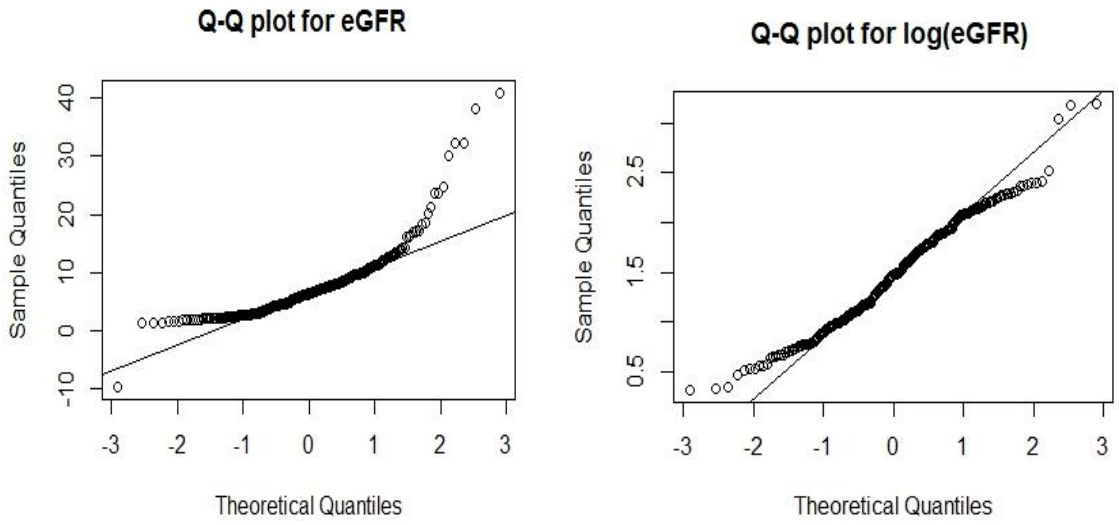

Figure 1. Q-Q plot of the eGFR measurements and log (eGFR) measurements over time

Figure 1 depicted that the Q-Q plot for eGFR measures of original and logarithm transformed data. The Q-Q plot for original eGFR showed that the normality assumption violated whereas the logarithm transformed attained normality proximally. Thus, the analysis of this study use the logarithm transformed eGFR data.

4.2.1. Exploring the Data

Exploratory data analysis was conducted in order to investigate various associations, structures and patterns exhibited in the data set. In addition, the individual profile plots and mean structure plots were obtained in order to gain some insights of the data (Verbeke and Molenberghs, 2000). Figure 2 visualized the pattern change of the overall individual plots estimated glomerular filtration rate measurements of patients over time.

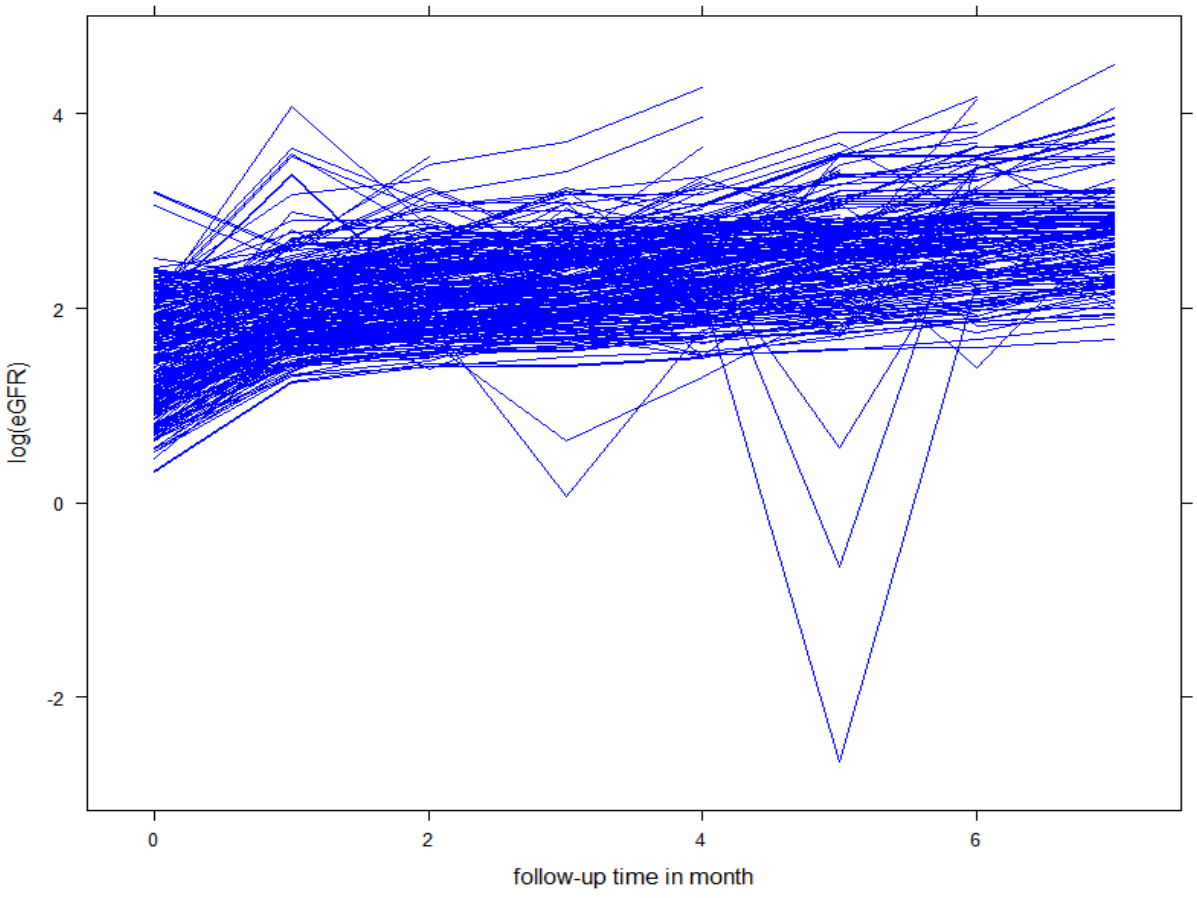

Figure 2. Individual Profile Plot

Figure 2 demonstrates the variability (within and between patients) in eGFR measurements of renal failure patients. Since the measurements were not equally spaced across the different subjects and data is not balanced, loess smoothing technique was used instead.

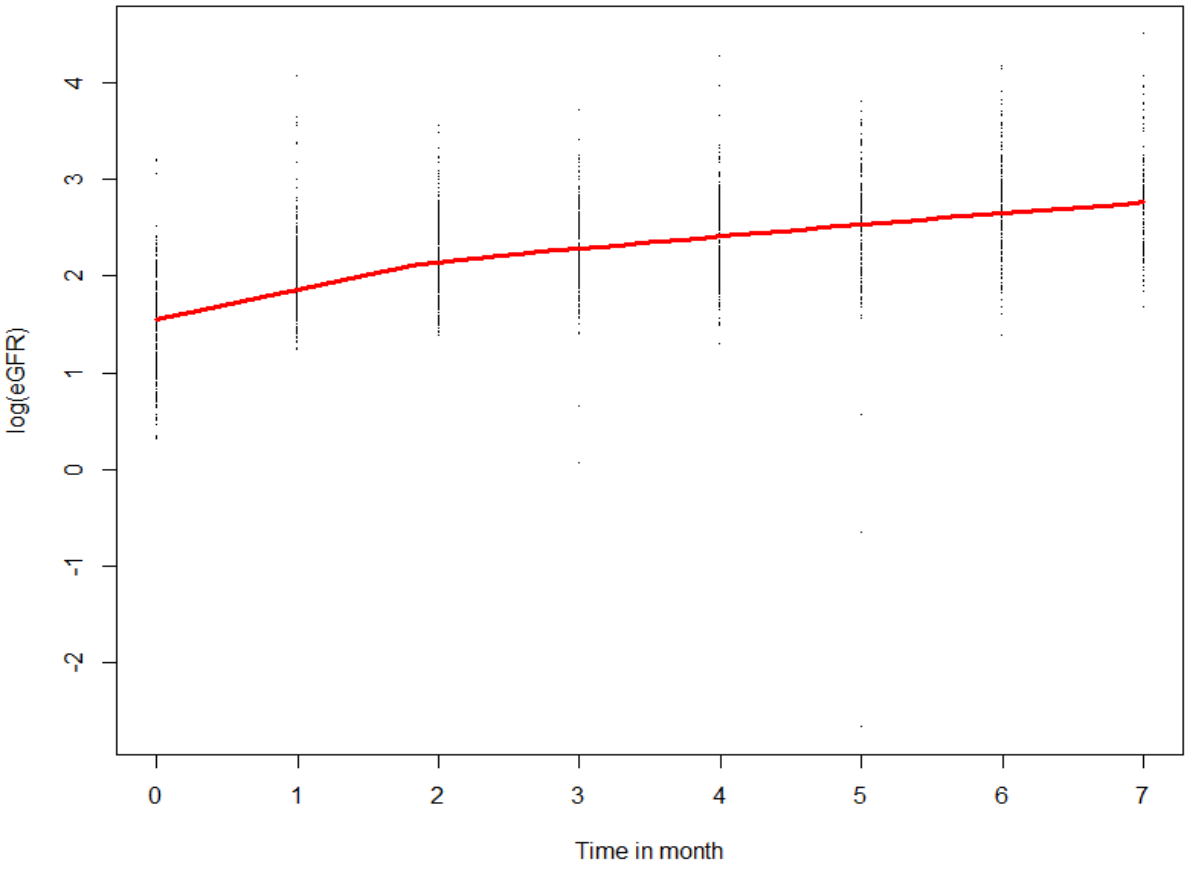

Figure 3 Loess smoothing plot of the mean structure

The line loess smoothing technique suggested that the mean structure of the variable is nearly linear (i.e., the relationship between eGFR and time seems to be linear). The plotted profiles tend to generate a linearly increasing pattern which rationalizes the use of Linear Mixed Effects model to analyze the trajectory of eGFR. Also figure 2.1 in appendix depicted that the longitudinal estimated glomerular filtration rate measurements were linearly increased over time.

4.2.2. Linear Mixed Effect Model

After exploring the data examine whether the assumption of heterogeneous within-subject variance for the eGFR is supported and also identify the random effects (random intercepts, random slope and random intercept random slope) to be included in the model. **Table 3** and **Table 4** below showed that coefficient and standard error for the parameter in fixed effect and variance, standard deviation and 95% confidence interval for standard deviation in random effect of the three models.

Table 3 Comparison of random intercept (RI), random slope (RS) and random intercept random slope model (RI RS) in fixed effect

Fixed effect	RI	RS	RI,RS
(Intercept)	1.97 *	1.98 *	1.97 *
	(0.06)	(0.04)	(0.06)
Sex(male)	0.37 *	0.39 *	0.38 *
	(0.04)	(0.03)	(0.04)
Age	-0.01 *	-0.01 *	-0.01 *
	(0.00)	(0.00)	(0.00)
Hypertension(yes)	-0.07	-0.05	-0.06
	(0.05)	(0.03)	(0.04)
Diabetes(yes)	-0.06	-0.00	-0.03
	(0.05)	(0.03)	(0.04)
Cardiac(yes)	-0.00	-0.00	-0.00
	(0.04)	(0.03)	(0.04)
Anemia(yes)	-0.08 *	-0.07 *	-0.07
	(0.04)	(0.03)	(0.04)
Obstime	0.18 *	0.19 *	0.19 *
	(0.00)	(0.01)	(0.01)
AIC	1814.61	1918.41	1719.82
BIC	1869.39	1973.18	1785.56
Log Likelihood	-897.31	-949.20	-847.91

* $p < 0.05$

Table 4. Comparison of RI, RS, and RI RS in random effect

Random effect		Variance	StDev	95% CI(StDev)
Random intercept	(Intercept)	0.091	0.301	(0.271, 0.335)
	Residual	0.124	0.352	(0.340, 0.365)
Random slope	Obstime	0.0065	0.081	(0.072, 0.092)
	Residual	0.1355	0.368	(0.355, 0.382)
Random intercept	(Intercept)	0.087	0.295	(0.258, 0.338)
Random slope	Obstime	0.004	0.065	(0.055, 0.078)
	Residual	0.103	0.321	(0.309, 0.334)

Here the results of the three models are quite similar to each other among the fixed effect. Under the random effect the above table tells us there is subject-specific variation. Hence it supports the assumption of heterogeneous variance for the repeated eGFR measurements. Also, the great reduction in the AIC for the model incorporating subject-specific variances is an evident that subject-specific eGFR variances must be considered in the analysis.

We fit the random-intercept-and-random-slope version of the model. The fixed effect and random effect estimates from the separate longitudinal model for change in log(eGFR) are presented below. It contained estimated regression parameters, 95% confidence intervals (95% CI), standard errors (SE), p-values (p) and percentage relative effects (RE %) in separate longitudinal analysis of the hemodialysis data set. RE % corresponding to an estimate $\hat{\beta}$, expressed as expected percentage change in eGFR, calculated as (exp($\hat{\beta}$)-1) *100.

Table 5. Parameter estimation of Random intercept-Random slope Linear mixed Model

Fixed effect								
Parameters	$\hat{\beta}$	Se($\hat{\beta}$)	DF	t-value	P	RE%	95% CI	
							Lower	Upper
(Intercept)	1.974	0.063	1498	31.359	<0.0001	-	1.851	2.097
Sex(Male)	0.381	0.042	262	9.147	<0.0001	46.37	0.299	0.463
age	-0.012	0.002	262	-7.535	<0.0001	-1.19	-0.016	-0.008
Hyperetension(yes)	-0.057	0.044	262	-1.277	0.2026	-5.54	-0.143	0.029
Diabetes(yes)	-0.075	0.041	262	-1.842	0.0666	-7.23	-0.155	0.005
Cardiac(yes)	0.0001	0.041	262	-0.003	0.9977	0.00	-0.080	0.080
Anemia(yes)	-0.028	0.045	262	-0.620	0.5358	-2.76	-0.116	0.060
Obstime	0.187	0.006	1498	2.620	<0.0001	20.56	0.175	0.199

Random effect				AIC	BIC	logLik
Parameters	StdDev	95% CI(Sd)		1863.436	1945.506	-916.718
		Lower	Upper			
(Intercept)	0.29	0.2578	0.3379			
obstime	0.07	0.0551	0.0780			
Residual	0.32	0.3091	0.3345			

The analysis of the longitudinal data were based on linear mixed effect model incorporating patient specific eGFR variability. In this model, among the six predictors included in the model sex, age, and follow-up time were statistically significant with 0.05 significance level.

Under the random effect model, the estimated patient specific variability was significant which supports the assumption of heterogeneous variances for the repeated eGFR measurements.

4.3. Separate Analysis of Survival Data

4.3.1. Kaplan-Meier Estimates and Log-rank Tests

The Kaplan-Meier estimator was applied to estimate the survival curves for categorical predictors. Plot of the Kaplan-Meier estimates for only two selected categorical covariates; Gender and Diabetes is displayed below. The remaining are presented in appendix (**fig. 2.1**).

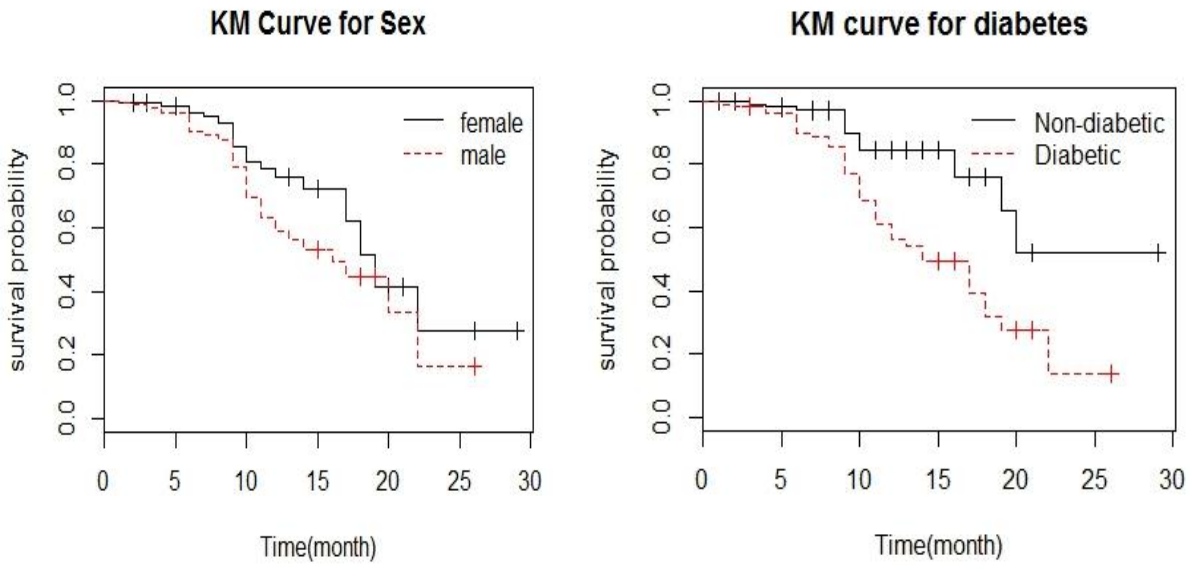

Figure 4. Plot of Kaplan-Meier Estimates for Sex and Diabetes

Plot of the left side panel showed that female patients have higher probability of survival throughout the three years dialysis treatment period than male patients. And the right side panel indicates non-diabetic patients have higher probability of survival.

For comparing the survival experiences between groups, the *log-rank tests* are applied to all categorical variables.

Table 6. Log rank test for categorical independent variables

Variable	Test statistics	Df	P -Value
Sex	4.1	1	0.042
Hypertension	14	1	0.000178
Diabetes	13.3	1	0.000266
Anemia	1.8	1	0.18
Cardiac	3.5	1	0.069

In the above table the Log-rank tests showed that there is no significant difference in the death rates between groups of anemic and cardiac patients. All the other categorical covariates have a significant difference in the death rates of their groups. So in this study hypertension and diabetes were significant factors.

4.3.2. Variable Selection and Cox PH Assumption

To determine the variables to be included in the survival model, an automatic variable selection method stepwise was used. Age, hypertension, diabetes, and cardiac were the variables that included in the model.

The proportional hazards assumption asserts that the hazard ratios are constant overtime. That means the risk of failure must be the same no matter how long subjects have been followed. In order to test this assumption GLOBAL test and schonfield residual was used.

Table 7. Proportional hazard model assumption

Variable	rho	Chisq	P
age	-0.1226	0.7342	0.3915
Hypertension(yes)	-0.254	5.196	0.0226
Diabetes(diabetes)	0.0227	0.0374	0.8466
Cardiac(yes)	-0.0414	0.1208	0.7281
GLOBAL	NA	7.7879	0.0997

From the above table, it is clear to see that the p- value of GLOBAL is not significant. This indicate that the PH assumption is not violated. Graphically schoenfield residual plots are presented in appendix (**fig. 2.3**) showed that the scaled Schoenfeld residuals are randomly distributed and a loess smoothed curve do not exhibit much departure from the horizontal line suggest that the proportional hazards assumption not violated.

4.3.3. Cox proportional hazards model

After checking the assumption of proportional hazard, the survival data was analyzed based on Cox proportional hazard model. The results are presented below.

Table 8. Cox proportional hazards model for the selected variable

Variables	$\hat{\beta}$	HR	Se($\hat{\beta}$)	Z	Pr(>|z|)	95% CI	
						Lower	Upper
Age	0.04228	1.043	0.007382	5.727	0.0003	0.027811	0.056749
Hypertension(yes)	0.556071	1.744	0.298734	1.861	0.0627	-0.02945	1.14159
Diabetes(yes)	0.610603	1.842	0.306894	1.990	0.0466	0.009091	1.212115
Cardiac(yes)	0.545455	1.725	0.244839	1.2.228	0.0259	0.065571	1.025339

In this model, among the six covariates included in the model, sex, hypertension and anemia are not statistically significant with 0.05 value of significance level.

4.4. Joint Model Analysis

4.4.1. Joint Model of Survival and Longitudinal Analysis

Cox proportional model was not including unobserved true eGFR. In fact eGFR were significantly associated with time to death. To alleviate these problems, we fit joint model longitudinal and survival analysis.

The result of joint model could obtain by combining the selected random-intercept-and-random-slope mode and Cox-proportional hazard model. Table 10 below revealed that the parameter estimates with other related statistic for the join models.

Table 9. Results for joint model of longitudinal and survival analysis

	Fixed effect	$\hat{\beta}$	Se($\hat{\beta}$)	Z-value	p-value	RE%	95% CI	
							Lower	Upper
Longitudinal sub model	Intercept	1.9236	0.056	34.24	<0.0001	-	1.8138	2.0334
	Sex(male)	0.3947	0.034	11.56	<0.0001	48.39	0.3281	0.4613
	Age	-0.0108	0.001	-8.75	<0.0001	-1.07	-0.0128	-0.0088
	Hypertension(yes)	-0.0342	0.036	-0.95	0.3434	-3.36	-0.1048	0.0364
	Diabetes(yes)	-0.0743	0.033	-2.22	0.0263	-7.16	-0.1390	-0.0096
	Cardiac(yes)	0.0085	0.034	0.25	0.8011	0.85	-0.0581	0.0751
	Anemia(yes)	-0.0040	0.036	-0.11	0.9135	-0.40	-0.0746	0.0666
	Obstime	0.2380	0.005	42.46	<0.0001	26.87	0.2282	0.2478
	Random effect	Variance	StDev					
	(Intercept)	0.0784	0.28					
	obstime	0.0064	0.08					
	Residual	0.1024	0.32					
	Fixed effect	$\hat{\beta}$	Se($\hat{\beta}$)	Z-value	p-value	HR	95% CI (HR)	
							Lower	Upper
Survival sub model	Age	0.09	0.01	12.32	<0.0001	1.094	1.0729	1.1158
	Hypertension(yes)	1.31	0.30	4.34	<0.0001	3.706	2.0585	6.6725
	Diabetes(yes)	1.02	0.35	2.87	0.0040	2.773	1.3965	5.5069
	Cardiac(yes)	0.37	0.32	1.17	0.2405	1.448	0.7732	2.7107
	Assoc	-1.21	0.11	-11.46	<0.0001	0.298	0.2404	0.3699

In longitudinal sub-model, Males were found to have 48.4% higher expected eGFR than females. Kidney function was found to decrease with increasing age at study start [$\hat{\beta}$ =-0.0108, 95% CI= -0.0128, -0.0088] and was found to increase with increasing time under observation [$\hat{\beta}$ = 0.2380, 95% CI= 0.2282, 0.2478]. Diabetic patients were found to have 7.16% lower expected eGFR than non-diabetic patients. A 1 year increase in age at study start was associated with a relative decrease of 1.1% in expected eGFR. Similarly, a 1 month increase in time under observation was associated with a relative increase of 26.87% in expected eGFR.

And in the survival sub-model all variables except cardiac were associated with the hazard of death. The risk of death for 1 year older in age is increased by 9.4% than 1 year younger in age. The estimated risk of death for hypertensive patients was 3.7 times higher than those patients who were not hypertensive patients. Regarding to the diabetic patients the hazard of death was 2.7 times higher than those patients who were not diabetic. In addition, the significant model association parameter revealed a negative association between the eGFR and hazard of death, which means that death is less likely to occur in patients with higher eGFR (HR = 0.298, P < 0.001).

4.4.2. Separate and joint model

The longitudinal sub-model was consistent with the results from the separate longitudinal analysis. The differences in magnitudes of the parameter estimates were negligible and there were some parameter difference in terms of statistical significance.

The results from the separate and joint analysis were quite similar to each other. In the longitudinal sub model sex, age, diabetes and follow up times were statistically significant predictors. But diabetes was not statistically significant.

Table 10. Comparison of separate and joint model

	Parameter	Separate model				Joint model			
Longitudinal sub model	***Fixed effect***	$\hat{\beta}$	Se($\hat{\beta}$)	t-value	p-value	$\hat{\beta}$	Se($\hat{\beta}$)	Z-value	p-value
	Intercept	1.974	0.063	31.359	<0.0001*	1.94	0.056	34.24	<0.0001*
	Sex(Male)	0.381	0.042	9.147	<0.0001*	0.39	0.034	11.56	<0.0001*
	Age	-0.012	0.002	-7.535	<0.0001*	-0.01	0.001	-8.75	<0.0001*
	Hypertension(yes)	-0.057	0.044	-1.277	0.2026	-0.03	0.036	-0.95	0.3434
	Diabetes(yes)	-0.075	0.041	-0.620	0.5358	-0.08	0.033	-2.22	0.0263*
	Cardiac(yes)	0.0001	0.041	-0.003	0.9977	0.01	0.034	0.25	0.8011
	Anemia(yes)	-0.028	0.045	-1.842	0.0666	-0.04	0.036	-0.11	0.9135
	Obstime	0.187	0.006	2.620	<0.0001*	0.23	0.005	42.46	<0.0001*
	Random effect	Variance	StDev			Variance	StDev		
	(Intercept)	0.087	0.29			0.078	0.28		
	obstime	0.0043	0.07			0.0064	0.08		
	Residual	0.1034	0.32			0.1024	0.32		
Survival sub model	***Fixed effect***	$\hat{\beta}$	Se($\hat{\beta}$)	Z-value	P-value	$\hat{\beta}$	Se($\hat{\beta}$)	Z-value	p-value
	Age	0.042	0.007	5.72	<0.0003*	0.09	0.01	12.32	<0.0001*
	Hypertension(yes)	0.556	0.29	1.86	0.0627	1.31	0.30	4.34	<0.0001*
	Diabetes(yes)	0.611	0.31	1.99	0.0466*	1.02	0.35	2.87	0.0040*
	Cardiac(yes)	0.545	0.24	2.23	0.0259*	0.37	0.32	1.17	0.2405
	Assoc	-	-	-	-	-1.21	0.11	-11.46	<0.0001*
	AIC	2336.2054				2102.661			

* $p < 0.05$

In the survival sub-model of the joint model, except hypertension all predictors included in the model were significantly associated with the hazard of death.

In general, the estimated parameters of the two model (separate and joint models) were quite similar to each other but not identical. However, the estimates of the association parameters in the joint analysis were significantly different from zero, providing evidence of association between the two sub-models. The estimate of association (β=-1.21) indicating that the higher eGFR fluctuation is associated with the lower hazard of death.

When evaluating the overall performance of both the separate and joint models in terms of model parsimony and goodness of fit, the joint model was performed better. As a result the joint model was preferred as it has a smaller total AIC than the separate model. Also, the statistical significance of both the association parameters was also evidence that the joint model was better than the separate models (Seid *et al*., 2014).

4.4.3. Joint model diagnostics

The joint models was fitted, the next step is to verify if all the necessary model assumptions are valid. Standard types of residuals plots can be used to validate the assumptions behind mixed models and relative risk models.

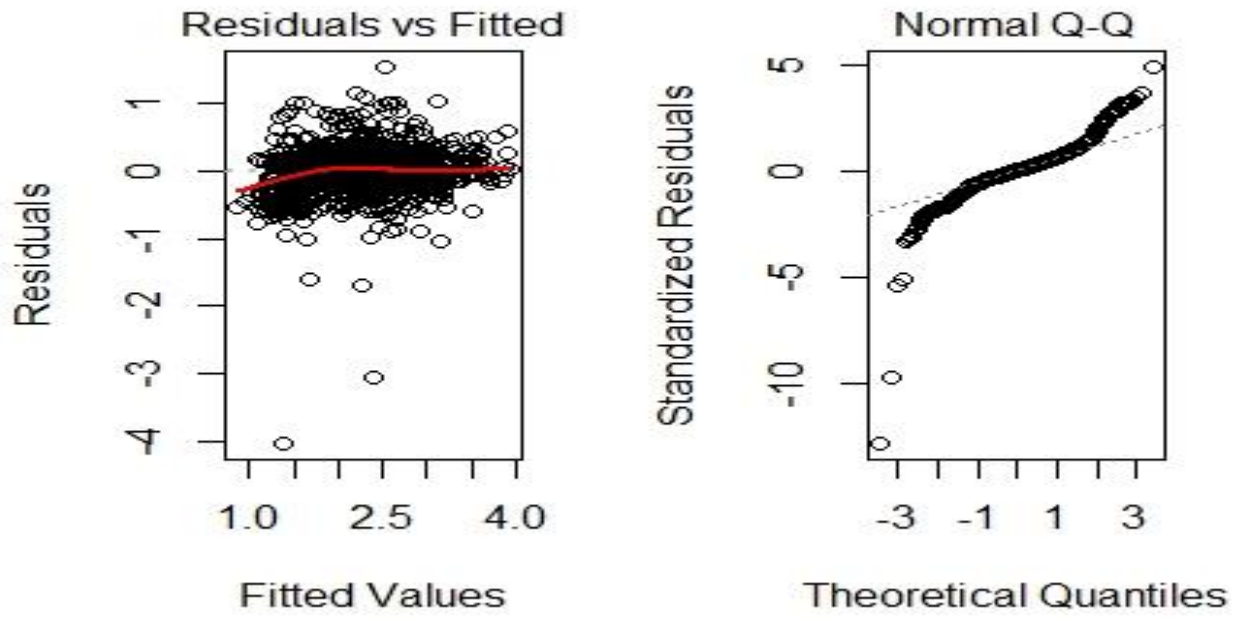

Figure 5.Residual plot against fitted values and Q-Q plot

The distributional assumption is checked by comparing theoretical quantiles. The qqplot shows that the response variables of longitudinal sub model (log(eGFR)) are normally distributed because the pointes are scattered on the line as well as the residual against fitted value plot didn't show any systematic pattern and no evidence of non-constant variance and the fitted LOWESS curve is close to 0. Hence, log(eGFR) is linear to the parameter and the error variance of longitudinal sub model are constant.

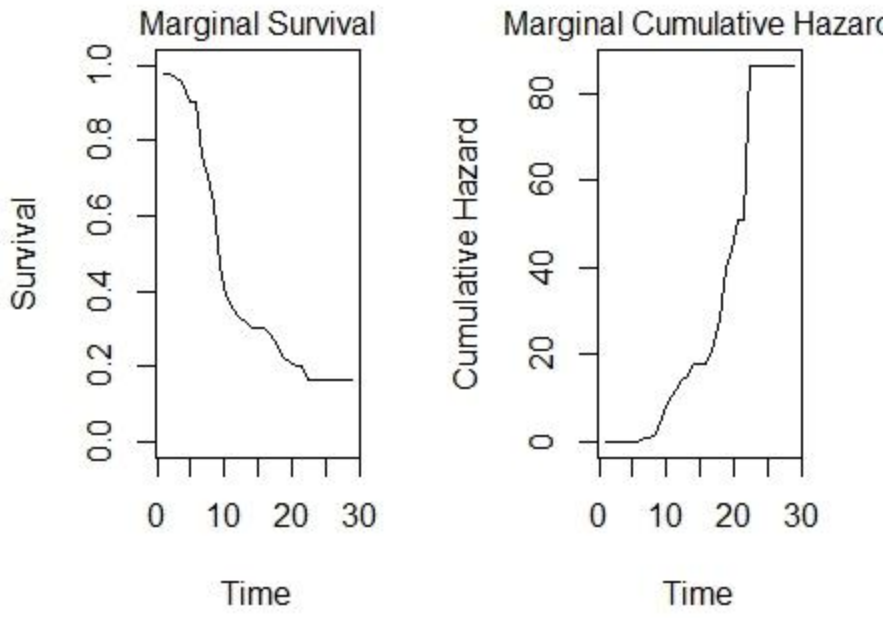

Figure 6. Marginal survival plot and marginal cumulative hazard plot

The Marginal survival plot and marginal cumulative hazard plot showed that the survival probability (not developing death) come down and the probability of developing death come up to one respectively when the follow up time was increased.

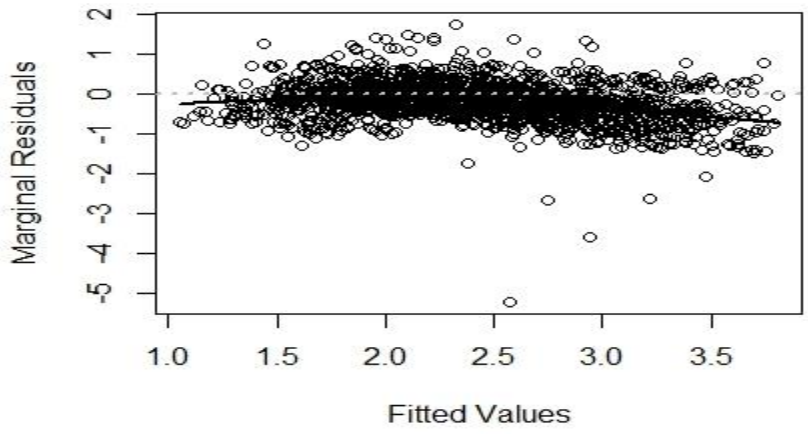

Figure 7.Marginal residual versus fitted values plot

A marginal residual versus fitted values plot of the standardized residuals for longitudinal process which is almost coinciding with the reference line passing through the origin and hence, is validating our assumption of normality of the error term in the longitudinal sub-model.

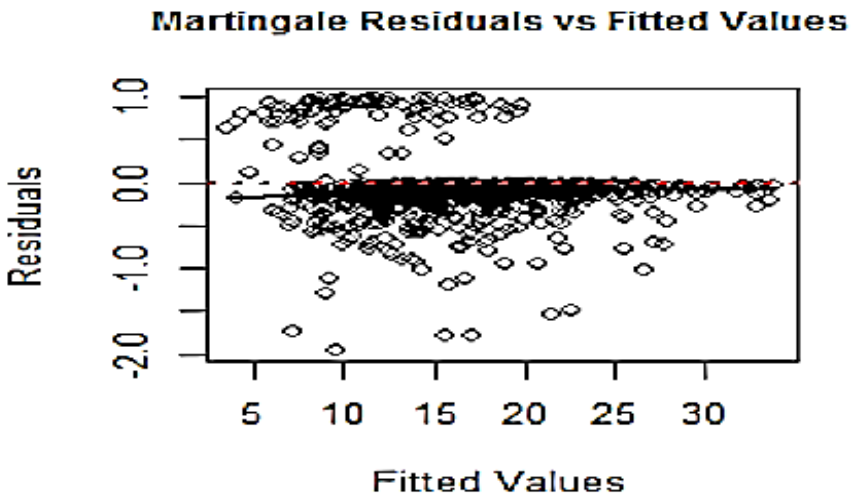

Figure 8.Martingale residuals versus the subject-specific fitted values

The estimated martingale residuals versus the subject-specific fitted values of the survival process shows no much deviations from the null horizontal line. This indicate the survival process model fits the data well.

Chapter Five

5. Discussion of the Results

In this study, three different models were explored, the linear mixed effects model, Cox proportional hazards model for each outcome independently, and joint modeling of the two outcomes together.

In the separate analysis of the longitudinal data, first the eGFR measurements are checked for normality using scatter plot. The plots indicates that there is a deviation from normality and needs some transformation. After a log transformation of the eGFR, the mean response of the longitudinal log of eGFR is determined to be normal in time. Then, the data were analyzed using random intercept, random slope, and random intercept-random slope model. The parameter estimates of the three models are very close to each other. But, the estimated patient specific variability is significant which supports the assumption of heterogeneous variances. Also, the random intercept –random slope model has a smaller AIC than the other models. As a result, sex, age and follow-up times are significantly associated with the progression of estimated glomerular filtration rate.

In the separate analysis of the survival data, the variables to be included in the survival model are determined using an automatic variable selection method using R. next to variable selection proportional hazard assumption was checked. As a result age, diabetes, and cardiac are significantly associated with the hazard of death.

After the most suitable separate model have been decided for the data, the proposed joint model were applied to the data, with the aim of investigating the effects of repeated eGFR measurements on time to death. In the longitudinal sub model sex, age, diabetes, and follow-up time are statistically associated with the progression of eGFR. Males were found to have higher expected eGFR than females and a 1 year increase in age at study start was associated with a relative decrease in expected eGFR. This result conform the study conducted by Asar et al., (2015). In this study the determinant prognostic factor affect the progression rate of GFR was diabetes mellitus. It conform the study conducted by Chielle (2015).

In the survival sub-model age, hypertension, and diabetes are significantly associated with the hazard of death. The risk of death for diabetic patients is higher than those patients who were not

diabetic. This results conform the study conducted by Mohammad (2016). The main known risk factor for survival in renal failure patients are diabetes. Age is important too, with a poorer prognosis for older patients. This result is agree with the study conducted by Sá Carvalho, Henderson, Shimakura, & SOUSA, (2003). In this study hypertension is also the significant factors associated with the hazard of death.

CHAPTER SIX

6. CONCLUSION AND RECOMMENDATIONS

6.1. CONCLUSION

From the result we conclude that the Kaplan-Meier survival curves and log-rank tests showed that the survival experience of different groups of renal failure patients on sex, histories of hypertension and histories of diabetes were statistically significant. Accordingly female patients has higher survival experience than male. Patients with no diabetic and no hypertensive having better probability of survival than the other groups.

From this specific study the joint model was the better fit than the separate survival and longitudinal models. The separate and longitudinal sub-model showed that: Sex, Age, and follow-up time of the patients were significant predictors of the progression change of GFR whereas diabetes incidence of the patients were significance predictors of the progression change of eGFR in longitudinal sub model. The variables follow-up time and age were positively and negatively associated with the progression change of eGFR respectively. Also, the separate and survival sub-model analysis showed that: Age, diabetes, hypertension, unobserved true eGFR were statistical significant predictors of the time to death. The variables unobserved true eGFR were negatively associated with the time to death.

When evaluating the overall performance of both the separate and joint models in terms of model parsimony, goodness of fit, smaller total AIC, and the statistical significance of both the association parameters, the joint model performs better. Thus, authors concluded that the joint model was preferred for simultaneous analyses of repeated measurement and survival data.

6.2. RECOMMENDATIONS

Based on the findings of the study the following recommendations are forwarded:

- Renal failure patients who had diabetes, and hypertension are especially vulnerable to death. In order to address this problem, continuous health checkup and timely medical care should be devised so as to minimize the risk of death.
- Health professionals give attention for this silent killer disease to minimize the risk of death.

- Health professionals are recommended to give more attention for end stage renal disease patients.
- Health professionals give attentions when a patient estimated glomerular filtration rates are decreased through follow-up and diabetic and older age patients.
- Health professionals, Governments and non-governmental organization promote and allocate budget in adequate amount for hemodialysis treatment for renal failure patients to minimize the risk level of death.
- Finally, it is recommended that further studies of this nature include other important covariates that were not included in this study.

REFERENCE

AROGUNDADE, F., ABD-ESSAMIE, M. & BARSOUM, R. 2005. Health-related quality of life in emotionally related kidney transplantation: deductions from a comparative study. *Saudi Journal of Kidney Diseases and Transplantation,* 16**,** 311.

AROGUNDADE, F. A. & BARSOUM, R. S. 2008. CKD prevention in Sub-Saharan Africa: a call for governmental, nongovernmental, and community support. *American Journal of Kidney Diseases,* 51**,** 515-523.

ASAR, Ö., RITCHIE, J., KALRA, P. A. & DIGGLE, P. J. 2015. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International journal of epidemiology,* 44**,** 334-344.

BARSOUM, R. 1991. Nephrology and African ecology. An overview. *Artif Organs,* 14**,** 235.

BARSOUM, R. S. 2012. History of Dialysis in Africa. *Dialysis.* WORLD SCIENTIFIC.

BARSOUM, R. S. 2013. Burden of chronic kidney disease: North Africa. *Kidney international supplements,* 3**,** 164-166.

BARSOUM, R. S., KHALIL, S. S. & AROGUNDADE, F. A. 2015. Fifty years of dialysis in Africa: challenges and progress. *American Journal of Kidney Diseases,* 65**,** 502-512.

BETHESDA, M. 2003. US Renal Data System: USRDS 2003 Annual Data Report: Atlas of End-Stage Renal Disease in the United States. *National Institute of Health, National Institute of Diabetes and Digestive and Kidney Diseases*.

BLEYER, A. J., TELL, G. S., EVANS, G. W., ETTINGER, W. H. & BURKART, J. M. 1996. Survival of patients undergoing renal replacement therapy in one center with special emphasis on racial differences. *American journal of kidney diseases,* 28**,** 72-81.

BLOEMBERGEN, W. E., STANNARD, D. C., PORT, F. K., WOLFE, R. A., PUGH, J. A., JONES, C. A., GREER, J. W., GOLPER, T. A. & HELD, P. J. 1996. Relationship of dose of hemodialysis and cause-specific mortality. *Kidney international,* 50**,** 557-565.

BRESLOW, N. E. & CLAYTON, D. G. 1993. Approximate inference in generalized linear mixed models. *Journal of the American statistical Association,* 88**,** 9-25.

CHIELLE, E. O., RIGON, K. A., ARCARI, I. A., STEIN, V. & SANTOS, G. A. D. 2015. Influence of hemodialysis on the plasma concentration of adenosine deaminase in patients with chronic kidney disease. *Jornal Brasileiro de Patologia e Medicina Laboratorial,* 51**,** 153-157.

CLARK, W. R., MUELLER, B. A., KRAUS, M. A. & MACIAS, W. L. 1998. Quantification of creatinine kinetic parameters in patients with acute renal failure. *Kidney international,* 54**,** 554-560.

COLLINS, A. J., FOLEY, R. N., HERZOG, C., CHAVERS, B. M., GILBERTSON, D., ISHANI, A., KASISKE, B. L., LIU, J., MAU, L. W. & MCBEAN, M. 2010. Excerpts from the US renal data system 2009 annual data report. *American journal of kidney diseases: the official journal of the National Kidney Foundation,* 55.

COX, D. R. & OAKES, D. 1984. *Analysis of survival data,* CRC Press.

DEVI, T. R., GUNASEKARAN, S., HUDSON, J. W. & JOYBELL, I. S. A. 2009. Analysis on renal failure patients blood samples: characterization and efficacy study. *Indian Journal of Science and Technology,* 2**,** 46-50.

DORGALALEH, A., MAHMUDI, M., TABIBIAN, S., KHATIB, Z. K., TAMADDON, G. H., MOGHADDAM, E. S., BAMEDI, T., ALIZADEH, S. & MORADI, E. 2013. Anemia and thrombocytopenia in acute and chronic renal failure. *International journal of hematology-oncology and stem cell research,* 7**,** 34.

ECKARDT, K.-U., CORESH, J., DEVUYST, O., JOHNSON, R. J., KÖTTGEN, A., LEVEY, A. S. & LEVIN, A. 2013. Evolving importance of kidney disease: from subspecialty to global health burden. *The Lancet,* 382**,** 158-169.

FAUCI A, KASPER D, HAUSER S, JAMESON J & J, L. 2008. Harrison's Principles of Internal Medicine, Kidney & Urinary Tract Disorders. 17.

FIELD, M. J., POLLOCK, C. & HARRIS, D. 2011. *The Renal System E-Book: Systems of the Body Series*, Elsevier Health Sciences.

FOUNDATION, N. K. 2002. *Clinical practice guidelines for chronic kidney disease: evaluation, classification and stratification*, National Kidney Foundation.

GRASSMANN, A., GIOBERGE, S., MOELLER, S. & BROWN, G. 2005. ESRD patients in 2004: global overview of patient numbers, treatment modalities and associated trends. *Nephrology Dialysis Transplantation,* 20**,** 2587-2593.

GUO, X. & CARLIN, B. P. 2004. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The american statistician,* 58**,** 16-24.

HENDERSON, R., DIGGLE, P. & DOBSON, A. 2000. Joint modelling of longitudinal measurements and event time data. *Biostatistics,* 1**,** 465-480.

HSIEH, F., TSENG, Y. K. & WANG, J. L. 2006. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics,* 62**,** 1037-1043.

HSIEH, R.-L., LEE, W.-C., HUANG, H.-Y. & CHANG, C.-H. 2007. Quality of life and its correlates in ambulatory hemodialysis patients. *Journal of nephrology,* 20**,** 731-738.

ISEKI, K., IKEMIYA, Y. & FUKIYAMA, K. 1997. Risk factors of end-stage renal disease and serum creatinine in a community-based mass screening. *Kidney international,* 51**,** 850-854.

JAFAR, T., ISLAM, M. & POULTER, N. 2006. Chronic kidney disease in the developing world. *N Engl J Med,* 354**,** 998-9.

JAFFA, M. A., GEBREGZIABHER, M. & JAFFA, A. A. 2015. Analysis of multivariate longitudinal kidney function outcomes using generalized linear mixed models. *Journal of translational medicine,* 13**,** 192.

KAPLAN, E. L. & MEIER, P. 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association,* 53**,** 457-481.

KIMMEL, P. L. 2001. Psychosocial factors in dialysis patients. *Kidney international,* 59**,** 1599-1613.

KIMMEL, P. L. & PETERSON, R. A. Depression in end-stage renal disease patients treated with hemodialysis: tools, correlates, outcomes, and needs. Seminars in dialysis, 2005. 91-97.

KRZESINSKI, J.-M., SUMAILI, K. E. & COHEN, E. 2006. How to tackle the avalanche of chronic kidney disease in sub-Saharan Africa: the situation in the Democratic Republic of Congo as an example. Oxford University Press.

KUO, H.-W., TSAI, S.-S., TIAO, M.-M. & YANG, C.-Y. 2007. Epidemiological features of CKD in Taiwan. *American journal of kidney diseases,* 49**,** 46-55.

LEVEY, A., STEVENS, L., SCHMID, C., ZHANG, Y. & CASTRO, A. 2009. 3 rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. *Ann Intern Med,* 150**,** 604-12.

LEVEY, A. S., BOSCH, J. P., LEWIS, J. B., GREENE, T., ROGERS, N. & ROTH, D. 1999. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Annals of internal medicine,* 130**,** 461-470.

LEVEY, A. S., CORESH, J., GREENE, T., MARSH, J., STEVENS, L. A., KUSEK, J. W. & VAN LENTE, F. 2007. Expressing the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate with standardized serum creatinine values. *Clinical chemistry,* 53**,** 766-772.

MODI, G. & JHA, V. 2006. The incidence of end-stage renal disease in India: a population-based study. *Kidney international,* 70**,** 2131-2133.

MONTASERI, M., CHARATI, J. Y. & ESPAHBODI, F. 2016. Application of Parametric Models to a Survival Analysis of Hemodialysis Patients. *Nephro-urology monthly,* 8.

MOUSAVI, S. S. B., HAYATI, F., VALAVI, E., REKABI, F. & MOUSAVI, M. B. 2015. Comparison of survival in patients with end-stage renal disease receiving hemodialysis versus peritoneal dialysis. *Saudi Journal of Kidney Diseases and Transplantation,* 26**,** 392.

NISHA, R., KANNAN SR, S. & JAGATHA, P. 2017. Biochemical evaluation of creatinine and urea in patients with renal failure undergoing hemodialysis. *Journal of Clinical Pathology and Laboratory Medicine,* 1.

RATCLIFFE, S. J., GUO, W. & TEN HAVE, T. R. 2004. Joint modeling of longitudinal and survival data via a common frailty. *Biometrics,* 60**,** 892-899.

RIZOPOULOS, D. 2010. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software (Online),* 35**,** 1-33.

RIZOPOULOS, D. 2012a. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis,* 56**,** 491-501.

RIZOPOULOS, D. 2012b. *Joint models for longitudinal and time-to-event data: With applications in R*, CRC Press.

RIZOPOULOS, D., VERBEKE, G. & LESAFFRE, E. 2009. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 71**,** 637-654.

RUSUL ARIF AA & S, H. 2014. A study of some biochemical changes in patients with chronic renal failure undergoing hemodialysis. *Int J Curr Microbiol App Sci,* 3**,** 581-586.

SÁ CARVALHO, M., HENDERSON, R., SHIMAKURA, S. & SOUSA, I. P. S. C. 2003. Survival of hemodialysis patients: modeling differences in risk of dialysis centers. *International Journal for Quality in Health Care,* 15**,** 189-196.

SHIBIRU, T., GUDINA, E. K., HABTE, B., DERIBEW, A. & AGONAFER, T. 2013. Survival patterns of patients on maintenance hemodialysis for end stage renal disease in Ethiopia: summary of 91 cases. *BMC nephrology,* 14**,** 127.

SHIGIDI, M., FAROUK, N., ABULIKAILIK, R. I., ALSIR, R. T. & ABU-AISHA, H. 2012. Active tuberculous infection among adult Sudanese patients on long term peritoneal dialysis. *Arab journal of nephrology and transplantation,* 5**,** 135-140.

SILINS, J., FORTIER, L., MAO, Y., POSEN, G., UGNAT, A.-M., BRANCKER, A., GAUDETTE, L. & WIGLE, D. 1989. Mortality rates among patients with end-stage renal disease in Canada, 1981-86. *CMAJ: Canadian Medical Association Journal,* 141**,** 677.

SINGER, J. & WILLETT, J. 2003. Applied longitudinal data analysis: Modeling change and event occurrence New York. *NY Oxford*.

SMELTZER, S., BARE, B., HINKLE, J. & CHEEVER, K. 2010. Textbook of medical surgical nursing Brunner and suddarth. *China.: Lippinicott Williams and Wilkins***,** 889.

SON, Y.-J., CHOI, K.-S., PARK, Y.-R., BAE, J.-S. & LEE, J.-B. 2009. Depression, symptoms and the quality of life in patients on hemodialysis for end-stage renal disease. *American journal of nephrology,* 29**,** 36-42.

SOUSA, I. 2011. A review on joint modelling of longitudinal measurements and time-to-event. *Revstat Stat J,* 9**,** 57-81.

STEVENS, L. A., CORESH, J., GREENE, T. & LEVEY, A. S. 2006. Assessing kidney function—measured and estimated glomerular filtration rate. *New England Journal of Medicine,* 354**,** 2473-2483.

SWANEPOEL, C. R., WEARNE, N., DUFFIELD, M. S. & OKPECHI, I. G. 2012. The evolution of our knowledge of HIV-associated kidney disease in Africa. *American Journal of Kidney Diseases,* 60**,** 668-678.

TOMÁS, I., MARINHO, J., LIMERES, J., SANTOS, M., ARAÚJO, L. & DIZ, P. 2008. Changes in salivary composition in patients with renal failure. *archives of oral biology,* 53**,** 528-532.

TSIATIS, A., DEGRUTTOLA, V. & WULFSOHN, M. 1995. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association,* 90**,** 27-37.

TUSO, P. J. 2009. SERVE Ethiopia| pdf. *TPJ,* 13.

UEDA, H., ISHIMURA, E., SHOJI, T., EMOTO, M., MORIOKA, T., MATSUMOTO, N., FUKUMOTO, S., MIKI, T., INABA, M. & NISHIZAWA, Y. 2003. Factors affecting progression of renal failure in patients with type 2 diabetes. *Diabetes Care,* 26**,** 1530-1534.

VAHEDI, M., MAHMOODI, M., MOHAMMAD, K., OSSAREH, S. & ZERAATI, H. 2016. What Is the Best Parametric Survival Models for Analyzing Hemodialysis Data? *Global journal of health science,* 8**,** 118.

VERMA, M., KHADAPKAR, R., SAHU, P. S. & DAS, B. R. 2006. Comparing age-wise reference intervals for serum creatinine concentration in a "reality check" of the recommended cut-off. *Indian Journal of Clinical Biochemistry,* 21**,** 90-94.

VOS, T., ALLEN, C., ARORA, M., BARBER, R. M., BHUTTA, Z. A., BROWN, A., CARTER, A., CASEY, D. C., CHARLSON, F. J. & CHEN, A. Z. 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet,* 388**,** 1545-1602.

WEISBORD, S. D., FRIED, L. F., MOR, M. K., RESNICK, A. L., KIMMEL, P. L., PALEVSKY, P. M. & FINE, M. J. 2007. Associations of race and ethnicity with anemia management among patients initiating renal replacement therapy. *Journal of the National Medical Association,* 99**,** 1218.

WULFSOHN, M. S. & TSIATIS, A. A. 1997. A joint model for survival and longitudinal data measured with error. *Biometrics***,** 330-339.

YASSIN, M., LUBBAD, A. M., ABU TAHA, A. & SAADALLAH, N. 2014. Homocysteine and hematological indices in hemodialysis patients. *Ibnosina Journal of Medicine and Biomedical Sciences,* 6**,** 173-179.

YOUNESPOUR, S., FOROUSHANI, A. R., MARAGHI, E., ROSTAMI, Z., EINOLLAHI, B., ESHRAGHIAN, M. R. & MOHAMMAD, K. 2016. Longitudinal Serum Creatinine Levels in Relation to Graft Loss Following Renal Transplantation: Robust Joint Modeling of Longitudinal Measurements and Survival Time Data. *Nephro-urology monthly,* 8.

Appendix

Implementation in R

1. Separate Analysis of Linear Mixed Effect Model

1.1. Linearity assumption

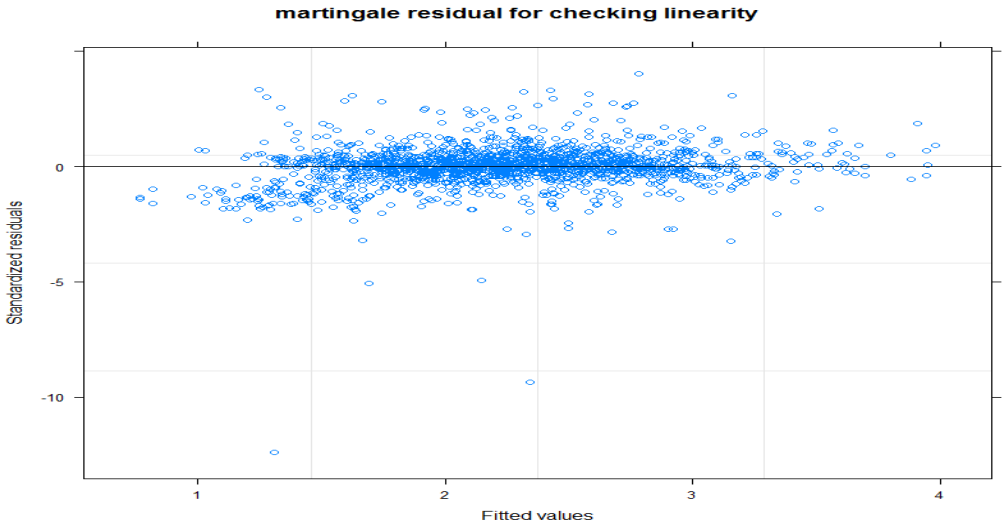

Figure 1.1 martingale residual

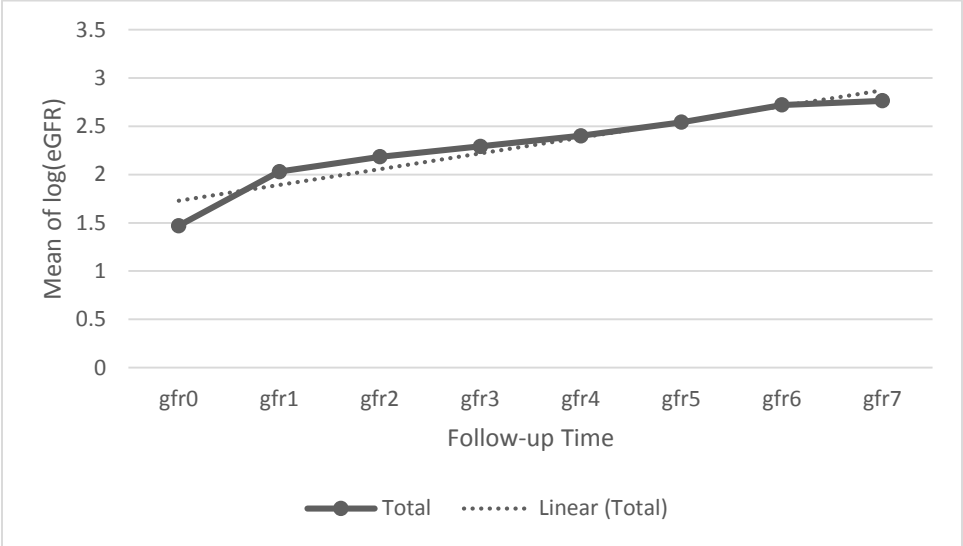

Figure 1.2 Mean structure of log(eGFR) over time

1.2. Comparison of linear mixed effect model

```
> #random-intercept model
> library(nlme)
> #random-intercept model
```

```
>  lme1 <- lme(eGFR2 ~ sex+age+hyp+diabetes+cardiac+anemia+obstime,
+                    random = ~ 1 | patient,method = "ML",data=x)
> #random-slopes model
>  lme2 <- lme(eGFR2 ~ sex+age+hyp+diabetes+cardiac+anemia+obstime,
+                    random = ~ obstime-1 | patient,method = "ML",data=x)
> #random-intercepts and random-slopes model
>  lme3 <- lme(eGFR2 ~ sex+age+hyp+diabetes+cardiac+anemia+obstime,
+                    random = ~ obstime | patient,method = "ML",data=x)
> library(texreg)
> screenreg(list(lme1,lme2,lme3),single.row = FALSE,stars = c(0.05),custom.mo
del.names = c("RI","RS","RI,RS"))

================================================
                RI          RS          RI,RS
------------------------------------------------
(Intercept)      1.97 *      1.98 *      1.97 *
                (0.06)      (0.04)      (0.06)
sex1             0.37 *      0.39 *      0.38 *
                (0.04)      (0.03)      (0.04)
age             -0.01 *     -0.01 *     -0.01 *
                (0.00)      (0.00)      (0.00)
hyp1            -0.07       -0.05       -0.06
                (0.05)      (0.03)      (0.04)
diabetes1       -0.06       -0.00       -0.03
                (0.05)      (0.03)      (0.04)
cardiac1        -0.00       -0.00       -0.00
                (0.04)      (0.03)      (0.04)
anemia1         -0.08 *     -0.07 *     -0.07
                (0.04)      (0.03)      (0.04)
obstime          0.18 *      0.19 *      0.19 *
                (0.00)      (0.01)      (0.01)
------------------------------------------------
AIC             1814.61     1918.41     1719.82
BIC             1869.39     1973.18     1785.56
Log Likelihood  -897.31     -949.20     -847.91
Num. obs.       1768        1768        1768
Num. groups      269         269         269
================================================
* p < 0.05
```

1.3. Final linear mixed effect model (Random Intercept Random Slope)

```
> summary(lme3)
Linear mixed-effects model fit by maximum likelihood
 Data: x
       AIC      BIC    logLik
  1719.824 1785.555 -847.9119

Random effects:
 Formula: ~obstime | patient
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 0.29518258 (Intr)
obstime     0.06554043 -0.25
Residual    0.32154043
```

```
Fixed effects: eGFR2 ~ sex + age + hyp + diabetes + cardiac + anemia + obstim
e
                 Value  Std.Error   DF  t-value p-value
(Intercept)  1.9736222 0.06293570 1498 31.35934  0.0000
sex1         0.3806911 0.04161834  262  9.14720  0.0000
age         -0.0115739 0.00153594  262 -7.53535  0.0000
hyp1        -0.0566787 0.04437092  262 -1.27738  0.2026
diabetes1   -0.0278277 0.04488560  262 -0.61997  0.5358
cardiac1    -0.0001183 0.04079426  262 -0.00290  0.9977
anemia1     -0.0749278 0.04066916  262 -1.84237  0.0666
obstime      0.1866893 0.00572324 1498 32.61951  0.0000
 Correlation:
          (Intr) sex1   age    hyp1   dibts1 cardc1 anemi1
sex1      -0.250
age       -0.530 -0.120
hyp1      -0.103 -0.130 -0.373
diabetes1 -0.220  0.157 -0.366  0.085
cardiac1  -0.242 -0.057  0.011  0.100 -0.094
anemia1   -0.323 -0.056 -0.052  0.090  0.146 -0.080
obstime   -0.172 -0.008  0.002  0.015  0.003  0.002 -0.007

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-12.68855641  -0.35927488   0.02442392   0.39223921   4.82163880

Number of Observations: 1768
Number of Groups: 269
```

2. Separate Analysis of Survival Model

2.1. Plot of Kaplan-Meier Estimates for categorical variable

```
> kmsurvival<-survfit(Surv(time,status)~sex,data = y)
> summary(kmsurvival)
> plot(kmsurvival,conf.int=FALSE,mark.time=TRUE, col=c("black","red"),lty=1:2
,xlab = "Time(month)"
+      ,ylab = "survival probability")
> legend("topright",c("female","male"),lty = 1:2,col = c("black","red"), bty
= "n")
> title(main='KM Curve for Sex')
```

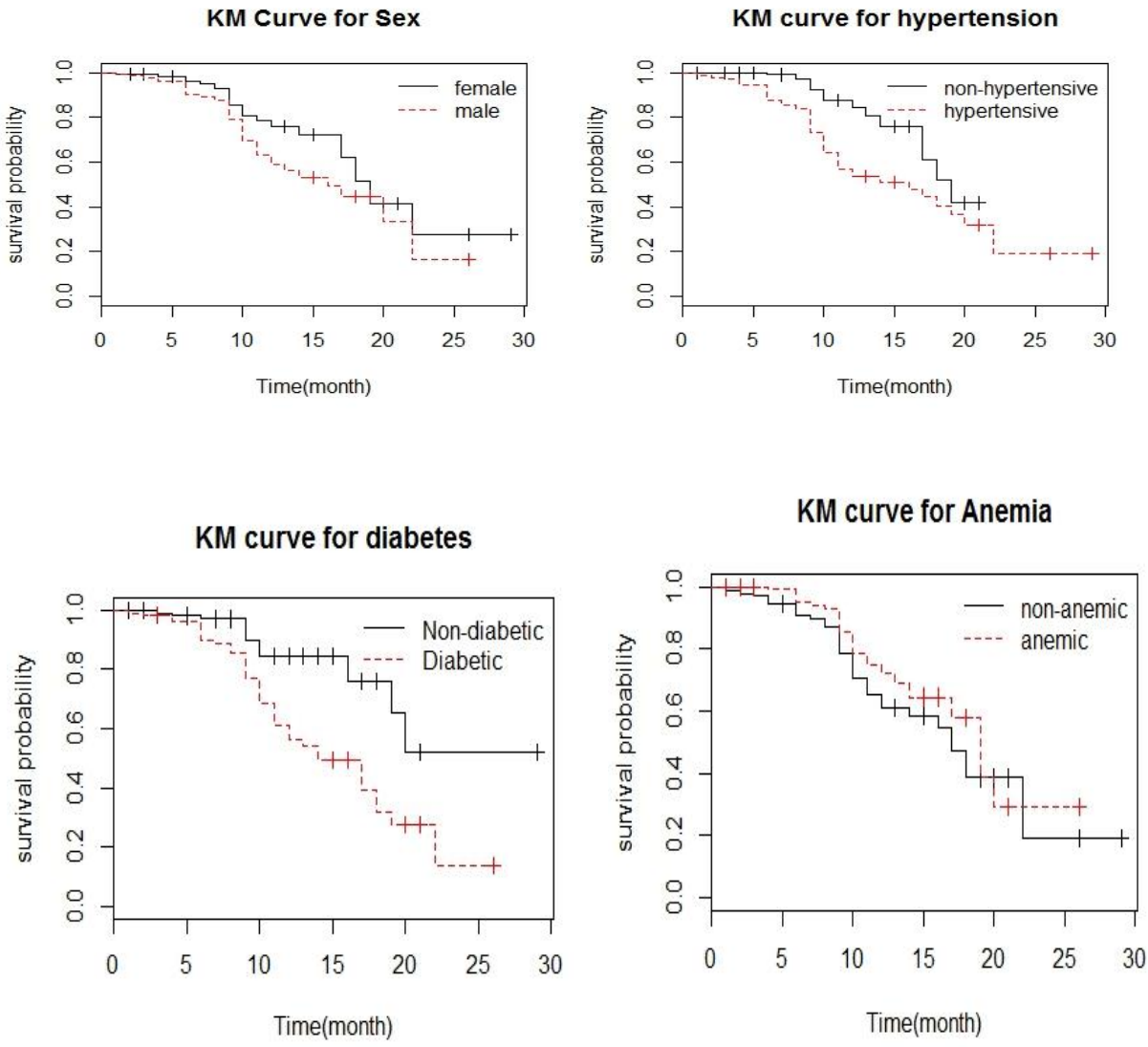

Figure 2.1 KM curve for each category

2.2. Perform the log rank test

```
> mysurv<-Surv(y$time,y$status)
> fit<-survdiff(mysurv~y$sex)
> fit1<-survdiff(mysurv~y$diabetes)
> fit2<-survdiff(mysurv~y$hyp)
> fit3<-survdiff(mysurv~y$cardiac)
> fit4<-survdiff(mysurv~y$anemia)
```

2.3. Test PH assumption

```
> z<-coxph<-coxph(Surv(time,status)~+age+hyp+diabetes+cardiac,data = y,x=TRUE
,method = "breslow")
> h<-cox.zph(z)
```

```
             rho  chisq      p
age       -0.1226 0.7342 0.3915
hyp1      -0.2540 5.1961 0.0226
diabetes1  0.0227 0.0374 0.8466
cardiac1  -0.0414 0.1208 0.7281
GLOBAL         NA 7.7879 0.0997
```

```
> plot(h)
> par(mfrow=c(2,2))
> plot(h)
```

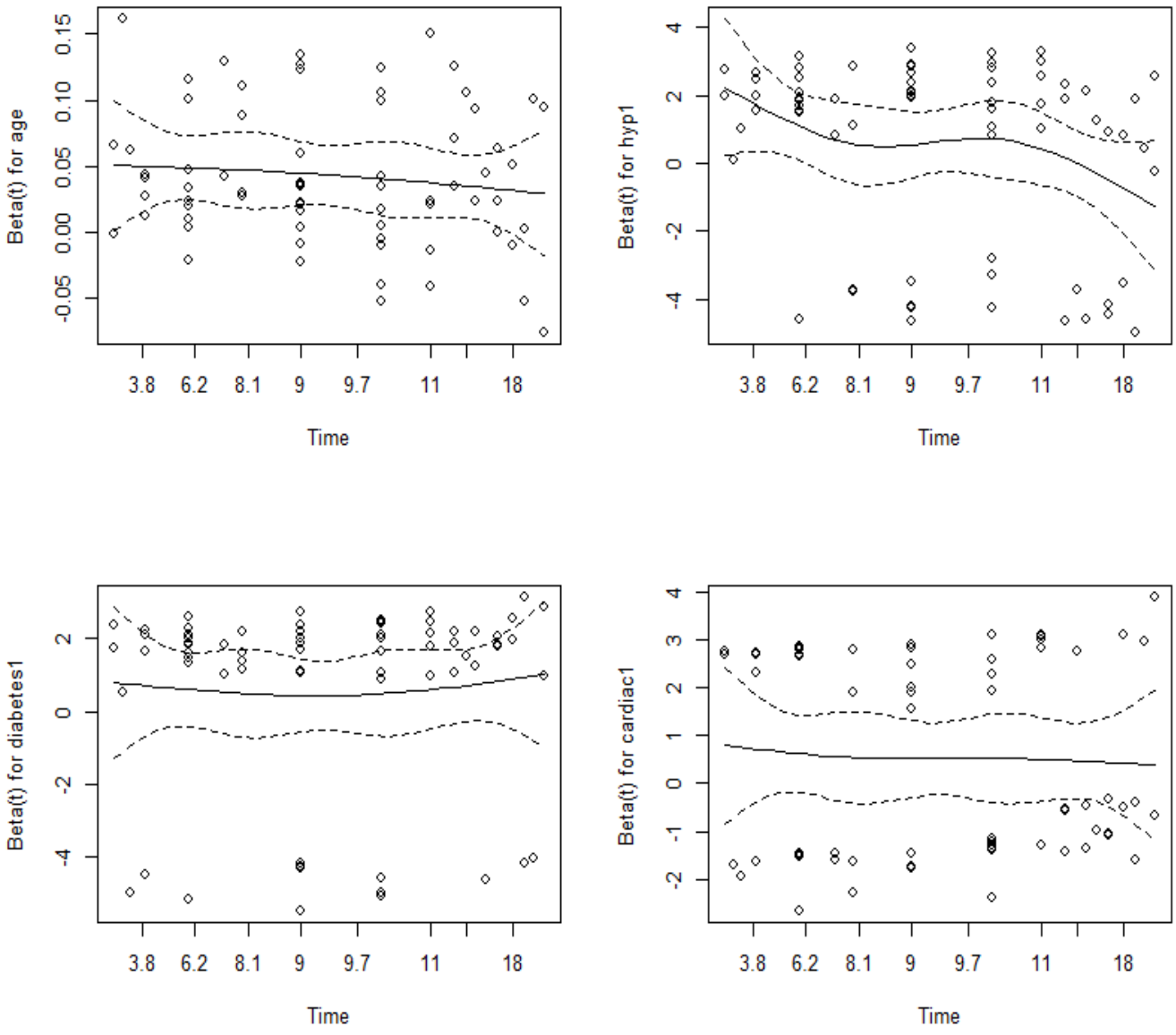

Figure 2.3 Schoenfeld residuals for Ph assumption

2.4. Cox proportional hazard model, coefficient and hazard rate

```
> coxph<-coxph(Surv(time,status)~sex+age+hyp+diabetes+anemia+cardiac,data = y
,method = "breslow")
> summary(coxph)
> #stepwise variable selection
> coxph<-coxph(Surv(time,status)~sex+age+hyp+diabetes+anemia+cardiac,data = y
,method = "breslow")
> step(coxph,direction = "both")
> coxph<-coxph(Surv(time,status)~age+hyp+diabetes+cardiac,data = y,x=TRUE,met
hod = "breslow")
> summary(coxph)
Call:
coxph(formula = Surv(time, status) ~ age + hyp + diabetes + cardiac,
    data = y, x = TRUE, method = "breslow")

  n= 269, number of events= 72

              coef exp(coef) se(coef)     z Pr(>|z|)
age       0.042280  1.043186 0.007382 5.727 1.02e-08 ***
hyp1      0.556071  1.743807 0.298734 1.861   0.0627 .
diabetes1 0.610603  1.841541 0.306894 1.990   0.0466 *
cardiac1  0.545455  1.725393 0.244839 2.228   0.0259 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

          exp(coef) exp(-coef) lower .95 upper .95
age           1.043     0.9586     1.028     1.058
hyp1          1.744     0.5735     0.971     3.132
diabetes1     1.842     0.5430     1.009     3.361
cardiac1      1.725     0.5796     1.068     2.788

Concordance= 0.812  (se = 0.041 )
Rsquare= 0.207   (max possible= 0.917 )
Likelihood ratio test= 62.29  on 4 df,   p=9.553e-13
Wald test            = 59.61  on 4 df,   p=3.504e-12
Score (logrank) test = 66.34  on 4 df,   p=1.341e-13
```

3. Joint model Analysis

```
> library(nlme)
> library("JM")
> lmeFit.renal<- lme(eGFR2 ~ sex+age+hyp+diabetes+cardiac+anemia+obstime,rand
om = ~ obstime| patient,method = "ML", data = x)
> coxphFit.renal<-coxph(Surv(time,status)~age+hyp+diabetes+cardiac,data = y,x
=TRUE,method = "breslow")
> jointFit.renal<- jointModel(lmeFit.renal, coxphFit.renal,timeVar = "obstime
", method = "Cox-PH-aGH")
> summary(jointFit.renal)
Call:
jointModel(lmeObject = lmeFit.renal, survObject = coxphFit.renal,
    timeVar = "obstime", method = "Cox-PH-aGH")

Data Descriptives:
Longitudinal Process           Event Process
Number of Observations: 1768   Number of Events: 72 (26.8%)
Number of Groups: 269
```

```
Joint Model Summary:
Longitudinal Process: Linear mixed-effects model
Event Process: Relative risk model with unspecified baseline risk function

Parameterization: Time-dependent

  log.Lik      AIC      BIC
 -1034.33 2102.661 2163.771

Variance Components:
             StdDev    Corr
(Intercept)  0.2838  (Intr)
obstime      0.0820 -0.1175
Residual     0.3157

Coefficients:
Longitudinal Process
              Value Std.Err z-value p-value
(Intercept)  1.9236  0.0562 34.2442 <0.0001
sex1         0.3947  0.0341 11.5655 <0.0001
age         -0.0108  0.0012 -8.7498 <0.0001
hyp1        -0.0342  0.0361 -0.9475  0.3434
diabetes1   -0.0743  0.0334 -2.2219  0.0263
cardiac1     0.0085  0.0338  0.2519  0.8011
anemia1     -0.0040  0.0369 -0.1086  0.9135
obstime      0.2380  0.0056 42.4646 <0.0001

Event Process
           Value Std.Err z-value p-value
age       0.0906  0.0074 12.3241 <0.0001
hyp1      1.3083  0.3013  4.3423 <0.0001
diabetes1 1.0180  0.3539  2.8766  0.0040
cardiac1  0.3756  0.3199  1.1738  0.2405
Assoct   -1.2143  0.1059 -11.4613 <0.0001

Integration:
method: (pseudo) adaptive Gauss-Hermite
quadrature points: 5

Optimization:
Convergence: 0
```

3.1. Joint model diagnostics

```
> par(mfrow = c(2, 2))
>  plot(jointFit.renal)
#Residuals for the Longitudinal Part
>  # marginal residuals
>  resMargY.renal <- residuals(jointFit.renal, process = "Longitudinal",type
= "Marginal")
>  # marginal fitted values
>  fitMargY.renal <- fitted(jointFit.renal, process = "Longitudinal",type = "
Marginal")
>  # function to produce scatteplots with superimposed smooth line
>  plotResid <- function (x, y, col.loess = "black", ...) {
+    plot(x, y, ...)
+    lines(lowess(x, y), col = col.loess, lwd = 2)
```

```
+    abline(h = 0, lty = 3, col = "grey", lwd = 2)
+  }
>  # scatteplot of marginal residuals vs marginal fitted values
>  plotResid(fitMargY.renal, resMargY.renal, xlab = "Fitted Values",
+            ylab = "Marginal Residuals")
>  #Residuals for the Survival Part
>  # martingale residuals
>  martRes <- residuals(coxph.renal,process = "Event")
>  # subject-specific fitted values for the longitudinal outcome
>  mi.t <- fitted(jointFit.renal, process = "Longitudinal",type = "EventTime"
)
>  # scatterplot of martingale residuals vs subject-specific fitted values
>  plotResid(mi.t, martRes, col.loess = "black",
+            xlab = "Fitted Values",ylab = "Residuals")
```

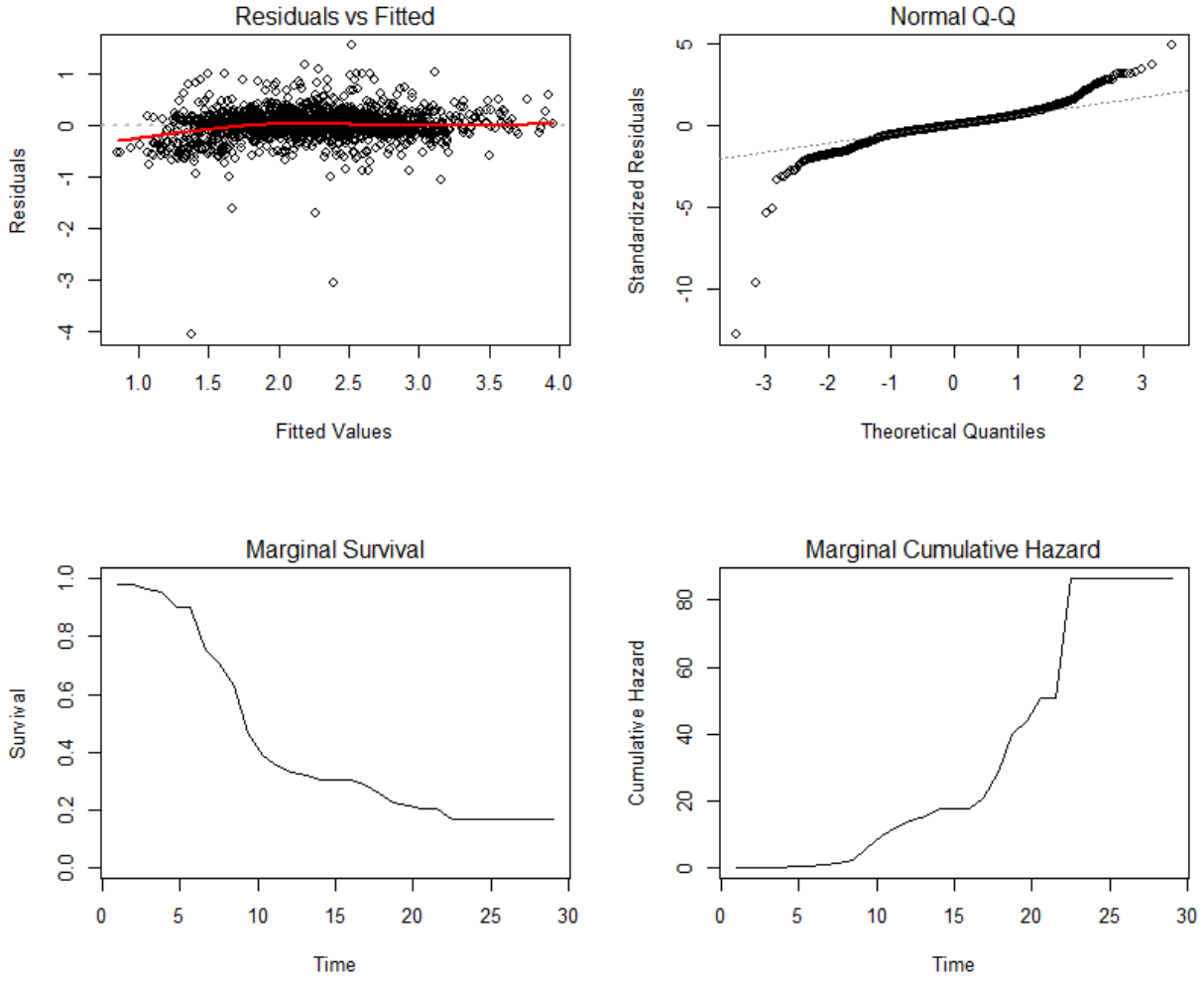

Figure 3. Default Plots for Joint Model

Approval Sheet I

ሳይንስ ኮሌጅ
የድህረ ምረቃ፣ምርምርና ማህበረሰብ
አገልግሎት ም/ዲን
ባሕር ዳር ዩኒቨርሲቲ
ባሕር ዳር - ኢትዮጵያ

Science College
The Graduate, Research
& Community Services V/Dean
Bahir Dar University
Bahir Dar – Ethiopia

Wisdom at the source of the Blue Nile

✉ 79
251 (582) 226 6597
ፋክስ Fax: 251 (582) 220- 20- 25
e-mail: nega7719@gmail.com
website: www.bdu.edu.et

ቁጥር: Rc/1012/10
ቀን : 13/07/10

Ethical Clearance Approval Form

Applicant's Name: Hiwot Abel

Research Title	Joint modeling on longitudinal glomerular filtration rate (GFR) measurement and time-to-death of renal failure patients treated under hemodialysis: a comparison of separate and joint model in case of St. Paulos Hospital, Addis Ababa
Researcher (s) Name (s)	Hiwot Abel

Thank you for submitting your application for ethical clearance, which was considered at the College of Science Research Ethics Committee meeting on 22 March 2018. The committee has reviewed your ethical application, issues pertaining to participants, consent form, debriefing, and relevant questionnaires.

The researcher should keep the confidentiality of the identity of research participants and data that will be obtained from them. Any serious adverse events or significant changes which occur in connection with this study and /or which may alter its ethical consideration must be reported immediately to the committee for a possible ethical amendment.

We are therefore pleased to inform you that the College's Ethical Clearance Committee has approved your study from an ethical point of view.

With kind regards

Nega Tassie (PhD)
The Graduate, Research and Community Services V/Dean
College of Science

CC//
- Dean office
- The Graduate, Research and Community Services V/Dean
- Department of Statistics

College of Science

Approval Sheet II

Ref. No. PM 23/312
Date: 09/04/2018

Institutional Review Board (IRB) of St. Paul's Hospital Millennium Medical College (SPHHMC)

Ethical Clearance

Research Title: - Joint modeling on longitudinal glomerular filtration rate measurement and time-to-death of renal failure patients treated under hemodialysis: a comparison of separate and joint model

Principal Investigator: - Hiwot Abel

The IRB of SPHMMC has reviewed the above mentioned research proposal and made the following decision:

1. Approved:- ✓
2. Approved with recommendation:- _____
3. Approved on condition :-_____
4. Disapproved:- _____

The decision is valid for 12 months and the research should be conducted in compliance with the protocol/proposal approved by the IRB of SPHMMC. Any subsequent revision/amendment of the protocol/proposal needs approval before conduct of the research. The researcher should also submit written summaries of the research status to the IRB every 03 months. Upon the conclusion of the study, manuscripts and thesis work to the final/completed research project needs to be submitted to the IRB.

IRB Chair: Prof. MarkosTesfaye
Signature: _____
Date: April 18/2018

Cc:

- Vice Provost for Academic and Research
- IRB
- Hiwot Abel
  PHMMC