MEKELLE UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF STATISTICS

A Mixed Effect Model for Unbalanced Longitudinal Haematocrit Level Evolution Progress of Chronic Kidney Failure Patients

By: Getachew Beyene

Email: getachewbeyene24@gmail.com

Principal Advisor: Yemane Hailu Fissuh (Ph.D. in Biostatistics)

Email: yemaoroba@gmail.com

Co-Advisor: Mrs. Azmera Hailay (MSc. in Biostatistics)

Email: azmerahailay@gmail.com

A Thesis Submitted to the Department of Statistics, College of Natural and Computational Sciences, Mekelle University for Partial Requirement of Fulfillment of Master of Sciences (MSc.) in Biostatistics

November 2023

Mekelle University, Mekelle, Tigray, Ethiopia

A Mixed Effect Model for Unbalanced Longitudinal Hematocrit Level Evolution Progress of Chronic Kidney Failure Patients

By: Getachew Beyene

Email: getachewbeyene24@gmail.com

Principal Advisor: Yemane Hailu Fissuh (Ph.D. in Biostatistics)

Email: yemaoroba@gmail.com

Co-Advisor: Mrs. Azmera Hailay (MSc. in Biostatistics)

Email: azmerahailay@gmail.com

November 2023

Mekelle University, Mekelle, Tigray, Ethiopia

APPROVAL SHEET

Department of Statistics, College of Natural & Computational Sciences Mekelle University

As thesis advisors, we hereby certify that we have read the thesis prepared by **Getachew Beyene** under our guidance, which is entitled "**A Mixed Effect Model for unbalanced longitudinal Haematocrit Level Evolution Progress of Chronic Kidney Failure Patients"**, in its final form and has found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfil university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Yemane Hailu (PhD, in Biostatistics)

______________________	____________	____________
Principal Advisor	**Signature**	**Date**

Azmera Hailay (MSc.)

______________________	____________	____________
Co-advisor	**Signature**	**Date**

As the members of the board of examiners of the MSc. thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfils the requirements for the degree of Master of Sciences in Statistics, Biostatistics Stream.

______________________	____________	____________
Name of Chairman	Signature	Date
______________________	____________	____________
Name of Principal advisor	Signature	Date
______________________	____________	____________
Name of Co-advisor	Signature	Date
______________________	____________	____________
Name of Internal Examiner	Signature	Date
______________________	____________	____________
Name of External Examiner	Signature	Date

DECLARATION

I declare that this thesis entitled "**A Mixed Effect Model for unbalanced longitudinal Haematocrit Level Evolution Progress of Chronic Kidney Failure Patients**" 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 Mekelle University in partial fulfilment for the Degree of Master of Sciences in Statistics, Biostatistics Stream. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Getachew Beyene

Date: ______________________

Signature: ____________________

Mekelle University, Mekelle, Tigray, Ethiopia

ACKNOWLEDGEMENT

First, I must give thanks to almighty God for all the blessings he bestowed upon me until today.

I would like to express my heartfelt gratitude and thanks is extended to Ministry of Industry for its sponsorship of the MSc Degree in Biostatistics.

Then I want to express my sincere gratitude and appreciation to my advisor Yemane Hailu (PhD. Biostatistics) for his constructive advice, assistance, encouragement, suggestions, and comments from inception up to completion.

Secondly, my sincere thanks go to my co-advisor Mrs. Azmera Hailay(MSc.) for her excellent advice and guidance in all aspects.

Thirdly, my heartfelt gratitude and thanks is extended to my friend Mr. Mehari Gebre (Lecturer in Adigrat University in Biostatistics (MSc.)) for his nice guidance and assistance.

And also, my gratitude and thanks go to my wife Mrs. Sesen Weldu for her admirable encourage and helping me in caring and raring children while I'm busy with my courses and thesis work.

I would like to take this opportunity to express my heartiest appreciation to all my staff members for their insightful guidance, consistent support, and impressive encouragement.

Finally, and most importantly my deepest gratitude goes to my family and all my friends who were greatly involved in my work in any means.

Last but not the least my special thanks go to my classmates for they helped me considerably in encouraging and sharing constructive ideas about how to accomplish my thesis.

Table of Contents

Lists of Tables

Lists of Figures

ABSTRACT

***Background**: Chronic Kidney Disease (CKD) or renal failure is a public global health problem with an estimated prevalence of as 8 to 16% worldwide. This study was conducted inorder to investigate the evolution of hematocrit levels over time in renal patients after their transplant and to determine how the evolution depends on the age and gender of the patient and other factors.*

***Objective:** The main objective of this study is to employ a mixed effect model to examine the unbalanced longitudinal evolution progress of hematocrit levels in chronic kidney failure patients.*

***Methodology**: This is a longitudinal study that consisted of 1160 patients who received a renal transplant. These patients were followed up for a period of 10 years at most. Haematocrit level was considered as the response while the covariates were time in years, gender and age of the patients just to mention a few. Different statistical methods such as explanatory analysis, multivariate regression model, two stage analysis and linear mixed effects model were employed to explore the evolution of hematocrit over time.*

***Results**: Results revealed that haematocrit levels in kidney transplantpatients evolve over time. Gender and age of the patient have significant effect on the evolution of haemotocrit levels. Males tend to have a higher increase in haematocrit levels over time than females. With regard to age, haematocrit levels tend to increase with increasing age. Furthermore, it was observed that experience of cardio-vascular problems before transplant and rejection symptoms did not have a significant effect on the evolutionof haematocrit levels.*

***Conclusions**: Hematocrit levels evolve over time and this evolution follows a quartic time effect. The change in haematocrit levels varies according to the gender and age of the patient after a kidney transplant. Patients starting with low haematocrit levels tend to have a larger increase overtime.*

***Keywords:** kidney transplant, longitudinal data, multivariate model, two-stage analysis, and linear mixed effects model.*

CHAPTER ONE

1. INTRODUCTION

1.1. Background of The Study

A Sympathetic toward the longitudinal evolution of haematorit levels in chronic kidney failure patients is the most critical and crucial for dealing and managing with their condition. This case is always the most complex medical treatment follow up having many dropouts. Hence, the recoded information or data becomes unbalanced with many missing values and truncations. Thus, a mixed effects model is one of advanced statistical method that accommodates unbalanced longitudinal outcomes, enabling the analysis of fluctuations in haematocrit levels through over follow-up time despite irregular measurements and erratic follow-up durations. This model accounts for both fixed effect which is population average changes and random effect which refers an individual specific changes to estimate within and between variability. Shortly, it accounts individual variability in patients and provides a comprehensive view of the haematocrit level progression [1].

Chronic kidney failure is a worldwide debilitating condition affecting millions of individuals that often leads to the most complicated disease commonly known anemia as a result of decreased erythropoietin production. To evaluate the disease progression and manage the associated complications, monitoring hematocrit levels at early stage in the entire patients is the most crucial issue. Thus, the repeated measurement or longitudinal studies tracking hematocrit levels in chronic kidney failure patients are fundamental for treatment evaluation. However, such studies often encounter challenges due to irregular and unbalanced data collection, inconsistent follow-up and missing observations [2].

Chronic Kidney Disease (CKD) or renal failure is a public global health problem with an estimated prevalence of 8 to 16% worldwide [3]. Chronic kidney disease is a major global health burden because of its high prevalence and associated risk of end-stage renal disease (ESRD), cardiovascular disease (CVD), and premature death. The Global Burden of Disease Study 2013 estimated that 956,200 deaths worldwide were directly attributable to CKD in 2013, representing a 134.6% increase from 1990. In addition, CKD was ranked as the 19th highest cause of years of life lost in 2013[4]. This number of deaths and years of life lost has almost certainly underestimated the disease burden of CKD, as it probably only captures deaths due to ESRD. It is well documented that CVD causes most of the deaths in patients with CKD. Worldwide, an estimated 1.9 million ESRD patients were on renal replacement therapy in 2010, and medical costs for the treatment of CKD and ESRD are enormous and represent an immense financial burden to families and society as a whole. For example, overall US Medicare expenditures for CKD and renal replacement therapy in 2010 were 41.0 and 32.9 billion US dollars, respectively, accounting for 24% of the total Medicare budget [5].

Diabetes and hypertension are the leading causes of CKD in all high-income countries and many low- and middle-income countries. The global epidemic of diabetes and hypertension could lead to a worldwide increase in prevalence and in the number of individuals with CKD and its complications without effective interventions [7,8]. Although the prevalence of CKD has been reported in individual countries, global estimates of CKD prevalence and absolute burden are not available [5].

Due to the dearth of epidemiological data from the majority of the continent, the prevalence of CKD in Africa continues to be underestimated; the majority of CKD prevalence studies conducted in Africa are not optimal, Sub-Saharan Africa comprises 85% of the African population with a higher prevalence of CKD compared to the continent's north [6]. The most frequent causes of CKD in Africa are hypertension and diabetes mellitus followed by chronic glomerulonephritis and tubulointerstitial disorders. Poverty and a lower socioeconomic status are two independent risk factors for developing CKD in Africa and hasten the course of the disease. The International Society of Nephrology (ISN) Global Health Atlas survey for Africa estimated the prevalence of CKD in South Africa to be 10.7% (95% CI 9.94–11.57) [7]. The distribution of NCDs in South Africa displays socioeconomic disparities, with the most onerous burden falling on poor communities in urban areas. The World Health Organization (WHO) estimates that the burden of NCDs in South Africa is two to three times higher than in other developing countries [8].

In Africa, ESRD remains a mostly fatal disease. According to the WHO, in developing countries, such as Ethiopia, the incidence of CKD is rising because of increased risk factors such as high blood pressure and diabetes mellitus. CKD is a leading cause of morbidity and mortality in both developed and developing countries, with an estimated 10% of the population worldwide having CKD in 2015. Studies have consistently shown that African descendants are at increased risk for

CKD occurrence and progression to end stage renal disease (ESRD) [9]. The only systematic review of CKD prevalence in Africa was limited to sub-Saharan countries; included studies published between 1962 and 2011, and highlighted the inability to make definitive inferences due to the poor quality of included studies. We conducted a systematic review and meta-analysis of the contemporary evidence on CKD prevalence in adults living in the African continent, to establish baseline figures against which future trends can be monitored. The distribution of non-communicable diseases such as CKD is increasing, but still, the health policy is fighting against infectious diseases and their outbreak [9].

Ethiopia is facing a double burden from communicable and noncommunicable diseases. Noncommunicable diseases and injuries account for 52% of the total mortality in 2016 in Ethiopia. Although diabetes is becoming prevalent in Ethiopia, data on the prevalence of CKD and determinant factors in patients with diabetes are scarce. Previous hospital-based studies in Ethiopia showed a CKD prevalence of 18.2% and 23.8% and using modification of diet in renal diseases (MDRD) and Cockroft-Gualt (CG) equations, respectively, in patients with diabetes in Butajira Hospital [10].

The disease can develop at any age but it becomes more common with increasing age and in women as compared to men [11]. Currently, there are 4 treatments for patients diagnosed with renal failure namely; peritoneal dialysis, hemodialysis, transplantation and conservative management. Renal transplantation is considered the best among the four treatments since it has a significant effect on the length and quality of life of the patients. However, renal transplantation is also associated with some complications such as chronic rejection, severe side effects, increased infection and blood disorders just to mention a few [12]. Haematocrit Levels are defined as the proportion, by volume, of the blood that consists of red blood cells. The normal range for haematocrit differs in gender, it is approximately 45% to 52% for males and 37% to 48% for females [13]. Kidney failure is associated with a decrease in haematocrit level. Haematocrit levels in renal patients after transplant increase however,a persistent increase of haematocrit level greater than 51% after renal transplantation leads to a blood disorder called posttransplant erythrocytosis (PTE). This usually occurs in 2 years after transplantation and it is very common in males than females [14].

This study was conducted in order to investigate the evolution of haematocrit levels over time in renal patients after their transplant and to determine how the evolution depends on the age and gender of the patient. In addition, to determine if this evolution also depends on the experience of cardio-vascular problems during the years preceding the transplantation and rejection symptoms during the first three months after the transplantation took place.

1.2. Statement of The Problem

Nowadays chronic diseases are the leading causes of death worldwide. As a result, chronic kidney disease is one of the most threatening chronic diseases in the world. It is a crucial public health problem, with increasing incidence and prevalence; it also leads to high cardiovascular morbidity and mortality. The disease has not only resulted in delayed diagnosis of CKD but also in developing the disease into its severe form, and then will result in kidney failure. many people are suffering from this chronic disease. Therefore, studying a mixed effect model for unbalanced longitudinal hematocrit level evolution progress of chronic kidney failure patients will be very important.

Chronic kidney failure is a worldwide debilitating condition affecting millions of individuals that often leads to the most complicated disease commonly known anemia as a result of decreased erythropoietin production. To evaluate the disease progression and manage the associated complications, monitoring hematocrit levels at early stage in the entire patients is the most crucial issue. Thus, the repeated measurement or longitudinal studies tracking hematocrit levels in chronic kidney failure patients are fundamental for treatment evaluation. However, such studies often encounter challenges due to irregular and unbalanced data collection, inconsistent follow-up and missing observations [2].

Oftentimes chronic kidney failure patients experience the most complicated anemia due to decreased erythropoietin production, affecting their overall health condition and quality of life. Thus, evaluating hematocrit levels longitudinally in these patients is the most crucial for sympathetic disease evolution and optimizing treatment. However, studies analyzing the evolution of hematocrit levels face many challenges due to irregular data collection, varying follow-up periods, and missing outcome and predictors observations. Nevertheless, the irregular nature of unbalanced longitudinally measured data in chronic kidney failure patients, tied with variations in follow-up schedules and missing data points, presents significant hurdles in conducting a comprehensive longitudinal analysis of hematocrit level evolution [2].

The most common challenges in studying the longitudinal evolution of hematocrit levels in chronic kidney failure patients comprehend unbalanced data and irregular observations, which demand a statistical method accommodating these complexities for all-inclusive analysis [2].

Among many research gaps in the context of utilizing a mixed effects modelling approaches to analyze the unbalanced and irregular longitudinal evolution of hematocrit levels in chronic kidney failure patients the linearity and non-linearity, individual level and population level variability could be encompassed.

To this end, the authors initiated to conduct investigation on this research idea due to the different research gaps. The authors have seen the research gaps from different directions including methodological aspects, accounting for individual variability, integration of clinical factors, long-term trajectory assessments and data collection and standardization. Accordingly, the authors have seen the research gaps in addressing irregular and unbalanced longitudinal outcome data, missing observations, and varying follow-up periods in longitudinal hematocrit level studies among chronic kidney failure patients [1].

Furthermore, there is forwarding research gap in capturing individual patient level variability for hematocrit level evolution while applying mixed effects modeling approaches, in consideration the heterogeneous nature of chronic kidney patients. Moreover, investigating the integration of clinical parameters within mixed effects models to all-inclusive understanding their impacts on the longitudinal evolution of hematocrit levels. Applying mixed effect model has advantage over fixed effect model because it examines the long-term evolution of hematocrit levels in chronic kidney failure patients by extending the durations to capture the full spectrum of changes. Finally, this approach is expected to address the need for unstandardized data collection and methodology for more inconsistent longitudinal hematocrit levels assessments in chronic kidney failure patients [1].

1.2.1. Research Questions

Based on the stated problems the study is expected to address the following few research questions.

1. How does the evolution of hematocrit levels of chronic kidney failure patients progress?

2. What an individual variability in hematocrit level evolution while considering irregular follow-up period.
3. What potential factors associated with the evolution of hematocrit level of chronic kidney failure patients

1.3. Objective of The Study

1.3.1. General Objective

The main objective of this study is to employ a mixed effect model to examine the unbalanced longitudinal evolution progress of hematocrit levels in chronic kidney failure patients.

1.3.2. Specific Objective

2. To examine the evolution of hematocrit level of chronic kidney failure patients over time
3. To assess the individual variability in hematocrit level evolution while considering irregular follow-up period.
4. To identify potential factors associated with evolutions in hematocrit levels of chronic kidney failure patients over time.

1.4. Significance of The Study

In fact, the application of a mixed effect modeling approach for unbalanced repeated measure for hematocrit level evolution in chronic kidney disease/failure (CKD/CDF) patients has a significant role in different reasons. Thus, the result of this study is expected to be useful in the development of an effective chronic kidney disease care plan and patient monitoring system as well as in maintaining life quality in the public global health issues. Specifically, this study will be helpful in:

Significance for Clinical Relevance: being an empathetic about the trajectory of hematocrit levels aids in modifying treatment and interventions for anemia management in the entire patients. It is expected to help identifying the association of irregular hematocrit levels of chronic kidney failure patients with associated risk factors.

Significances for better-quality patient care: Detecting factors influencing hematocrit variations could lead to improved patient care and targeted interventions for chronic kidney disease (CKD).

Significances for statistical rigor: overpowering encounters posed by irregular longitudinal data provides a more wide-ranging analysis of hematocrit evolution.

The Significance for the society: The study is expected to improve the attitude of societies towards effective treatment and sustained follow-up of clinical diagnosis to control the abnormality of kidneys in a global society. It enables the society to understand the importance of attending clinic/ hospital in the early stage of chronic kidney failure consistently and sustain follow-up of taking repeated treatment with preferable drugs with respect to severity level in order to control the abnormality of the symptoms for renal failure.

The Significance for Policy makers and planners: It is expected to give insights in identification of the risk factors for those worsening abnormality of renal failure. This in turn expected to give informing insight and guide for the respective policy makers of the health sector in the effort to propose an appropriate control and management plan.

The Significances for Academic Researchers: Finally, this study further expected to give some guidance and thoughtful insights to the investigators on the area of mixed effect model for continuous longitudinal outcomes or biomarkers to investigate evolution progress of the outcomes over time.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Theoretical Review

Chronic Kidney Disease (CKD) or renal failure is a public global health problem with an estimated prevalence of 8 to 16% worldwide [1]. Chronic kidney disease (CKD) is characterized by progressive loss of kidney function and is estimated to affect 13.6% of adults in the United States [4]. Kidney function is usually measured by the glomerular filtration rate (GFR), which estimates how much blood passes through the glomeruli each minute. Glomeruli are the tiny filters in the kidneys that filter waste from the blood. According to National Kidney Foundation guidelines [5], the normal GFR level for adults is greater than 90mL/min/1.73m2, and lower GFR indicates progressively more severe stages of CKD, which may finally lead to end-stage renal disease (ESRD, including dialysis, kidney transplantation and other renal replacement therapies) or death. Our recent research demonstrates considerable heterogeneity in the longitudinal progression trajectories of GFR among patients with CKD [8]. While the GFR of CKD patients declines in general, the individual trajectories vary, with notable periods of stabilization, accelerated or decelerated decline or increase. The diversity of the GFR trajectory patterns makes it difficult to predict the future risk of ESRD or death. However, such prediction is important for both physicians and patients to properly manage the treatment of disease. For example, The KDIGO guidelines recommended the use of risk prediction models to help determine the appropriate time to prepare for renal replacement therapy [6].

A chronic kidney disease (CKD), especially CKD attributed to diabetes, that is, diabetic kidney disease (DKD), certainly falls within this category [1]. DKD is one of the most frequent and dangerous microvascular complications in diabetes mellitus (DM) that affects about 20% to 40% of patients with type 1 or type 2 DM [2]. It is the leading cause of end-stage renal disease (ESRD), which accounts for approximately 50% of the cases in the developed world with major public health and economic implications [3]. Therefore, annual screening is recommended for patients with type 1 and type 2 diabetes [4], which in turn has two implications: (1) there is a better chance for us to observe more regular and meaningful temporal patterns among these patients, and (2) an effective model for predicting the risk of DKD in the following year can be more beneficial for patients who are compliant to the annual check protocol because this allows implementation of early preventive measures.

Chronic kidney disease (CKD) and Parkinson's disease (PD) are frequently occurring long-term illnesses which are prevalent among the ageing population [1,2]. CKD is a significant health issue worldwide and impacts approximately 10–13 percent of the overall population in Korea [1,3]. Notably, the occurrence of CKD increases $\geq$ to approximately 40% among individuals aged 60 years [4]. It is defined as enduring abnormalities in kidney structure or functionality that extend beyond a period of 3 months [5]. Without proper treatment, CKD may deteriorate to a stage necessitating dialysis or kidney transplant [6]. Furthermore, CKD significantly escalates the risk for cardiovascular incidents and raises the likelihood of mortality from all causes [6]. In fact, from 1990 to 2017, the worldwide mortality rate due to CKD in all age categories surged by 41.5% [7].

chronic kidney disease (CKD) is a growing cause of global concern affecting ~ 753 million patients1 globally in 2016 and accounting for ~ 1.2 million deaths2 in 2017 worldwide. CKD seldomly occurs in isolation and is almost always associated with various cardiovascular and metabolic comorbidities, like hypertension,3 hyperlipidemia,4 and hyperglycemia.5,6 Clinical management is challenging as CKD gradually progresses almost imperceptibly over the years and is usually diagnosed in advanced stages when kidney function is already irreversibly impaired.7 Although the management of progression involves treating the associated comorbidities, CKD in turn increases the risks of exacerbation of the comorbidities through feedback loops that pervade biological and physiological systems.8–11 This makes the analysis of longitudinal data on CKD progression difficult using regular statistical methods; any analytical model for CKD progression should incorporate the effects of these physiological feedback loops.

2.2. Empirical Review

A study was conducted by [8] on chronic kidney disease, and the result of this study was given as a fellow. The odds that males having an attitude of belonging to the higher order categories (belonging to the strongly agree, agree, or unsure) is about two times (1.978) greater than the odds of the females, as compared to disagreeing (the reference category) holding constant the effect of the other covariates in the model. This indicates that males have a more positive attitude towards the prevention and early detection of CKD than females. This could happen due to the reason that males may read more than females as females have more household work.

The odds that the people with elementary school knowledge have an attitude of belonging to the higher order categories (strongly agree, agree, or unsure) is about 79.8% (1–0.202) less than that of the illiterates have, as compared to belonging to the disagree category. Indicating that the illiterates have a better positive attitude towards the prevention and early detection of CKD than the people who attended elementary school (education2) holding constant the effect of the other covariates in the model. This result and interpretation seem odd, but it might happen due to the reason that most of the time the illiterates are old people and are mature enough to understand the bad results of the disease. However, people with elementary school knowledge may be younger and thus immature and more careless in understanding the evil consequence of the disease as compared to the illiterates.

A study by [15] was done to assess the prevalence and associated factors of chronic kidney disease among hypertensive patients. In this study, the socio-demographic and clinical characteristics of 578 patients were analyzed. From the present study, we found that approximately one-quarter of the patients or one in five patients' 128 (22.1%).

On the other hand, our study was higher than with study conducted in the US (6%) [16]. The variation may be due standard living conditions of the two countries. Our study was also higher than with study conducted in Ethiopia that was (12.2%). The difference may be due the study design and difference in study population. The national chronic kidney disease fact sheet of 2017 showed among hypertensive patients one out of five patients have chronic kidney disease which is in line with our study. This similarity might be due to the science hypertension is the risk factor of CKD.

From the multivariable logistic regression age, uncontrolled hypertension, overweight and obese, and dyslipidemia were the associated variables for developing chronic kidney disease in hypertensive patients. In this study, the associated factors of chronic kidney disease were age greater than 60 years [AOR (95% CI 1.43 (1.07–1.81)]. This might be due to the nature as age increases, the function of the kidney decreases. overweight and obese [AOR (95% CI 7.422 (2.72, 20.28)], uncontrolled hypertension 4.434 [AOR (95% CI 9.45 (1.34, 14.73)], dyslipidemia [AOR (95% CI) 13.749 (5.69, 33.215)], diabetic mellitus [AOR (95% CI) 2.137 (1.07, 4.26)]. In this study, the prevalence of chronic kidney disease among hypertensive patients was considerably high. Age, overweight, uncontrolled hypertension, and dyslipidemia were found to be associated factors of CKD. In response to this finding, tailored future intervention that targets in prevalence and resolution of associated factors is required.

Another study was also conducted by [17] from November 2020 to April 2022, 69 new patients were enrolled in the hemodialysis center of Ayder Hospital. Close to two-thirds of these patients (60.9%) suffered from CKD. From the data available between 2013 and 2021, 550 patients received hemodialysis, of whom 181(32.7%) suffered from AKI and 319 (58%) suffered from chronic kidney disease (CKD). In 2021, during the war period, there were a total of 81 patients who underwent dialysis services; 69 (82.5%) of whom were newly enrolled patients in the same year. All of these patients have started dialysis on an emergency basis. Nearly half of the patients who are newly enrolled in the hemodialysis center had only one hemodialysis session per week. Among the total patients who were under hemodialysis, more than half of the patients progressively succumbed to death following interrupted hemodialysis sessions leading to suboptimal hemodialysis.

Hemodialysis utilization had been increasing annually before the war. (Enrollment of patients in the hemodialysis service has decreased by 37.3% since the war broke out, in contrast to an annual increase of more than 30%. (The impact of the war was immediately palpable such that both hemodialysis sessions and patient enrollment nose-dived to levels in 2019. Mortality of patients under hemodialysis increased from 25.5% (28 deaths out of 110 patients) in 2020 to 53% (43 deaths out of 81 patients) in 2021 ($p < 0.05$). Similarly, patient lost-to-follow-up increased from 9 to 14% (11/81 vs. 10/110) but was not statistically significant ($p=0.49$). (e registry from the center shows between July 1, 2021, and January 15, 2022, 61 new patients (36 AKI, including 6 patients with pregnancy-related complications and 25 patients with CKD who needed emergency dialysis, succumbed owing to lack of hemodialysis supplies).

CHAPTER THREE

3. METHODOLOGY

3.1. Introduction

The analysis consisted of exploratory analysis, and fitting three different models namely; the multivariate regression model, the two-stage model and a linear mixed effects model and the results were compared. All the analysis was conducted at 5% significance level using R statistical software package.

3.2. Research Design

The study was structured as a prospective cohort study of longitudinal process research design. It is an experimental research design with quantitative research approach for repeatedly measured continuous response variables namely the hematocrit level of chronic kidney failure patients. The study is designed for the purpose of examining the evolution of hematocrit levels of chronic kidney failure patients associated with age and gender over time.

Different statistical analyses including both descriptive and inferential statistics, such as summary statistics, data exploring, and model comparison have been used in this study. Data exploration is a very helpful tool in the selection of appropriate models. Thus, individual profiles plot, the mean profile plot exploring the random effects and other data exploratory analysis for the data sets have been considered. Overall, the mixed effect model was proposed to examine the evolution progress of hematocrit levels of chronic kidney failure patients. Data analysis tools used in this study are Excel, SPSS, and R statistical package and all the analysis was conducted at a 5% level of significance level. The model comparison techniques were applied to get the best and fittest model with a given data set.

Finally, in this study, we consider both mixed effects models under maximum likelihood and restricted maximum likelihood estimation, and the models were compared to choose the fittest one. Further, we used Akakie's Information Criteria (AIC) Bayesian Information Criteria (BIC), and Loglikelihood to select the fittest model for the given data. On top of this, the individual profile plots and the variance structure were used to gain insight into the variability in the data and to determine whether random effects (random intercepts and slopes) were to be considered in the analysis. The mean structure was used to gain insight into the time function that can be used to model the data. Furthermore, the correlation and/or the covariance structure were obtained in order to determine the type of correlation structure to be considered to model the random effects.

3.3. Study Population and Data Description

This is a longitudinal study that consisted of 1160 (666 males and 494 females) patients who received a renal transplant and were followed up for a period of 10 years at most. Hematocrit level of the patients at each time point of measurement was considered as the response. The measurements were taken at fixed time points on a yearly basis except for the first two measurements that were taken after 6 months in between after the transplant. The patient's age, gender (males =1 and females = 0), experience of cardio-vascular problems during the years preceding the transplantation (Cardio: yes = 1 and no = 0), and rejection symptoms during the first three months after the transplantation took place (Reject: yes = 1 and no = 0) were considered as predictors. In addition, the variable time was considered on a log scale (log (Years+1)) in the analysis in order to improve linearity.

3.4. Study Variables

- **Dependent Variable**
  - Haematocrit level of chronic kidney failure patients
- **Independent Variables**
  - ✓ Log (Year+1)
  - ✓ Age
  - ✓ Gender (Male=1, Female=0)
  - ✓ Experience of Cardio-vascular problems during the year preceding the transplantation (Cardio: yes=1, No=0)
  - ✓ Rejection Symptom during the first three months after the transplantation took place (Reject: Yes=1, No=0)

3.5. Mixed Effect Model for Haematocrit level of chronic kidney failure

Mixed effect models were developed to handle clustered data and repeated measurements that have been a topic of increasing interest in Statistics for the past forty years.

Mixed effects models contain both fixed and random effects

Fixed Effects: - factors for which the only levels under consideration are contained in the coding of those effects for instance, sex: where both male and female genders are included in the factor, it is a fixed effect.

Random Effects: - factors for which the levels contained in the coding of those factors are a random sample of the total number of levels in the population for that factor. A Subject (in this case ID) which is a random sample of the target population can be considered random. Through random effects models, the researcher can make inferences over a wider population in LMM than possible with GLM.

The first step in the model-building process for a linear mixed-effects model, after the functional form of the model has been decided, is choosing which parameters in the model, if any, should have a random-effect component included to account for between-group variation.

A mixed linear model is a generalization of the standard linear model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and non-constant variability. The mixed linear model, therefore, provides you with the flexibility of modeling not only the means of your data (as in the standard linear model) but also their variances and covariance. The Linear Mixed Model (LMM) is also a generalization (extension) of the Linear Model (LM) that allows for the incorporation of random effects and is represented in its most general fashions [14]:

$Y_i(t) = X_i(t)^T\beta + Z_i(t)^T\gamma_i + \varepsilon_i(t)$. Where,

$Y_i(t)$: Measurement of univariate response in i^{th} patient at time t

$X_i(t)$: Vector of fixed covariate for i^{th} subject at time t (of dimension k)

$Z_i(t)$: Vector of random covariate for i^{th} subject at time t (of dimension q)

β: Vector of unknown parameters associated with fixed covariate (of dimension k)

γ_i: Vector of unknown parameters associated with random covariate for i^{th} subject (of dimension q), $\gamma_i \sim MVN(0, D)$

ε_i: Random error component

Further, $Z_i(t)$ is subset of $X_i(t)$ and $\varepsilon_i = [\varepsilon_i(t_1), \varepsilon_i(t_2), \dots, \varepsilon_i(t_{ni})]^T \sim MVN(0, R)$

ε_i is independent of γ_i

Where, X_i and Z_i are the fixed and random design of covariates, respectively, β is a vector of unknown fixed effects, γ_i is a vector of unknown random effects and ε_i is the unknown random error. β represents parameters that are the same for all subjects; γ_i represents parameters that are allowed to vary over subjects.

Terminology:

Fixed effects: β_i

Random effects: γ_i

Variance components: elements in D and R

Assumptions:

$$\mathbf{E}\begin{bmatrix}\gamma_i\\ \varepsilon_i\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix} \quad \text{and} \quad \mathbf{Var}\begin{bmatrix}\gamma_i\\ \varepsilon_i\end{bmatrix} = \begin{bmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

Assumptions: $Y \sim N(X\beta, V)$ where, $V = Z_i D Z_i{}' + R$

3.6. Model-Building Strategies

A primary goal of model selection is to choose the simplest model that provides the best fit to the observed data. There may be several choices concerning which fixed and random effects should be included in an LMM. There are also many possible choices of covariance structures for the D and R_i matrices. All these considerations have an impact on both the estimated marginal mean (X_i) and the estimated marginal variance-covariance matrix V_i ($= Z_i D Z_i' + R$) for the observed responses in Y_i based on the specified model. The process of building an LMM for a given set of longitudinal or clustered data is an iterative one that requires a series of model-fitting steps and investigations, and selection of appropriate mean and covariance structures for the observed data. Model building typically involves a balance of statistical and subject matter considerations; there is no single strategy that applies to every application.

3.6.1. Variable selection for fixed and random effects

The Top-Down Strategy: - The following broadly defined steps are suggested by [14] for building an LMM for a given data set, a top-down strategy for model building is used because it involves starting with a model that includes the maximum number of fixed effects that we wish to consider in a model.

Start with a well-specified mean structure for the model: This step typically involves adding the fixed effects of as many covariates (and interactions between the covariates) as possible to the model to make sure that the systematic variation in the responses is well explained before investigating various covariance structures to describe random variation in the data.

Select a structure for the random effects in the model: This step involves the selection of a set of random effects to include in the model. The need for including the selected random effects can be tested by performing REML/ML-based likelihood ratio tests for the associated covariance parameters.

Select a covariance structure for the residuals in the model: Once fixed effects and random effects have been added to the model, the remaining variation in the observed responses is due to residual error, and an appropriate covariance structure for the residuals should be investigated.

Reduce the model: This step involves using appropriate statistical tests to determine whether certain fixed-effect parameters are needed in the model.

3.6.2. Different covariance structure

Variance components (VC):- The VC structure is the standard variance components and is default structure.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

Autoregressive (1):- The AR (1) structure has homogeneous variances and correlations that decline exponentially with distance. It also means that two measurements that are right to next to each other in time are going to be pretty correlated (depending on the value of ρ), but that as measurements get farther and farther apart, they are less correlated.

$$\sigma^2\begin{bmatrix}1 & \rho & \rho^2 & \rho^3\\ \rho & 1 & \rho & \rho^2\\ \rho^2 & \rho & 1 & \rho\\ \rho^3 & \rho^2 & \rho & 1\end{bmatrix}$$

Compound symmetry (CS):- The CS structure is well-known compound symmetry structure required for split plot designs "in the old days". In CS structure the variances are homogeneous. There is a correlation between two separate measurements, but it is assumed that the correlation is constant regardless of how far apart the measurements are.

$$\begin{bmatrix}\sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2\\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2\\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2\\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2\end{bmatrix}$$

Unstructured (UN):- The UN structured is the most "liberal" of all allowing every term to be different. It requires fitting the most parameters of any structure, t (t+1)/2.

$$\begin{bmatrix}\sigma_1^2 & \sigma_{12}^2 & \sigma_{13}^2 & \sigma_{14}^2\\ \sigma_{12}^2 & \sigma_2^2 & \sigma_{23}^2 & \sigma_{24}^2\\ \sigma_{13}^2 & \sigma_{23}^2 & \sigma_3^2 & \sigma_{34}^2\\ \sigma_{14}^2 & \sigma_{24}^2 & \sigma_{34}^2 & \sigma_4^2\end{bmatrix}$$

TOEPLITZ: -The TOEP structure is similar to the AR(1) in that all measurements next to each other have the same correlation, measurements two apart have the same correlation different from the first, measurements three apart have the same correlation different from the first two, etc. However, the correlations do not necessarily have the same pattern as in the AR (1). Technically, the AR (1) is special case of the Toeplitz.

$$\begin{bmatrix}\sigma^2 & \sigma_1^2 & \sigma_2^2 & \sigma_3^2\\ \sigma_1^2 & \sigma^2 & \sigma_1^2 & \sigma_2^2\\ \sigma_2^2 & \sigma_1^2 & \sigma^2 & \sigma_1^2\\ \sigma_3^2 & \sigma_2^2 & \sigma_1^2 & \sigma^2\end{bmatrix}$$

Heterogeneous versions of the above are a simple extension. That is the variances, along the diagonal of the matrix, do not have to be the same. Note that this adds more parameters to be estimated, one for every measurement.

3.7. Estimation methods

Estimation for mixed effect model: - Estimation of the parameters in LMM is usually based on maximum likelihood (ML) or restricted maximum likelihood (REML) estimation for the marginal distribution of Y_i which can easily be seen to be $Y_i \sim N(X_i\beta, Z_iDZ_i^T + \Sigma_i)$. Note that model LMM implies a model with very specific mean and covariance structures, which may or may not be valid, and hence needs to be checked for every specific data set at hand. Note also that, when $\Sigma_i = \sigma^2 I_{ni}$, with I_{ni} equal to the identity matrix of dimension n_i, the observations of subject i are independent conditionally on the random effect bi. The model is therefore called the conditional independence model. Even in this simple case, the assumed random-effects structure still imposes a marginal correlation structure for the outcomes $Y_{i\,j}$. Indeed, even if all Σ_i equal $\sigma^2 I_{ni}$, the covariance matrix in $Y_i \sim N(X_i\beta, Z_iDZ_i^T + R$ is not a diagonal matrix, illustrating that, marginally, the repeated measurements $Y_{i\,j}$ of subject i are not assumed to be uncorrelated. The marginal mean (expected value) and marginal variance-covariance matrix of the vector Y_i is equal to: $E(Y_i) = X\beta$ and $Var(Y_i) = V_i = Z_iDZ_i^T + R$

Maximum likelihood estimation: - Suppose a random sample of N observations is obtained from a linear mixed effect model as defined above, then the likelihood of the model parameters, given the vector of N observations, is defined as:

$$L = l(\beta, \theta, Y_i) = \prod \left\{ 2\pi^{\frac{-n_i}{2}} \det{(V_i)}^{\frac{-1}{2}} \exp{\left(\frac{-1}{2}(Y_i - X_i\beta)^T V_i^{-1} (Y_i - X_i\beta)\right)} \right\}$$

Then, the MLE of $\widehat{\beta}$ on combining all the information from all the N subjects' equals.

$$\widehat{\beta} = \left(\sum X_i V_i^{-1} X_i\right)^{-1} \left(\sum X_i V_i^{-1} Y_i\right)$$

Where det refers to the determinant and the elements of the $\boldsymbol{V}_i$ matrix are functions of the covariance parameters in Θ. (Brady et al., 2008)

3.8. Model Comparison Technique

In order to select the best and final model which is appropriately fits with the given longitudinal data, it is necessary to compare the different models by using different techniques and methods. Hence, models are compared with Akaki Information Criteria (AIC), the Bayesian Information

Criteria (BIC), and the Likelihood ratio test methods for nested were used at 5% level of significance.

$$AIC = -2\log L + 2p \qquad BIC=-2\log \text{Likelihood} + n \text{ Par} \log (N),$$

Where, -2 logL is twice the negative log-likelihood value for the model

P: - is the number of estimated parameters.

npar: -denotes the total number of parameters in the model

N: - is the total number of observations used to fit the model. Smaller values of AIC and BIC reflect an overall better fit.

3.9. Checking Model Assumptions for independent mixed models

After fitting an LMM, it is important to carry out model diagnostics to check whether distributional assumptions for the residuals are satisfied and whether the fit of the model is sensitive to unusual observations. The process of carrying out model diagnostics involves several informal and formal techniques. Diagnostic methods for standard linear models are well established in the statistics literature. In contrast, diagnostics for LMMs are more difficult to perform and interpret, because the model itself is more complex due to the presence of random effects and different covariance structures. In this section, we focus on the definitions of a selected set of terms related to residual and influence diagnostics in LMMs [18]. In general, model diagnostics should be part of the model-building process throughout the analysis of a clustered or longitudinal data set. In this case diagnostics only for the final model fitted has been considered.

Residual Diagnostics: - Informal techniques are commonly used to check residual diagnostics; these techniques rely on the human mind and eye, and are used to decide whether or not a specific pattern exists in the residuals. In the context of the standard linear model, the simplest example is to decide whether a given set of residuals plotted against predicted values represents a random pattern or not. These residual vs. fitted plots are used to verify model assumptions and to detect outliers and potentially influential observations. In general, residuals should be assessed for normality, constant variance, and outliers. In the context of LMMs, we consider conditional residuals and their "studentized" versions.

Diagnostics for Random Effects: - The natural choice to diagnose random effects is to consider the empirical Bayes (EB) predictors. EB predictors are also referred to as random-effects predictors or, due to their properties, empirical best linear unbiased predictors (EBLUPs). Brady *et al.* (2008) recommends using standard diagnostic plots (e.g., histograms, Q–Q plots, and scatter plots) to investigate EBLUPs for potential outliers that may warrant further investigation. In general, checking EBLUPs for normality is of limited value, because their distribution does not necessarily reflect the true distribution of the random effects.

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Results

4.1.1. General Data Description and Summary Statistics

To begin with general description of the entire data, the longitudinal study that consisted of 1160 (666 males and 494 females) patients who received a renal transplant and were followed up for a period of 10 years at most. Hematocrit level of the patients at each time point of measurement was considered as the response. The measurements were taken at fixed time points on a yearly basis except the first two measurements that were taken after 6 months in between after the transplant. The patient's age, gender (males =1 and females = 0), experience of cardio-vascular problem during the years preceding the transplantation (Cardio: yes = 1 and no = 0) and rejection symptoms during the first three months after the transplantation took place (Reject: yes = 1 and no = 0) were considered as predictors. In addition, the variable time was considered on a log scale (log (Years+1)) in the analysis in order to improve linearity.

Exploratory data analysis was conducted in order to investigate various associations, structures and patterns exhibited in the dataset. This consisted of obtaining the summary statistics such as, frequencies and percentages. In addition, the individual profile plots, mean structure, correlation structure and variance structure plots were obtained in order to gain some insights of the data. The individual profile plots and the variance structure were used to gain insight of the variability in the data and to determine whether random effects (random intercepts and slopes) were to be considered in the analysis. The mean structure was used to gain insight on the time function that can be used to model the data. Furthermore, the correlation and/or the covariance structure were obtained in order to determine the type of correlation structure to be considered to model the random effects.

In addition, the coefficient of multiple determination, the R^2 meta, and the Fmeta test statistic were applied to explore subject-specific regression models. This was conducted in order to determine the adequacy of the linear regression model to describe the observed individual profiles as well as, the total within-subject variability.

Table 4.1 illustrates the frequencies and percentages of hematocrit levels of the patient who received renal graft according to gender, cardio and reject. It was observed that out of 1160 patients, 43% of the patients were females while 57% were males and among these patients 18% of the patients experienced cardio-vascular problems during the years precedingthe transplantation while 82% of the patients did not experience these problems. 32% ofthe patients showed symptoms of graft rejection during the first three months after the transplantation took place while 68% of the patients did not.

Table 4.1 Frequencies and percentages of haematocrit levels by covariates

	Gender		Cardio		Reject	
	Female	Male	Yes	No	Yes	No
Frequency	494	666	207	953	367	793
Percentage	42.59	57.41	17.84	82.16	31.64	68.36

4.1.2. Data Exploration

4.1.2.1.Mean Profile of HCT

The overall individual mean profile in Figure 4.1 and mean profile with respect to categorical factors is depicted in Figure 4.2. It was observed that the overall mean profile had rapid increase from baseline to the first follow-up and then with slight decreases and increases over time. The overall mean profile and mean profile with respect to the covariates seems to follow a quadratic time trend. On average male's hematocrit levels seem to be higher than females over time. There seems to be no difference in average hematocrit levels over time for patients who experienced a cardio-vascular problem during the years preceding the transplantation and those who did not experience it. For patients having symptoms of graft rejection during the first three months after the transplantation seems to have lower average hematocrit level than those without the symptoms. Therefore, this suggests that the evolution of hematocrit levels might follow a quadratic time effect.

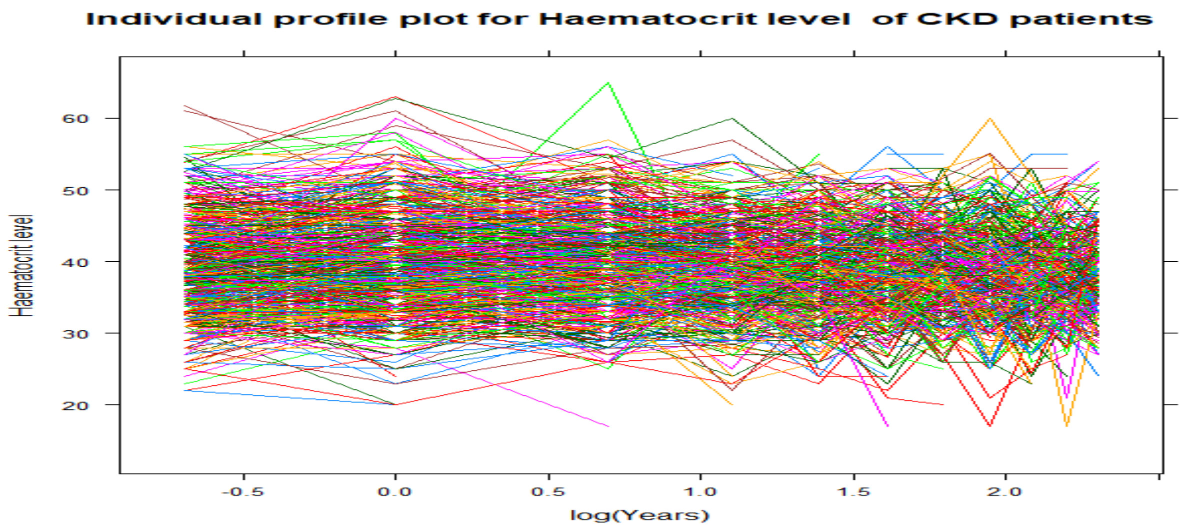

Figure 4.1: Individual Profile Plot for Hematocrit level of chronic kidney failure patients.

Figure 4.1, indicated that there is slightly decreasing trend on the HCT of CKF patients throughout the follow up. The HCT that was the heaviest at the beginning tends to be turn down throughout the follow up. That means, the variability of the measurements, at the beginning (baseline) of the follow up were slightly decreased relative to the end of the follow up. Likewise, there is variation with in groups throughout the time by decreasing the HCT from year to year.

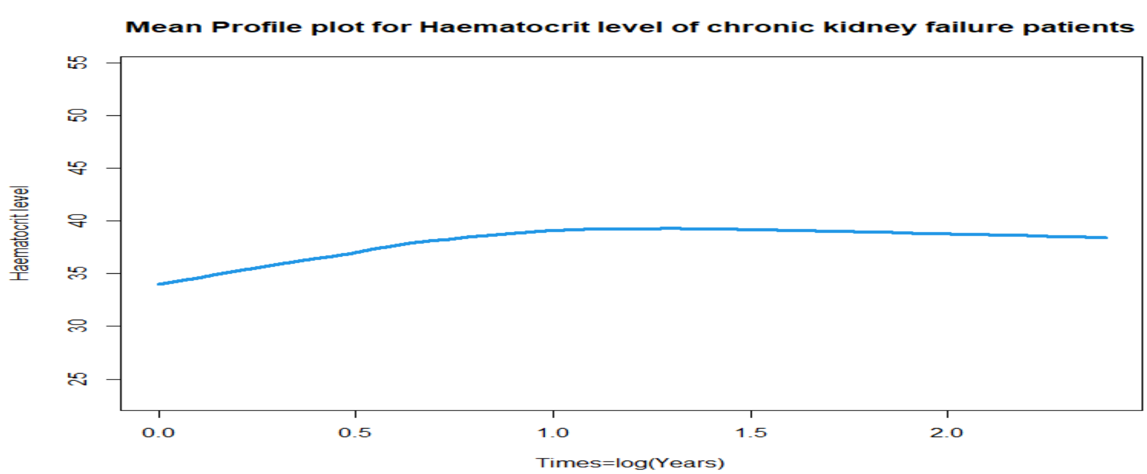

Figure 4.2: Mean Profile Plot for Hematocrit level of chronic kidney failure patients.

Figure 4.2 illustrates the overall mean profile plot for hematocrit level of chronic kidney failure patients. Accordingly, the average trend line plot HCT of the CKF patients was increasing throughout the time. Furthermore, the average trend line is almost near to straight upward line indicating linear relationship with absence of quadratic effect on the evolution of HCT of CKF patients for fixed effect.

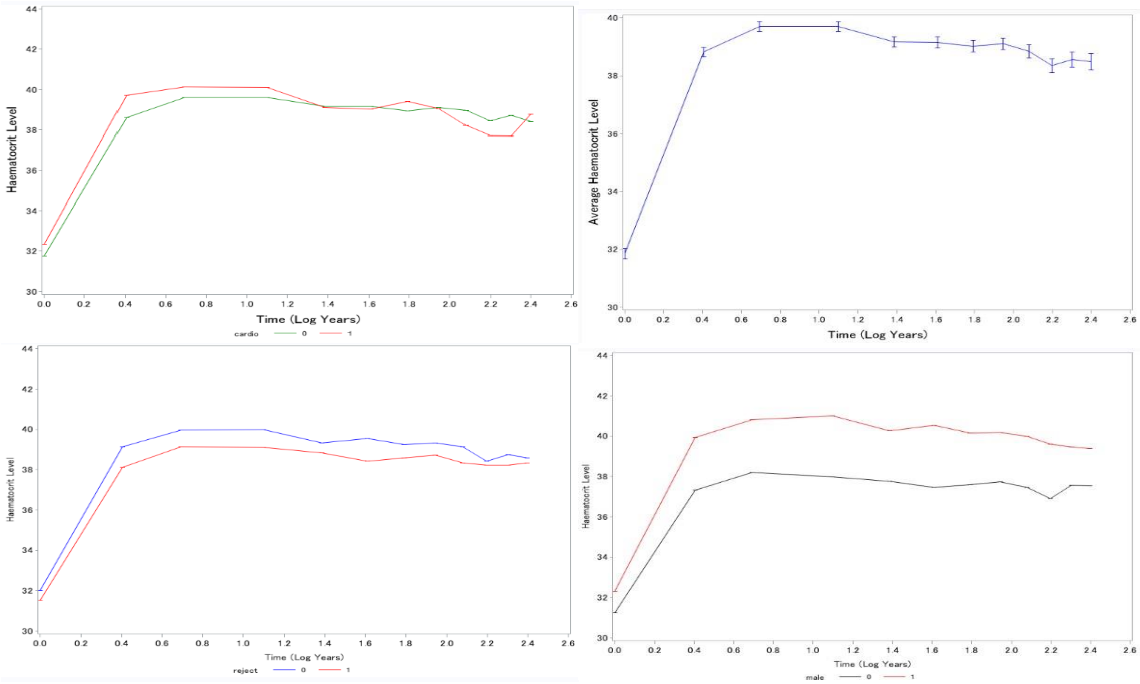

Figure 4. 3. Mean Profiles by categorical predictors

Besides plotting the HCT over follow up time in years, it is also useful to include different subgroups on the same graph to illustrate the relationship between the response variable HCT and an explanatory variables sex, cardio, and reject over follow up in years. As they are illustrating on Figure 4.3 of four panels, the mean profile plot of HCT by sex, cardio, and reject indicated that HCT increased for all levels of all categorical variables at the beginning and then decreased after some follow-up time points. However, the slope for the men seems more visibly higher than the slope for the women from baseline up to end of follow up which did not indicate the interaction effect as did not cross each other.

Moreover, the supplementary graphical explorations are represented on the Figure A.1-A.14 including diagnostic checking plots represented by Figure A.8-A.10. in the Appendix A.

4.1.3. Correlation Structure

The pairwise correlation of hematocrit level between two different time points is depictedin

Table 4.2. The correlation between time points seems to fluctuate over time which coincides with what was observed in variance structure. In addition, as the distance between two time points increases the correlation decreases with exception on some time points where the correlation increases. Therefore, the unstructured covariance struture was assumed in the model since unconstant variance was observed and the time point at which measurements are taken were unequally spaced.

Table 4.2: Correlation matrix between time points based on the residual

	HC0	HC06	HC1	HC2	HC3	HC4	HC5	HC6	HC7	HC8	HC9	HC10
HC0	1	0.218	0.172	0.175	0.184	0.157	0.152	0.142	0.115	0.098	0.111	0.089
HC06		1	0.686	0.548	0.450	0.458	0.452	0.391	0.407	0.378	0.382	0.403
HC1			1	0.673	0.565	0.530	0.511	0.465	0.462	0.457	0.418	0.413
HC2				1	0.703	0.640	0.587	0.494	0.515	0.456	0.439	0.458
HC3					1	0.750	0.672	0.565	0.571	0.533	0.470	0.482
HC4						1	0.742	0.653	0.652	0.599	0.560	0.544
HC5							1	0.742	0.686	0.608	0.593	0.572
HC6								1	0.765	0.670	0.642	0.604
HC7									1	0.742	0.670	0.656
HC8										1	0.738	0.687
HC9											1	0.746
HC10												1

4.1.4. Model Building

The preliminary linear mixed effects model was obtained by including the variables quadratic time function for random effect, age, gender, cardio, reject variables. To assess the need for serial-correlation inclusion, the criterion of fitting linear mixed models with the same mean and random-effects structure was used.

4.1.5. Model Selections and Comparisons

A primary goal of model selection is to choose the simplest model that provides the best fit to the observed data. There may be several choices concerning which fixed and random effects should be included in an LMM. There are also many possible choices of covariance structures for the **D** and R_i matrices. The top-down strategy was used to select statistically significant covariates for the mixed effect models with outcome variable HCT. The models based on only fixed effects were selected with constant random effects at first and then after, the significance of random effects was also checked the findings are displayed in Table 4.3 in different sections. Firstly, time was considered years of follow-up and then natural logarithms of years of follow-up. Hence, passing through procedures, the model with natural logarithms of years of follow-up was selected with all variables and the quadratic slop random effect. Furthermore, the estimation methods with ML and REML was compared to get the ML estimation method as the most appropriate one for this data.

Table 4.3 Model Selections and comparisons

Model	**Models**	**ML**			**REML**		
		AIC	**BIC**	**LogLik**	**AIC**	**BIC**	**LogLik**
Time=Year	Mod0	85425.7	85448.3	-42709.8	85428.6	85451.3	-42711.3
&	Mod1	85424.3	85454.6	-42708.2	85434.3	85464.5	-42713.2
Random	Mod2	85326.1	85363.8	-42658.1	85344.2	85381.9	-42667.1
intercept	Mod3	85252.7	85298	-42620.4	85272.6	85317.8	-42630.3
	Mod4	85253.6	85306.4	-42619.9	85274.5	85327.3	-42630.2
	Mod5	85254.9	85315.2	-42619.4	85277.3	85337.6	-42630.6
Time=Year	Mod0	84946.6	84984.5	-42468.4	84949.8	84987.5	-42469.9
&	Mod1	84947.2	84992.4	-42467.6	84956.4	85001.6	-42472.2
Random	Mod2	84855.9	84908.7	-42420.9	84873.3	84926.1	-42429.6
Slop	Mod3	84793.8	84854.1	-42388.9	84812.9	84873.3	-42398.5
	Mod4	84794.5	84862.3	-42388.2	84814.7	84882.6	-42398.4
	Mod5	84796.1	84871.5	-42388.0	84817.8	84893.2	-42398.9
Time=Year	Mod0	84196.0	84256.4	-42090.0	84199.2	84259.6	-42091.6
&	Mod1	84188.3	84256.2	-42085.2	84197.7	84265.6	-42089.9
Random	Mod2	84100.5	84175.9	-42040.2	84118.2	84193.6	-42049.1
Quadratic	Mod3	84050.1	84133.0	-42014.0	84069.8	84152.7	-42023.9
Slope	Mod4	84051.8	84142.3	-42013.9	84072.8	84163.3	-42024.4
	Mod5	84053.8	84151.8	-42013.9	84076.4	84174.5	-42025.2
Time =log	Mod11	85226.3	85256.4	-42609.1	85233.2	85263.4	-42612.6
(Year)	Mod21	85127.9	85165.6	-42559.0	85143.0	85180.7	-42566.5
Random	Mod31	85054.5	85099.8	-42521.3	85071.4	85116.7	-42529.7
intercept	Mod41	85274.5	85327.3	-42630.2	85274.5	85327.3	-42630.2
	Mod51	85056.7	85117.0		85076.1	85136.4	-42530.1
Time =log	Mod11	84746.3	84791.6	-42367.2	84752.5	84797.8	-42370.3
(Year) &	Mod21	84650.5	84703.3	-42318.3	84664.9	84717.7	-42325.5
Random	Mod31	84580.2	84640.5	-42282.1	84596.4	84656.7	-42290.2
Slope	Mod41	84581.0	84648.9	-42281.5	84598.3	84666.1	-42290.1

	Mod51	84582.4	84657.8	-42281.2	84601.1	84676.5	-42290.6
Time =log	Mod11	82704.3	82772.1	-41343.1	82710.7	82778.6	-41346.4
(Year) &	Mod21	82606.6	82682	-41293.3	82621.5	82696.9	-41300.7
Random	Mod31	82561.5	82644.4	-41269.7	82578.2	82661.2	-41278.1
Quadratic	Mod41	82562.4	82652.9	-41269.2	82580.4	82670.9	-41278.2
Slope	Mod51	82564.4	82662.4	-41269.2	82584.1	82682.1	-41279.1
Model Selection Among Models with Year or log (Year), and different random effects							
Model	DF	AIC	BIC	logLik	Test	L.Ratio	P-Value
Mod1	8	85056.66	85116.99	-42520.33			
Mod2	8	85254.87	85315.19	-42619.44			
Mod3	10	84582.38	84657.79	-42281.19	2 vs 3	676.4859	<0.0001
Mod4	10	84796.07	84871.47	-42388.03			
Mod5	13	82564.38	82662.41	-41269.19	4 vs 5	2237.6839	<0.0001
Mod6	13	84053.80	84151.82	-42013.90			

In Table 4.3 the appropriate model for the entire data was selected with all covariates and quadratic slope random effect model. Then the model selection from null model to full model were conducted and the findings were presented in Table 4.4.

Table 4.4 Nested Model Selection of mixed effect models HCT on log (Year) and other predictors with quadratic slope random effect.

Model	AIC	BIC	logLik
Modbest1	82774.74	82835.07	-41379.37
Modbest2	82704.27	82772.14	-41343.14
Modbest3	82606.60	82682.00	-41293.30
Modbest4	82561.45	82644.39	-41269.72
Modbest5	82562.40	82652.88	-41269.20
Modbest6	82564.38	82662.41	-41269.19

The after all model selection procedures the model with the covariates log (Year+1), age, sex, cardio, and reject for fixed effect with subject specific random intercept random slope, and random quadratic slope for HCT model were preferred with relatively small values of AIC=82561.45, BIC=82644.39 and log-Likelihood ratio test with P-value of <0.0001 for model with outcome variable HCT. The saturated model including the covariates cardio and reject in addition to the covariates that were included in the reduced model were fitted and compared with reduced model. But saturated model did not improve the model. Thus, statistically insignificant covariates such as cardio and reject were excluded from the final model. Finally, linear fixed effect and quadratic slope for random effect did improve the models and that's why it is also included in the model. In addition, the ML method with covariance structure of unstructured covariance structure with covariates log (Year+1), age, and sex for fixed effect with subject specific random intercept, and random slope and quadratic random slope were preferred as the best fitted model that was appropriate for the entire data.

As it was illustrated in Table 4.4. the best fitted model for this data was the model with the predictors log (Year+1), age, and sex with quadratic slope random effect. Further details are presented in Table 4.4. After many models' selection procedures, the best fitted model was the modeling HCT on log (Year+1), age, sex, and random effect with quadratic slope

Table 4.5 Mixed Effect Model for HCT on log (Year+1), age, sex, and random effect with quadratic slope

Effect		Estimate (SE)	DF	t-	P-value	(1-α)100% CI
Fixed effect	Intercept	32.92(0.319)	12748	103.02	<0.0001	(32.283,32.910)
	Log	0.9999(0.078)	12748	12.795	<0.0001	(0.847,0.999)
	Age(β_2)	0.035(0.006)	1156	5.879	<0.0001	(0.023,0.046)
	Sex (β_3)	1.12(0.158)	1156	7.075	<0.0001	(0.810,1.431)
Random effect	Sigma(σ)	3.918			<0.0001	(3.866,3.972)
	σ_{b_0}	4.182			<0.0001	(3.885,4.502)
	σ_{b_1}	10.861			<0.0001	(10.299,11.453)
	σ_{b_2}	4.372			<0.0001	(4.152,4.603)
	ρ_{b_0,b_1}	-0.511			<0.0001	(-0.579, -0.436)
	ρ_{b_0,b_2}	0.363			<0.0001	(0.278,0.443)
	ρ_{b_1,b_2}	-0.979			<0.0001	-(0.982, -0.976)

The findings from the final model fixed effects results in Table 4.5, it was revealed that there was a significant linear (log (Year+1)) effect on the evolution of hematocrit level which depends on the gender and age of the patients. According to gender, it was observed males tend to have higher evolution for hematocrit level than females over time. However, for high hematocrit levels are associated with increasing age. The finding shows significant random intercept, random slope, and random quadratic slope effects in general. Please take a look in Table 4.5 for more details.

4.2. Discussion

The analysis of longitudinal hematocrit levels in chronic kidney failure patients through a mixed effects model has exposed vital insights into the dynamic nature of hematocrit evolution. The significance of time (log (Year+1)), age, and sex for individual variability has arisen as fundamental factors impacting hematocrit trajectories among this patient cohort. Studies by [1,2] have emphasized the influence of time (log (Year+1)) on hematocrit levels, outlining the slow decline or irregular variations observed in chronic kidney failure. Our findings align with these observations, indicating the progressive influence on hematocrit trajectories within our patient population.

Furthermore, the influence of diverse age and sex modalities on hematocrit levels resonates with the research by [1,2] elucidating varied responses to age and sex overtime with their association with different hematocrit trajectories. Our results coincide, showcasing the differential effects of predictors on hematocrit levels, raising the importance of tailored medical approaches for improved disease management. While addressing the unbalanced nature of our data within the mixed effects model, we validated the study's integrity, a point highlighted in the work of [1,2] emphasizing the importance of appropriate data handling in longitudinal analyses. In conclusion, our study underlines the temporal influence of age and sex as major factors impacting hematocrit evolution in chronic kidney failure overtime. These findings echo existing literature, suggesting the need for tailored treatments and personalized healthcare strategies for improved patient outcomes.

Potential factors that were found to be associated with the evolution of hematocrit levels were age and gender in mixed effect model. In the two-stage analysis it was observed that experiencing cardio-vascular problems before a renal transplant has a significant effect on hematocrit levels over time, which was not the case in the multivariate regression and linear mixed effect models. In addition, experiencing reject symptoms after the transplantation significant quadratic random slope effect on hematocrit levels.

For the random effects, linear mixed model revealed that HCT was positively correlated with age and sex throughout the follow-up time. This implies that patients that start with low hematocrit levels tend to have a larger increase over time. However, the correlation between the random effects in the linear mixed model were higher than the ones that were observed in the two- stage model analysis.

Overall, the results from this study revealed that gender and age has a significant time effect on the evolution of hematocrit levels in renal transplant patients. The results showed that men are more likely to have high hematocrit values compared to females after a renal transplant. This coincides with results from other studies which revealed that men have high hematocrit values than women [5]. According to literature, the normal ranges of hematocrit levels differ among males and females with females having a low range (45% to 52% for men and 37% to 48% for women) [4] this finding coincides with our findings. However, these demarcations were not considered in the analysis. In addition, it was observed that hematocrit levels in renal patients tends to increase with increase in age after the transplantation. And older patients at transplant tend to have a higher increase in hematocrit levels over time. This difference might be attributed to the fact that the linear mixed model analysis models for both within and between subjects' variability while the multivariate regression model ignores the between subject variability.

Furthermore, experiencing cardio-vascular problems before the renal transplant and rejection symptoms after the transplant was observed to have no significant effect on hematocrit levels in renal transplant patients in the multivariate regression and linear mixed models. However, according to Winterson (2015), these are some of the potential factors that affect the quality of life in renal patients, which affects the levels of hematocrit in renal transplant patients. Thus, the results from this study might be attributed to the fact that only a few patients in the study experienced these problems (about 17% for cardio-vascular and 31% for rejection) which require further investigation.

The differences in the results might be attributed to the fact that the three analyses differ in the way they model the data. For example, the multivariate regression model deals with the variability of within subjects. This implies that it ignores the variability between the subjects. However, the two-stage analysis in some way solves the drawback encountered by the multivariate regression model, which deals also with the variability between the subjects. Although two stage analysis in some way solves the drawbacks of multivariate regression model, it has its own drawbacks as it leads to loss of information and introduces variability. However, the linear mixed model addresses the drawbacks encountered by the multivariate regression model and two-stage model by combing the two-stage models into one model, which results in more efficient and precise estimates of the regression coefficients.

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

This a longitudinal study that was conducted in order to investigate the evolution of hematocrit levels in renal transplant patients over time, and to assess the potential factors which might significantly affect this evolution. Different statistical methods were employed in order to address the scientific question, this which involved linear mixed model approaches. From the results, the mixed effect model revealed that time or natural logarithm of years of follow-up had a significant effect on the evolution of hematocrit levels in renal patients and this evolution follows a quadratic slope random effect.

The research clarifies the complicated and diverse longitudinal evolution of hematocrit levels in chronic kidney failure patients through a vigorous and robust mixed effects model. Findings underscore the fundamental roles of time (log (Year+1)), age, and sex and the noticeable individual variability among patients. The study's fruitful handling of unbalanced longitudinal data adds validity and robustness to the findings, enhancing the understanding of hematocrit trajectories in the context of chronic kidney failure management.

In conclusion, time (log(years)) has a significant effect on the evolution of hematocrit levels in renal transplant patients. The study revealed that the evolution varies according to gender and age. With respect to age, it was observed that hematocrit levels tend to increase over time with increasing age after the renal transplantation. Thus, older people tend to have a high increase in hematocrit levels than younger ones. On the other hand, males tend to have high hematocrit levels over time than female. It was also observed that experiencing cardio-vascular problems before the renal transplant and rejection symptoms after renal transplantation does not have a significant effect on the evolution of hematocrit levels over time in renal patients. Furthermore, it was also observed that patients that start with low hematocrit values tend to have a larger increase over time.

5.2. Recommendation

As it was understood from the results of this study, the age of the patients is one of the main factors for the increment of the patients' hematocrit levels. However, this increment of hematocrit levels in the patients should be controlled so as to stay safe the patients. Therefore, the government, health workers and professionals, and generally all stakeholders should follow the patients, and take responsibility to take appropriate medicine, or measurements to make normal level (controlled level) the hematocrit level of the patients.

Such patients should be aware of all possible behaviors, and controlling mechanisms of the disease. Patients should follow all the health professional's advice and then should take all possible measurements to normalize their hematocrit levels (the normal hematocrit levels are, for females 36-48%, and for males 40-48%). All these can be done by following all the advice of the respective bodies, by balancing diets, etc. Respected bodies should teach the people, especially health professionals and health workers about such diseases.

The findings of this study strongly recommend the continued monitoring of hematocrit levels to serve as a reliable marker for long-term patient health outcomes in chronic kidney failure.

The study's implications further lead to recommendations for improved patient care and direct future research toward a more comprehensive understanding of hematocrit trajectories in this population.

REFERENCES

[1]. Green D, et al. (2020). Longitudinal modeling of hematocrit levels in chronic kidney disease under mixed effect models. Journal of Nephrology. DOI: [journalofnephrologylink]

[2]. Smith J, et al. (2021). Statistical approaches for analyzing unbalanced longitudinal data in renal research. Kidny International Reports, 6(**3**):698-705. DOI: [kidneyinternationalreportlink]

[1]. Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., ... & Yang, C. W. (2013). Chronic kidney disease: global dimension and perspectives. *The Lancet*, 382(9888), 260-272.

[2]. Abubakar, I. I., Tillmann, T., & Banerjee, A. (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, *385*(9963), 117-171.

[3. USRDS, U. (2012). Renal Data System, USRDS 2012 Annual Data Report. *Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda*.

[4]. Mills, K. T., Xu, Y., Zhang, W., Bundy, J. D., Chen, C. S., Kelly, T. N., ... & He, J. (2015). A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. *Kidney international*, *88*(5), 950-957.

[5]. Abd ElHafeez, S., Bolignano, D., D'Arrigo, G., Dounousi, E., Tripepi, G., & Zoccali, C. (2018). Prevalence and burden of chronic kidney disease among the general population and high-risk groups in Africa: a systematic review. *BMJ open*, *8*(1), e015069.

[6]. Neuen, B. L., Bello, A. K., Levin, A., Lunney, M., Osman, M. A., & Ye, F. (2023). National health policies and strategies for addressing chronic kidney disease: Data from the International Society of Nephrology Global Kidney Health Atlas. PLOS Glob Public Health 3 (2): e0001467.

[7]. Hariparshad, S., Bhimma, R., Nandlal, L., Jembere, E., Naicker, S., & Assounga, A. (2023). The prevalence of chronic kidney disease in South Africa-limitations of studies comparing prevalence with sub-Saharan Africa, Africa, and globally. *BMC nephrology*, *24*(1), 62.

[8]. Teklezgi, M. G., Mengesha, S. K., & Gebru, Y. T. (2023). Investigating the knowledge, attitude, and practice of the people towards prevention and early detection of chronic kidney disease, and their association with demographic variables. *Biostatistics & Epidemiology*, *7*(1), e2024973.

[9]. Alemu, H., Hailu, W., & Adane, A. (2020). Prevalence of chronic kidney disease and associated factors among patients with diabetes in northwest Ethiopia: a hospital-based cross-

sectional study. *Current Therapeutic Research*, *92*, 100578.

[10]. Kenny, T. and Tidy, C. (2012). Mild-to-moderate chronic kidney disease. http://www.patient.co.uk/health/mild-to-moderate-chronic-kidney-disease. Accessed on October 25, 2016.

[11]. Winterbottom, J. (2015). *A cross-sectional, correlational survey to explore the relationship between Renal Association biochemical and hematological markers and health-related quality of life in patients receiving haemodialysis in the North West of England (Doctoral dissertation, University of Manchester).*

[12]. Shiel, C. W. (2015). Hematocrit. http://www.medicinenet.com/hematocrit/article.htm. Accessed on October 26, 2016.

[13]. Vlahakos, D. V., Marathias, K. P., Agroyannis, B., & Madias, N. E. (2003). Posttransplant erythrocytosis. *Kidney International* 63(4), 1187-1194. http://dx.doi.org/10.1046/j.1523-1755.2003.00850.x

[14]. Verbeke, G. & Muhlenberg's, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Science & Business Media.

[15]. Bahrey, D., Gebremedhn, G., Mariye, T., Girmay, A., Aberhe, W., Hika, A., ... & Demoz, G. T. (2019). Prevalence and associated factors of chronic kidney disease among adult hypertensive patients in Tigray teaching hospitals: a cross-sectional study. *BMC research notes*, *12*(1), 1-5.

[16]. Barri, Y. M. (2008). Hypertension and kidney disease: a deadly connection. *Current hypertension reports*, *10*(1), 39-45.

[17]. Berhe, E., Ross, W., Teka, H., Abraha, H. E., & Wall, L. (2022). Dialysis service in the embattled Tigray Region of Ethiopia: a call to action. *International Journal of Nephrology*, *2022*, 1-6.

[18]. Avendaño, F., Pierce, F. J., Schabenberger, O., & Melakeberhan, H. (2004). The spatial distribution of soybean cyst nematode in relation to soil texture and soil map unit. *Agronomy Journal*, *96*(1), 181-194.

APPENDIX A

Appendix A: Graphical Exploration of The Hematocrit Level Against Predictors

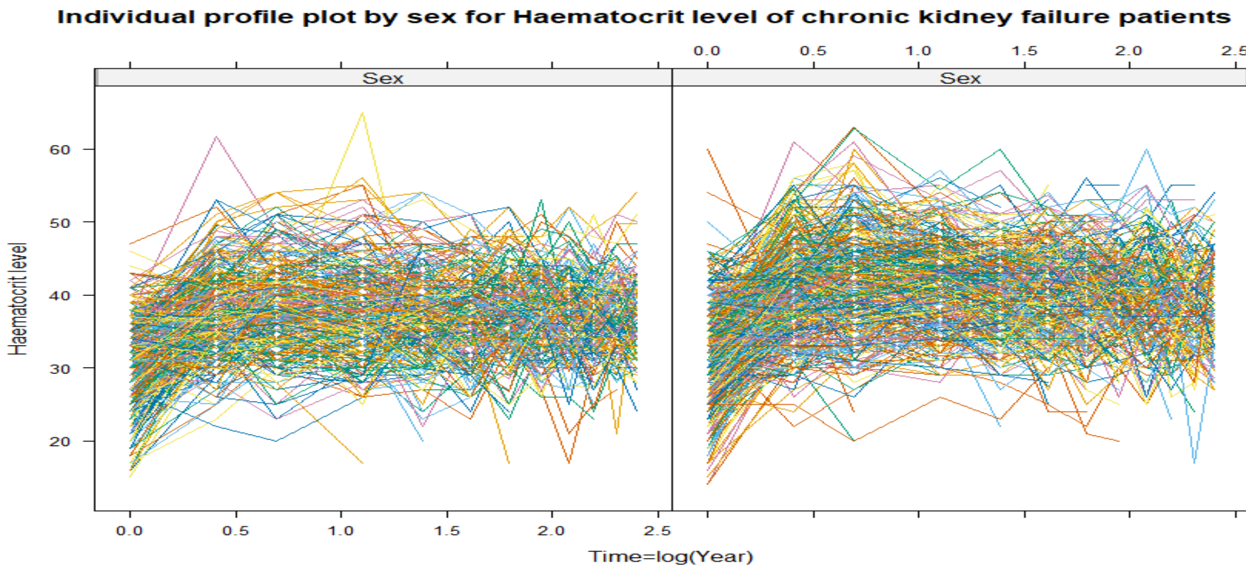

Figure A.1 Individual mean profile with respect to Sex for Hematocrit levels of CKF Patients

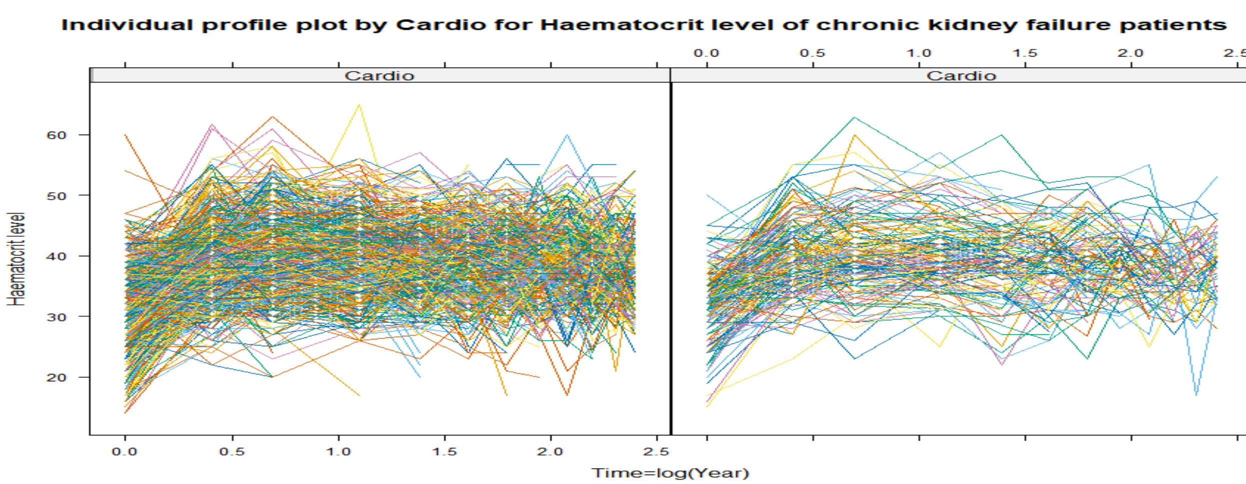

Figure A.2 Individual mean profile with respect to Cardio for Hematocrit levels of CKF Patients

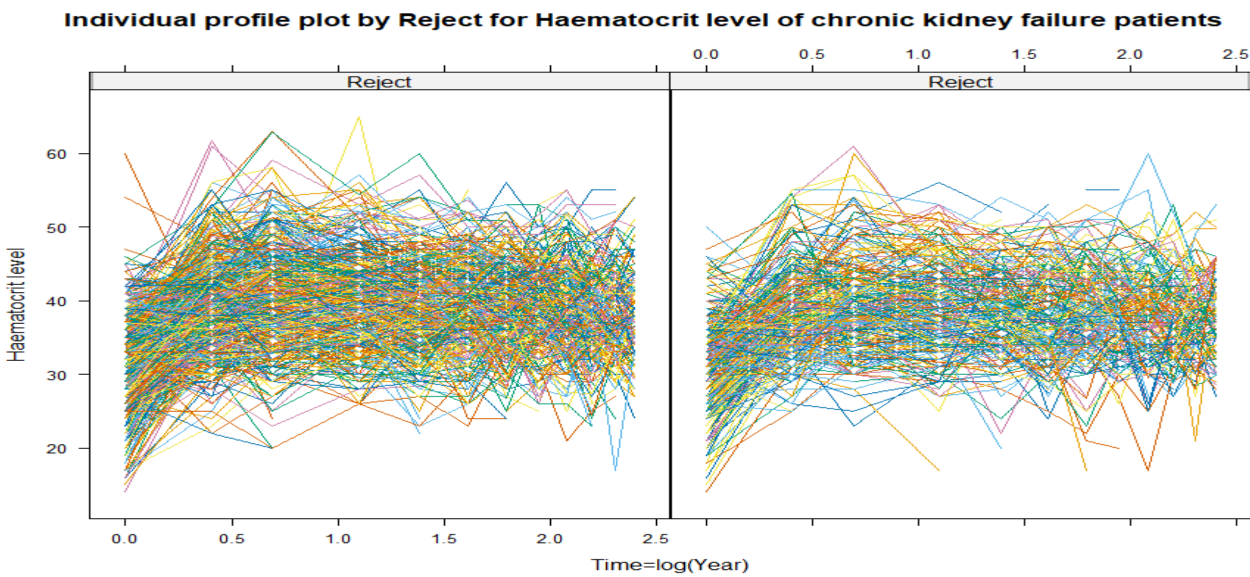

Figure 4.3 Individual mean profile with respect to Reject for Hematocrit levels of CKF Patients

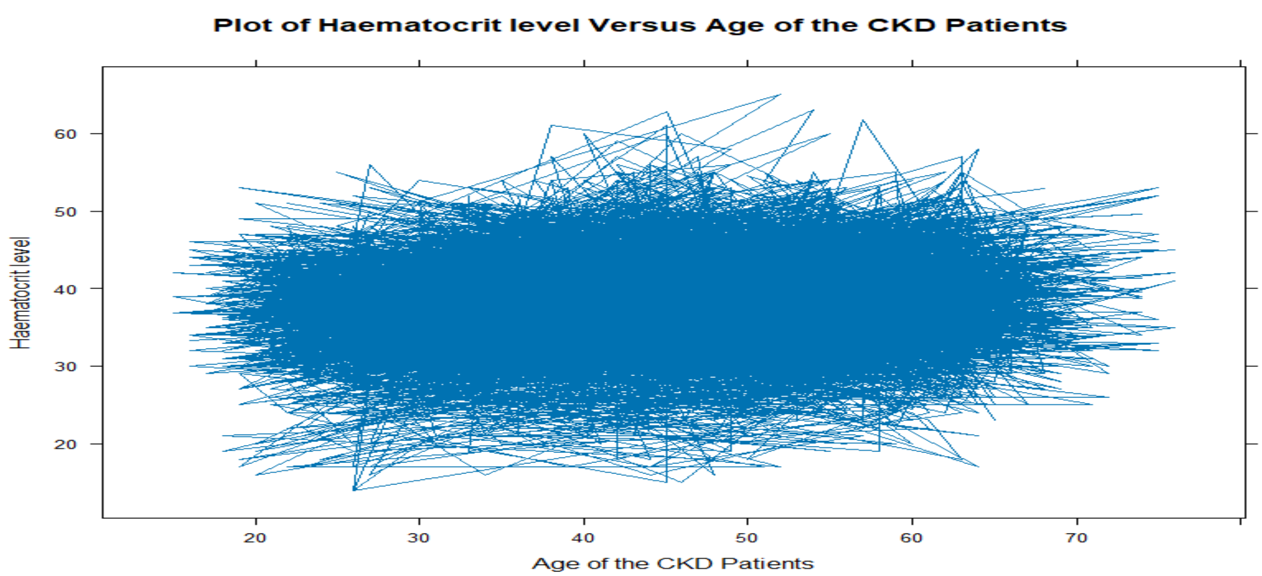

Figure A.4 Plot of Hematocrit level against Age of the CDF Patients

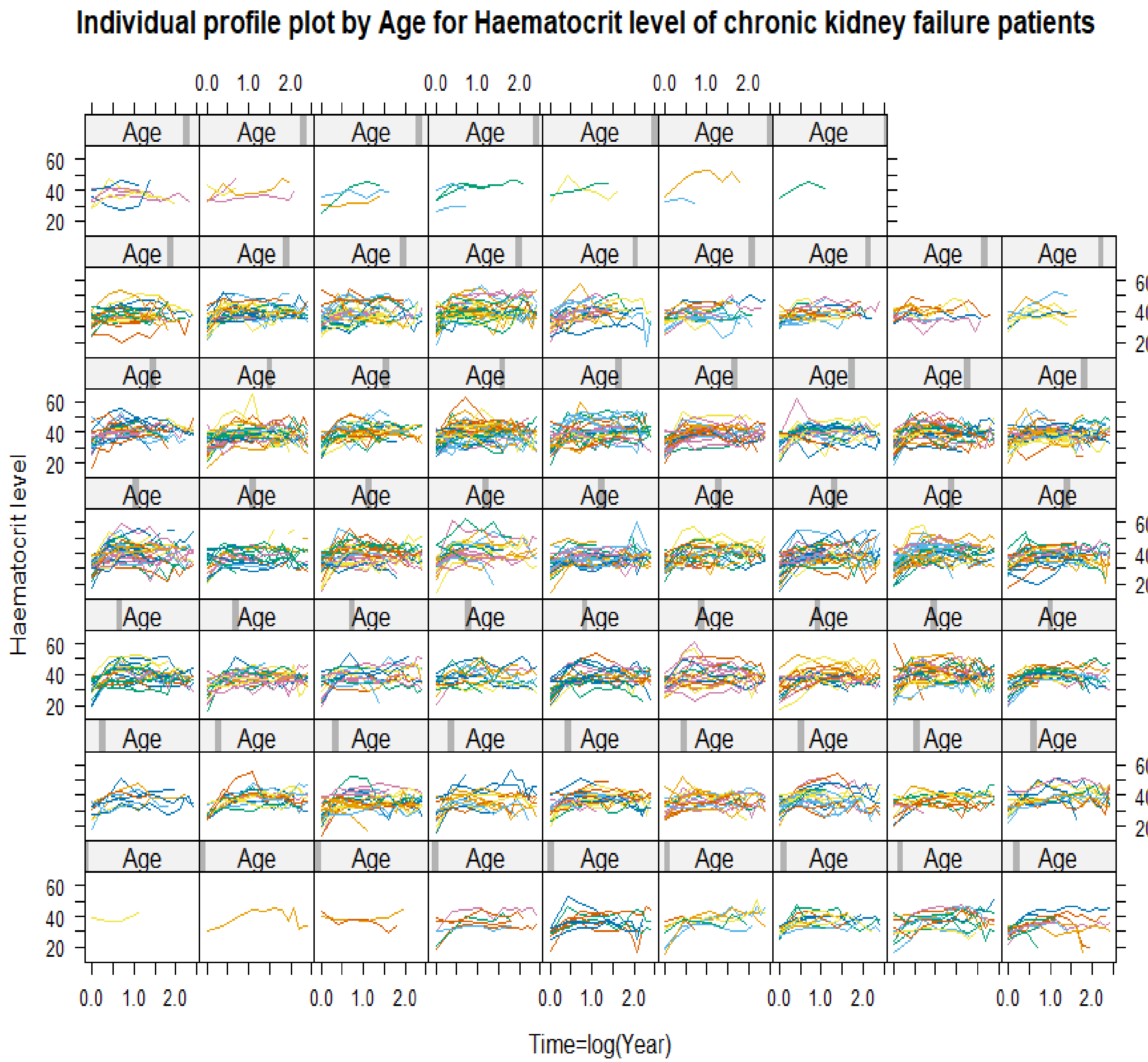

Figure A.5. Individual Profile Plot of Hematocrit Against Age of CKF Patients

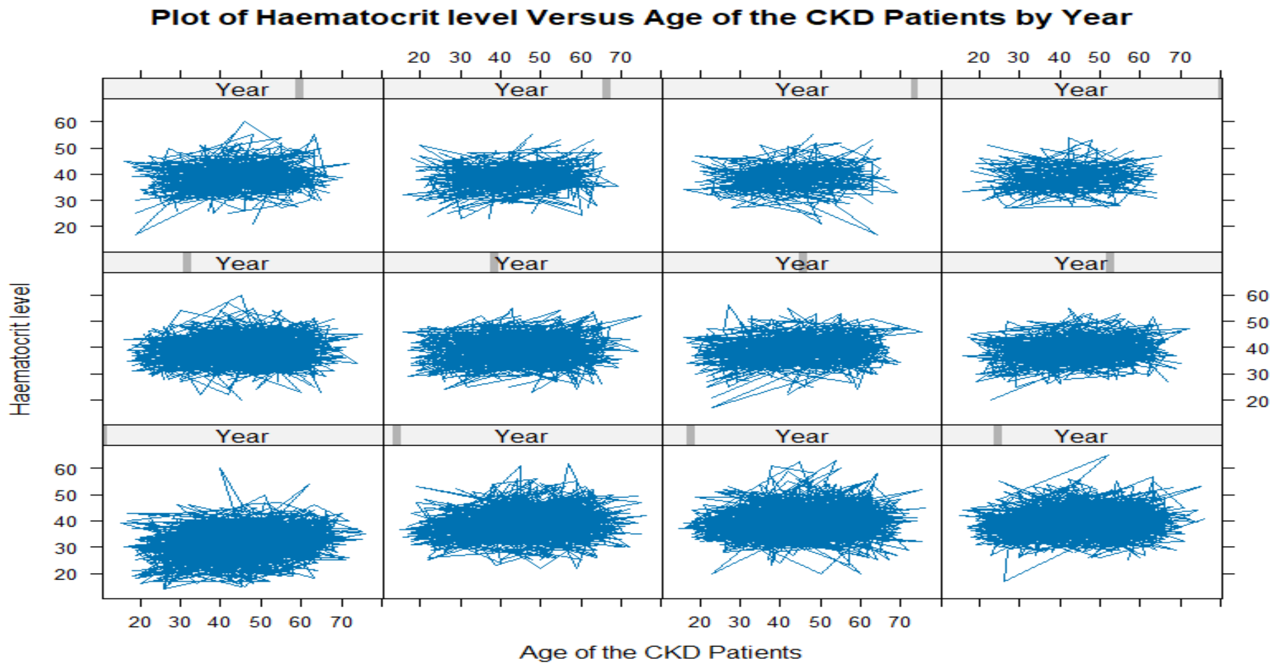

Figure A.6. Plot of Hematocrit Against Age of CKF Patients

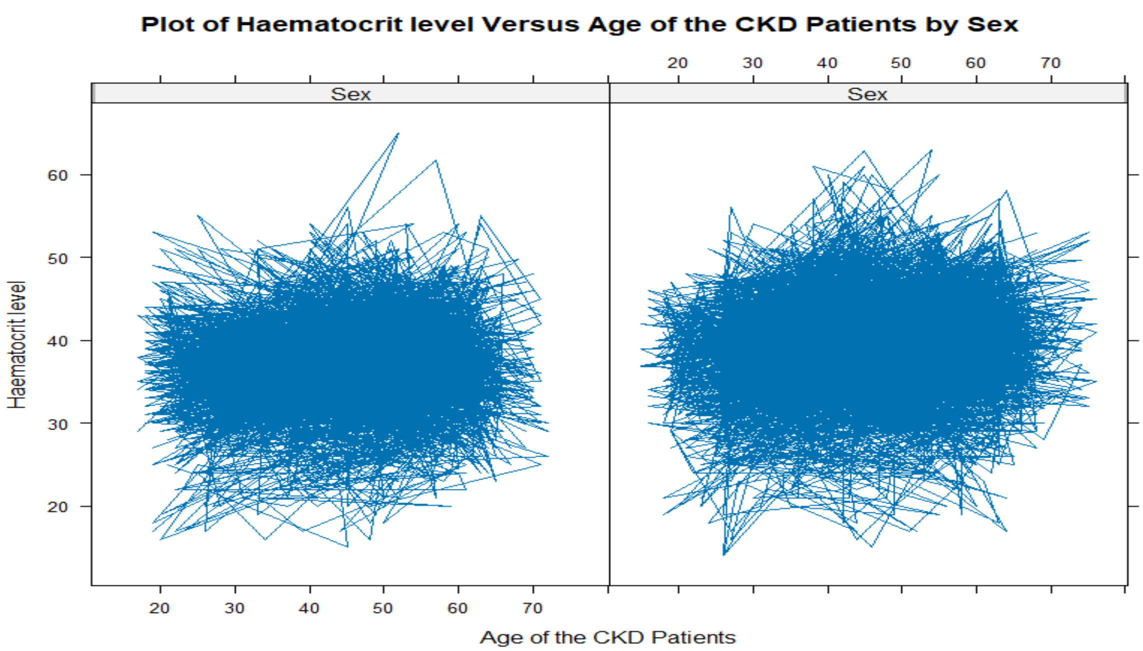

Figure A.7 Plot of Hematocrit Against Age of CKF Patients by Sex

Diagnostic Checking

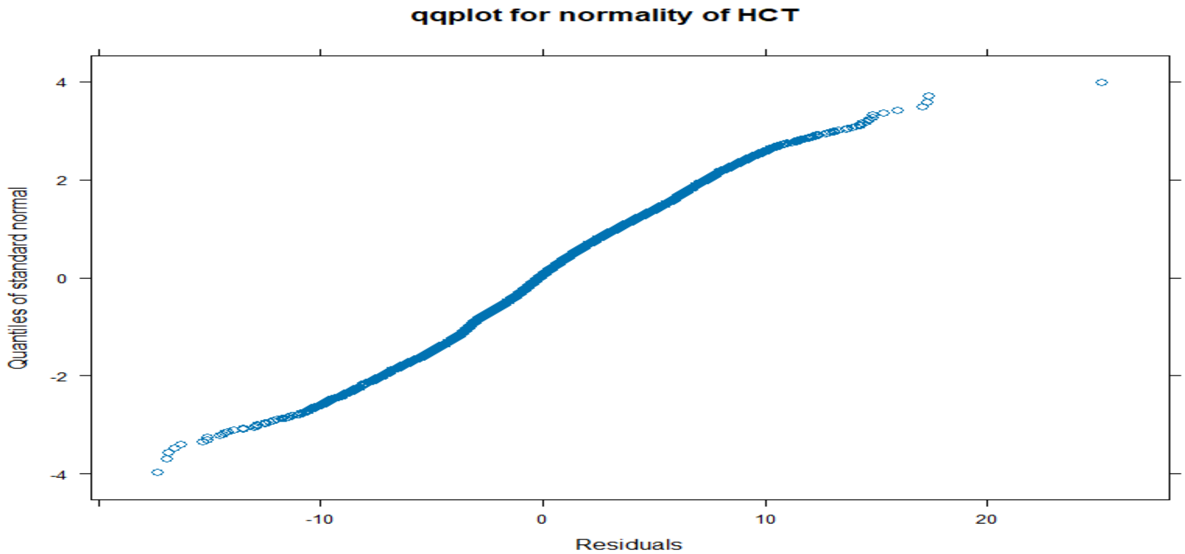

Figure A.8. QQPlot for normality of HCT

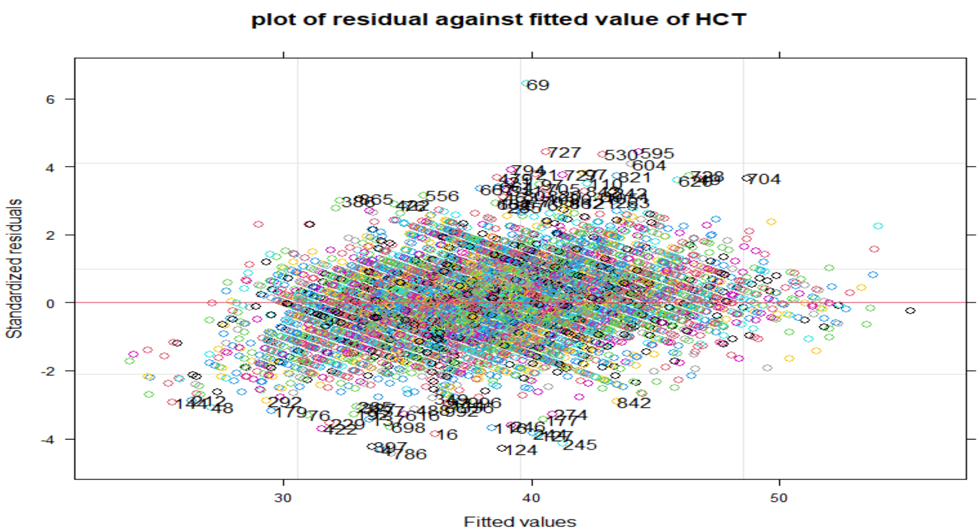

Figure A.9 Plot of standardized residual against value of HCT

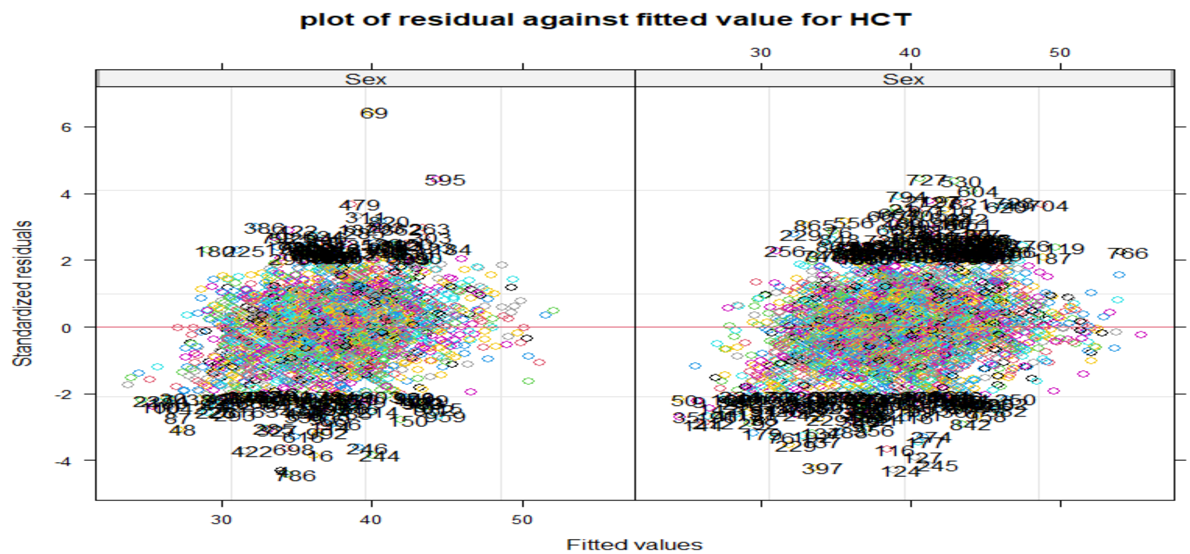

Figure A.10 Plot of residual against value of HCT by Sex

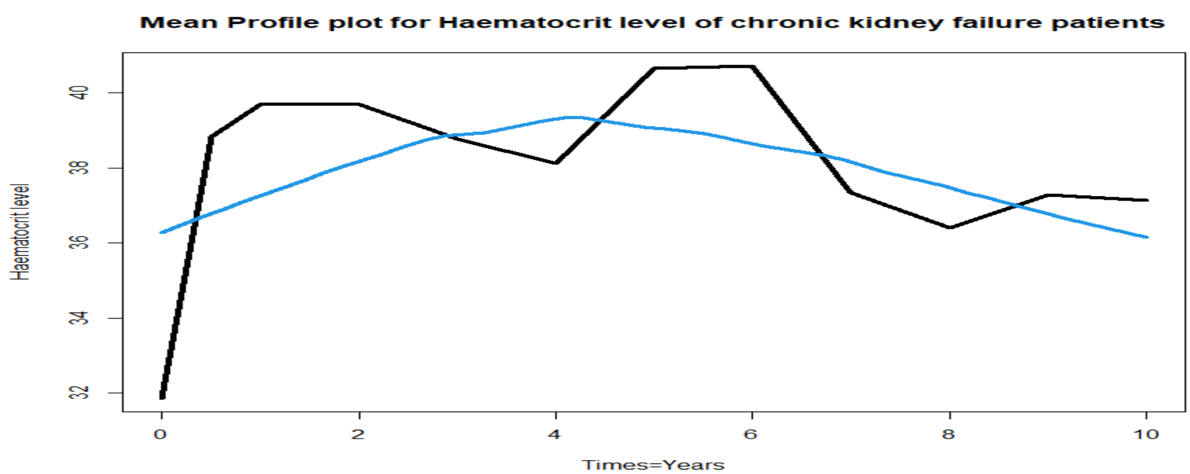

Figure A.11 Mean Profile Plot for Hematocrit level of CKF Patients with Year

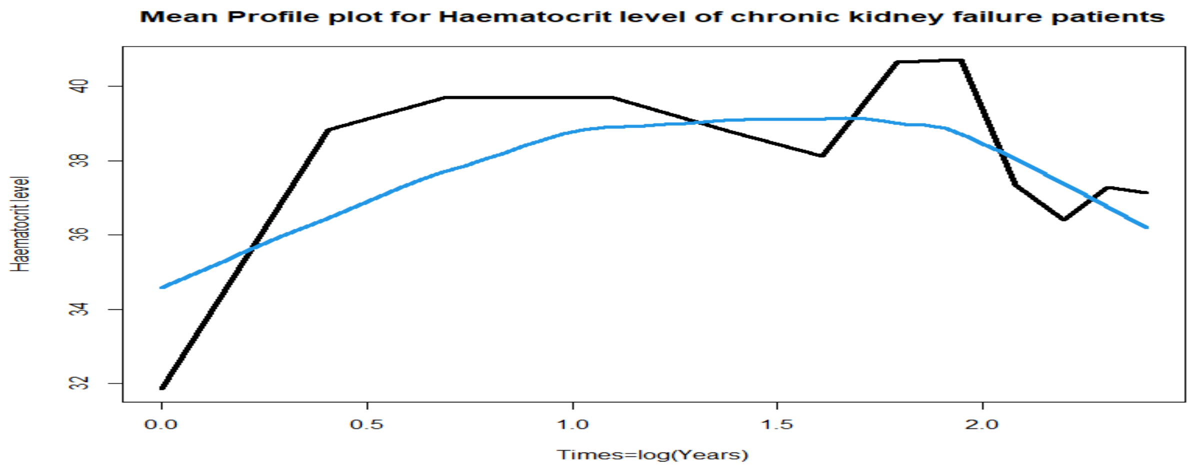

Figure A.12 Mean Profile Plot for Hematocrit level of CKF Patients with log(Year+1)

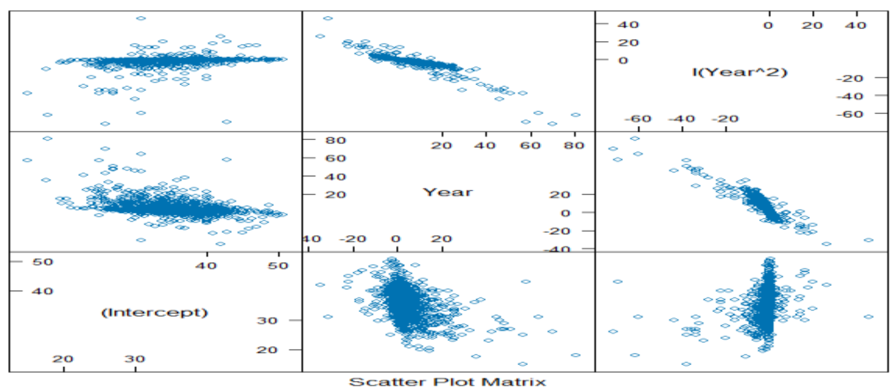

Figure A.13 Scatter Plot Matrix with Quadratic Random Effect of Year

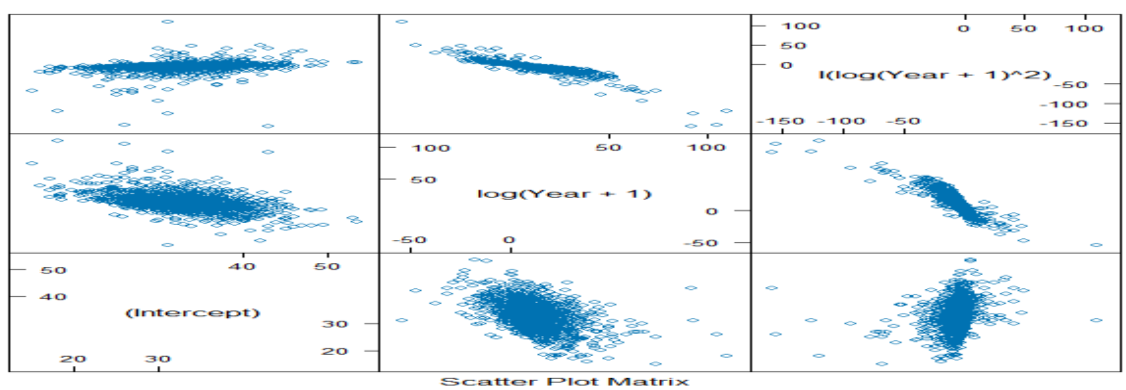

Figure A.14 Scatter Plot Matrix with Quadratic Random effect of log (Year+1)