APPLICATION OF MULTISTATE MARKOV MODEL ON THE PROGRESSION OF CHRONIC KIDNEY DISEASE PATIENTS AT TIKUR ANBESSA SPECIALIZED HOSPITAL, TASH, ADDIS ABABA, ETHIOPIA.

MSc. THESIS

BY:

ZELALEM TOLOSA

HAWASSA, ETHIOPIA

October, 2023

APPLICATION OF MULTISTATE MARKOV MODEL ON THE PROGRESSION OF CHRONIC KIDNEY DISEASE PATIENTS AT TIKUR ANBESSA SPECIALIZED HOSPITAL, TASH, ADDIS ABABA, ETHIOPIA.

MSc. THESIS

BY: ZELALEM TOLOSA

ADVISOR: SELAMAWIT S. (PhD)

CLINICAL ADVISOR: Dr. BEKA ABERRA

A THESIS SUBMITTED TO COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES DEPARTMENT OF STATISTICS HAWASSA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN STATISTICS (BIOSTATISTICS)

October, 2023

HAWASSA, ETHIOPIA

Approval Sheet I

This is to certify that, the thesis prepared by Zelalem Tolosa entitled "Application of Multistate Markov Model on the Progression of Chronic Kidney Disease patients at Tikur Anbessa Specialized Hospital, TASH, Addis Ababa, Ethiopia." and submitted in partial fulfillment of the requirements for the Degree of Masters of Science in Statistics (Biostatistics), the graduated program of the department of statistics natural and computational science college, Hawassa university carried out under our supervision. Therefore, we recommend that the student has fulfilled the requirements and hereby can submit the thesis to the department.

Advisor 's name	Signature	Date
Clinical Advisor 's name	Signature	Date

Approval Sheet II

We, the undersigned, members of the Board of Examiners of the final open defence by Zelalem Tolosa have read and evaluated his thesis "Application of Multistate Markov Model on the Progression of Chronic Kidney Disease patients at Tikur Anbessa Specialized Hospital, TASH, Addis Ababa, Ethiopia." and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Masters of Science in Biostatics.

Name of Department head	Signature	Date
Name of Advisor	Signature	Date
Name of Clinical Advisor	Signature	Date
Name of External Examiner	Signature	Date
Name of Internal Examiner I	Signature	Date
Name of Internal Examiner II	Signature	Date
School of Graduate committee	Signature	Date

Final approval and acceptance of the thesis is contingent upon the submission of the final copy of the thesis to the School of Graduate Studies (SGS) through the School Graduate Committee (SGC) of the candidate's department.

Stamp of SGS **Date: _____**

Declaration

I hereby declare that this MSc thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis / dissertation have been duly acknowledged.

Name: _____

Signature: _____

Acknowledgement

First and foremost, I would like to acknowledge the one and the true God for blessing me with the ability and guidance to succeed my MSc study over the past years.

My special gratitude goes to my advisor Selamawit Serka (PhD) and my clinical advisor Dr. Beka Aberra for their constructive comments in numerous ways and shared their expertise and research insight, whose guidance, encouragement, patience, and time from the beginning to the final stage enabled me to write this thesis.

I also extend my sincere gratitude to my instructors, staff of the Statistics Department, and friends for their unreserved knowledge sharing and cooperation.

I am highly grateful to all staff members of outpatient's clinic at Tikur Anbessa Specialized Hospital for their unreserved support during data collection. My gratitude also goes to Madda Walabu University who sponsored my study.

Finally, I would like to express my deepest gratitude to my mother, my father, all my sisters and brothers; I know you are always sacrifices in educating and preparing me for my success.

Table of Contents

Contents

List of tables

List of figures

List of Abbreviations

MSM: Multi-State Model

CKD: Chronic Kidney Disease

GFR: Glomerular Filtration Rate

KDOQI: Kidney Disease Outcomes Quality Initiative

NCD: Non-Communicable Disease

WHO: World Health Organization

CVD: Cardio Vascular Disease

BMI: Body Mass Index

TASH: Tikur Anbessa Specialized Hospital

CKD-EPI: Chronic Kidney Disease Epidemiological Collaboration

Abstract

***Background**: Chronic kidney disease (CKD) is a serious issue for public health. According to the WHO report for 2022, 17 million people will die from NCD diabetes (2.0 million including kidney disease deaths caused by diabetes) before the age of 70, with low- and middle-income countries providing for 86% of these premature deaths. 77% of NCD-related deaths occur in low- and middle-income countries including Ethiopia. is one of the low- and middle-income nations in the sub-Saharan region. This study was aimed to estimated the effect of covariates on progression between different stages of CKD among patients under follow up treatment at Tikur Anbessa Specialized Hospital TASH, Addis Ababa, Ethiopia, using Multistate Markov Model.*

***Method**: The study was carried out using a retrospective cohort study design on 267 CKD patients age greater than 18 randomly selected at nephrology clinic of TASH who start follow up in May 2018 up to April 2023 for five years. The five stages of CKD disease defined based on the Kidney Disease Improving Global Outcome (KDIGO) guidelines with make only forward transition among different transient stage continuously considered in the Multi-state Markov Model to estimate the transition conditional probabilities, transition intensity rate, total length of stay in different CKD patient's stage.*

***Result**: From the total number of patients included in the study (267 CKD patients), 153 (57.3%) were males and 114 (42.7%) were females. Patients in stages 1, 2.3A, 3B, and 4 had an estimated probability of 94% (0.94), 93% (0.93), 93% (0.93), 96% (0.96), and 98% (0.98) of staying in the same stage, respectively, after one month. Estimated sojourn times for states 1, 2, 3A,3B and 4 were 16.5, 14.5, 15.5, 30.3 and 53.8 months respectively.*

***Conclusion**: Prognostic factors like being male, having a history of Diabetes, having a history of Hypertension, and having a history of heart disease were the factors that had a higher risk of progressing to severe stages in CKD patients, and Age, Haemoglobin, and Potassium were positively (or harmfully) associated with the progression of eGFR or CKD stages. Whereas Phosphate, Sodium, and Urea were negatively associated with the progression change of eGFR or CKD stages. The transition probability from a given good stage to the next worse stage increases with time, reaches its optimum (peak) at a time, and starts to decline as time goes on. stage 4 CKD had the longest estimated mean duration, followed by stage 3B, while the expected mean duration of stage 2 CKD was the shortest.*

***Keywords**: Multi-State Markov Model, Chronic Kidney Disease (CKD), Transition Probability.*

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Chronic kidney disease (CKD) is caused by a variety of diverse disease processes that, over the duration of months or years, significantly affect the kidney's functioning. A chronic decrease in kidney function and structural failure to the kidneys are requirements for the diagnosis of CKD (Romagnani et al., 2017). Glomerular filtration rate (GFR), which measures the total amount of fluid filtered through all of the functional glomeruli in a unit of time, is the best indicator currently known of kidney function as a whole. The classification and diagnosis of CKD have changed with time, but according to current international guidelines, CKD is defined as having a GFR of less than 60 mL/min per 1.73 m^2 or indicators of kidney damage, or both, for at least three months (Webster et al., 2016).

Chronic kidney disease (CKD) is a long-term, sequentially advanced type of disease that occurs in the kidneys and causes poor functioning of the aforementioned activities. The progression of CKD is typically divided into five phases or stages based on the estimated glomerular filtration rate (eGFR) or KDOQI guidelines (Chen et al., 2019). The five stages of chronic kidney disease are: Stage 1: Normal or high GFR (GFR > 90), Stage 2: Mild CKD (GFR = 60 - 89). Stage 3A Mild-Moderate CKD (GFR = 45 - 59), Stage 3B Moderate-severe CKD (GFR = 30 - 44), Stage 4 Severe CKD (GFR = 15-29), and Stage 5 End Stage of CKD (GFR below 15) in mL/min per 1.73 m^2 units (Chen et al., 2019). Based on their severity, the five stages of CKD are divided into three categories namely normal (improved GFR) GFR of greater than or equal to 60 mL/min per 1.73 m^2, unimproved GFR of 15 to 60 mL/min per 1.73 m^2, and ESRD below 15 mL/min per 1.73 m^2 (To, 2022).

Chronic kidney disease CKD, is a condition that decreases kidney function and can finally leads to renal failure. When the kidneys fail, dialysis or a kidney transplant is needed to support life and people can live for decades with dialysis and/or kidney transplants. Numerous diseases can develop to CKD such as heart attack or stroke, diabetes, high blood pressure, and hypertension, over use of painkiller and drug use and antibiotic allergy Trauma, accidents, hepatitis, HIV/AIDS, congestive heart failure, and sickle cell anemia might typically be present (Josephine, 2018)

Globally, chronic kidney disease (CKD) is a serious issue for public health. Noncommunicable diseases (NCDs) kill 41 million people each year, equivalent to 74% of all deaths worldwide. Almost 800 million people worldwide are affected by chronic kidney disease, which is a progressive condition (Kovesdy, 2022). Chronic kidney disease is more common in older persons, women, African Americans, and those with diabetes and high blood pressure. Low- and middle-income nations, which are least prepared to handle its effects, suffer a particularly heavy burden from chronic kidney disease (George et al., 2017). One of the major killers in the world, chronic kidney disease is one of the few non-communicable diseases that have had an increase in deaths related to them over the last two decades. The large number of people affected and the serious consequences of chronic kidney disease should spur further efforts for better prevention and treatment. The high number of affected individuals and the significant adverse impact of chronic kidney disease should prompt enhanced efforts for better prevention and treatment (Kovesdy, 2022).

Millions of people worldwide suffer from the serious medical condition CKD, which is progressive and irreversible and can result in ESRD. It is a typical illness characterised by a long-term decline of renal function. Age, gender, race/ethnicity, and other comorbidities are some of the many variables that affect the complicated epidemiology of CKD (Wani et al., 2023). To improve care for patients with CKD and lower the risk of complications, numerous management guidelines have been created over time. According to the most recent CKD therapy recommendations, the condition should be detected early and treated to delay or stop its progression. This entails regular monitoring of renal function, blood pressure management, and lifestyle changes such adopting a balanced diet, engaging in regular exercise, and quitting smoking. The guidelines also recommend using medications to manage complications such as anemia, bone disease, and cardiovascular disease(Elendu et al., 2023).

A study conducted on the prevalence of chronic renal disease throughout the continent of Africa show that the prevalence of CKD stages 1–5 was estimated to be 15.8% in the general adult population of adults living on the African continent, according to a review of 98 studies including 98,432 participants. Furthermore, they demonstrated that serious to moderate reductions in kidney function occur in 4.6% of individuals living in Africa (i.e., CKD stages 3 to 5). The prevalence of

CKD was higher in sub-Saharan Africa than North Africa, and nearly two times higher in high-risk populations than in general populations. (Kaze et al., 2018).

With a total of 1,014,697 individuals, the systematic review sought to compare the prevalence of chronic kidney disease (CKD) in South Africa with that of sub-Saharan Africa, Africa, and the world from 2013 to 2021. The estimated pooled prevalence of CKD in South Africa was 13.9%, with sub-Saharan Africa having the greatest incidence (19.7%) and Europe and North America having the lowest (8.5%). Additionally, the prevalence rose with age, sex, and concomitant conditions including diabetes and hypertension. The study came to the conclusion that CKD is a substantial public health issue in Sub-Saharan Africa and other areas.(Hariparshad et al., 2023)

According to the WHO report for 2022, 17 million people will die from an NCD before they age 70, with low- and middle-income countries providing for 86% of these premature deaths.77% of NCD-related deaths occur in low- and middle-income countries including Ethiopia. The majority of NCD deaths, or 17.9 million people per year, are caused by cardiovascular diseases, which are followed by cancers (9.3 million), chronic respiratory diseases (4.1 million), and diabetes (2.0 million including kidney disease deaths caused by diabetes (WHO, 2022).

In Ethiopia, CKD disease become abundant and need attention. The severity of the disease affects people of all ages, with males being more likely to develop CKD than females, and it is more prevalent in men than in women. Participants with a history of kidney infection had a greater prevalence of CKD, indicating necessitates urgent attention. According to a cross-sectional study, the prevalence of CKD in Ethiopia is estimated to be 12.2% and has risen in recent years along with the rise in diabetes and hypertension cases. The prevalence of CKD is as high as 41.0% in ages < 35 years and 62% in males (Kore et al., 2018). In Ethiopia also CKD disease become abundant and need attention.

Age, sex, BMI, blood pressure, haemoglobin level, fasting blood glucose level, length of chronic illness, and usage of nephrotoxic medications are important risk factors for CKD in patients with chronic illnesses. CKD is very common in Ethiopian patients who have chronic illnesses, particularly among those with diabetes (28.1%) and hypertension (29.8%). 4,535 patients were included in 12 studies from 10 regions as part of a systematic review and meta-analysis that intended to provide an evidence overview on the prevalence of chronic kidney disease (CKD) and related factors among patients with chronic illness in Ethiopia (Animaw et al., 2022).

The results of the cross-sectional study on the prevalence and awareness of chronic kidney disease among adult patients with diabetes in Northeast Ethiopia indicate that these individuals had high prevalence and little awareness of CKD. The existence of clinically significant CKD in a diabetic outpatient is underestimates by current screening techniques, such as blood creatinine or urine dipstick for albuminuria (Fiseha & Tamir, 2020).

Most of the studies in Ethiopia focused on prevalence and risk factors associated with survival time to kidney failure. CKD progression has different five phases (stages) progression according to KDODI guidelines. This progression is a complex process with different intermediate events. Therefore, the risk factors associated with survival times of different transitions CKD patients can be identified using a multi-state model.

1.2 Statement of problem

Most of the time in medical investigations, survival analysis is a well-developed statistical method, which explores time to single event analysis. Standard survival approaches such as the Kaplan-Meier method or Cox proportional hazards model are sufficient to handle the simple survival settings with no intermediate events (Hussein et al., 2017, Workie et al., 2022).

However, such methods may not sufficiently confine the process of any disease as the progression of disease may occupy intermediate events of interest. Chronic kidney disease progression is an example of a complex process with different intermediate events (Jackson, Sharples, Thompson, Stephen, et al., 2003). Thus, the multi-state model is an efficient way to handle complex processes. The subjects can be at one state at the beginning of the study, further pass-through different states and eventually end up in a final state. These transitions of a subject can be modelled and the risk factors associated with survival times of different transitions can be identified using a multi-state model (Shih et al., 2007).

Multistate Markov model based on computing eGFR has been extensively used to evaluate the progression of CKD patients. (Taha & Mohammad, 2023) used a 5-staged (merge stage 3A and stage 3B) forward transitions Markov model based on Kidney Disease Improving Global Outcome (KDIGO) scale. Another study developed 4 states with first three states reversible Markov model based on based on the Kidney Disease Improving Global Outcome (KDIGO) score (Lintu et al., 2022). Grover et al. employed homogeneous continuous-time hidden Markov model HMM on estimation of misclassification probabilities of chronic kidney stages (GROVER et al., 2019).

According to current update guidance of Chronic Kidney assessment and management (Pearce, 2021) based on computing eGFR of patients we have develop six CKD stages and progression between CKD stages was irreversible and only forward transition.

However, studies have been done on the model progression of CKD patients; the effect of cofactors for people living with CKD on the progression of CKD disease and the conditional transition probability of CKD patients in different stages to predict the future clinical stages are not well studied yet. To the best of our knowledge, no study on MSMM on the progression of CKD has been carried out in Ethiopia.

Therefore, this study aims to model CKD disease progression and the estimate effect of cofactors on the progression between different states of chronicity among CKD patients receiving treatment during their follow-up at Tikur Anbessa Specialized Hospital TASH using a multistate Markov model to create awareness about CKD disease in society.

Consequently, the study was motivated to determine the risks that are correlated with the progression of CKD as well as prevalent in a number of other chronic diseases. Generally, the purpose of this study is to address the following key research questions related to the clinical progression of CKD:

- What is the effect of covariates on the progression of CKD patients?
- What is the likelihood that a CKD patient currently in stage i will be in the next subsequent worse stage j of the disease?
- What is the average length of time that patients stay in the same CKD stages?

1.3 Objective of Study

1.3.1 General Objective the Study

The main objective of this study was to estimate the effect of covariates on progression between different stages of CKD among patients under treatment at Tikur Anbessa Specialised Hospital (TASH) using a multistate Markov model.

1.3.2 Specific Objective

- To estimate the likelihood that CKD patients are in different stages of the disease.
- To estimate the total length of stay in different CKD patients' stages.

1.4 Significance of the Study

The results of this study are useful in many ways.

- To inform relevant information for society, CKD patients, health professionals, and policymakers about the factors that affect the progression of CKD stages in order to provide intervention and set appropriate plans to reduce the progression of CKD stages.
- The study also serves as a reference for future research on CKD in Ethiopia.

1.5. Scope of the study

This study was focused on "Application of the Multistate Markov Model on the Progression of CKD among Patients Under Treatment at Tikur Anbessa Specialised Hospital TASH". The study targets all CKD patients aged greater than 18 who are under treatment at Tikur Anbessa Specialised Hospital (TASH) from May 2018 to April 2023 G.C.

1.6 Operational Definitions

Chronic: means ongoing (persistent or long-term). It does not mean 'severe' as some people think. You can have a mild chronic disease. Many people have mild CKD (Kovesdy, 2022).

Absorbing: An absorbing state is a stage from which there is improvement condition or final stage (Lintu et al., 2022).

Chronic kidney failure: is defined as either pathological abnormalities or markers of damage, including abnormalities in blood or urine tests or eGFR less than 60 mL/min per 1.73 m^2 for greater than three months (Jha et al., 2023).

Estimated glomerular filtration rate (eGFR): is measures how much blood these filters clean every minute based on your body size and measures patients' level of kidney function and stages. If patients eGFR number is low, patients' kidneys may not be working as well as they should. People with a lower eGFR are at increased risk of having chronic kidney disease (CKD) progress to kidney failure (National Kidney, 2021).

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview of the Progression of Chronic Kidney Disease

Chronic Kidney Disease is defined as abnormalities of kidney structure or function, present for greater than 3 months, with implications for health. CKD is a general term for heterogeneous disorders affecting kidney structure and function with variable clinical presentation, in part related to cause, severity and the rate of progression. The concept of CKD evolved after the recognition of the contribution of disordered kidney structure and function on the health of individuals across a wide range of severity (Milik & Hrynkiewicz, 2014).

The incidence and prevalence of end-stage renal disease vary widely, according to CKD epidemiology. More than 80% of patients receiving treatment for ESKD reside in countries with a large elderly population with access to affordable health care. In high income nations like the USA and Australia, the prevalence of CKD is regularly stated to be over 11%. The risk of developing progressive CKD is 60% higher for those in the lowest socioeconomic quartile than for those in the highest. The development of CKD and the progression of the disease are more likely in Black and Asian people in the UK, Hispanic people in the USA, and Indigenous people in Australia, New Zealand, and Canada (Webster et al., 2017).

The global CKD progression is heterogeneous, as shown by epidemiological research from low- to middle-income nations in sub-Saharan Africa, Asia, and Latin America (Orantes et al., 2014). In sub-Saharan Africa, diabetes, hypertension and HIV appear to account for only a portion of the significant CKD burden, especially in urban settings, where many risk factors remain undetermined. even the estimated prevalence of CKD in Democratic Republic Congo is highest 12.4 % (Sumaili et al., 2009), in Tanzania 7% (Stanifer et al., 2015), Senegal 6.1% (Seck et al., 2014) and Ghana 4.7% (Eastwood et al., 2010).

According to (Animaw et al., 2022) a systematic analysis of studies on the prevalence of chronic renal disease and its related variables among patients, Ethiopia has a pooled prevalence of 21.71% for CKD. Oromia, with a prevalence of 32.55%, has the greatest prevalence of chronic kidney disease among individuals with other chronic conditions. The stage of CKD, the pooled prevalence

of individuals with stage one is 26.16%, Stage 2 is 30.76%, Stage 3 is 13.88%, Stage 4 is 1.19%, and Stage 5 is reported to have the lowest pooled prevalence (0.4%).

2.2 Review on prognostic Factors Affecting CKD progression

GROVER et al. applied the continuous time homogeneous multistate model based on Markov processes to progression of chronic kidney disease CKD. They referred a retrospective study of 117 patients suffering from CKD during the period March 2006 to October 2016. Progression of CKD into different stages is irreversible hence probability an individual in state k at a time s will be in state l at the time t. The results of the study showed that Mean sojourn time for stage 1 is highest 10.2 years when compared with other stage 2,3,4 (8.38, 8.70,1.17) years respectively. This proves the slow movement of progression of CKD in earlier stages. The mean sojourn time for stage 4 (1.1733) years which is quite low in comparison to other stages. the prognostic factors like age, hypertension, diabetes, haemoglobin, urea, serum creatinine are significant factors for the progression of CKD into different stages. The progression of chronic kidney disease is very slow at the initial stage of the disease but is quite fast in severe stages. The probability of transition from stage 3 to stage 4, stage 3 to stage 5 and from stage 4 to stage 5 is very high. Finally, they concluded that the mean sojourn times along with p-next probabilities provide more intuitive parametric information of continuous time multistate model based on Markov processes than crude transition intensities (GROVER et al., 2019).

Lintu et al. study conducting on a multi-state model for kidney disease progression by using data obtained from 225 patients to see their kidney disease progression under under colistin therapy follow up from January 2016 to December 2017 in Kasturba Hospital. The results of the study showed that the median length of hospital stay was 21 days and the median survival time was 38 days. A total of 83 (36.89%) patients died in the hospital. The prognostic factors such as gender, hypertension, sepsis, and surgery are significant factors affecting the progression of kidney disease in different stages (Lintu et al., 2022).

The study conducted south-eastern Indian state from Shar Hospital aimed to develop a multistate model to analyses the progression of chronic kidney disease (CKD) and the effect of risk factors on it (Taha & Mohammad, 2023). They used retrospective study of 153 CKD patients seen between February 2015 and May 2022. The progression of CKD modelled and analysed using multi-state models based on Markov processes and found that age, gender, diabetes, hypertension,

and proteinuria were significant risk factors for CKD progression. They also found that the risk of CKD progression increased with increasing proteinuria and age. Moreover, the proposed model can provide valuable insights into the progression of CKD, help clinicians identify patients at high risk of developing end-stage renal disease, and improve patient care (Taha & Mohammad, 2023).

The study conducted on progression of CKD to estimates the transition rates and mean sojourn times for 117 patients suffering from CKD during the period March 2006 to October 2016 by using Hidden Markov model (HMM) with continuous time. Result indicates that the estimated transition intensity corresponding to transition from stage 1 to stage 2 is 0.0405 and reverse transition intensities are zero. In addition to that the estimated mean sojourn time corresponding to stage 1, stage 2, stage 3 and stage 4 are 15.923 years, 11.976 years, 7.936 years and 2.890 years respectively. The probability of a CKD patient with stage 1 of disease will be misclassified as a patient of stage 2 is 0.211 (Grover, Sabharwal, et al., 2018).

2.3 Equations used in calculating GFR

Globally, creatinine became known for the first time in 1847, and it was suggested as a filtration marker in 1926. A few equations that are currently used worldwide and are considered to be rather good estimators of GFR have recently experienced modifications. Creatinine has been used in numerous estimation equations. variance in serum creatinine tests and variance in creatinine synthesis based on muscle mass and diet have been the main constraints. The equations are inappropriate despite the standardisation of serum creatinine assays and the inclusion of age, sex, race, and body size as substitutes for creatinine production. The Cockcroft and Gault equation, which had been suggested by the FDA guidance to industry in 1998, and the MDRD study and CKD-EPI equations, which are recommended by more recent healthcare guidelines, are just some of the equations that have been properly established and implemented in application (Levey & Inker, 2017).

In clinical practise, the glomerular filtration rate (GFR) is considered as the most important overall indicator of kidney function. The strongest overall indicator of kidney function in clinical settings is the glomerular filtration rate (GFR) (Inker & Titan, 2021). The 2009 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine equation is advised by current recommendations for the initial assessment of renal function. The serum creatinine, age, sex, and race factors are included in this equation. Inker et al. validated a new CKD-EPI creatinine equation

in 2021 that used the same regression function as the existing equations but did not include race as an explanatory variable (eGFRcr [2021, ASN]). A new eGFR CKD-EPI creatinine equation that has been modified without the African/Black race coefficient was suggested in a report provided by the National Kidney Foundation (NKF) and the American Society of Nephrology (ASN) Task Force on September 24, 2021 (Inker et al., 2021).

The 2021 CKD-EPI creatinine equation, which takes into account factors for age and sex but not race group, was suggested for immediate use by clinical laboratories by the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Diseases, which reached the conclusion that race should not be included in GFR estimating equations. Additionally, they advised measuring cystatin C more frequently to confirm the estimated GFR obtained from creatinine when clinically necessary because cystatin C and creatinine work together to produce a more precise eGFR. The Task Force suggested using the creatinine-cystatin C equation from the 2021 CKD-EPI because it does not include a term for race group (National Kidney, 2021).

2.4 Review of Markov and Multistate Model in Progression of Chronic Kidney Disease.

A multi-state model (MSM) is a model for a continuous time stochastic process allowing individuals to move among a finite number of states. In biomedical applications, the states might be based on biological markers (e.g., eGFR in scaling stage of CKD using Kidney Disease Outcomes Quality Initiative (KDOQI). A change of state is called a transition, or an event. States can be transient or absorbing, if no transitions can emerge from the state (for example, death) (Milik & Hrynkiewicz, 2014).

Shih et al., define that most of the time the status of the disease in epidemiology expressed as a dichotomous state i.e., disease and non disease. In reality, most chronic diseases have a multi-state characteristic that allows them to progress dynamically from the early stage to the late-stage while being influenced by a variety of internal and external risk factors. The progression of chronic diseases is increasingly being modelled using multi-state models. These models are helpful for studying the development and natural history of the related disease (Shih et al., 2007).

Jackson, Sharples, Thompson, & Duffy, (2003) describes that Chronic disease multistate models based on Markov processes have a natural interpretation in terms of staged progression and an attempted way of estimating rates of transition between stages of disease. A number of people are observed during the stage Si(t) at arbitrary times t, which may differ between people. A

homogeneous continuous time Markov process could be used to simulate the stages of disease. Sometimes the disease under study is thought to be irreversible, in which case the transition intensities corresponding to recovery are taken to be 0 (Jackson, Sharples, Thompson, & Duffy, 2003). In addition to that Grover, Sabharwal, et al., discuss that hidden model is the unobserved true states follow a Markov process with transition matrix Q, and that the observed states are generated from the latent states through a misclassification probability matrix (Grover, Sabharwal, et al., 2018).

Time-to-event outcomes, progression-free survival, or full response are frequent clinical endpoints in oncology studies analysed using standard methodology for survival analysis. These techniques are suitable and adequate to analyses the effect of treatment on a single final outcome regardless of intermediate occurrences (as in intent-to-treat analyses). But nevertheless, multi-state models provide a powerful and adaptable tool to examine treatment impacts on different intermediate events that patients may experience in these complicated conditions. Using a multi-state model, one can examine how a treatment affects each state transition. To present a more thorough picture of the therapy effect and to better comprehend the illness process, multi-state models can be applied in addition to conventional survival analysis methods. Despite the fact that a multi-state model can become complex with many states and transitions, the number of patients at risk for these transitions can become too small to draw any useful conclusions. It's important to balance complexity with usefulness, though, as it is with any statistical models (Le-Rademacher et al., 2018).

CHAPTER THREE

METHODOLOGY

3.1. Study Area

The study was conducted at nephrology clinic of Tikur Anbessa Specialized Hospital (TASH). TASH is the largest general public hospital, which is located at Addis Ababa, Ethiopia, and the largest teaching hospital under the administration of Addis Ababa University in Ethiopia. The hospital was established in 1972. The Tikur Anbessa Specialized Hospital is now the main teaching hospital for both clinical and preclinical training of most disciplines. It is also an institution where specialized clinical services that are not available in other public or private institutions are rendered to the whole nation.

The TASH has 200 doctors, 379 nurses and 115 other health professionals dedicated to providing health care services. The various departments, faculties and residents under specialty training in the School of Medicine provide patient care in the hospital. The hospital also has 950 permanent and contract administrative staff to support the hospital activities. In addition, almost all regional and federal hospitals in Addis Ababa are affiliated to the School of Medicine as clinical services and training sites (http://www.aau. edu.et/chs/tikur-anbessa-specialized-hospital/background-of-tikuranbessahospital/2018).

The hospital has more than 800 beds providing diagnostic and treatment service for about 370,000 to 400,000 and about >600 CKD patients per year. The renal clinic has nephrologists, nurses, and pharmacists. It provides treatment to different types of renal disease and its complications.

3.2 Study design

Retrospective cohort study design was used to carried out on adults CKD patients, who were under the subsequent treatment follow up at Renal clinic, Tikur Ambessa Specialized hospital TASH from May 2018 to April 2023 G.C.

3.3 Study population, Study Period and Source of Data

The study population for this study was all CKD patients who were receiving follow-up treatment at nephrology clinic of TASH from May 2018 G.C to April 2023 G.C. Here there were more than 5,500 CKD patients during the study period. Patients are check their eGFR at least every year to get kidneys function and stages. Healthcare team do a simple blood test creatinine level in urine

and blood at the same time to find out patients eGFR. An eGFR test, this rate is not measured directly mathematically derived entity based on a patient's serum creatinine level, age and sex using CKD-EPI equations formula. The results of estimated GFR used for know stage of kidney disease patients and function based on guidelines of KDOQI CKD classification. to check eGFR level of CKD patients repeat test at least in 3 months. When patients have previously been diagnosed with kidney disease or had abnormal eGFR tests, healthcare team doing one or more repeat tests may aid in diagnosing and monitoring patients' condition during follow up. The source of data in this study was secondary data obtained by reviewing of the patient's chart under the follow-up period.

3.4 Sampling Techniques

We consider all patients age greater than 18 under follow up as target population. One of the most fundamental probability selection techniques is simple random sampling (SRS), which involves choosing a predefined number of units at random from a population list, giving each unit an equal chance of being chosen for the sample. The documented lists of patients who started receiving follow-up treatment at TASH and who have a unique identifying number are taken into consideration as the sampling frame in this study. Therefore, simple random sampling procedure for selection participant included in our study was applied.

3.5. Inclusion and Exclusion Criteria

Inclusion criteria: All CKD patients registered with full information including study variables of interest in the registration card will be considered to be eligible for the study. Patients who started their follow up and whose serum creatinine measured at least once in three months during their follow up TASH were included in the study.

Exclusion Criteria: The patients who had incomplete information regarding to the study variables on the registration card were not eligible for the study. Also, the CKD patients lost from the study without starting any treatment was not included in this study.

3.6 Sample Size Determination

Sample size determination is the process of choosing the right number of observations or participants from a larger group to use in a sample. The appropriate sample size determination formula used in multistate data were (Cochran, 1977) sample size determination formula based on

previously study (Mengesha & Ferede, 2019). Based on the above related literature sample size determination formula for this research as follows:

$$n_0 = \frac{Z^2{}_{\alpha/2}\, p(1-p)}{d^2} = \frac{Z^2{}_{\alpha/2}\, pq}{d^2}$$

$$if\ \frac{n_0}{N} < 5\%\ then\ n_0 = n\ unless\ we\ calculate\ n\ using\ the\ formula$$

$$n = \frac{\frac{Z^2{}_{\alpha/2}\, pq}{d^2}}{1 + \frac{1}{N}\left[\frac{Z^2{}_{\alpha/2}\, pq}{d^2} - 1\right]} = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

$$n_0 = \frac{(1.96)^2(0.5 * 0.5)}{(0.06)^2} \approx 267$$

$$\frac{n_0}{N} = \frac{267}{5{,}500} = 0.048\ldots$$

$$then\ \frac{n_0}{N}\ is\ greater\ than\ 5\%.\ so\ n = n_0$$

Therefore, total sample for this study was **267**.

Were,

- $\boldsymbol{Z^2{}_{\alpha/2}}$ is standard normal distribution with α =0.05 significance level, which is $Z^2{}_{\alpha/2}$=1.96.
- **d** is the degree of precision that is mostly selected by the investigators.
- **p** represents proportion of chronic kidney failure among patients. When the proportion p is not known previously, it is common to use p = 0,5 to maximize sample size (optimum sample size).
- n is desired sample size for population >5,500;

3.7 Study Variables

The variables we consider in this study were selected based on related literatures.

3.7.1 Response Variable

The states of disease progression adopted the time lapse between state transitions was determined using the difference (in months) between the eGFR or level of eGFR. Thus, the main response variable is the change in measurement of eGFR stage of patients in an interval of time. According to KDOQI clinical practise guidelines, current chronic kidney disease (CKD) was identified and classified based on the cause of GFR category stage 1, stage 2, stage 3A, stage 3B, stage 4, and stage 5 depending on the degree to which kidney function has deteriorated (To, 2022). These stages are as follows:

Table 3.1: Chronic Kidney Disease Staging.

Stage (level eGFR)		**Description**	**eGFR (ml/min/1.73 m^2)**
1		Kidney damage with normal ↓ in GFR	> 90
2		Kidney damage with mild ↓ in GFR	60-89
3	A	Mild-Moderate ↓ in GFR	45-59
	B	Moderate-severe ↓ in GFR	30-44
4		Severe ↓ in GFR	15-29
5		Kidney failure	< 15 (or dialysis)

The five stages of CKD are stage 1, stage 2, stage 3 (A, B), stage 4 and stage 5. All the stages 1 to 4 are transient states and stage 5 is an absorbing state. A patient may make forward transition only among different transient states continuously (Chen et al., 2019).

GFR is measured using the 2021 CKD-EPI Creatinine equation (National Kidney, 2021) as: -

$$\mathbf{eGFR_{cr} = 142 * min(Scr/k)^{\alpha} * max(Scr/k)^{-1.200} * 0.99380^{age} * 1.012sex}$$

Were,

- **eGFR** denotes estimated glomerular filtration rates measured by **ml/min/1.73 m^2**
- **SCr** denotes standardized serum creatinine measured by mg / dl,
- **Sex** coded as male=0 and female=1,
- k= 0.7 (females) or 0.9 (males)
- α = -0.241 (female) or -0.302 (male)
- min (S_{cr}/k, 1) is the minimum of S_{cr}/k or 1.0
- max (S_{cr}/k, 1) is the maximum of S_{cr}/k or 1.0

- **Age** (years)

3.7.2 Predictor (Explanatory) Variables

The explanatory variables that may have influence on the prediction of the response variable were listed below in the table.

Table 3.2: Demographic and clinical Factors that Influence the progression on CKD.

Variables	**Coding /category**
Age	Continues
Gender	0= female,1 =male
Heart disease (HD)	0= no, 1= yes
Diabetes (TD)	0= no, 1= yes
Hypertension (HTN)	0= no, 1= yes
Serum creatinine, (mg/dl)	Continues
Urea protein (mg/dl)	Continues
Phosphorus (mg/dl)	Continues
Haemoglobin (Hgb), (gm/dL)	Continues
Sodium, (mEq/L)	Continues
Potassium, (mmol/L)	Continues

Description of clinical variables

Creatinine is a fairly reliable indicator of kidney function. Elevated creatinine level signifies impaired kidney function or kidney disease. Normal levels of creatinine in the blood are approximately 0.6 to 1.2 milligrams per decilitres (mg/dL) in adult males and 0.5 to 1.1 milligrams per decilitres in adult females (Tahir Mahmood Javaid Asad et al., 2014).

The Phosphorus - serum measure the amount of phosphate in the blood. This test is ordered to see how much phosphorus is in your blood. Kidney, liver, and certain bone diseases can cause abnormal phosphorus levels. Normal values range from adults: 2.8 to 4.5 mg/dL(Nadkarni & Uribarri, 2014).

The sodium blood test measures the concentration of sodium in the blood. Sodium can also be measured using a urine test. The normal range for blood sodium levels is 135 to 145 milliequivalents per litre (mEq/L) (Cole et al., 2019).

Potassium is an electrolyte, which is a mineral in the blood that can be measured by a blood test and critical to the function of nerve and muscle cells, including those in your heart. blood potassium level is normally 3.6 to 5.2 millimoles per litre (mmol/L). Having a blood potassium level higher than 6.0 mmol/L can be dangerous (Hyperkalaemia) (Mckay & Parker, 2016).

Urea is produced by the liver and transported by the blood to the kidneys, where it is filtered out and excreted in urine. When the kidneys are not functioning properly, urea accumulates in the blood and causes high blood urea nitrogen BUN levels. Urea should range from 42 to 131 milligrams per decilitre (mg/dl) (C.I.D, 2014).

Haemoglobin level is expressed as the amount of haemoglobin in grams per decilitre (g/dL) of whole blood. The normal ranges for haemoglobin depend on age and gender for adult males: 14 to 18 gm/dL and adult women: 12 to 16 gm/dL (Tahir Mahmood Javaid Asad et al., 2014).

Data Structure and layout for a multi-state event. Data arrangement is an important step in fitting model. Using the above variables, we are properly organizing the data. Individual subjects are easily categorized according to the assigned identification numbers. Therefore, when capturing the data, care should be taken to ensure that all information pertaining to a single patient is included under the same unique ID. Table 3.3 presents Data Structure and format for a Model with Multiple States.

Table 3.3: Data Structure for a Model with Multiple States

ID	Age	Sex	HTN	Diabetes	Heart D.	No. months	Cr.	eGFR	Stages	K	The else variables
382	46	1	0	0	0	1	0.96	98.72	Stage1	5.23	
382	46	1	1	1	0	8	1.00	94.00	Stage1	5.10	
382	47	1	1	1	0	13	1.30	68.19	Stage2	4.92	

382	47	1	1	1	0	14	1.40	62.38	Stage2	4.92	
382	47	1	0	0	0	16	1.20	75.06	Stage2	4.77	
382	47	1	1	1	0	19	1.90	43.24	Stage3B	4.56	
382	49	1	1	1	0	30	1.90	42.71	Stage3B	4.50	
382	50	1	1	1	0	46	2.00	39.91	Stage3B	4.40	
639	66	0	0	0	0	1	0.96	87.17	Stage2	4.00	
639	67	0	0	0	0	13	1.30	60.21	Stage2	4.92	
639	67	0	0	0	0	16	1.20	66.28	Stage2	4.77	

The progression of CKD into different stages is irreversible in nature. A patient may make forward transition only among different transient states continuously. The arrows show the possible transition between stages. The time of movement between the different states and the state occupancy in between the observation times are not known.

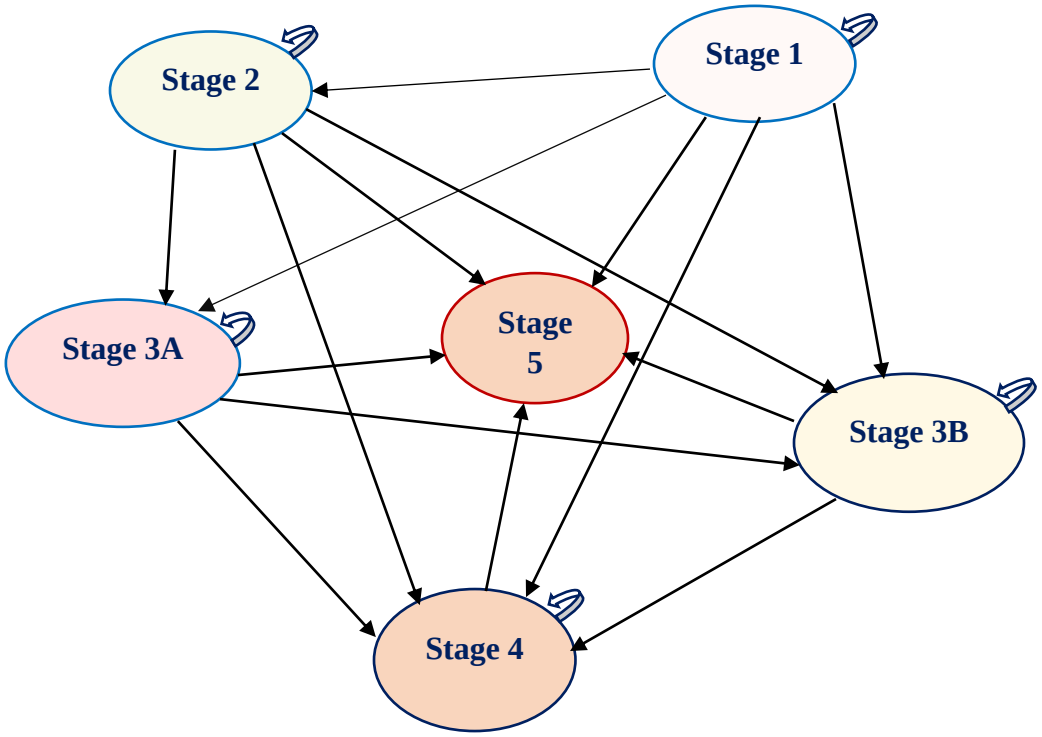

Figure 3.1: State Transition Diagram

3.9 Data collection procedure

The data collection procedures were based on patient's chart and electronic database system (i-care) under the follow-up period. The data collection process was started by reviewing their

medical registration number (ID) of patients from database for selection of patient charts and collected necessary information socio demography and clinical variables. Data were collected by a dedicated nurse who works at renal clinic of TASH under supervision of principal investigators. The longitudinal data would be extracted from the patient's chart (i-care) which contains socio-demographic and clinical information of all CKD patients under the follow-up. Estimated GFR of patients was recorded longitudinally, each patient eGFR was observed during the study period. stages of CKD patients classified by 2021 CKD-EPI Creatinine equation from eGFR of CKD patients. The longitudinal Successive eGFR measurements of CKD patients are recorded with at least 3 months interval.

3.10 Ethical Considerations

Ethical clearance and approval to conduct the research were obtained from Research and Ethical Committee, College of Natural and Computational Science, Hawassa University, and the Tikur Ambessa Specialized Hospital; internal medicine department granted permission for data collection. Privacy of the patients and cultural norms were respected properly.

3.10 Method of Data Analysis

The baseline characteristics of the study participants will report using descriptive statistics. A multistate model will apply to assess the relationship between the clinical conditions of patients and to determine the effects of covariates on individual states. The data were first entered into excel, SPSS and exported to R statistical software and Mendeley Reference Manager used for insert references. The mstate package in R software, which was developed by (Jackson, 2011), was used to carry out all analyses and generate graphs. The frequency distribution table was used to summarize the data based on the study variables.

3.11 Modelling the Clinical Progression of CKD Disease using Multi-State Model

multi-state model (MSM) is modelling time for event data where all the individuals start in one or more states, and eventually may end up in one or several absorbing state(s). It has also been defined as a process in which an individual move through a series of states in continuous time.

The progression of CKD into different five states can be shown by Markov process permitting only the forward transitions from one state to another state over time. A process is Markovian if the future is solely dependent on the present. diagram 1 depicts the model. The progression of the CKD disease is continuous in time and the time of transitions are random in nature but the state

spaces are discontinuous. Due to the fact that disease progression is a continuous function of time and that the transition probability between states is time dependent and thus independent of the time at which the transition actually occurs, we have considered a Homogeneous continuous time multistate model based on Markov processes is the suitable model to describe the course of progression of CKD.

3.12 **Homogeneous continuous time Multistate Markov Model**

A homogeneous continuous-time multistate Markov model is an appropriate model for describing the progression of CKD. In this model, the observed states are precisely the same as true states of disease. States of Markov process is calculated on the basis of eGFR values.

Markov process is defined as a multi-state model where the multi-state model is defined as a model for a stochastic process ($X(t)$, $t\epsilon T$) with a finite space S in which the future knowledge of the process is only provided by the current state of the process.

Let X = {X_1, X_2, ..., X_N} be a random process of random variables taking on values in a discrete state space E = {1, 2, ..., e} and X_t be the state of the process of an individual at time t. Now, let the realisation of the entire history of the process up to and including time t be

$$\{X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_1 = x_1\}$$

where $x_t, x_{t-1}, \dots, x_1$ is a sequence of states at different time points. A random process is classified as a Markov Chain if it satisfies the following condition:

$$P(X_{t+1} = x_{t+1} | X_t = x_t\ , X_{t-1} = x_{t-1}, \dots, X_1 = x_1) = P(X_{t+1} = x_{t+1} | X_t = x_t)$$

for every sequence x_1, x_t, x_{t+1} of the elements in E and every time point t ≥ 1.

In the homogeneous continuous time Markov context, we define the transition probability, $P_{kl}(t)$, to be the conditional probability of entering state k at time t, given the system was in state k at time 0. This definition of $P_{kl}(t)$ implies that after leaving state k the process could have migrated to other states, say l_1, . . ., l_{i-1} before finally entering state l at time t.

For a continuous time, stochastic process {X(t), t ≥ 0} whose state space is S, we say it has the Markov property if

$$P(X(t) = j | X(s) = i, X(t_{n-1}) = i_{n-1}, \dots, X(t_1 = i_1) = P(X(t) = j | X(s) = i$$

where$0 < t_1 \leq \cdots \leq t_{n-1} \leq s \leq t$ is any non decreasing sequence of $n + 1$ state occupation times and $i_1, \ldots, i_{n-1}, i, j \in S$. In other words, states that the state of the process at time t depends only on the most recent state occupied prior to time t.

3.12.1 Transition intensity matrix

The intensity between two states k and l, can be defined as the rate of change of the probability p_{kl} in a small-time interval δt. Let S(t)=k be the function of current state of a patient at time t then the transition intensity with which the patient moves to the state l during the time interval (t+δt) is given by

$$\lambda_{kl}(t) = \lim_{\delta t \to 0} \frac{P(X_{t+\delta t} = l | X_t = k)}{\delta t}$$

where $\lambda_{kl}$is the instantaneous risk of moving from state k to state l. From the schematic diagram 1 the transition intensity matrix for the progression of CKD is given by: All possible intensities between possible states are collected in the transition intensity matrix denoted by $Q(\lambda)$ and given by

$$Q(\lambda) = \begin{bmatrix} -(\lambda_{12} + \lambda_{13A} + \lambda_{13B} + \lambda_{14} + \lambda_{15}) & \lambda_{12} & \lambda_{13A} & \lambda_{13B} & \lambda_{14} & \lambda_{15} \\ 0 & -(\lambda_{23A} + \lambda_{23B} + \lambda_{24} + \lambda_{25}) & \lambda_{23A} & \lambda_{23B} & \lambda_{24} & \lambda_{25} \\ 0 & 0 & -(\lambda_{3A3B} + \lambda_{3A4} + \lambda_{3A5}) & \lambda_{3A3B} & \lambda_{3A4} & \lambda_{3A5} \\ 0 & 0 & 0 & -\lambda_{3B4} + \lambda_{3B5} & \lambda_{3B4} & \lambda_{3B5} \\ 0 & 0 & 0 & 0 & -\lambda_{44} & \lambda_{45} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Q(λ) is 6x6 matrix denote transition intensity matrix with λ_{kl} elements of a multi-state process. λ_{kl} denotes the transition intensity or instantaneous rate from k state to l where k =1,2,3A,3B,4,5 and l = 1,2,3A,3B,4,5. The transition intensity matrix again is also used to calculate the transition probability matrix.

Property of transition intensity matrix are;

i. The sum of elements of each row of the transition matrix $Q(\lambda)$is zero or if no way for k to change to l, the entry (kl) is zero.
ii. The diagonal element of $Q(\lambda)$ become:

$$\lambda_{kk} = \lambda_{k.} = -\sum_{k \neq l} \lambda_{kl}(t)$$

Therefore, the rates on the diagonal represent states that subjects remain stationary and the off-diagonal values contain rates in which the subject moves to other states

3.12.2 The Transition Probability

A multistate process is a stochastic process (X(t), t ∈ T) with finite state space S = {1, 2, 3 . . . N} where T = [0, t] is the period of observation. Consider,

$$P_{kl}(s,t) = P\big(X_{(t)} = k \big| X_{(s)} = l\big)\ for\ k,l\ \in S\ and\ s,t \in T$$

where $P_{kl}(s,t)$ can be interpreted as the probability of entering state k at time t, conditional on being in state l at time s.

Let $P(s,t)$ be the transition probability matrix with (k,l) th entry has $p(s,t)$ and also let Q be the transition intensity matrix with (k,l) th entry has λ_{kl}. The transition probability matrix can be calculated from the transition intensity matrix by using the forward Kolmogorov equations [19], which leads to

$$P(s,t) = \exp\big((t-s)Q\big),\ \ s < t \quad OR$$

let $P(s,t)$ denotes the transition probability matrix with transition probabilities $P_{kl}(s,t), k = 1,2,\dots,5\ and\ l = 1,2,\dots,5$ where $P_{kl}(s,t)$ denotes Pr {an individual in state k at time s will be in state l at time t}. The transition probability matrix $P(s,t)$ is given by:

$$P(s,t) = \begin{bmatrix} p_{11}(s,t) & p_{12}(s,t) & p_{13A}(s,t) & p_{13B}(s,t) & p_{14}(s,t) & p_{15}(s,t) \\ 0 & p_{22}(s,t) & p_{23A}(s,t) & p_{23B}(s,t) & p_{24}(s,t) & p_{25}(s,t) \\ 0 & 0 & p_{3A3A}(s,t) & p_{3A3B}(s,t) & p_{3A4}(s,t) & p_{3A5}(s,t) \\ 0 & 0 & 0 & p_{3B3B}(s,t) & p_{3B4}(s,t) & p_{3B5}(s,t) \\ 0 & 0 & 0 & 0 & p_{44}(s,t) & p_{45}(s,t) \\ 0 & 0 & 0 & 0 & 0 & p_{55}(s,t) \end{bmatrix}$$

Hence the progression of CKD into different five states is irreversible. All the probabilities in the transition probability matrix must be greater than or equal to zero, that is $p_{kl}(s,t) \geq 0, \forall k,l\ \in \{1,2\dots,5\}$ and each row must sum to one $\sum_1^5 p_{kl} = 1, \forall k,l\ \in \{1,2\dots,5\}$.

The above matrix defined as the transition probability matrix with its elements providing the probability of being in state k at time t + 1, conditional on being in state l at time t. The transition probability matrix is time dependent and is therefore denoted as P(t) instead of P. In time homogeneous Markov models, the dependency of t is omitted.

3.12.3 Distribution of Mean Time Length

When the process enters state k, the time it spends there before moving to another state is called the holding time in state k or the sojourn time. The sojourn time of a process X in a subset of states will be an integer-valued random variable if X is a chain or real-valued one in the case of a continuous-time process. The sojourn time of a continuous Markov process that is in state k is an independent and exponentially distributed random variable with mean $\frac{-1}{\lambda_{kk}}$ (Gillings et al., 1974). The remaining elements in the k^{th} row of the transition intensity matrix is proportional to the probabilities that govern the next state after state k to which the individual makes a transition. The probability that the next transition is from state k to state l is $\frac{-\lambda_{kl}}{\lambda_{kk}}$. The new state and the sojourn time are only dependent on state k and not on the history of the process prior to time t. Therefore, the sojourn time and the new state are independent of each other, given that the current state is state k. The mean sojourn time describes the average time period in a single stay in a state (Kotzé, 2019).

3.13 Model assumption

3.13.1 Markov model assumption

The Markov assumption state that the future progress only depends on the current state not on the past states. This means that the transition times from each state are independent of the history of the process prior to entry to that state. The past history of a system plays no role in its future evolution, which is usually known as the "memoryless property of a Markov process (Kalbfleisch & Lawless, 1985).

3.13.2 Time homogeneous Markov model assumption

In time homogeneous Markov models, all transition intensities are assumed to be constant as functions of time, independent of time t (Ali et al., 2014). This assumption can be assessed with a likelihood ratio test. When intensities are treated as being time homogeneous then the dependency on time can be removed. The transition probability matrix and transition intensity matrix form the building block of Kolmogorov equations that are used to yield unique solutions for probability matrix P (t).

Let X(t), to be $t \geq 0$ be a continuous-time multi-state process with a finite state $S = \{1,2,3 \dots, K\}$. Define the transition probability from state k at time s to state l at time t as

$$P_{kl}(s,t) = P(X(s) = k|X(t) = l), s < t$$

The Kolmogorov equations state that

$$\frac{d}{dt}p(t) = p(t)Q(\lambda)$$

which yield a unique for $p(t)$ and condition on $p(0) = I$.

$$P(T) = e^{Q(\lambda)} \text{ only valid with time homogeneous intensities.}$$

$Q(\lambda)$is the transition intensity matrix therefore $p(t)$can be found from $Q(\lambda)$ using Kolmogorov equations.

3.14 The Effect of Covariates on the Transition Rate

The covariate effects on transition intensities were considered using the proportional hazards regression model. It used to describe how explanatory variables (covariate vector Z) affect an individual's transition intensity at time t. Suppose we have a baseline covariate vector $\boldsymbol{Z} = \{Z1, Z2, \dots, Zp\}$. A time-homogenous multi-state Markov model used to estimate the effect of covariates on the transition rate with the following form:

$$\lambda_{kl}(\boldsymbol{Z}|\boldsymbol{\beta}) = \lambda_{kl,0}exp(\boldsymbol{\beta}_{kl}{}^{T}\boldsymbol{Z}) = exp((\boldsymbol{\beta}_{kl,0} + \boldsymbol{\beta}_{kl}{}^{T}\boldsymbol{Z})$$

$\lambda_{kl,0} = exp(\boldsymbol{\beta}_{kl,0})$ is called the baseline intensity for the transition from state k to state l, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_{kl,0}, \boldsymbol{\beta}_{kl}; k = 1,2, \dots, 5;\ l = 1,2, \dots, 5; l \neq 1)$ which represents all the parameters associated with the multi-state model.

3.15 Methods Estimation

3.15.1 Maximum Likelihood Estimation

The method of maximum likelihood estimation enables the unknown parameters in the model to be estimated. The maximum likelihood estimate is the number of transitions from state k to state l divided by number of overall transitions from state k to other states calculated from the transition probability matrix. Maximum likelihood estimates for a particular class of a model can be computed from transition probability matrix P (s, t) with (kl) entry through the Kolmogorov relationship $P(s,t) = \exp\ (tQ(\lambda))$

The likelihood function of transition intensities is the product of probability of transition between observed states over all individuals and observation times. The optimization method “BFGS” from the msm package from R was used for a fast derivative-free method (Jackson, 2007)

3.16 Handling Missing Data

Missing data can frequently occur in a longitudinal data analysis. Most statistical analyses exclude observations with any missing variable values from the analysis. This method is called the complete case (CC) method. While applying the CC method is simple, it results in a loss of information that resides in the incomplete cases as well as a potential for inducing bias(Lou et al., 2017). Missing data are commonly classified into three categories. Those are: Missing completely at random (MCAR), missing at random (MAR) and missing not at random (NMAR). In this study, mean Imputation missing data handling mechanisms were used. mean Imputation (substitute) fill in the values that were not recorded with imputed values. Once a filled-in data set has been constructed, standard methods for complete data can be applied. The Mean substitution method primarily works for missing continuous variables, and it imputes the missing values with the sample mean or its conditional mean calculated from the observed data.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Descriptive analysis

The study used data from Tikur Anbessa Specialized Hospital TASH, with 267 CKD patients on different therapy who were observed 1647 times during follow up. The study was included 153 (57.3%) males and 114 (42.7%) females. The Table 4.1, shows the demographic and clinical characteristics of CKD patients and provides the frequency and percentage of categorical variables, such as sex, hypertension, heart disease, and diabetes, and the mean and standard deviation of continuous variables, such as age, haemoglobin, creatinine, potassium, sodium, phosphate, and urea. The mean age of the patients was 54.74 years, with a standard deviation of 14.64 years. This means that most of the patients were in older age group and majority of patients were male (56.3%). The most common comorbidities among the patients were hypertension (67.5%), heart disease (50.9%), and diabetes (69.0%). The mean values of the laboratory tests for haemoglobin, creatinine, potassium, sodium, phosphate, and urea were 12.35 g/dL, 2.37 mg/dL, 4.57 mmol/L, 138. mEq/L, 4.08 mg/dL, and 67.34 mg/dL, respectively. These values indicate that the patients had varying degrees of anemia, renal impairment, electrolyte imbalance, and uraemia, which are common features of CKD.

Table 4.1: Demographic and Clinical Characteristics.

Categorical Variables	
Variables	**In percent.**
Sex-male (%)	153 (57.3)
HTN-yes (%)	67.5%
Heart Disease-yes (%)	50.9%
TD-yes (%)	69.0%
Continues variable	
Variables	**mean (SD)**
Age in year (SD)	54.74 (14.64)
Hgb (SD)	12.35 (2.39)
Cr (SD)	2.37 (1.45)

Potassium (SD)	4.57 (0.75)
Sodium (SD)	138.09 (3.76)
Phosphate (SD)	4.08 (3.518)
Urea (SD)	67.34 (38.00)

4.2 Continuous-time homogeneous multi-state Markov analysis of progression of CKD

This study considered that progression of CKD patients stages which is based estimated glomerular filtration rate (eGFR). Since the progression of CKD stages is naturally irreversible deterioration in renal function the transitions were not reversible. There was more than one transition of the same type in the same patient. If there is an improvement on serum creatinine or on protein urea, the patient has not recovery from the initial eGFR stages but stay on initial stages else patients can transit to reach a next other severe stage. The transition of the patient in different state occurs at any time.

This Table 4.2 shows the number of patients who progressed from one stage of chronic kidney disease (CKD) to another over a certain period of time. The rows represent the initial stage (state i) of CKD and the columns represent the final stage (state j) of CKD and main diagonal shows the number of individual observations stay at their corresponding states. For example, 35 patients who started with stage 1 CKD progressed to stage 2 CKD, while 4 patients who started with stage 1 CKD progressed to stage 3A CKD and there were 46 observations are observed from severe decreasing eGFR to kidney damages or kidney failure observations from severe states individuals, 100 patients transited to stage 3B from stage 3A and 82 are got kidney damage in hospital. Total number of transition frequency between hypertension states

Table 4.2: Transition frequency between CKD patients' stages

From	**To**					
	Stage1	Stage2	Stage3A	Stage3B	Stage4	Stage5
Stage1	66	35	4	4	3	2
Stage2	0	130	64	30	4	1
Stage3A	0	0	140	100	10	3
Stage3B	0	0	0	308	121	3

Stage4	0	0	0	0	224	46
Stage5	0	0	0	0	0	82

4.3 Transition intensity and probability matrix

Initially, the Markov model without covariate has been used to study the overall disease progression. The estimates of transition intensities (λ_{kl}) are presented in Table 4.3. The estimated intensity transition matrix indicates that the rate of transiting from good states to the worst state is decreasing. The rows represent the current state of the patients, and the columns represent the possible next states. The number of transitions from higher stages to lower stages is zero since CKD is irreversible. The transition intensities moving from the stage 1 to stage 2 was 0.052 or the rate of change of the probability of CKD patients currently entry in stage 1 move to stage 2 is 0.052 per a month. The highest rate (0.063) of moving from state 3A is to state 3B, followed by state 5 (0.001), and state 4 (0.0003).

The values of the main diagonal provide the rate of change of the probability of staying in the same stage of CKD in one month. A patient in stage 1 has a negative rate (-0.061) means that the probability of CKD patients staying in stage 1 decreases by 0.061 per a month and CKD patients have high risk progress to higher stages. The risk rate of stage 5 (end stage) is 0 means that there is no possibility of recovery or improvement and patient needs dialysis or a kidney transplant to survive.

Table 4.3: Transition intensity matrix

	Stage 1	Stage 2	Stage 3A	Stage 3B	Stage 4	Stage 5
Stage 1	-0.061	0.052	0.001	0.002	0.003	0.001
Stage 2	0.00	-0.068	0.057	0.009	0.001	0.0007
Stage 3A	0.00	0.00	-0.064	0.063	0.0003	0.001
Stage 3B	0.00	0.00	0.00	-0.033	0.032	0.0001
Stage 4	0.00	0.00	0.00	0.00	-0.018	0.018
Stage 5	0.00	0.00	0.00	0.00	0.00	0.00

Based on Table 4.3: results, we can see that the transition intensities of CKD are not linear or uniform. Some stages have higher rates of change of the probabilities of advancing to the next

stage than others. Stage 3B has a transition intensity of 0.063 of moving to stage 4, while stage 4 has a transition intensity of 0.018 of moving to stage 5. This means that stage 3B is more likely to progress to stage 4 than stage 4 is to progress to stage 5. In addition to that Table 4.4 shows that most patients with CKD tend to stay in the same stage or progress to the next stage over time. The risk of progression increases as the stage of CKD worsens, and that the risk of reaching end-stage renal disease (ESRD) or stage 5 CKD is highest for patients who were in stage 4 CKD.

Table 4.4: The transition probability matrix P(t) over an interval of t = 1 months.

	Stage 1	Stage 2	Stage 3A	Stage 3B	Stage 4	Stage 5
Stage 1	0.94	0.048	0.002	0.003	0.003	0.0014
Stage 2	0.00	0.93	0.05	0.01	0.001	0.0007
Stage 3A	0.00	0.00	0.93	0.06	0.001	0.001
Stage 3B	0.00	0.00	0.00	0.96	0.032	0.0004
Stage 4	0.00	0.00	0.00	0.00	0.98	0.018
Stage 5	0.00	0.00	0.00	0.00	0.00	1.00

The Table 4.4 shows the probability of moving from one state to another in one month for CKD patients at under treatment follow up. Each stage represents a stage of chronic kidney disease, from stage 1 (good stage) to 5 most severe (kidney damage or failure). A patient in stage 1 has a very high probability 94% (0.94) of staying in the same stage after one month. The result describes the probability of patients developing CKD disease gets increased as the eGFR of patients gets declined or smaller.

The probability of CKD patients currently being in stage 5 have 100% chance of kidney failure stage. This implies Clinically, kidney function is very low and the patients needs dialysis or a kidney transplant to survive and associated with a high risk of mortality and morbidity, as well as a reduced quality of life.

Table 4.5: Transition probability matrix over different time t (months).

Transition	t=12 months	t=24 months	t=36 months	t=48 months	t=60 months
Stage 1 → stage 1	0.483	0.233	0.114	0.055	0.026
Stage 1 → stage 2	0.287	0.265	0.185	0.114	0.064
Stage 1 → stage 3A	0.105	0.187	0.192	0.157	0.113
Stage 1 → stage 3B	0.065	0.176	0.267	0.303	0.297
Stage 1 → stage 4	0.039	0.093	0.164	0.239	0.304
Stage 1 → stage 5	0.019	0.043	0.078	0.128	0.193
Stage 2 → stage 2	0.437	0.192	0.084	0.036	0.016
Stage 2→ stage 3A	0.308	0.277	0.187	0.112	0.062
Stage 2 → stage 3B	0.194	0.347	0.389	0.358	0.296
Stage 2 → stage 4	0.046	0.144	0.251	0.336	0.386
Stage 2 → stage 5	0.012	0.039	0.087	0.156	0.239
Stage 3A→ stage 3A	0.460	0.212	0.097	0.044	0.020
Stage 3A→ stage 3B	0.425	0.482	0.415	0.321	0.234
Stage 3A→ stage 4	0.097	0.245	0.357	0.416	0.430
Stage 3A→ stage 5	0.016	0.059	0.129	0.218	0.314
Stage 3B→ stage 3B	0.673	0.452	0.304	0.204	0.137
Stage 3B→ stage 4	0.290	0.428	0.473	0.466	0.433
Stage 3B→ stage 5	0.037	0.119	0.222	0.328	0.429
Stage 4→ stage 4	0.800	0.638	0.509	0.407	0.327
Stage 4→ stage 5	0.199	0.361	0.490	0.595	0.672
Stage 5→ stage 5	1.00	1.00	1.00	1.00	1.00

As you could easily see from Table 4.5, the probability of a patient staying in the same state, state 1 is decreasing from year to year. A patient in stage 1 is 48 % likely to remain in the same state and 5% chance of being kidney failure at the end of one year. patients in state 4 have a probability of 0.800, 0.669, 0.548, 0.448 and 3.27 of staying in the same stages after 12, 24, 36, 48, 60 months respectively, and this probability also decreases as the time interval increases. A patient in state 4

has 32.7% and 67.2% chances of being in states 4 and 5 respectively at the end of fifth year. Generally, as t increases, the probability of the patient transiting to a next worse state is increasing while the probability to remain in the same state is decreasing.

We can calculate transition probabilities by using the stacked.data.msm () function to over time.

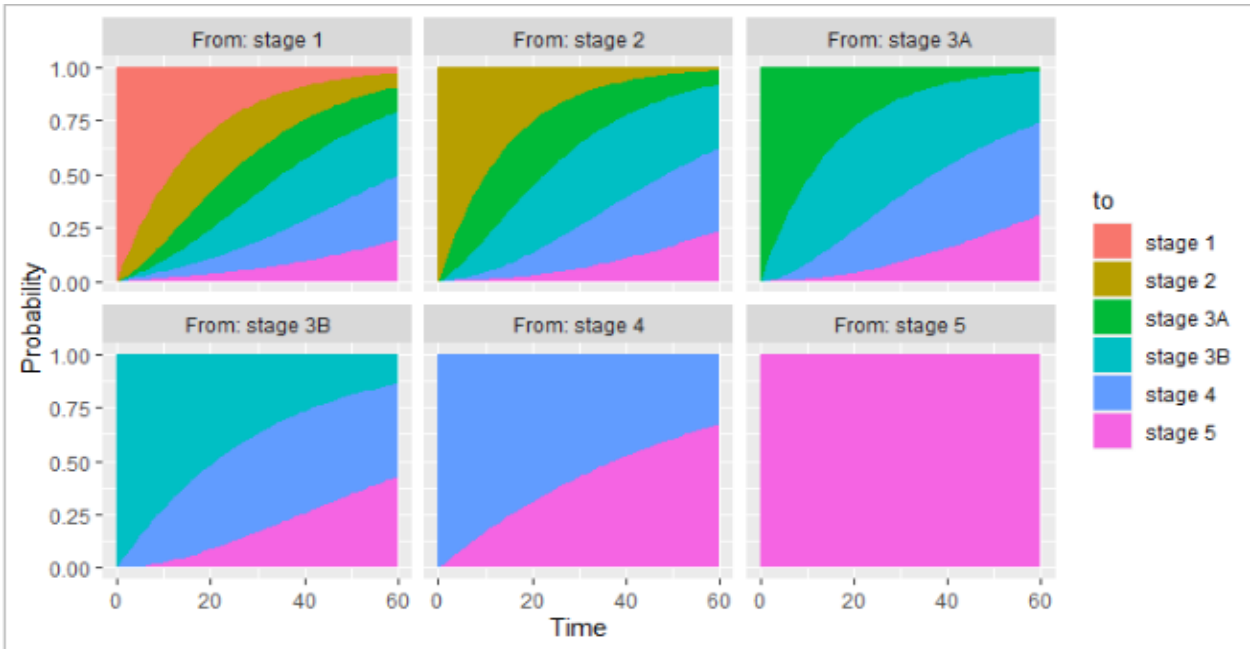

Figure 4.1: Estimated transition probability over five years.

4.4 Sojourn time

Sojourn time describes the average time an individual spends in each stage in a single stay before patients makes a transition to another stage. The sojourn times of a continuous-time Markov process in a stage j are independent, exponential random variables with mean $(\frac{-1}{\lambda_{kk}})$. From the results, if an individual is in stage1 patient spends 16.5 [$\frac{-1}{-0.0605} = 16.5$ Table 4.6] months in that stage before making a transition to other more severe stages. Stage 3 has the mean sojourn time of 14.5 months, meaning that patients is expected to stay in this stage for 14.5 months on average before moving to another stage. The standard error of this estimate is 1.14, and the 95% CI (12.5, 16.9). From the results, an individual who were in stage 3B he/she spends 30.3 months in that state before making a transition to other states while a patient spends 53.8 months in Stage 4 before transiting to other stages. These two stages (stage 3B and 4) have the highest sojourn time. The table suggests that patients tend to stay longer in stage 3B and 4 than in stage 1,2 and 3A. Patients

whose are in stage 4 has an estimated average time stay of 53.8 months before progress to next stage, means that individual in stage 4 are less likely to progress to stage 5 and the same information with transition intensity and probability of CKD patients on Table 4.3 and Table 4.4 respectively.

Table 4.6: Estimated Mean Time length stay in single stage of CKD patients.

Stages	Estimated Coefficient	Standard error (SE)	95% CI	
			L	U
Stage 1	16.5	2.32	12.5	21.7
Stage 2	14.5	1.14	12.5	16.9
Stage 3A	15.5	1.12	13.4	17.8
Stage 3B	30.3	2.02	26.5	34.5
Stage 4	53.8	5.64	43.8	66.1

4.5 Markov model with Effect of Covariates on Transition Intensities

Multi-state analysis was carried out using the msm package in order to learn how different factors influence the strengths of the transitions. The Table 4.7 shows the hazard ratios and 95% confidence intervals for socio-demographic variables (sex female as reference and age) for different transitions of stages of chronic kidney disease (CKD). The stages range from 1 to 5, with higher stages indicating more severe kidney damage and lower estimated glomerular filtration rate (eGFR). The hazard ratio is a measure of how often a particular event happens in one group compared to how often it happens in another group, over time. A hazard ratio of 1 means that there is no difference between the groups. A hazard ratio greater than 1 means that the event is more likely to happen in the first group than reference group and hazard ratio less than 1 means that the event is more likely to happen in the reference group.

For the transition from stage 1 to stage 2, the hazard risk for sex is 0.62, which means that male CKD patients were 38 % less likely to progress to stage 2 than female CKD patients, with a 95% CI (0.32, 1.132). The transition found high risk with male patients from stage 1 to stage 3A, stage 1 to stage 1, stage 2 to stage 5, stage 3A to stage 3B,4 and 5 than female CKD patients. Male CKD patients were about 1.61 times more likely to move from stage 1 to stage 3A than those of female

patients (HR = 1.61, 95% CI = 0.86, 2.97). Similarly, the hazard of male patients moving from stage 4 to stage 5 (kidney failure CKD stage) was about 27% lower than for female patients (HR = 0.73, CI =0.48, 1.108).

The hazard risk for age effect on progressing from stage 1 to stage 2 was 1.13, which means that for every one-year increase in age of CKD patients, the risk of progression from stage 1 to stage2 increases by 13% with a 95% CI (1.09,1.169).

Table 4.7: The Hazard ratios and 95% CI for socio-demographic variables.

Transitions of stages	**Sex (female as reference)**	**Age**
Stage 1 → Stage 2	0.62 (0.33, 1.15)	1.13 (1.09,1.169)
Stage 1 → Stage 3A	1.61 (0.86, 2.97)	1.10 (1.06,1.140)
Stage 1 → Stage 3B	0.68 (0.04, 9.68)	1.09 (1.05,1.137)
Stage 1 → Stage 4	0.88 (0.08, 9.36)	1.15 (1.06,1.232)
Stage 1 → Stage 5	1.46 (0.10, 20.63)	1.19 (1.08,1.312)
Stage 2→ Stage 3A	0.72 (0.47, 1.10)	1.01 (0.99,1.020)
Stage 2 → Stage 3B	0.80 (0.15, 4.15)	1.04 (1.02,1.059)
Stage 2 → Stage 4	0.68 (0.04, 9.68)	1.01 (0.99,1.040)
Stage 2 → Stage 5	1.13 (0.11, 11.99)	1.03 (0.98,1.069)
Stage 3A→ Stage 3B	1.14 (0.84, 1.53)	1.02 (1.00,1.035)
Stage 3A→ Stage 4	1.38 (0.90, 2.10)	1.02 (1.00,1.038)
Stage 3A→ Stage 5	1.24 (0.24, 6.39)	1.01 (0.99,1.038)
Stage 3B→ Stage 4	0.95 (0.75, 1.22)	1.07 (1.04,1.104)
Stage 3B→ Stage 5	0.88 (0.65, 1.13)	0.98 (0.97,1.004)
Stage 4→ Stage 5	0.73 (0.48, 1.108)	1.01 (0.98,1.022)

The fourth row shows that the hazard ratio of sex for the transition from stage 1 to stage 3A is 2.6296, with a 95% CI (0.03,178.207). This means that males had a higher hazard rate than females for this transition. The hazard risk for age of patients was 1.09, which means that for every one-year increase in age of CKD patients, the risk of progressing to stage 3A increases by 9%, with a 95% CI (1.05,1.137).

The general pattern for the effect of age was positively associated with most of the transitions of CKD stages of patients, except for Stage 3B to stage 5. This means that as the age of individual increases in years, the risk of moving from one stage to another also increases, except for moving from stage 3B to stage 5.

Table 4.8 Hazard ratios and 95% confidence intervals for clinical Laboratory variables.

Transitions	Haemoglobin (HGB)	Potassium in blood (K)	Phosphate in blood (P)	Sodium (Na)	Urea in blood
1→ 2	0.91 (0.72,1.14)	1.25 (0.71, 2.20)	0.93 (0.81, 1.06)	1.00 (0.89,1.12)	0.99 (0.96,1.02)
1→ 3A	1.12 (0.52,2.38)	0.96 (0.06, 13.71)	0.94 (0.83, 1.07)	0.94 (0.78,1.12)	0.98 (0.93,1.04)
1→ 3B	1.03 (0.57,1.84)	0.93 (0.17, 4.90)	0.95 (0.53, 1.70)	0.82 (0.67,1.00)	0.97 (0.88,1.06)
1→ 4	0.99 (0.62,1.57)	1.28 (0.28, 5.69)	0.97 (0.28, 3.33)	0.86 (0.70,1.07)	0.98 (0.87,1.10)
1→ 5	0.93 (0.61,1.39)	1.25 (0.30, 5.15)	0.97 (0.32, 2.97)	0.84 (0.69,1.01)	0.98 (0.89,1.08)
2→ 3A	1.06 (0.93,1.20)	0.99 (0.71, 1.38)	0.96 (0.91, 1.01)	1.02 (0.94,1.10)	0.99 (0.98,1.01)
2→ 3B	1.37 (0.94,1.98)	0.98 (0.04, 24.02)	0.96 (0.65, 1.41)	0.98 (0.87,1.10)	1.01 (0.98,1.03)
2→ 4	1.02 (0.56,1.84)	0.84 (0.08, 7.92)	0.97 (0.29, 3.25)	0.93 (0.81,1.06)	1.01 (0.94,1.06)
2→ 5	1.05 (0.64,1.71)	0.96 (0.24, 3.82)	0.99 (0.03,28.39)	0.93 (0.75,1.16)	0.99 (0.78,1.26)
3A→ 3B	0.93 (0.85,1.01)	1.12 (0.89, 1.41)	0.92 (0.76, 1.11)	0.96 (0.90,1.02)	1.01 (1.00,1.02)
3A→ 4	1.18 (0.67,2.06)	1.01 (0.01,114.41)	0.98 (0.64, 1.52)	1.01 (0.89,1.14)	0.99 (0.96,1.03)
3A→ 5	1.17 (0.83,1.67)	1.02 (0.16, 6.32)	0.98 (0.50, 1.94)	0.99 (0.88,1.11)	1.01 (0.97,1.06)
3B→ 4	1.01 (0.94,1.08)	1.12 (0.90, 1.40)	1.11 (0.95, 1.29)	0.91 (0.87,0.95)	1.00 (0.99,1.01)
3B→ 5	0.92 (0.31,2.70)	0.98 (0.01,432.80)	0.84 (0.68, 1.04)	1.01 (0.85,1.17)	1.00 (0.99,1.01)
4→ 5	0.93 (0.80,1.08)	1.56 (1.09, 2.23)	1.21 (1.11, 1.32)	0.86 (0.81,0.92)	0.99 (0.98,0.99)

The results show Table 4.8 the hazard ratios and 95% confidence intervals for the effect of five laboratory continuous variables (haemoglobin in blood by (gm/dl), potassium in blood (mmol/l), phosphate in blood (mg/dl), sodium in blood (mmol/l) and urea) on the transition of chronic kidney disease (CKD) stages for patients.

We can look at each transition and each variable separately. For example, for the transition from state 1 (CKD stage 1) to state 2 (CKD stage 2), the hazard ratio for haemoglobin was 0.91 with a 95% CI (0.72,1.14). This means that for every unit increase in haemoglobin, the risk of progressing from CKD stage 1 to CKD stage 2 decreases by 9%. The hazard ratio for potassium was 1.25 with a 95% CI (0.71, 2.20). This means that for every unit increase in potassium, the risk of progressing from CKD stage 1 to CKD stage 2 increases by 25% and the same fashion for another variables.

However, some general patterns can be observed from results Haemoglobin has a negative or protective effect on most transitions, except for the transition from state 2 and 3A (CKD stage 2&3A) to other more severe state, where it has a positive or harmful effect. This means that higher levels of haemoglobin are associated with lower risks of progressing to more advanced stages of CKD. Sodium has a positive or harmful effect on some transitions and a negative or protective effect on others.

The most prognostic factor results are for potassium, phosphate and urea, which show positive associations with CKD progression for stage 4 to stage 5 CKD transitions and for urea stage 3A to stage 3B.

Table 4.9: Hazard ratios and 95% confidence intervals for comorbidity variables.

Stages	**Heart disease (HD)** [no HD as reference]	**Diabetic (TD)** [no TD as reference]	**Hypertension (HTN)** [no HTN as reference]
Stage 1 → Stage 2	1.18 (0.64,2.17)	1.88 (0.84, 4.18)	1.95 (0.99, 3.82)
Stage 1 → Stage 3A	0.85 (0.46, 1.54)	0.53 (0.24, 1.18)	0.51 (0.26, 1.01)
Stage 1 → Stage 3B	1.83 (0.15,21.52)	2.02 (0.19, 20.86)	1.64 (0.18, 14.64)
Stage 1 → Stage 4	0.89 (0.07, 10.82)	1.40 (0.29, 6.63)	1.26 (0.30, 5.19)
Stage 1 → Stage 5	0.55 (0.05, 6.46)	1.16 (0.13, 9.85)	1.01 (0.16, 6.24)
Stage 2→ Stage 3A	1.36 (0.94, 1.96)	1.17 (0.72, 1.87)	1.08 (0.71, 1.65)
Stage 2 → Stage 3B	0.33 (0.04, 2.40)	0.85 (0.53, 1.38)	0.92 (0.60, 1.40)
Stage 2 → Stage 4	1.83 (0.15, 21.52)	0.49 (0.04, 5.01)	0.60 (0.06, 5.41)
Stage 2 → Stage 5	1.12 (0.1, 13.4)	0.71 (0.15, 3.36)	0.79 (0.19, 3.24)
Stage3A→Stage3B	0.87 (0.64, 1.19)	0.86 (0.63, 1.17)	0.91 (0.68, 1.21)
Stage 3A→ Stage 4	0.74 (0.51, 1.06)	0.86 (0.10, 7.30)	0.98 (0.16, 6.11)
Stage 3A→ Stage 5	0.33 (0.50, 2.40)	0.91 (0.05, 16.85)	1.09 (0.06, 19.21)
Stage 3B→ Stage 4	1.02 (0.80, 1.29)	0.64 (0.49, 0.83)	0.72 (0.57, 0.92)
Stage 3B→ Stage 5	1.34 (0.84, 1.54)	1.16 (0.85, 1.58)	1.09 (0.08, 1.46)
Stage 4→ Stage 5	1.05 (0.71, 1.55)	1.69 (1.07, 2.67)	1.95 (1.17, 3.22)

The Table 4.9 shows the hazard ratios and 95% confidence intervals for comorbidity variables (heart disease, diabetic, and hypertension) for different transitions of stages of chronic kidney disease (CKD).

For the transition from stage 1 to stage 2, the hazard ratio for heart disease was 1.18, which means that patients with heart disease are 18% more likely to progress to stage 2 than patients without heart disease, with a [HR=1.18, 95% CI = 0.64, 2.17]. The hazard ratio for diabetic is 1.88, which means that patients with diabetic are 88% more likely to progress to stage 2 than patients without diabetic, [HR= 1.88, 95% CI = 0.84, 4.18]. Hypertension had effect noticeable impact on the shift from stage 4 to renal damage CKD stages (HR = 1.95). The hazard of hypertensive patients moving from stage 4 to stage 5 (kidney failure CKD stage) was about 1.92 times more likely than do not have Hypertension conditions patients [HR = 1.95, 95% CI = 1.17, 3.22].

4.6 Survival probability plot

Survival probability plot of CKD patients with no cofactor, with the red line representing patients from stage 1, the green line representing patients from stage 2, the blue line representing patients from stage 3A, the purple line representing patients from stage 3B, and the pink line representing patients from stage 4 was fitted on Figure 4.2.

According to the plot (Figure 4.2) the survival probability for CKD patients with severe CKD (stage 4) was starts from 1 at initial month and ends at 0.3 at month 60. Means that CKD patients enter stage 4 had survival probability 1 at beginning of the study and after 60 months 30% of CKD patients was alive with stage 4 and the else are progressed to next stage.

The survival probability of CKD patients with stage 4 decreases faster than the survival probability of CKD patients with stage 3B over time. In addition to that the lines show a general decrease in survival probability over time, with the highest probability for CKD patients from stage 1 and the lowest probability for patients from stage 4.

According to the plot the survival probability of CKD patients decreases as the stage of the disease progresses and CKD patients with advanced stages of the disease have a worse prognosis than CKD patients with early stages of the disease and the difference in survival probability between the stages becomes more pronounced over time.

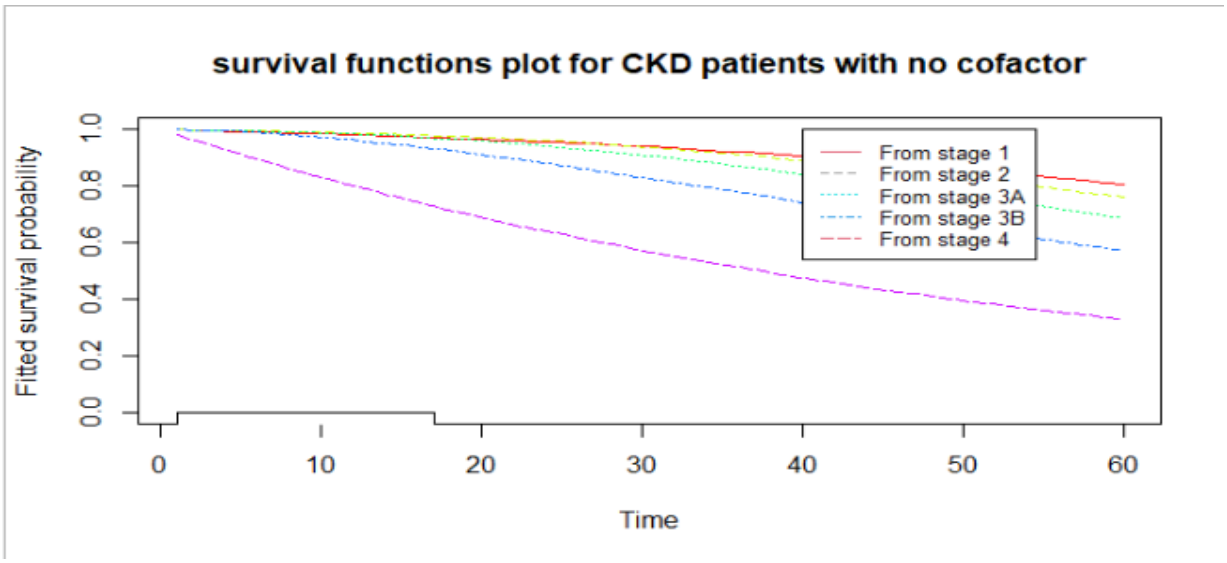

Figure 4.2: Reliability or survival probability of CKD patients at 60 months.

4.7 Prediction of Clinical CKD Progression.

The conditional probability of a patient making changes in disease stages given individual current stage status is computed. The Figure 4.3 shows conditional probability that an CKD patient who is currently in stage 1 has been in the subsequent "worse" stage 2, 3A, 3B, 4, 5 after 60 months.

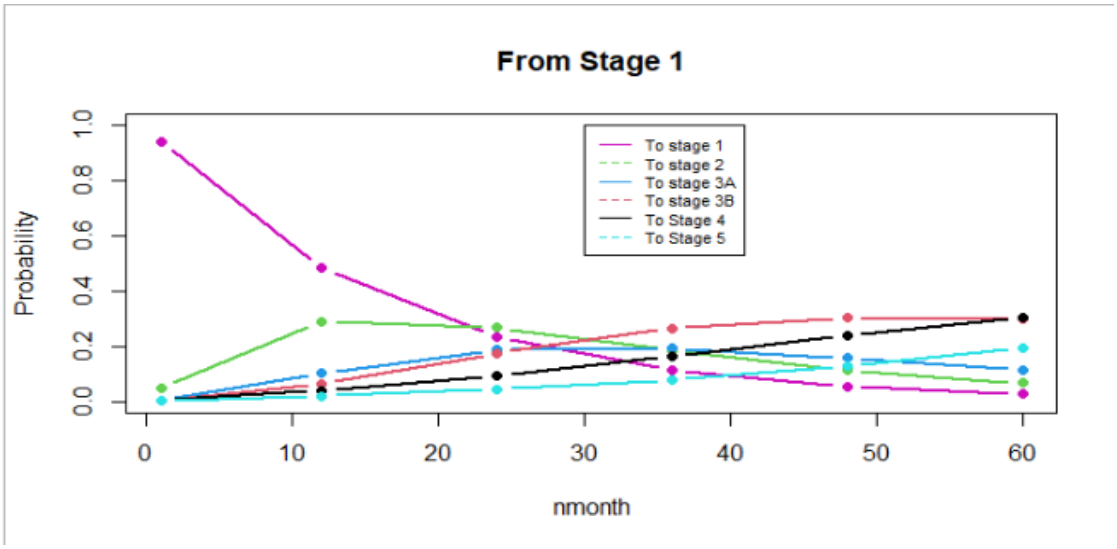

Figure 4.3: Conditional probabilities of being in next stage after a month t given the stage 1.

The probability of a patient starting from stage 1 at time 1 month enters to stage l ∈ {1, 2, 3A, 3B, 4, 5} after 12 months, are estimated to be 0.48, 0.3, 0.11, 0.07, and 0.04 respectively, and after 60 months are estimated to be 0.03, 0.06, 0.11, 0.3, and 0.2 respectively. The transition probability from stage 1 to the next worse state increases with time and gets optimum probability at a time,

then decreases with increasing time. This means that there is some period of time when such probability is highest for a patient to transit to a worse state of the disease.

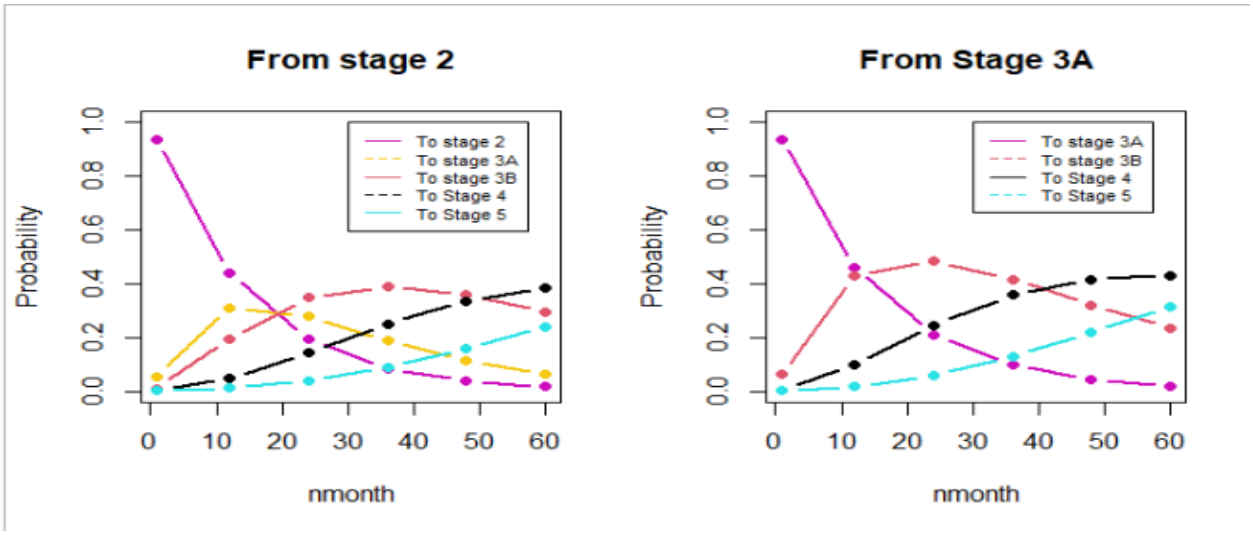

Figure 4.4: Conditional probabilities of being in next worse stage after a month t given the stage 2 and 3A respectively.

Figure 4.4 shows conditional probability that an CKD patient who is currently in stage 2 and 3A has been in the subsequent "worse" stage 2, 3A, 3B, 4, 5 and stage 3A, 3B, 4, 5 respectively after 60 months. probability that a patient starting from stage 3A at initial time progress to next stage 3A, 3B, 4, 5 after 36 months are estimated to be 0.09, 0.42, 0.35, and 0.13 respectively. Stage 3A gets optimum probability at 12 months, stage 3B gets optimum probability at 36 months and shows decreasing probability as time increase, stage 4 and 5 increases probability as time increases. The same fashion interpretation for figure 4.6.

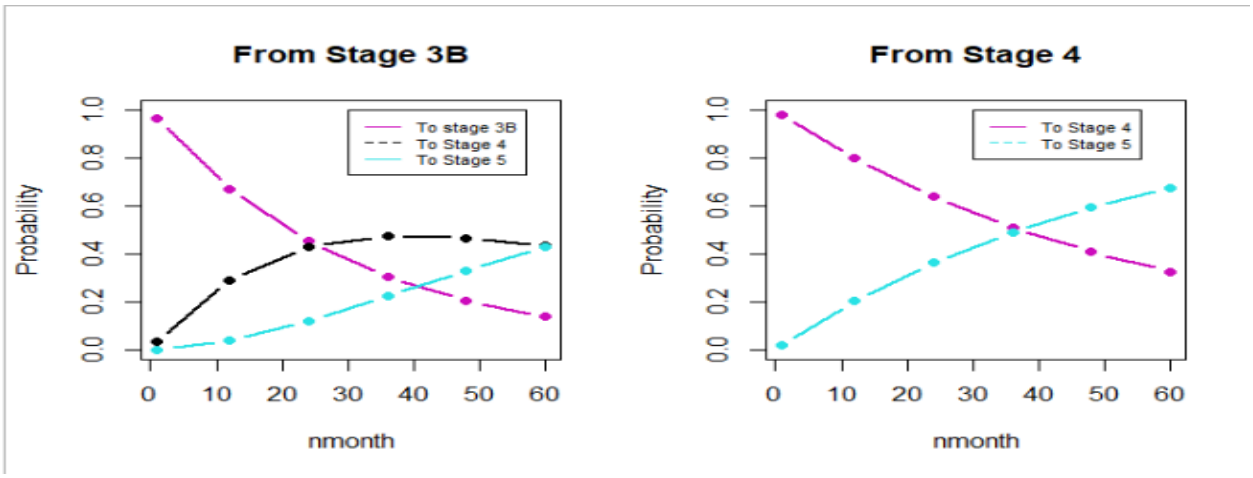

Figure 4.5: Conditional probabilities of being in next worse stage after a month t given the stage 3B and 4 respectively.

4.8 Model Comparison

The table 4.10 shows the results of fitting different models with different covariates to the data. The models are compared using the log likelihood ratio test, the degrees of freedom (df), and the p-value. The log likelihood ratio test is a measure of how much the model improves compared to the null model (no covariate), with higher values indicating better fit.

Table 4.10: Likelihood Ratio Test Statistics and P-Values of the Selected Models.

Models	**-2 Log likelihood**	**-2 Log likelihood ratio test**	**Df**	**P-value**
No covariate	2080. 422	-	-	-
All as a covariate	1584.13	496.2919	165	P < 0.0001

Multi state Markov Model with null model (without covariates), model with single covariates one by one and then all as covariates was fitted. the likelihood ratio test in tests between nested models fitted in msm can be performed using the function lrtest.msm to compare the model that were fitted. The nested models fitted in msm package were compared using a Likelihood ratio test contrast to the model without covariates (null model) with all as covariates (full model) (Taha & Mohammad, 2023). Therefore, full model has $LRT = \left(L_0(\theta) - L_1(\theta)\right) = 496.292$ with df of 165, a p-value of < 0.0001. where $L_0(\theta)$ is the -2 Log likelihood null model (without covariates) and $L_1(\theta)$ is the -2 Log likelihood general model (full model). The results show that the full model does significantly fit better than null models.

4.9 Model assessment

A time homogeneous multistate Markov model assumes that the transition rates between states are constant and do not depend on time or covariates. The plot of observed and expected prevalence used for assessing time homogeneous multistate Markov model assumption and tool indication of the goodness of fit of a multi-state Markov model.

The expected prevalence is the proportion of individuals who are predicted to be in a certain stage at a given time, based on the fitted model and the initial stage distribution. The plot can be used to

compare the observed and expected prevalence for each state and check if they are close to each other. If the plot shows a good agreement between the observed and expected prevalence, it suggests that the model fits the data well and can be used for forecasting and inference. The implication of the plot is that it can help to evaluate the validity and accuracy of the multi-state Markov model and its assumptions.

According to Figure 4.6, we look that the observed and forecast prevalence for stage 1, stage 2, stage 3A, stage 3B, and stage 5 was good agreement between the observed and expected prevalence, which suggests that the model fits the data well. However, stage 4 where the observed prevalence is slightly higher than the expected prevalence after 50 months. This is because of the number of patients in stage 4 tends to increase as time increase, see from Table 4A, Appendix I. It indicates that the model underestimates the transition rate to stage 4 or overestimates the transition rate from stage 4 to other states. The assumption transition rates between states constant practically always may not true. Thus, such problem was improved by fit the model adding covariates. Hence plot seems that the plot was fit the assumption the of homogeneity of transition rate for most of the stages through the specified time is satisfied and Multistate Markov model fits the data well.

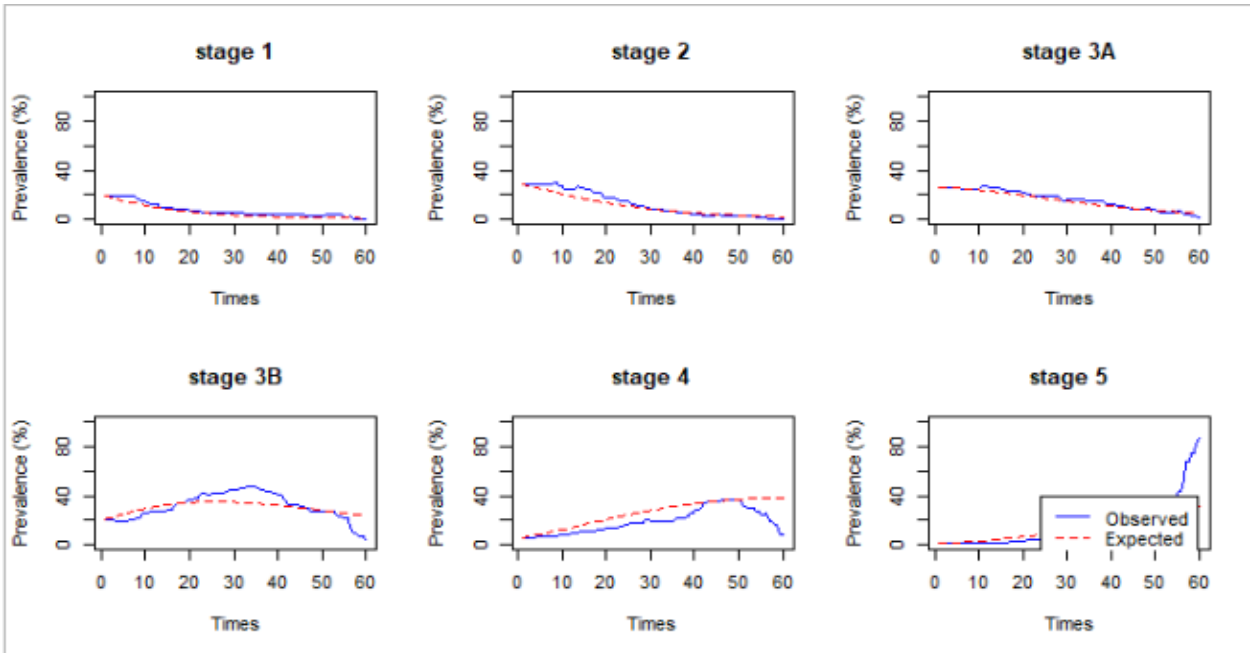

Figure 4.6: A comparison of observed and expected prevalence in the fitted model.

4.10 Discussions

The aim of the study was to model the progression of Chronic Kidney disease CKD among patients under follow up at Tikur Anbessa Specialized Hospital, TASH, Addis Ababa, Ethiopia using Multistate Markov Model. A Multistate Markov Model is fitted with and without covariates and comparison of these two nested models (Null and full model or all as covariates) using likelihood ratio test. Analysis Results shows that the model with All as a covariate is the best model fits. Assumption of transition rate on specific time were approximately satisfied using the observed and expected plot.

Multistate models based on Markov processes offer a convenient framework for the analysis and modelled the progression of chronic disease processes frequently (Cook & Lawless, 2014), (Josephine, 2018), (Meenaxi & Singh, 2022).

The probability of a patient transiting from states 1, 2, 3A, 3B, and 4 to state 5 after one month is 0.0014, 0.0007, 0.001, 0.0004, and 0.0018, respectively. The estimated sojourn times for stages 1, 2, 3A, 3B, and 4 are 16.5, 14.5, 15.5, 30.3, and 53.8 months, respectively. The results show patients that speed limits the progression of CKD from stages 1, 2, and 3A to severe stages and stays more in stages 3B, and 4 before progressing to kidney failure stages. Since the results of the transition count show that the number of patients within stages 1, 2, and 3A was small, a large number of participants started follow-up in stages 3B and 4. Then, after regular monitoring and appropriate management of CKD, progress slowly and more slowly. This result was not similar to the studies done by (Taha & Mohammad, 2023) and (GROVER et al., 2019b). Grover et al., 2019 found that the probability of patients moving from stage 1 to the most severe CKD stages is more likely that CKD patients of stage 1 will remain in stage 1 over five years of time than those moving to higher stages, and the result showed that the mean sojourn time for stage 1 was the highest. Stage 2 and Stage 3 also have considerably high sojourn times. This proves the slow progression of CKD in earlier stages.

Sex has a high effect on the transitions from the good stage to the severe stage, with males having higher risks than females. Male patients in state 1 are at a higher risk [HR: 1.61; 95% CI (0.86–2.97)] of disease progression compared to females at other stages. Additionally, male patients' stage 3A have a higher risk of regress than other stages. This is in accordance with the most recent study that used a multi-state model to understand the progression of chronic kidney disease (Lintu

et al., 2022). The highest hazard ratios for the sex effect are observed for the transition from stage 1 to stage 3B, indicating that this is the most critical transition in terms of risk factors.

This study shows that age was an important socio-demographic factor that is positively associated with the progression of CKD patients with most of the transitions in stages, except for 3B to 5. This means that as age increases, the likelihood of moving from one stage to another also increases. These findings contradict with study conducted by (Chang et al., 2019). The study showed that age as a risk factor found that in CKD patients, every 10 years rise in age will decrease the risk of dialysis (HR: 0.95, 95% CI: 0.91–0.99).

In our study, hypertension was the most common comorbidity (67.5%). Similar findings were reported by other studies: 91.1% (Kefale et al., 2018) and 80% (Berhe et al., 2023) in Ethiopia. The hazard ratio for CKD progression was higher for patients with hypertension than those without hypertension in most comparisons.

The study indicated that diabetes patients with stage 3B disease and stage 4 disease had a higher risk of progressing to kidney failure (stage 5 CKD) when compared with CKD stage 3A patients without diabetes. Other studies also documented similar findings (Zhang et al., 2022). A study discovered that people with diabetes had significantly higher CKD G3b and G4 patients had a higher risk of progression to ESRD when compared to people without CKD G3a. The result implied that most of the hazard ratios are greater than 1, indicating that diabetes increases the risk of moving to a higher stage of CKD. This is consistent with the finding of (Taha & Mohammad, 2023), who also reported that diabetes was a significant factor in the progression of CKD, with a higher risk of moving from stage 1 to stage 2 and from stage 3B to renal failure.

This study also found that lower haemoglobin, higher urea, higher potassium, higher phosphate, and lower sodium were associated with a higher risk of CKD progression, and the effect of these variables varied depending on the initial and final CKD stages. The study showed that potassium had positive associations with CKD progression from stage 4 to stage 5 CKD. These findings are correlated with recent studies (RK & Nagdeve, 2023). The study found that as the stages of CKD get worse, the serum potassium value is also found to be higher.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study evaluated the progressions of chronic kidney disease CKD longitudinally and its possible predictors via time continuous homogenous Markov model. Model with and without covariates have been compared using the LRT and, the model with full model exhibited the best fit model.

Prognostic factors like being male, having a history of Diabetes, having a history of Hypertension, and having a history of heart disease were the factors that had a higher risk of progressing to severe stages in CKD patients, and Age, Haemoglobin, and Potassium were positively (or harmfully) associated with the progression of eGFR or CKD stages. Whereas Phosphate, Sodium, and Urea were negatively associated with the progression change of eGFR or CKD stages.

The transition probability from a given good stage to the next worse stage increases with time, reaches its optimum (peak) at a time, and starts to decline as time goes on. This means that there is some period of time when the probability is highest for a patient to transit to a worse state of the disease. Intervention for patient care may minimize such an effect.

According to the reliability (survival probability) analysis, all survival probabilities of CKD patients' stages are declining over time, which suggests that improving patient circumstances (implication) is necessary to keep the probability of survival as high as possible.

The estimated mean duration of stage 4 CKD patients was the longest, followed by stage 3B, while the expected mean duration of stage 2 CKD was the shortest. The probability of a patient with chronic kidney disease (CKD) being in that stage of the disease decreases as time increases.

5.2 Recommendations

Based on the findings of the study, we recommend the following:

- The Ministry of Health and policymakers should work on raising awareness about the risk factors for CKD progression because it is the most serious disease in the world.
- Health workers should be careful to follow up on patients with a history of diabetes, hypertension, or heart disease to reduce the risk of CKD patients' progression.
- Patients need to regularly check their eGFR on the appropriate day in order to know their disease stage, improve survival probability, and pay attention to the implications of electrolyte level.
- Researcher should be done in the area using the most developed and flexible methodologies, including covariates like body mass index (BMI), educational level, albumin (Alb), cardiovascular disease, and smoking status of patients.

5.3 Limitation of the Study

Since this study was based on retrospective secondary data from chart review (electronic database i-care), some important or potential predictor variables (sociodemographic and clinical) like Body mass index (BMI), educational level, Albumin (Alb), cardiovascular disease and Smoking status of patients which predict the evolution of chronic kidney disease which change the progression of disease were not available on patient's chart.

Most of research were done about prevalence of CKD, Predictors of Glomerular Filtration Rate and Time to Develop End-stage for CKD and on few on progression CKD using Markov process by merge stage of kidney disease, so there was scarce of literature on our country specifically in the study area related to progression of CKD using Kidney Disease Improving Global Outcome (KDIGO) stages i.e., stage1, stage2, stage 3A, stage 3B, stage4, stage 5.Therefore the authors faced regarding to discuss with another study or scarcity of literature on this study, especially in our country.

References

Ali, Z., Mahmood, M., Kazem, M., Hojjat, Z., Mostafa, H., & Kourosh, H. N. (2014). Assessing Markov and time homogeneity assumptions in multi-state models: Application in patients with gastric cancer undergoing surgery in the Iran Cancer Institute. Asian Pacific Journal of Cancer Prevention, 15(1), 441–447. https://doi.org/10.7314/APJCP.2014.15.1.441

Animaw, Z., Walle Ayehu, G., & Abdu, H. (2022). Prevalence of chronic kidney disease and associated factors among patients with chronic illness in Ethiopia: A systematic review and meta-analysis. SAGE Open Medicine, 10, 1–9. https://doi.org/10.1177/20503121221089442

Asena, T. (2021). Semi-Markov Modelling of HIV / AIDS Disease Progression.

Berhe, T., Tegegne, G. T., & Berha, A. B. (2023). Quality of life and associated factors among chronic kidney disease patients at Zewditu Memorial and Tikur Anbessa Specialised Hospitals, Ethiopia: a cross-sectional study design. BMJ Open, 13(6), e069712. https://doi.org/10.1136/bmjopen-2022-069712

Bozorgnezhad, M. (2018). Journal of Biostatistics and Epidemiology. J Biostat Epidemiol., 1(1), 37–44.

Centers for Disease Control and Prevention in the United States. (2021). Chronic Kidney Disease in the United States, 2021. Cdc, 1, 1–6. https://www.cdc.gov/kidneydisease/publications-resources/ckd-national-facts.html%0Ahttps://www.cdc.gov/kidneydisease/publications-resources/2019-national-facts.html

Chang, H., Wu, C., Lee, S., Chen, Y., Su, W., & Su, S. (2019). A predictive model for progression of CKD. 0(April).

Cook, R. J., & Lawless, J. F. (2014). Statistical Issues in Modeling Chronic Disease in Cohort Studies. Statistics in Biosciences, 6(1), 127–161. https://doi.org/10.1007/s12561-013-9087-8

Dessie, Z. G. (2014). Multi-State Models of Hiv/Aids By Homogeneous Semi-Markov Process. American Journal of Biostatistics, 4(2), 21–28. https://doi.org/10.3844/amjbsp.2014.21.28

Eastwood, J. B., Kerry, S. M., Plange-Rhule, J., Micah, F. B., Antwi, S., Boa, F. G., Banerjee, D., & Cappuccio, F. P. (2010). Assessment of GFR by four methods in adults in Ashanti, Ghana:

The need for an eGFR equation for lean African populations. Nephrology Dialysis Transplantation, 25(7), 2178–2187. https://doi.org/10.1093/ndt/gfp765

Fiseha, T., & Tamir, Z. (2020). Prevalence and awareness of chronic kidney disease among adult diabetic outpatients in Northeast Ethiopia. BMC Nephrology, 21(1), 1–7. https://doi.org/10.1186/s12882-020-01768-y

Gillings, D., Hoel, P. G., Port, S. C., & Stone, C. J. (1974). Introduction to Stochastic Processes. In Technometrics (Vol. 16, Issue 1, p. 137). https://doi.org/10.2307/1267505

Grover, G., Sabharwal, A., Kumar, S., & Kumar Thakur, A. (2018). On the estimation of misclassification probabilities of chronic kidney disease using continuous time hidden Markov models. Journal of Nephropharmacology, 8(1), 7–7. https://doi.org/10.15171/npj.2019.07

GROVER, G., SABHARWAL, A., KUMAR, S., & THAKUR, A. K. (2019a). A Multi-State Markov Model for the Progression of Chronic Kidney Disease. Turkiye Klinikleri Journal of Biostatistics, 11(1), 1–14. https://doi.org/10.5336/biostatic.2018-62156

GROVER, G., SABHARWAL, A., KUMAR, S., & THAKUR, A. K. (2019b). A Multi-State Markov Model for the Progression of Chronic Kidney Disease. Turkiye Klinikleri Journal of Biostatistics, 11(1), 1–14. https://doi.org/10.5336/biostatic.2018-62156

Grover, G., Swain, P. K., Goel, K., & Singh, V. (2018). Multistate markov modelling for disease progression of breast cancer patients based on CA15-3 marker. Thailand Statistician, 16(2), 129–139.

Hussein, M., Muleta, G., Seyoum, D., Kifle, D., & Bedada, D. (2017). Survival Analysis of Patients with End Stage Renal Disease the Case of Adama Hospital , Ethiopia. 6(6), 201–208. https://doi.org/10.11648/j.cmr.20170606.15

Inker, L. A., Eneanya, N. D., Coresh, J., Tighiouart, H., Wang, D., Sang, Y., Crews, D. C., Doria, A., Estrella, M. M., Froissart, M., Grams, M. E., Greene, T., Grubb, A., Gudnason, V., Gutiérrez, O. M., Kalil, R., Karger, A. B., Mauer, M., Navis, G., … Levey, A. S. (2021). New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race. New England Journal of Medicine, 385(19), 1737–1749. https://doi.org/10.1056/nejmoa2102953

Inker, L. A., & Titan, S. (2021). Measurement and Estimation of GFR for Use in Clinical Practice: Core Curriculum 2021. American Journal of Kidney Diseases, 78(5), 736–749. https://doi.org/10.1053/j.ajkd.2021.04.016

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. Journal of Statistical Software, 38(8), 1–28. https://doi.org/10.18637/jss.v038.i08

Jackson, C. H., Sharples, L. D., Thompson, S. G., & Duffy, S. W. (2003). <1467-9884.00351.Pdf>. 193–209.

Jackson, C. H., Sharples, L. D., Thompson, S. G., Stephen, W., Journal, S., Statistical, R., Series, S., Statistician, D. T., & Jackson, C. H. (2003). Multistate Markov Models for Disease Progression with Classification Error Duffy and Elisabeth Couto Published by : Wiley for the Royal Statistical Society Stable URL : http://www.jstor.org/stable/4128245 REFERENCES Linked references are available on JSTO. 52(2), 193–209.

Jeong, W. G., Choi, H., Chae, K. J., & Kim, J. (2022). Prognosis and recurrence patterns in patients with early stage lung cancer: a multi-state model approach. Translational Lung Cancer Research, 11(7), 1279–1291. https://doi.org/10.21037/tlcr-22-148

Josephine, X. L. (2018). STOCHASTIC MODEL APPLIED IN CHRONIC KIDNEY. 4(3), 13–17.

Kalbfleisch, J. D., & Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. Journal of the American Statistical Association, 80(392), 863–871. https://doi.org/10.1080/01621459.1985.10478195

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53(282), 457–481. https://doi.org/10.1080/01621459.1958.10501452

Kaze, A. D., Ilori, T., Jaar, B. G., & Echouffo-Tcheugui, J. B. (2018). Burden of chronic kidney disease on the African continent: A systematic review and meta-analysis. BMC Nephrology, 19(1), 1–11. https://doi.org/10.1186/s12882-018-0930-5

Kefale, B., Tadesse, Y., Alebachew, M., & Engidawork, E. (2018). Management Practice, and Adherence and Its Contributing Factors among Patients with Chronic Kidney Disease at Tikur

Anbessa Specialized Hospital: A Hospital Based Cross-Sectional Study. International Journal of Nephrology, 2018. https://doi.org/10.1155/2018/2903139

Kore, C., Tadesse, A., Teshome, B., Daniel, K., Kassa, A., & Ayalew, D. (2018). The Magnitude of Chronic Kidney Disease and its Risk Factors at Zewditu Memorial Hospital, Addis Ababa, Ethiopia. Journal of Nephrology & Therapeutics, 08(03), 8–12. https://doi.org/10.4172/2161-0959.1000313

Kotzé, L. (2019). Markov Modelling of Disease Progression in the Presence of Missing Covariates. April.

Kovesdy, C. P. (2022). Epidemiology of chronic kidney disease: an update 2022. Kidney International Supplements, 12(1), 7–11. https://doi.org/10.1016/j.kisu.2021.11.003

Le-Rademacher, J. G., Peterson, R. A., Therneau, T. M., Sanford, B. L., Stone, R. M., & Mandrekar, S. J. (2018). Application of multi-state models in cancer clinical trials. Clinical Trials, 15(5), 489–498. https://doi.org/10.1177/1740774518789098

Levey, A. S., & Inker, L. A. (2017). Assessment of Glomerular Filtration Rate in Health and Disease: A State of the Art Review. Clinical Pharmacology and Therapeutics, 102(3), 405–419. https://doi.org/10.1002/cpt.729

Lintu, M. K., Shreyas, K. M., & Kamath, A. (2022a). A multi-state model for kidney disease progression. Clinical Epidemiology and Global Health, 13(May 2021), 100946. https://doi.org/10.1016/j.cegh.2021.100946

Lintu, M. K., Shreyas, K. M., & Kamath, A. (2022b). A multi-state model for kidney disease progression. Clinical Epidemiology and Global Health, 13(May 2021), 100946. https://doi.org/10.1016/j.cegh.2021.100946

Mafu, T. J. (2014). Modelling of Multi-State Panel Data : The Importance of the Model Assumptions by. ELKHA, Jurnal Teknik Elektro, 5(December), 23–26.

Marshall, G., & Jones, R. H. (1995). Multi-state models and diabetic retinopathy. Statistics in Medicine, 14(18), 1975–1983. https://doi.org/10.1002/sim.4780141804

Matsena Zingoni, Z., Chirwa, T. F., Todd, J., & Musenge, E. (2019). HIV Disease Progression Among Antiretroviral Therapy Patients in Zimbabwe: A Multistate Markov Model. Frontiers in Public Health, 7. https://doi.org/10.3389/fpubh.2019.00326

Meenaxi, & Singh, D. (2022). a Multistate Markov Model Describing the Progression of Various Deteriorating Stages of Chronic Kidney Disease. JP Journal of Biostatistics, 21, 11–28. https://doi.org/10.17654/0973514322018

Mengesha, S. K., & Ferede, T. (2019). RESEARCH PAPERS APPLICATION OF MULTI-STATE SEMI-MARKOV MODELS ON HIV / AIDS DISEASE PROGRESSION By. 7(3), 2018.

Milik, A., & Hrynkiewicz, E. (2014). On translation of LD, IL and SFC given according to IEC-61131 for hardware synthesis of reconfigurable logic controller. IFAC Proceedings Volumes (IFAC-PapersOnline), 19(1), 4477–4483. https://doi.org/10.3182/20140824-6-za-1003.01333

National Kidney, F. (2021). Implementation of the 2021 CKD-EPI GFR equations. Tufts Medical Center.

Orantes, C. M., Herrera, R., Almaguer, M., Brizuela, E. G., Nuñez, L., Alvarado, N. P., Jackeline Fuentes, E., Bayarre, H. D., Amaya, J. C., Calero, D. J., Vela, X. F., Zelaya, S. M., Granados, D. V., & Orellana, P. (2014). Epidemiology of chronic kidney disease in adults of Salvadoran agricultural communities. MEDICC Review, 16(2), 23–30. https://doi.org/10.37757/mr2014.v16.n2.5

Reddy, T. (2011). The Application of Multistate Markov Models to HIV Disease Progression. 238.

Seck, S. M., Doupa, D., Guéye, L., & Dia, C. A. (2014). Epidemiology of chronic kidney disease in northern region of Senegal: A community-based study in 2012. Pan African Medical Journal, 18, 1–8. https://doi.org/10.11604/pamj.2014.18.307.3636

Shih, H. C., Chou, P., Liu, C. M., & Tung, T. H. (2007). Estimation of progression of multi-state chronic disease using the Markov model and prevalence pool concept. BMC Medical Informatics and Decision Making, 7. https://doi.org/10.1186/1472-6947-7-34

SOLOMON, K. M., GEBREGEWERGIS, A. G., TILAHUN, F., & CHERU, A. (2018). Application of Multi-State Semi-Markov Models on Hiv/Aids Disease Progression. I-Manager's Journal on Mathematics, 7(3), 30. https://doi.org/10.26634/jmat.7.3.14988

Stanifer, J. W., Maro, V., Egger, J., Karia, F., Thielman, N., Turner, E. L., Shimbi, D., Kilaweh, H., Matemu, O., & Patel, U. D. (2015). The epidemiology of chronic kidney disease in Northern Tanzania: A population-based survey. PLoS ONE, 10(4), 1–12. https://doi.org/10.1371/journal.pone.0124506

Stanifer, J. W., Muiru, A., Jafar, T. H., & Patel, U. D. (2016). Chronic kidney disease in low- and middle-income countries. Nephrology Dialysis Transplantation, 31(6), 868–874. https://doi.org/10.1093/ndt/gfv466

Sud, M., Tangri, N., Pintilie, M., Levey, A. S., & Naimark, D. (2014). Risk of end-stage renal disease and death after cardiovascular events in chronic kidney disease. Circulation, 130(6), 458–465. https://doi.org/10.1161/CIRCULATIONAHA.113.007106

Sumaili, E. K., Krzesinski, J. M., Zinga, C. V., Cohen, E. P., Delanaye, P., Munyanga, S. M., & Nseka, N. M. (2009). Prevalence of chronic kidney disease in Kinshasa: Results of a pilot study from the Democratic Republic of Congo. Nephrology Dialysis Transplantation, 24(1), 117–122. https://doi.org/10.1093/ndt/gfn469

Swain. (2013). A Multistate Markov Model Based on CD4 Cell Count for HIV/AIDS Patients on Antiretroviral Therapy (ART). International Journal of Statistics in Medical Research, 1998, 144–151. https://doi.org/10.6000/1929-6029.2013.02.02.08

Taha, A. A., & Mohammad, M. A. (2023). A Multistate Model for the Analyzing of Chronic Kidney Disease Progression with Detecting Risk Factors Effect. 13.

Webster, A. C., Nagler, E. V, Morton, R. L., & Masson, P. (2016). Seminar Chronic Kidney Disease. The Lancet, 6736(16), 1–15. https://doi.org/10.1016/S0140-6736(16)32064-5

Webster, A. C., Nagler, E. V., Morton, R. L., & Masson, P. (2017). Chronic Kidney Disease. The Lancet, 389(10075), 1238–1252. https://doi.org/10.1016/S0140-6736(16)32064-5

Workie, S. G., Zewale, T. A., Wassie, G. T., & Belew, M. A. (2022). Survival and predictors of mortality among chronic kidney disease patients on hemodialysis in Amhara region , Ethiopia ,. BMC Nephrology, 1–9. https://doi.org/10.1186/s12882-022-02825-4

Appendix I

Table 4A: Observed number of patients in each state evaluated at 10 equally spaced.

Time	Stage1	Stage2	Stage 3A	Stage 3B	Stage 4	Stage 5	Total
1	51	77	68	54	14	2	266
6.9	51	76	66	54	17	2	266
12.8	31	64	67	71	23	3	259
18.7	21	54	59	86	30	7	257
24.6	14	37	46	103	41	10	251
30.5	12	19	37	105	44	15	232
36.4	9	12	34	99	49	19	222
42.3	9	5	21	64	67	29	195
48.2	5	5	14	43	58	35	160
54.1	5	2	7	27	32	45	118
60	0	0	1	3	5	27	66

Table 4B: Expected number of patients in each stage evaluated at 10 equally spaced intervals

Time	Stage1	Stage2	Stage 3A	Stage 3B	Stage 4	Stage 5	Total
1	51.00	77.00	68.0	54.0	14.0	2.00	266
6.9	35.67	61.99	66.0	71.3	25.8	5.24	266
12.8	24.29	47.50	59.1	81.1	37.7	9.40	259
18.7	16.86	36.45	51.7	87.3	49.9	14.82	257
24.6	11.52	27.17	43.3	87.7	60.3	21.03	251
30.5	7.45	18.97	33.5	80.3	65.3	26.51	232
36.4	4.98	13.58	26.3	74.0	70.2	32.98	222
42.3	3.01	8.87	18.7	61.1	67.1	36.28	195
48.2	1.76	5.37	12.2	46.2	58.3	36.15	160
54.1	0.91	2.91	7.1	30.9	44.6	31.56	118
60	0.35	1.19	3.1	15.5	25.4	20.46	66

Table 4C: Observed percentages of patients in each state.

Time	State1	State2	State3	State4	State5	State6
1	19.17	28.95	25.56	20.30	5.26	0.75
6.9	19.17	28.57	24.81	20.30	6.39	0.75
12.8	11.97	24.71	25.87	27.41	8.88	1.16
18.7	8.17	21.01	22.96	33.46	11.67	2.72
24.6	5.58	14.74	18.33	41.04	16.33	3.98
30.5	5.17	8.19	15.95	45.26	18.97	6.45
36.4	4.05	5.41	15.32	44.59	22.07	8.56
42.3	4.62	2.56	10.77	32.82	34.36	14.87
48.2	3.12	3.12	8.75	26.88	36.25	21.87
54.1	4.24	1.69	5.93	22.88	27.12	38.14
60	0.00	0.00	1.52	4.55	7.58	89.36

Table 4D: Expected percentages of patients in each state.

Time	State1	State2	State3	State4	State5	State6
1	19.17	28.95	25.56	20.3	5.26	0.75
6.9	13.41	23.31	24.84	26.8	9.70	1.96
12.8	9.38	18.34	22.80	31.3	14.54	3.62
18.7	6.56	14.18	20.12	34.0	19.41	5.76
24.6	4.58	10.83	17.24	34.9	24.02	8.37
30.5	3.21	8.17	14.42	34.6	28.14	11.42
36.4	2.24	6.12	11.84	33.3	31.63	14.85
42.3	1.57	4.55	9.57	31.3	34.40	18.60
48.2	1.09	3.36	7.63	28.9	36.45	22.59
54.1	0.76	2.47	6.02	26.2	37.79	26.74
60	0.53	1.81	4.70	23.5	38.47	30.99

Appendix II

Annex 1: Data abstraction form

Title of research: Application of Multistate Markov on Progression of Chronic Kidney Disease CKD in case of Tikur Ambessa Specialized Hospital, Ethiopia.

Objective of the study: The objective of the study is to model and identify risk factors of progression of Chronic Kidney disease CKD of patients in Tikur Ambessa Specialized Hospital using multistate Markov processes from September 2015 to February 2022.

Data collection procedure: To achieve the above objective, this checklist was prepared for collecting information on socio-demographic, laboratory test, and comorbidity related risk factors of progression of Chronic Kidney disease CKD of patients. The data will be collected by health professionals after training will be given.

Risk or discomfort: As data were collected from records, there was no harm or discomfort to patients whose cards is reviewed.

Benefit: Although this study has no direct benefit to patients, the result of the study is helpful in the hospital care of CKD patients which helps in determining the risks and support early decision in developing treatment plan

Confidentiality: Trained health professionals would collect data and information will be stored without including the name and address of the patient. The information collected was kept secretly and revealed to principal investigators only.

Contact information

Principal investigator: Zelalem Tolosa

Cell phone: +2519 24194777

Major advisor: Selamawit Serka (PhD)

Question for gathering data from recorded document and background of individual patients.

1) Socio-demographic information of patients.

Medical number of patient ID	Sex of patient (male, female)	Age of patient in year

2) Did the patient have laboratory test during their follow up time (period)?

1: yes 2: no

If yes for # 02, then fill the following table for given variable values recorded?

Medical number of patient ID	Time of follow up	Serum creatinine, (mg/dl)	Urea protein (mg/dl)	Phosphorus (mg/dl)	haemoglobin (Hgb), (gm/dL)	Sodium, (mEq/L)	Potassium, (mmol/L)
0001	1						
	2						
	.						
0002	.						
	.						
	.						
	.						

3) Did the patient have comorbid illness? 1. Yes 2. No If yes for # 03, then which one?

Medical number of patient ID	history of diabetes (0=no, 1= yes)	Hypertension (HTN) (0=no, 1= yes)	Heart disease (0=no, 1= yes)

Annex 2: Ethical clearance.

ሀዋሳ ዩኒቨርሲቲ
ኮሌጅ ተፈጥሮና ቀመር ሳይንስ ኮሌጅ
የምርምር ስነ-ምግባር ክምኒ ኮሚቴ

Hawassa University
College of Natural and Computational Sciences
Research Ethics Review Committee

April 13, 2023

Name of Researcher(s): Zelalem Tolosa, Selamawit Serka (PhD)

Department: **Statistics**

Email: zelalemtolosa@gmail.com

Dear Zelalem

Protocol: **Application of Multistate Markov on the Progression of Chronic Kidney Disease (CKD) at Tikur Anbessa Specialized Hospital, Ethiopia**

RERC reference number: CNCS - RERC 022/23

The Research Ethics Review Committee (RERC) has considered the above mentioned application. The study is approved by a meeting held on 12 April 2023. This is valid for one year from 13 April 2023. To ensure uninterrupted approval of this study beyond the approval expiry date, an application for recertification must be submitted to RERC on the appropriate RERC from 2-3 months before expiry date.

1. Are the objectives, design, and methods are shown in line of scientific approach of research proposal?
   - 1.1 Is the research study conducted with qualified researchers? Yes ☑ No ☐
   - 1.2 Are the objectives of the study ethically achievable? Yes ☑ No ☐
   - 1.2 Appropriate research design is indicated? Yes ☑ No ☐
   - 1.4 Are the proposed research methods ethically sound? Yes ☑ No ☐
2. Is the recruitment of research participants equitable?
   - 2.1 Appropriate recruitment methods stated Yes ☑ No ☐
   - 2.2 Safeguard for vulnerable population Yes ☑ No ☐
3. Community Considerations
   - 3.1 The community will be benefited from this study Yes ☑ No ☐
   - 3.2 Consultation with community Yes ☑ No ☐
4. Are all principles considered
   - 4.1 Respect for persons Yes ☑ No ☐
   - 4.2 Risk benefit and beneficence Yes ☑ No ☐
   - 4.3 Justice Yes ☑ No ☐

Therefore, College of Natural and Computational Sciences of Hawassa University Research Ethics Review Committee (RERC) approved the proposal for implementation.

With regards,

ብዙነህ ይርጋ ገ/ማርያም(ዶ/ር)
በተፈጥሮና ቀመር ሳይንስ ኮሌጅ
የምርምርና ሥነ- ምግባር ጉዳዮች
ኮሚቴ ሊቀመንበር
Bizuneh Yirga G/Mariam(PhD)
College of Natural & Computational Sciences Research Review Committee Chairperson