College of Natural Sciences

Department of Statistics

Modeling the Progression of Chronic Kidney Disease Using the Multi-state Continuous Time Markov Model: A Case of Jimma University Medical Center

By: Tolesa Futasa Begna

A Research Thesis Submitted to Jimma University, College of Natural Sciences, Department of Statistics as a Partial Fulfillment of the Requirements for the Degree of Master of Science (MSc) in Biostatistics

June 2024

Jimma, Ethiopia

Jimma University

College of Natural Sciences

Department of Statistics

Modeling the Progression of Chronic Kidney Disease Using the Multi-state Continuous Time Markov Model: A Case of Jimma University Medical Center

A Research Thesis Submitted to Jimma University, College of Natural Sciences, Department of Statistics as a Partial Fulfillment of the Requirements for the Degree of Master of Science (MSc) in Biostatistics

By: Tolesa Futasa

Advisor: Akalu Banbeta (PhD)

Co-Advisor: Jaleta Abdisa (MSc)

June 2024

Jimma, Ethiopia

Declaration

I, the undersigned author of this thesis, declare that this thesis is the outcome of my own efforts and all sources of materials and information used for writing it have been properly acknowledged. I have submitted this thesis to Jimma University as a partial fulfillment for the Degree of Master of Science in Biostatistics. The library directorate of Jimma University is authorized to archive a copy of this thesis in the university library for accessibility to students and researchers. I declare that this thesis has not been previously submitted to any other institution for academic recognition. Limited use of brief quotations from this thesis is permissible with proper acknowledgment and citation. Any other usage requires explicit consent from the author.

Tolesa Futasa Begna		
Name of candidate	Signature	Date

June 2024

Jimma, Ethiopia

Approval Sheet

This is to certify that the thesis entitled "Modelling the Progression of Chronic Kidney Disease Using the Multi-state Continuous Time Markov: A Case of Jimma University Medical Center" submitted in partial fulfillment of the requirements for the degree of Master of science in Biostatistics, the graduate program of the department of Statistics, and has been carried out by Tolesa Futasa, under our supervision. Therefore, we recommend that the student has fulfilled the requirements and hence hereby can submit the thesis to the department.

Akalu Banbeta (PhD)	_____	_____
Name of Advisor	Signature	Date
Jaleta Abdisa (MSc)	_____	_____
Name of co-Advisor	Signature	Date

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

Tadele Akeba Diriba (PhD)	_____	_____
Name of Internal Examiner	Signature	Date
Leta Lencha Gemechu (PhD)	_____	_____
Name of External Examiner	Signature	Date
Kibrealem Sisay (MSc.)	_____	_____
Name of chair man	Signature	Date

Acknowledgment

First and foremost, I would like to extend my unshared thanks to the almighty God, "who is all knowing, all embracing". Praise be to him for guiding me and giving me the strength and wisdom to accomplish this goal. He has filled my life with goodness, grace, joy and mercy, the sum of which I cannot contain. May his name be forever praised.

I am deeply indebted to my advisor, Dr. Akalu Banbeta (PhD, Ass. professor) for his kind support, advice, guidance and constructive comments through this study, and for sharing his substantial experience to do this thesis in the expected way and giving wonderful personality throughout the time without any reservation time and will always be grateful for that. Likewise, I am sincerely grateful to my co-advisor, Mr. Jaleta Abdisa (MSc), for his valuable comments and suggestion. His comments have benefited me not only on my thesis, but also my daily life. I really appreciate his patience, encouragement and valuable suggestions.

I forward my greatest admire and acknowledgment to the class of 2022/23 "MSc Biostatistics students" for the valuable, friendly and unlimited information we have shared. My sincere appreciation and thanks also go to all the following doctors who taught me during the MSc degree; Dr. Tadele Akeba, Dr. Akalu Banbeta, Dr. Abdisa Gurmessa., Dr. Shiferaw Befkadu, Mr. Abiyot Negash (Ass. Prof), Mr. Kibrealem Sisay (MSc) and the rest staff of the department of Statistics, for their unreserved knowledge sharing and collaboration.

I would like to express my deepest gratitude to the esteemed staff members of Jimma University Medical Center, chronic department and all individuals affiliated with the general record office for providing me with the invaluable data for this study. My special appreciation and acknowledgment go to Dambi Dollo University for offering me the opportunity and financial support during my study at Jimma University.

Finally and most importantly, I take this opportunity to sincerely express my gratitude to my beloved family who are the source of pride and encouragement throughout my life.

Thank you all from the bottom of my heart.

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

ASN:	American Society of Nephrology
BP:	Blood Pressure
BMI:	Body Mass Index
CVD:	Cardio vascular Disease
CKD:	Chronic kidney Disease
CKD-EPI:	Chronic kidney Disease Epidemiology Collaboration
CI:	Confidence Interval
Cr:	Creatinine
GFR:	Glomerular filtration Rate
HR:	Hazard Ratio
HMM:	Hidden Markov Model
HGB:	Hemoglobin
HTN:	Hypertension
JUMC:	Jimma University Medical Center
KDIGO:	Kidney Disease Improving Global Outcome
KDOQI:	Kidney Disease Outcomes Quality Initiative
MST:	Mean Sojourn Time
MDRD:	Modification of Diet in Renal Disease
MSM:	multi-state Model
NCD:	Non-Communicable Disease
NKF:	National Kidney Foundation
RRT:	Renal Replacement Therapy
SCr:	Serum Creatinine
SSA:	Sub-Saharan Africa
SE:	Standard Error
WHO:	World Health Organization

Abstract

Background: Chronic kidney disease (CKD) is a condition that decreases kidney function and can finally leads to renal failure. The progression of CKD is a complex process with different intermediate events. This complex disease processes may be efficiently handled with a multi-state model rather than a simple survival model. Therefore, this study aims to model the progression of CKD using a continuous time multi-state Markov (MSM) model which evaluates the defined stages of the disease.

Methods: This study is a retrospective cohort study of 194 CKD patients who were visited 1506 times in JUMC starting from february 2019 to february 2024. The CKD progression can be analyzed using the multi-state model to assess the effect of different risk factors on the transitions rates between each stages. The MSM model is employed to estimate the transition probabilities, transition intensity and sojourn time along with p-next probabilities between the stages of the underlying disease.

Results: CKD is asymptomatic in its early stages, with only 15.9% of patients in stage 1, while 26.2% of patients are in stage 5. The transition probabilities of patients from any stable stages to worst stage are increasing over time, while the probabilities of remaining in the same stage is decreasing. The average duration patients spent in stages 1, 2, 3, and 4 before transiting to other stage was 4.33, 4.44, 5.79, and 4.57 months respectively. For transition from stage 3 to stage 4, serum creatnine, urea protein, sodium and hemoglobin were associated with risk of progression. Patients with diabetes were 2.58 times more likely to move from stage 1 to stage 2 than those with no diabetes (HR = 2.58, CI = 1.44, 4.62). The progression of CKD was significantly accelerated with predictor variables serum creatnine, urea protein, hemoglobin, sodium, and patient's co-infected with diabetes, hypertension, cardiovascular disease and anemia.

Conclusion: The mean sojourn time together with p-next probabilities in this study provides more perceptible parametric information of the multi-state continuous time model based on markov processes than transition intensity. This study concludes, chronic kidney disease predominantly affects an older, male, rural demographic patients; and is commonly accompanied by HTN, diabetes, CVD and anemia.

Key Words: Multi-state markov model, Chronic kidney disease, End stage renal disease, Glomerular filtration rate, Transition intensity.

1 INTRODUCTION

1.1 Background of the Study

Chronic kidney disease (CKD) is a condition that decreases kidney function and can finally lead to renal failure. It 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. 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. This chronic disease is a leading cause of global loss of life and one of the fastest-rising major causes of death over the globe [1], by affecting over 850 million people worldwide [2]. Many diseases can develop CKD, among them diabetes, high blood pressure, hypertension, hepatitis, HIV/AIDS, congestive heart failure, and sickle cell anemia might typically be present [3].

A higher percentage of people with CKD are older adults, and people who also have other health problems such as diabetes and hypertension [2]. Low- and middle-income countries, which are least prepared to handle its effects, suffer a particularly heavy burden from CKD [2]. In Sub-Sahara Africans (SSA), CKD is highly prevalent affecting young adults aged 20-50 years, and is primarily due to hypertension and glomerular diseases [4, 5]. Currently, in Ethiopia CKD has emerged as a major concern, with a notable impact on males and individuals with a history of kidney infections [6, 7]. Recent cross-sectional study by Kore et al., (2018), indicates the prevalence of CKD in Ethiopia is high and has increased in the last few years to be 12.2% with an increased prevalence of diabetes and hypertension which shows it is becoming one of the public health problems [6].

The increasing prevalence of CKD has emerged as a critical public health issue [2]. However, if kidney issues are caught early and treated, they can be prevented altogether [8]. Due to frequent hospitalizations of patients with renal disease, their physicians are interested in the insights gained from modeling outcomes, such as persistence, and disease progression [9]. Especially modeling the progression of CKD is more important, because of understanding it's progression is crucial for effective management and intervention strategies [10]. However, not much attention is given to the progression of CKD, and few studies have been conducted in our country regarding kidney failure.

The progression of chronic kidney disease is a complex process with different intermediate events of interests [11], that means CKD progresses through multiple states namely; initial, intermediate/transient and absorbing states. It has five phases (stages) based on the estimated glomerular filtration rate (eGFR) according to KDOQI score [12]. The KDOQI score guidelines describe CKD as kidney disease

or decreased kidney function that has been present for more than three months and is verified by abnormalities in blood or urine testing. The five stages of CKD, from 1 through 5, reflect the degree of severity of the disease [12].

In the analysis of time-to-event outcomes, such as progression-free survival, standard methodology for survival analysis is commonly used [13–15]. However, such methodology may not adequately capture the process of any disease as the progression of disease may occupy intermediate events of interest. Thus, multi-state model which is used for modelling time to multiple events of interest offers a valuable approach to investigate how treatments influence different intermediate events that patients may encounter in complex conditions [16]. These types of modelling approach is used particularly in medical applications in which levels of a disease are represented by the states in the model [17–23].

By employing a multi-state markov model, which is a well established and widely used model for describing an individual's movement between a series of states in continuous time, we can explore how a specific treatment influences transitions between different stages of chronic kidney disease, providing a more comprehensive insight into the treatment's effect and the progression of the disease [24]. When examining patient progression through consecutive stages of a chronic condition, multi-state continuous time markov is preferable to discrete-time markov chains because transitions may be slow, have smaller probability, and cannot be correctly characterized in discrete time units.

In this context, the present study aims to apply multistate markov models to investigate the progression of CKD, unraveling the intricate pathways through which the disease evolves. The model estimates the effect of different factors on complex disease processes; and estimates the transition rates between the stages of CKD in continuous time [24–27]. Thus, this study is aimed to identify factor that affect the progression of CKD patients using multi-state continuous-time markov models at Jimma University Medical Center (JUMC).

1.2 Statements of the Problem

Although CKD is a prevalent health issue, and the burden of CKD is still severe and several problems like prognostic factors influencing progression are hindering a comprehensive understanding of CKD progression in Ethiopia. Though different studies were done in Ethiopia on the prevalence of CKD [6,7], as far the investigator's knowledge there is no previous studies conducted on the progression of CKD with multiple transitions among it's states in Ethiopian context using MSM modelling. The transitions of a subject from one state to another can be modeled and the risk factors associated with the transitions can be identified using a multi-state model [11,28]. Hence, this study is motivated to identify prognostic

factors that either accelerate or decelerate the progression and development of CKD.

From the methodological point of view, different studies like Grover, Sabharwal, et al., (2018) and other studies have been done on CKD disease progression using MSM to model the pre-defined patients stage [10, 28, 29]. While these studies provided insights into the investigation of progression patterns of CKD, the inferences drawn about the effects of prognostic factors on the transition intensity were inconsistent. Therefore, to investigate consistent valuable insight into progression of CKD, it is essential to estimate the impact of different prognostic factors on the transition rates of CKD stages; and the duration that the CKD patient spend in each stages before transiting to other stages.

Furthermore, previous studies have primarily used reversible transition probabilities for CKD [29]. However, the progression of CKD stages is naturally irreversible deterioration in renal function [30]. Additionally, different studies conducted on progression of CKD finds only mean sojourn time for each stages to conclude the progression of the disease without finding the probability of transition between the stages [29, 31]. However, the transition probability along with sojourn time gives comprehensive insight into the idea of CKD disease progression [11, 24]; and MSM model is said to be completely specified if the transition probability matrix is completely specified along with the expected sojourn time of each stage [32]. Hence, it is essential to analyze the transitional probability between stages of CKD in addition to mean sojourn time to investigate valuable insight into the progression of CKD.

Generally, since the researcher did not yet find a study conducted on modeling the progression of CKD using the MSM model in Ethiopia and the cases under study is found to be really a predominant issue, it happened to be a reason to conduct this study. Therefore, this study aims to fill this literature gaps as well to model chronic kidney disease progression and determine prognostic factors that affect the transition between different states of chronicity among CKD patients receiving treatment through the application of continuous time MSM process in the case of Jimma University Medical Center.

Thus, this study addressed the following research questions regarding the progression of CKD:

- Which risk factors significantly accelerate or decelerate the progression of CKD?
- What is the probability of a patient moving from one state to another state in continuous time?
- What is the effects of prognostic factors on the transition between different stages of CKD?
- What is the average duration that CKD patients spend in each state?

1.3 Objectives of the Study

1.3.1 General Objective

The general objective of this study is to employ the multi-state continuous time markov processes for modelling the progression of chronic kidney disease among patients admitted to Jimma University Medical Center.

1.3.2 Specific Objectives

- To investigate the factors that either accelerate or decelerate the progression of CKD.
- To analyze the transition probabilities between the stages of CKD using the multistate markov model.
- Estimating the effects of the prognostic factors on the transition between the stages of CKD using the cox-proportional hazard.
- To estimate the mean sojourn times for CKD patients.

1.4 Significance of the Study

The results of this study is utilized to demonstrate the key elements that significantly influence the development of CKD. It may help in providing information for health professionals (clinicians), policy makers and other governmental and non-governmental organizations for facilitating and enhancing the awareness of society about CKD and to give attention for this silent killer disease and minimize the risk of CKD both nationally and in the study area. Especially for clinicians it will be beneficial since the clinicians will be able to see how patients progress from certain states and in which way clinical factors and demographic factors influence the disease progression. Ultimately, the importance of this study will be useful for Academician or Biostatisticians to build a platform to be able to do further research to enable building a forecasting tool to predict the progression of CKD. Additionally, modeling the evolution of the disease will assists in gaining an idea of the anticipated severity of the condition and future researchers may uses the study's findings as a source of information or uses as references.

1.5 Operational Definition

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

Absorbing: An absorbing state is a state from which there is no escape.

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.

Renal: Relating to the kidney.

End-stage renal disease (ESRD): Is the last stage of chronic kidney disease which is a world-wide public health problem and it is associated with adverse outcomes of kidney failure, cardiovascular disease and premature death.

Co-morbidity: Presence of any two or more of morbid condition or disease that occur in one person at the same time.

Homogeneous Markov chain: A Markov chain is called homogeneous if and only if the transition probabilities are independent of the time t.

1.6 Organization of the Study

This thesis is organized into five chapters. Chapter one contains the introduction of the research work, which describe the background of the study, statements of the problems, objectives of the study, significance of the study and operational definition for some terms. Chapter two provides the literature review which includes the overview of CKD, prognostic factors of CKD and the methodological review. Chapter three presents the detailed methodology used for the study whiles chapter four covered the data analysis, interpretation and discussion of the results. The final chapter (chapter five) is devoted to the conclusion and recommendations of the study depending on the results.

2 LITERATURE REVIEW

2.1 Definition and 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. It 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 [33]. 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. The utility of the concept is that recognition of CKD will have implications for the individual and their care [30].

Traditionally, kidney failure or damage has been perceived as the most severe consequence of CKD, characterized by structural or functional abnormalities beyond decreased GFR. Symptoms mainly arise from complications of reduced kidney function and, in severe cases, necessitate treatment through dialysis or transplantation [34]. The early stages of kidney disease often present no symptoms, typically discovered during the evaluation of concurrent conditions and may potentially be reversible. While certain rapidly progressive conditions can lead to kidney failure within months, many diseases progress over decades, and some patients show no advancement over several years of monitoring. In the 2015 global burden of disease study, CKD emerged as a leading cause of global years of life lost and one of the rapidly increasing major causes of death, with a 31.7% rise in overall mortality due to CKD from 2005 to 2015 [1].

The rates of end-stage renal disease incidence and prevalence exhibit significant variation in CKD epidemiology. A substantial majority, over 80%, of individuals undergoing ESKD treatment are concentrated in nations with sizable elderly populations and accessible healthcare [3]. In high-income countries such as the USA and Australia, CKD prevalence often surpasses 11%. Furthermore, the occurrence, prevalence, and progression of CKD demonstrate variations based on ethnicity and socioeconomic status within countries [35]. Individuals in the lowest socioeconomic quartile face a 60% higher risk of developing progressive CKD compared to those in the highest quartile [3]. Notably, the likelihood of both CKD development and disease progression is heightened among Black and Asian populations in the UK, Hispanic communities in the USA, and Indigenous groups in Australia, New Zealand, and Canada [36].

The majority of the world's population lives in Low medium income countries, where most of the population growth is also occurring [37]. The global CKD is heterogeneous, as shown by epidemio-

logical research from low- to middle-income nations in sub-Saharan Africa, Asia, and Latin America. 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% [38], in Egypt 10.6% [39], Tanzania 7% [40], Senegal 6.1% [41], Ghana 4.7% [42] and Kenya 4% [43].

According to 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 [44]. However, Ethiopia is one of the low- and middle-income nations in the sub-Saharan region. 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 [44].

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 [45]. Oromia, with a prevalence of 32.55%, has the greatest prevalence of chronic kidney disease among individuals with other chronic conditions [45]. The stage of CKD, the pooled prevalence of individuals with stage 1 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%) [45].

2.2 Review on Prognostic Factors Affecting CKD Progression

Risk factors associated with CKD progression, CVD and other CKD complications are highly interrelated [36], and hence so is their management. We use the term "CKD treatment and risk modification" to encompass the aim of CKD treatment, which is to impart meaningful beneficial effects on "CKD manifestations" and on "CKD outcomes". CKD manifestations include symptoms and clinical/laboratory abnormalities associated with CKD which confer health implications. These include increased BP, anemia, dyslipidemia, potassium disorders, severe acidosis, decreased fertility, and increased risk of complications of pregnancy [36]. The com-morbidities like CVD, HTN, and diabetes are highly associated with the development and progression of chronic kidney disease [46, 47]

Additionally, serum creatinine, urea, potassium, and sodium in blood are risk factors for CKD progression [48, 49]. Reducing the risk of CKD progression by targeting its underlying pathophysiology may have beneficial effects on a range of CKD manifestations and CKD-associated outcomes, whilst some complications may need specific targeted interventions. Healthcare systems should aim to provide safe and proven cost-effective therapies which achieve CKD treatment and risk modification and to

minimize limitations to access for people with CKD as their disease can substantially impact on quality of life and healthcare system resources. A key goal for healthcare providers should be to identify people at risk and start treatments early in the course of CKD in order to maximize potential benefits.

The study by Caldeira, T.C.M. et al. investigated the temporal trends in comorbidities of obesity with hypertension or diabetes over a 16-year span from 2006 to 2021. Their findings revealed a notable rise in incidence among male and elderly cohorts over time [50]. Additionally, Sun MX et al. conducted a study on the prevalence of various diseases in the Chinese population, assessing different comorbidity patterns and their associated risks of all-cause mortality. Their analysis identified hypertension-chronic kidney disease and hypertension-diabetes-chronic kidney disease as the most prevalent comorbidity pairs, with diabetes-chronic kidney disease carrying the highest mortality risk among them [51].

Sun, P. et al. developed a comprehensive simulation model depicting disease metastasis over the entire life cycle. They established a simulation framework for the life cohort based on a probabilistic model of chronic disease transition and validated its effectiveness using short and medium-term data [52]. Uddin S et al. utilized 24 years of administrative medical data spanning from 1995 to 2018 to investigate the progression of common chronic diseases and their primary comorbidities. By employing network analysis techniques, they ranked the comorbid conditions most likely to precede the onset of various diseases. Their findings indicated a heightened probability of cardiovascular disease and diabetes leading to other comorbidities, a likelihood that also notably increased with advancing time and age [53].

GROVER et al., (2019) applied the continuous time homogeneous MSM based on Markov processes to progression of CKD [28]. They referred a retrospective study of 117 patients suffering from CKD during the period March 2006 to October 2016. The results of their 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. They were also recorded prognostic factors such as gender, age, body mass index, diabetes, hypertension, hemoglobin, urea, serum creatinine, albumin and duration of the disease for each patient and found that 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 [28].

Lintu et al., (2022) in their study that is a study conducting on a MSM model for kidney disease progression by using data obtained from 225 patients to see their kidney disease progression under colistin therapy follow up from January 2016 to December 2017 in Kasturba Hospital, India. The

results of the study showed that a total of 83 (36.89%) patients died in the hospital [29]. The prognostic factors such as gender, hypertension, sepsis, and surgery are significant factors affecting the progression of kidney disease in different stages [29].

The study conducted by Taha & Mohammad (2023) in south-eastern Indian state from Shar Hospital aimed to develop a multi-state model to analyses the progression of chronic kidney disease (CKD) and the effect of risk factors on it [31]. They used retrospective study of 153 CKD patients seen between February 2015 and May 2022. Under their study each patient's gender, age, BMI, diabetes, hypertension, HIV/AIDS, Corona virus, smoking status, urea, serum creatinine, albumin, and disease duration were noted as predictive markers . 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 [31]. They also found that the risk of CKD progression increased with increasing proteinuria and age [31]. 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.

Another study conducted on progression of CKD by Grover, Sabharwal, et al., (2018) 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 [10]. 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 [10]. The probability of a CKD patient with stage 1 of disease misclassified as a patient of stage 2 is 0.211 [10].

2.3 Equations used in Calculating Glomelural Filtration Rate

Globally, creatinine became known for the first time in 1847, and it was suggested as a filtration marker in 1926 [54]. 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 standardization 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 [54].

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 [55]. 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 as a 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 of eGFRcr [55]. 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 [55].

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 [55]. 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. [12]

2.4 Overview of the Multi-state Markov Model

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)) [30]. 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).

A study by Shih et al., (2007) define that most of the time the status of the disease in epidemiology expressed as a dichotomous state i.e., disease and non disease [8]. 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 modeled using multi-state models. These models are helpful for studying the development and natural history of the related disease.

Jackson, et al. (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 [11]. A number of people are observed during the stage $S_i(t)$ 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 (or uni-directional), in which case the transition intensities corresponding to recovery are taken to be 0. In addition to that Grover, Sabharwal, et al., (2018) 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 [18].

Le-Rademacher et al., (2018) in their study of time-to-event outcomes, progression-free survival, or full response are frequent clinical endpoints in oncology studies analyzed using standard methodology for survival analysis [24]. 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 [24]. It's important to balance complexity with usefulness, though, as it is with any statistical models [24].

In a prospective cohort study conducted in the United Kingdom, involving 1,325 patients with CKD stages 3-5 and an average age of 65.1 years, findings revealed that 20% of the participants passed away (equivalent to 9.6 deaths per 100 patient-years) over a median follow-up period of 26 months [56]. Among them, 13% reached the endpoint of Renal Replacement Therapy (RRT) with a frequency of 5.1 events per 100 patient-years. The application of a Markov model for projections indicated a consistent decline in proportions for stages 3 and 4, a continuous rise in mortality and RRT occurrences, and a biphasic trend for stage 5 (non-RRT), featuring an initial plateau in the first 2 years succeeded by a gradual decline thereafter [56].

3 METHODOLOGY

3.1 Study Area

The study was conducted at Jimma University Medical Center (JUMC). Jimma University Medical Center is now the main teaching medical center 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. The various departments, faculties and residents under specialty training in the School of Medicine provide patient care in the center. Currently it is the only teaching and medical center in the southwestern part of the country, providing services for approximately 15,000 inpatient, 160,000 outpatient attendants, 11,000 emergency cases and 4500 deliveries in a year coming to the center from the catchment population of about 15 million people.

3.2 Study Design and Target Population

This study is a retrospective cohort analysis on CKD patients who were receiving follow-up treatment at JUMC. The population of this study was all patients of CKD who had been under follow up starting from february 2019 to february 2024. We consider all patients under follow up as target population. All the data was carefully reviewed from the registration logbook and patients' registration card; the inadequate information counters were checked and excluded from analysis.

3.3 Sample Size

The number of participants was defined after counting the number of patients that met the CKD criteria according to KDOQI and had an electronic medical record in JUMC. No formal sample size calculation was made in this study, since the entire population that met the inclusion criteria was included. The total number of patients considered in the study was 194 individual patients who visited 1506 times in the follow up period.

3.4 Inclusion and Exclusion Criteria

3.4.1 Inclusion Criteria

All CKD patients who had begun their follow-up and had their serum creatinine measured at least once every three months during their follow-up at JUMC were included in the study. The patients registered with appropriate information in the registration card, were considered eligible for the study.

3.4.2 Exclusion Criteria

The CKD patients who were lost from the study without starting any treatment were not included in this study. Additionally, the patients who had incomplete information regarding essential follow up data on clinical parameters on the registration card were not eligible for the study.

3.5 Data Collection Methods

The retrospective cohort data for this study was collected by trained two nurses and the principal investigator. The two nurses were trained regarding the appropriate use of data collection instruments, focusing on uniform interpretation of questions, strict adherence to study criteria, explanation of study objectives, and ensuring the confidentiality of the collected data.

3.6 Data Structure

Data arrangement is an important step in fitting a multi-state model. While documenting the data we have to make sure that all information from a single patient is included under the same unique ID.

Table 3.1: Sample data layout for a multi-state model..

ID	Duration	Age	Gender	Residence	SCr	...	States	Diabetes	HTN	...	Covid-19
1001	0	44	0	1	1.2	...	1	0	0	...	0
1001	3	44	0	1	1.3	...	1	0	0	...	0
1001	6	44	0	1	2.04	...	2	0	1	...	0
1001	8	44	0	1	4.8	...	5	0	1	...	1
1001	9	44	0	1	6.4	...	5	0	1	...	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1194	0	30	1	1	3.08		3	0	1		0
1194	3	30	1	1	3.20		3	0	1		0
1194	6	30	1	1	5.30		5	1	1		0
1194	8	30	1	1	9.50		5	1	1		0
1194	9	30	1	1	13.90		5	1	1		0
1194	10	30	1	1	8.75		5	1	1		0

ID= patient identification number, SCr= Serum creatnine, HTN= Hypertension

3.7 Study Variables

The variables that the researcher is considered in this study were selected based on related literatures. The models consider demographic characteristics, comorbidity conditions and clinical flags to describe

an individual.

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 estimated glomerular filtration rate (eGFR) or level of eGFR. Thus, the main response variable is the change in measurement of eGFR stage of patients in an interval of three months. CKD was diagnosed and classified in to five stages, i.e., stage 1, stage 2, stage 3, stage 4 and stage 5 according to KDOQI clinical practice guideline which are "irreversible" and describe the degree to which kidney function has deteriorated in nature [57], and they are described as follows;

Table 3.2: Chronic Kidney Disease Staging.

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

Many patients who have CKD, progresses through these stages. The stages of CKD are continuous in time and only make forward transition. That means, a patient may make forward transition only among different transient states continuously at any given time, not just at specific time points or intervals. Unfortunately, GFR lab tests are fairly expensive, so doctors typically estimate glomerular filtration rate (eGFR) to define the patient's disease stage [12]. For some of the patients we ourselves computed the eGFR to define the CKD patient's stage using the formula for eGFR; see the formula in Appendix A.

3.7.2 Explanatory Variables

The candidate explanatory variables that may have influence on the prediction of the response variables considered in the study were, demographic factors, clinical risk factors and co-morbidity factors. The list of the variables with their categories are given below.

Table 3.3: Prognostic factors that influence the progression on CKD and their coding

Variables	Coding
Age	Continuous
Gender	0 = female, 1 = male
Residence	0 = rural, 1 = urban
Urea protein (mg/dl)	Continuous
Serum creatinine (mg/dl)	Continuous
Hemoglobin level	Continuous
Potassium (K)	Continuous
Sodium (Na)	Continuous
Diabetes	0 = no, 1 = yes
Hypertension (HTN)	0 = no, 1 = yes
Cardiovascular disease	0 = no, 1 = yes
Smoking status	0 = no, 1 = yes
Heart Disease	0 = no, 1 = yes
HIV/AIDS	0 = no, 1 = yes
Anemia	0 = no, 1 = yes
Covid-19	0 = no, 1 = yes

Urea should range from 42 to 131 milligrams per deciliter (mg/dl), and serum creatinine should be less than 1.4 mg/dl for nephropathy to be diagnosed.

3.8 Progression Procedure of Chronic Kidney Disease

The procedure of CKD progression involves a structured series of stages based on changes in kidney function, typically defined by eGFR and the presence of kidney damage. These stages provide a framework for monitoring and managing the progression of CKD, guiding healthcare providers in assessing the severity of the disease and determining appropriate interventions. 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. The arrows on the diagram 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.

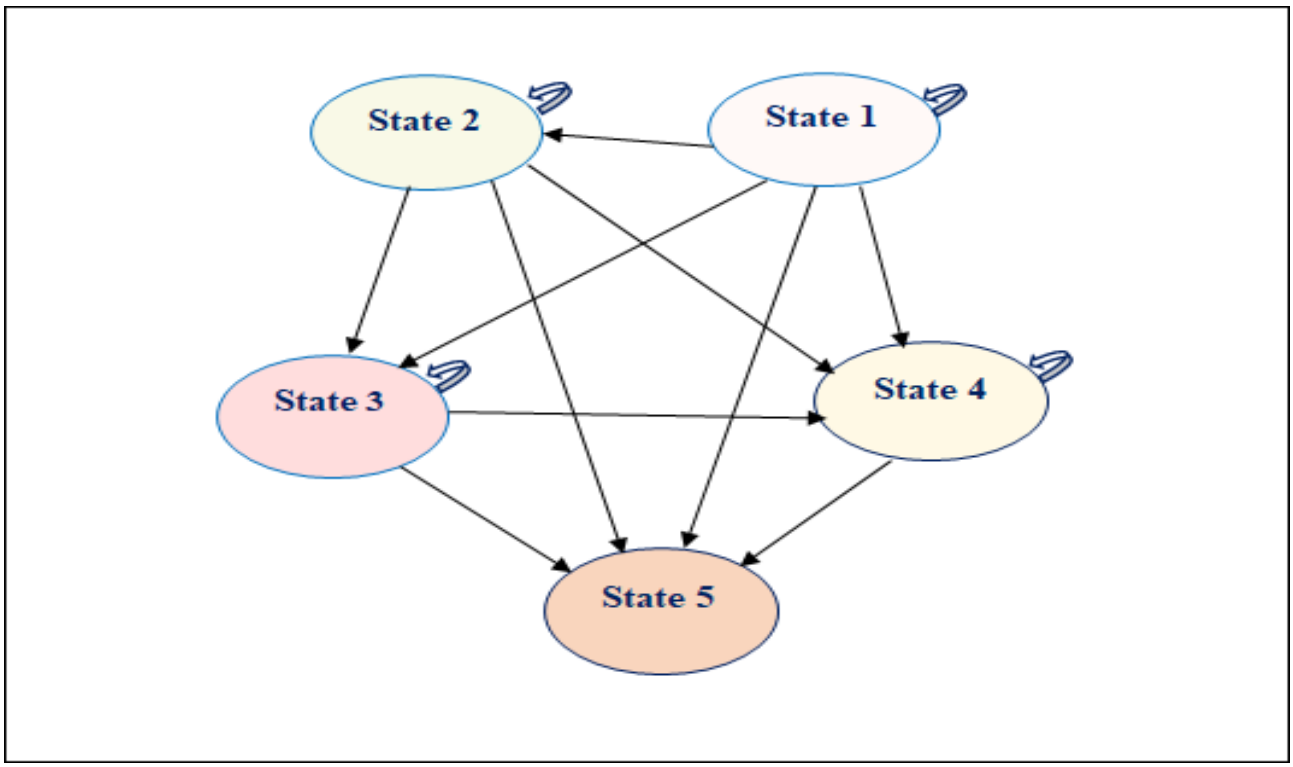

Figure 1: State Transition Diagram

3.9 Method of Data Analysis

The baseline characteristics of the study participants are reported using descriptive statistics. Descriptive analysis is used to describe the sociodemographic and clinical data. For continuous data, the mean and standard deviation (SD) are used, and for categorical data, frequencies and percentages are used. A multistate model is applied to assess the relationship between the clinical conditions of patients and to determine the effects of covariates on individual states (section 3.9.1). The data is first entered into SPSS and then exported to R statistical software version 4.2.2 for further analysis.

3.9.1 The Multi-State Model

3.9.1.1 Introduction

Multi-state models are basically a generalization of survival models. While a survival model considers time to some event, a multi-state model deals with a system of related events. This system of events is described by defining a set of states and viewing events as transitions between states. Many results from survival analysis can be extended to multi-state models since a multi-state model can be seen as a composite of single transitions. One of the most apparent differences is that in a multi-state model, a subject is often under observation for more than one event at a time. Thus, the Multi-state methods model the transitions between multiple states. Under the markov assumption, only transitions

between consecutive visits are included in the likelihood formulation. These methods are relatively new-to-use in CKD literature, and have been used only for the last decade. Some studies directly model transition probabilities using polytomous logistic regression, whereas some studies model transition intensities [58]. A multi-state model based on transition intensity is the focus of this study.

Multi-state models have been thoroughly examined and discussed across various contexts. For instance, Andersen and Keiding (2002) [59], Hougaard (1999) [60] and Putter et al. (2007) [61]. have provided insights into different approaches to multi-state modeling. In essence, a multi-state model represents a stochastic process where individuals occupy discrete states at any given time [60]. These states are categorized into initial, intermediate/transient, and final/absorbing states. The absorbing state serves as the endpoint, from which individuals cannot transition out once reached. Intermediate or transient states lie between the initial and absorbing states [61].

The multi-state model is utilized to represent the progression of time in event data, where individuals initiate in one or more states and may eventually transition to one or multiple absorbing states. Essentially, it captures a process where individuals traverse a sequence of states continuously. In the realm of Multi-State Models (MSMs), the primary objective is to explore the impact of various risk factors and assess the relationship between different predictors and the desired outcome or variable of interest. Transition intensities within MSMs signify the risks associated with transitioning from one state to another, enabling the calculation of the average duration spent in a particular state, as seen in the context of chronic kidney disease (CKD) progression [62].

The evolution of chronic kidney disease (CKD) across five distinct states is represented by a continuous-time homogeneous multi-state Markov model. In a Markovian process, future events are determined solely by the current state. The provided Diagram 1 illustrates this model. While CKD stages progress continuously over time, the state transitions occur abruptly. Given that disease advancement unfolds consistently over time and the transition probabilities between states are influenced by time but not the specific timing of the transition occurrence, we adopt a Markov process-based continuous-time homogeneous multi-state modeling approach.

3.9.1.2 Markov Process

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 realization of the entire history of the process up to

and including time t be: $X_t = x_t, X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \ldots X_1 = x_1$, where $x_t, x_{t-1}, \ldots x_1$ are a sequence of states at different time points. A random process is classified as a markov chain if it satisfies the following condition:

$$P((X_{t+1} = x_{t+1}|X_t = x_t, X_{t-1} = x_{t-1}, \ldots X_1 = x_1) = P(X_{t+1} = x_{t+1}|X_t = x_t), \tag{3.1}$$

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

3.9.1.3 The Transition Probability

The transitions of a markov model describe the likelihood of being in a particular state at some future time, given the present state. A multi-state process is a stochastic process $(X(t), t \in T)$ with finite state space $S = \{1, 2, 3, \ldots, N\}$ where T = [0, t] is the period of observation. Consider $P_{kl}(s,t) = P(X(t) = k|X(s) = l)$ 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) denote the transition probability matrix with transition probabilities $P_{kl}(s,t), k = 1, 2, \ldots, 5$ and $l = 1, 2, \ldots, 5$ 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:

$$\mathbf{P(s,t)} = \begin{pmatrix} p_{11}(s,t) & p_{12}(s,t) & p_{13}(s,t) & p_{14}(s,t) & p_{15}(s,t) \\ p_{21}(s,t) & p_{22}(s,t) & p_{23}(s,t) & p_{24}(s,t) & p_{25}(s,t) \\ p_{31}(s,t) & p_{32}(s,t) & p_{33}(s,t) & p_{34}(s,t) & p_{35}(s,t) \\ p_{41}(s,t) & p_{42}(s,t) & p_{43}(s,t) & p_{44}(s,t) & p_{45}(s,t) \\ p_{51}(s,t) & p_{52}(s,t) & p_{53}(s,t) & p_{54}(s,t) & p_{55}(s,t) \end{pmatrix}. \tag{3.2}$$

Hence the progression of CKD into different states is irreversible and hence $P_{kl}(s,t) = 0$ for all $k > l$. Stage 5 of CKD is an absorbing state and therefore $P_{55}(s,t) = 1$. All the probabilities in the transition probability matrix must be greater than or equal to zero, that is $p_{kl}(s,t) \geq 0, \forall k, l \in \{1, 2, \ldots, 5\}$ and each row must sum to one $\sum_{l=1}^{5} p_{kl} = 1, \forall k, l \in \{1, 2, \ldots, 5\}$.

The above matrix is 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 $P(t)$ is time dependent and is therefore denoted as $P(t)$ instead of P. In time homogeneous markov models, the dependency of t is omitted.

3.9.1.4 Transition Intensity Matrix

This sub-section is primarily concerned with time-homogeneous progressive markov models in which transition intensities are independent of time t. The intensity between two states k and l, can be defined

as the rate of change of the probability 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 $(t, t + \delta t)$ is given by:

$$\lambda_{kl}(t) = \lim_{\delta t \to 0} \frac{P(X_{t,t+\delta t} = l | X_t = k)}{\delta t}, \tag{3.3}$$

where λ_{kl} is the instantaneous risk of moving from state k to state l. Similar to the hazard function in survival models, the transition intensities measure the instantaneous hazard of transition from the current state k to another state l in the small time interval $(t, t + \delta t)$. From the schematic diagram on figure 1 above, the transition intensity matrix for the progression of CKD is given by:

$$\boldsymbol{Q(\lambda)} = \begin{bmatrix} -(\lambda_{12} + \lambda_{13} + \lambda_{14} + \lambda_{15}) & \lambda_{12} & \lambda_{13} & \lambda_{14} & \lambda_{15} \\ 0 & -(\lambda_{23} + \lambda_{24} + \lambda_{25}) & \lambda_{23} & \lambda_{24} & \lambda_{25} \\ 0 & 0 & -(\lambda_{34} + \lambda_{35}) & \lambda_{34} & \lambda_{35} \\ 0 & 0 & 0 & -\lambda_{45} & \lambda_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.4}$$

For a process with an arbitrary number of states, denoted by n, Q is the transition intensity matrix (n rows by n columns). The pace at which a system changes from one state to another can be calculated instantly using the transition intensity.

$Q(\lambda)$ is 5x5 matrix denote transition intensity matrix withλ_{kl} elements of a multi-state process. λ_{kl} denotes the transition intensity or instantaneous rate from k state to l where $k = 1, 2, 3, 4, 5$ and $l = 1, 2, 3, 4, 5$. 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(λ) 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:

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

When $k \in S$ Since state 5 is an absorbing state, there is no chance of leaving it. Each row of the transition matrix has a total value of zero. Finding the unknown transition intensities that optimize the likelihood of a multi-state model is called "fitting" the model [63].

Therefore, the off-diagonals in this matrix are rates at which the subjects move into other states and the diagonal elements are rates at which the subjects remain in their states . The transition probability matrix can be obtained by taking the matrix exponential of the scaled transition intensity matrix

$P(t) = exp(tQ(\lambda))$. This equation can be expressed in-terms of Taylor's theorem, see the formula in Appendix A. The transition intensity matrix $Q(\lambda)$ and transition probability matrix P can be obtained by maximising the likelihood, $L(Q(\lambda))$. After the construction of $L(Q(\lambda))$, the estimated intensity and transition probabilities will maximise $L(Q(\lambda))$ [64]. These all lead to the state occupation probability which is the quantity most of interest in multi-state modeling. It involves combining the hazards and probabilities of relevant transitions.

3.9.1.5 The Sojourn Time

Upon entering state k, the duration spent before transitioning to another state is referred to as the holding time or sojourn time in state k. This metric holds significant importance in disease modeling as it offers insights into the pace of disease progression. Longer sojourn times suggest a slower disease progression, while shorter times indicate a rapid advancement of the disease. The sojourn time of a process X within a subset of states is either an integer-valued random variable for a discrete chain or a real-valued variable for a continuous-time process. In a continuous Markov process residing in state k, the sojourn time is an independent, exponentially distributed random variable with a mean of $\frac{-1}{\lambda_{kk}}$ or rate given by $-\lambda_{kk}$ [65]. The other components in the k^{th} row of the transition intensity matrix are proportional to the probabilities governing transitions to subsequent states following state k. The likelihood of transitioning from state k to state l (i.e., the conditional probability of moving to state l from state k) is determined by:

$$\frac{-\lambda_{kl}}{\lambda_{kk}}. \tag{3.5}$$

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. Here we estimate the mean sojourn time for each state except absorbing state.

3.9.1.6 Multi-state Markov Model

A multi-state markov model describes the process in which a patient moves through a series of states. Kaplan & Meier, (1958) [66] give a thorough introduction into the theory of continuous-time markov chains. A single period of occupancy in state k has an exponential distribution with rate $-\lambda_{kk}$ (or mean $\frac{-1}{\lambda_{kk}}$) in a time-homogeneous continuous-time markov model. The MSM model in continuous time are often used in the progression of diseases. And it is said to be completely specified if the transition intensity between different states are known or the transition probability matrix is completely specified along with the expected sojourn time of each state [32].

3.9.1.7 Homogeneous Continuous Time Markov

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 $l_1, \ldots, l_{i-1}$ before finally entering state l at time t.

For a continuous time, stochastic process $[X(t), t \geq 0]$ whose state space is S, we say it has the markov property if:

$$P(X(t) = j | X(s) = i, X(t_{n-1}) = i_{n-1}, \ldots X(t_1 = i_1)) = P(X(t) = j | X(s) = i), \tag{3.6}$$

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 $i_1, \ldots, i_{n-1}, i, j \in S$.

3.9.2 Model Assumption

3.9.2.1 Markov Model Assumption

Different model assumptions can be made about the dependence of the transition rates on time [67]. Markov property and the homogeneity assumptions are strong assumptions which may lead to biased estimates if violated, consequently it is significant to assess and further examine msm model once it has been fitted. It is also important to get an overall idea of the goodness of fit and to recognize time intervals in which the model may greatly under or overestimate parameters of interest. 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. To put it in simple terms markov assumption simple means that to make the best possible prediction of what happens "tomorrow", we only need to consider what happens "today", as the "past" (yesterday) gives no additional useful information. 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 [68].

3.9.2.2 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 [69]. 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 $P(t)$.

3.9.2.3 Kolmogorov equations

The Kolmogorov equations are used to derive the correlation among the transition intensity matrix Q and the transition probability matrix P. In other words the transition probabilities can be calculated from the intensities by solving the kolmogorov differential equation. The relationship between the transition intensity and probability matrix involves canonical decomposition. The canonical decomposition was discussed by Kalbeisch and Lawless (1985) [68].

The Kolmogorov equations state that:

$$\frac{\partial}{\partial_t} P(t) = P(t)Q(\lambda), \tag{3.7}$$

which yield a unique/closed form solutions for $P(t)$ and condition on $P(0) = l$.

$$P(t) = e^{Q(\lambda)t} = \sum_{r=0}^{\infty} \frac{(Q(\lambda)t)^r}{r!}, \tag{3.8}$$

which is only valid with time homogeneous intensities.

$Q(\lambda)$ is the transition intensity matrix, therefore $p(t)$ can be found from $Q(\lambda)$ using the above kolmogorov equations. The solution for the transition probabilities in terms of the transition intensities can be found using eigenvalue decomposition for $Q(\lambda)$ but the solutions are complicated functions of the intensities and it is only practical to calculate them for simple models with small intensities

3.9.3 Parametric Maximum Likelihood Estimation

The method of maximum likelihood estimation enables the unknown parameters in the model to be estimated. Likelihood function is formed using transition probability interms of transition intensities. The likelihood estimates of model parameters are obtained from the likelihood function. Therefore 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. The 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

$$\boldsymbol{P(s,t) = exp(Q(\lambda)t)}. \tag{3.9}$$

The likelihood function of transition intensities is the product of probability of transition between observed states over all individuals and observation times. Let $S_k(t), k = 1, 2, \ldots 5$ denotes state of

each individual and known for all the times t of the entire period of study. The likelihood function is formed with the transition probabilities.

This likelihood function, $L(Q)$ is given by:

$$L(Q) = \prod_{k} L_k = \prod_{k,l} L_{k,l} = \prod_{k,l} P_{S(t_{kl})S(t_{k,l+1}-t_{k,l})}, \tag{3.10}$$

where $L_{k,l}$ is the element corresponding to $S(t_{kl})^{th}$ row and $S(t_{k,l+1})^{th}$ column of the matrix $P(t)$ evaluated at $t = t_{k,l+1} - t_{kl}$. The likelihood $L(Q)$ is maximized in term of $log(\lambda_{kl})$ therefore the estimates and standard error of λ_{kl} are obtained from $log(\lambda_{kl})$ using optimization technique.

3.9.4 Covariates

In the applications of the MSM model, our interest leis in computing the change in the transition rates in the presence of covariates. Covariates are the explanatory variables that are incorporated to the model through the transition intensities. Once the covariates are incorporated in the model, the interest is not only on the movement of subjects through different states but also on how these covariates influence this movement. The effect of covariate is measured by the regression coefficient. These effects are introduced as covariates in the model via the transition intensities. In multistate models it is the transition intensities that are going to be modeled, since one can directly allow for these covariate dependent intensities in the model.

The generalized hazard ratio is a statistical measure that is commonly used in survival analysis to assess the effect of covariates on the hazard rate or risk of an event occurring. In the context of transition states, the hazard ratio can be used to compare the risk of moving from one state to another (transition intensities) based on different covariates. The proportional hazards model expresses the hazard function for an individual l as a product of a baseline hazard function common for all individuals and a function of the covariate vector. The generalized hazard ratio is typically estimated using a cox proportional hazards model, which can be expressed mathematically as follows:

$$\boldsymbol{\lambda_{kl}[t|Z(t)] = \lambda_{kl}^{(0)} exp[\beta'_{kl} Z(t)]}, \tag{3.11}$$

where:

$\Rightarrow$ $\lambda_{kl}[t|Z(t)]$ represents the hazard rate at time t for a subject with covariate vector Z.

$\Rightarrow$ $\lambda_{kl}^{(0)}$ is the baseline hazard function that represents the hazard rate for a subject with all covariates equal to zero.

$\Rightarrow$ $exp[.]$ is the exponential function that transforms the coefficients into hazard ratios.

$\Rightarrow$ β_{kl} is the vector of regression coefficients, which are of primary interest.

$\Rightarrow$ $Z(t)$ covariate vector at time t.

Time-homogeneous models refer to models whose intensities change with time. Marshall & Jones, (1995) also described the likelihood ratio and Wald tests for selection of covariates and testing hypotheses [70].

If the proportional hazards assumption fails in a Cox proportional hazards model, alternative option is to use a time-varying Cox model, also known as a Cox model with time-dependent covariates extended cox model. This model allows for the effect of covariates to vary over time, which can address violations of the proportional hazard's assumption. The formula for the time-varying Cox model is similar to that of the standard cox model, but with some modifications to account for the time-varying nature of the covariates.

3.9.5 Model Comparison Criteria

The Akaike Information Criterion (AIC) value has been used to make comparisons between the various stochastic models for CKD stage development. The Akaike Information Criterion (AIC) is a measure of how well one statistical model compares to others. It is calculated using the formula

$$\text{AIC} = 2K - 2ln(\hat{L}), \tag{3.12}$$

where K is the total number of free parameters in the model and $\hat{L}$ is the highest value of the likelihood function. Reduced AIC model is favored over other models. The likelihood ratio test provides a second criterion for model selection. It's a way to see how well two different models match the data, with the null model serving as an example of an alternative model. It relies on the likelihood ratio, which shows how much more likely one model is to explain the data than another model. The LR statistic, or log likelihood ratio statistic, is used in hypothesis testing. To calculate the LR statistic:

LR Statistic = -2(ln (reduced model)) - (-2(ln (current model)).

A higher likelihood ratio value in general, the existing model performs better as $(-2lnL)$ increases. A larger value suggests that the present model is preferable to the reduced model, as indicated by the smaller p-value. If there are many observations, the LR statistic will have the form of a chi-square test with K degrees of freedom, where K is the total number of predictors [28].

3.10 Statistical Software

The statistical software we used in our analysis is R statistical software, which is a popular statistical software that can be used for multi-state markov modeling, including for studying the progression of chronic kidney disease (CKD). In R, the researcher can leverage various packages and functions to define different states of CKD progression, model transitions between these states, and perform statistical analyses to understand factors influencing disease progression.

The "msm" package in R software, which was developed by Jackson, (2011) [71], can be used to provides tools for fitting multistate models and allows us to specify transition intensities between CKD states. This package supports the estimation of transition probabilities, model selection, and visualization of results, making it a valuable tool for studying CKD progression using MSM modeling in R. The "msm" function itself implements maximum-likelihood estimation for general multi-state markov models in continuous time. The researcher used this package to estimate parameters such as transition probabilities and hazard ratios, and assess the impact of covariates on CKD progression.

3.11 Ethical Consideration

The need to observe ethics of research was highly taken into consideration in this study. All information gathered in this study whether in physical or digital form, is treated with the highest level of confidentiality where patient's identity is not disclosed anywhere whatsoever and the data is securely stored in a locked cabinet and on a password-protected computer. Patients names were assigned codes, which was used throughout the analysis process, without any personal information attached. Hence privileges of the patients whose information were used were treated with confidentiality.

3.12 Dissemination

The final result of this study will be submitted to Jimma University Medical Center Health Research and Emergency Management Directorate. The summarized results of the present study and evidence based strategy to monitor the progression of chronic kidney disease will be presented for health concerned body. Additionally, this research paper will be made available to the Library Directorate of Jimma University for access to students and researchers who use the University library. It will also published on journals for scientific community and all concerned bodies engaged on reducing the progression of CKD in the study area and in Ethiopia at large.

4 RESULT AND DISCUSSION

A total of 290 medical registration numbers corresponding to CKD patients were identified from the registration logbooks at Jimma University Medical Center, chronic departments. Upon further assessment, only 194 patients met the inclusion criteria for the study and were deemed suitable for the present study. Consequently, the present study is a retrospective cohort study of 194 CKD patients who were undergoing treatment and follow-up care during the period february 2019 to february 2024. In this study, various methods related to continuous time MSM modelling of longitudinal data with an inclination towards its application to progression of CKD based on level of eGFR were addressed.

4.1 Descriptive analysis

The descriptive statistics shown on table 4.1 describes all the demographic, clinical and co-morbidity characteristics of CKD patients by providing the frequency and percentage of categorical variables, and the mean and standard deviations of continuous variables. Of all 194 CKD patients on different therapy, 106(54.6%) were male and the mean age of the patients was 50.8 years, with a standard deviation of 14.9 years (Table 4.1). This means that most of the patients in this dataset were in their older age and majority of them were males. Having percentage of 58.8% large number of the patients were from rural area.

The most common co-morbidities among the patients were hypertension (92.7%), diabetes (90.2%)and anemia (85.1%). While only 37.6% of the patients had been diagnosed with heart disease, 55.6% had some cardiovascular disease. With an incidence rate of 25.8%, covid-19 was the least frequently observed comorbidity. The mean values of the laboratory tests for creatinine, urea, potassium, hemoglobin, and sodium were 4.6 mg/dL, 176.4 mg/dL, 4.6 mmol/L, 9.3 g/dL and 137.1 mmol/L respectively. These values indicate that the patients had varying degrees of renal impairment, uremia, electrolyte imbalance, and anemia, which are common features of CKD. The result reveals serum creatnine and urea protein have 4.58 & 176.38 mean respectively and standard deviation 3.99 & 114.83. This shows that CKD patients in our study have higher serum creatinine and urea protein, leading to various other dangerous diseases.

Table 4.1: Demographic and Clinical Characteristics of CKD patients.

Predictor Variables		**n(%)** or **Mean (SD)**
Gender	Male(%)	106(54.6%)
	Female (%)	88(45.4%)
Residence	Urban (%)	80 (41.2%)
	Rural (%)	114(58.8%)
Age (SD)		50.8 (14.9)
Serum Creatnine (SD)		4.58 (3.99)
Urea Protein (SD)		176.38 (114.83)
Potassium (SD)		4.64 (1.86)
Hemoglobin (SD)		9.27 (1.97)
Sodium (SD)		137.12 (7.50)
Diabetes-yes (%)		175 (90.2%)
Hypertension-yes (%)		180 (92.7%)
Heart Disease-yes (%)		73 (37.6%)
Smoking-yes (%)		52 (26.8%)
Cardiovascular-yes (%)		108 (55.6%)
HIV/AIDS(%)		67 (34.5%)
Anemia-yes (%)		165 (85.1%)
Covid-19 -yes (%)		50 (25.8%)

Source: Jimma University Medical Center, Ethiopia, from February 2019 to February 2024.

The distribution of chronic kidney disease patients across five states based on numbers of visited times during the follow up period is shown on figure 2. Of all 1506 times visited patients, state 2 is the least common state, with only 234 (15.54%) and 240 (15.94%) patients in state 1. State 5 has 395 (26.22%) patients, which is the most common state among all states. 305(20.25% patients are in severe decreasing in eGFR or state 4, while 332(22.0%) times visited patients are in moderate level of eGFR and 395(26.22%) patients are under final states kidney disease damage or end state renal disease (ESRD). The implication of this result is that, CKD is asymptomatic at its earlier stages with 240(15.94%) times visited patients at the kidney damage with normal and most worst at end stage renal disease.

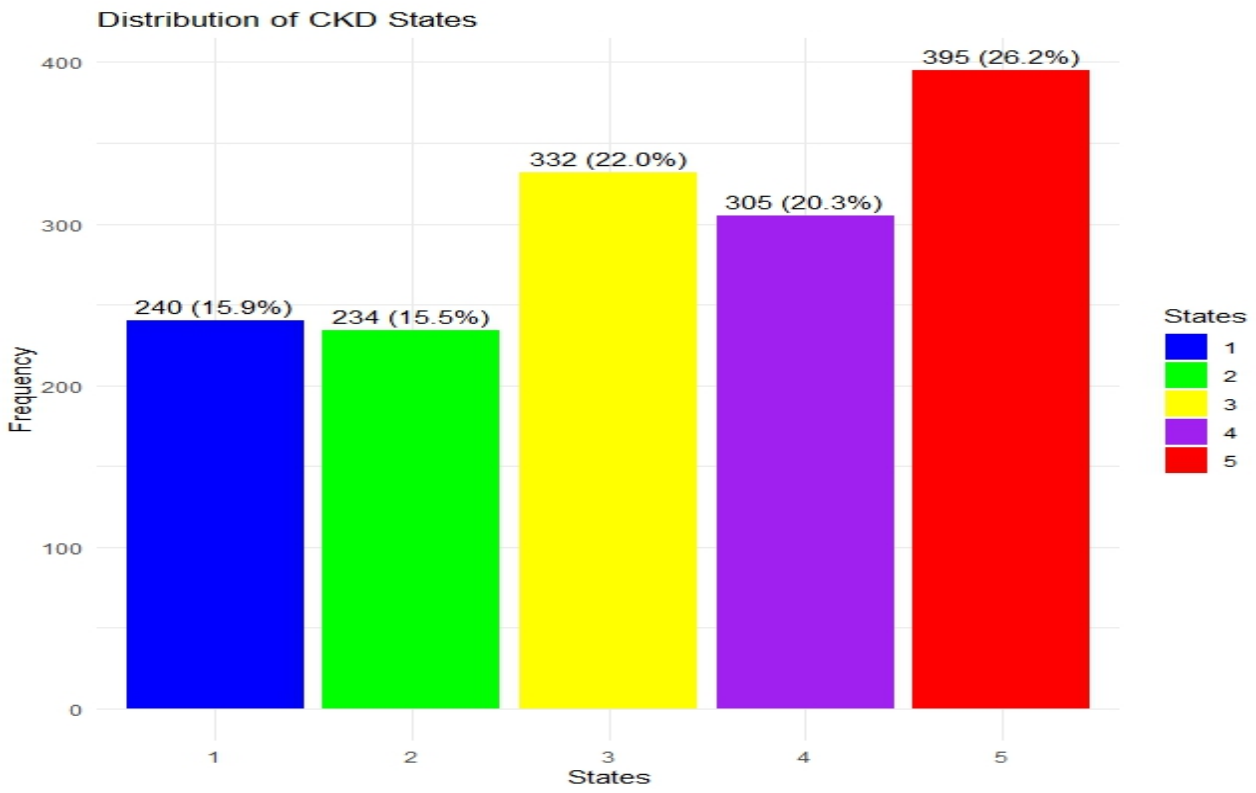

Figure 2: Plot of CKD patients distribution across five states during the study period.

4.2 Multi-State Analysis of Clinical Progression of Chronic Kidney Disease

With the growing access to longitudinal and survival data, it is essential to employ appropriate statistical techniques to analyze longitudinal data and create suitable methods. From a longitudinal data of CKD patients at JUMC, chronic department, the continuous time multi-state markov models have been used to estimate the transitional probability, mean sojourn time along with p-next probability in-addition to forecasted total length of stay in each state and hazard rates at which individuals transits between predetermined states. Additionally this model allows us to identify the prognostic factors which either accelerate or decelerate the progression of chronic kidney disease.

4.2.1 Analysis of Transition Count and Transitional Probability

This study considered that an infected patient can move based on serum creatinine using 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 patients in different state occurs at any time and the number of transition counts for CKD patients is presented based on the severity of their condition.

Moreover, a continuous time-homogeneous MSM model was fitted to the data to evaluate the effectiveness of the treatment by comparing the forward transitions and the result summarizes the multi-state data by counting the transitions between states for each patient over their observation period and reveals that transition counts from state i to j are higher when $i = j$. For instance, the total number of transitions of state 1 of CKD patient to state 1 is 122 for all the patients under study (Table 4.2). It means that there were 122 instances when a patient of state 1 remained in state 1 in his subsequent visits. Reverse (backward) transition is not possible so none of the patients had moved from a higher state to lower state in the subsequent visit. During the follow-up period, 240, 233, 324, 288, and 227 transitions were occurred from states 1, 2, 3, 4, and 5, respectively.

The transitions counts to state 5 from states 1, 2, 3, 4, and 5 were 11, 13, 24, 120, and 127, respectively (Table 4.2). This indicates that as the severity of the disease increases the transition to absorbing state (state 5) is increasing. As CKD progresses slowly, the number of transitions along the diagonal (i.e., patients remaining in the same state) is relatively high. Additionally, Due to the rapid decline in kidney function that characterizes CKD state 4, a disproportionately large number of patients (120) progress from this state to state 5 and 227 CKD patients are got kidney damage in hospital (Table 4.2). Similar interpretation will be done for the rest transition counts. A total of 1312 transitions among states of CKD were occurred during the study period.

Table 4.2: Transition counts

	To					
From	State 1	State 2	State 3	State 4	State 5	Total transitions
State 1	122	68	20	19	11	240
State 2	0	119	83	18	13	233
State 3	0	0	200	100	24	324
State 4	0	0	0	168	120	288
State 5	0	0	0	0	227	227
Total	122	187	303	305	395	1312

Source: Jimma University Medical Center, Ethiopia, from February 2019 to February 2024.

Table 4.3 shows the estimated transition probabilities at time $t = 1$ month. According to the result, a CKD patient in state 1 will continue in state 1 with a probability of 0.794 in 1 month and 0.143, 0.018, 0.032 and 0.012 are the probabilities of him traveling from state 1 to state 2, to state 3, to state 4

and to state 5, respectively. Over the course of 1 month, the CKD patient in state 1 is more likely to stay in state 1 than to progress to higher states (Table 4.3).

Similarly, as presented on the result of transitional count table, the probabilities that patients stay in their respective states are higher in comparison to transiting to other states, which is indicated by 0.794 for state 1, 0.798 for state 2, 0.84 for state 3, 0.803 for state 4 and 1 for state 5 (Table 4.3). The individual patient from state 1, 2, 3 and 4 transits to state 5 with probabilities 0.012, 0.018, 0.021 and 0.197 respectively (Table 4.3). This transitions shows that as the severity of the disease increases (levels of eGFR decreased), the probability transitions to ESRD is increasing. The transition probability for state 5 is exact probability due to state 5 is an absorbing state.

Table 4.3: Estimated Transition Probability at $t = 1$ month

	State 1	State 2	State 3	State 4	State 5
State 1	0.794	0.143	0.018	0.032	0.012
State 2	0.000	0.798	0.167	0.017	0.018
State 3	0.000	0.000	0.841	0.138	0.021
State 4	0.000	0.000	0.000	0.803	0.197
State 5	0.000	0.000	0.000	0.000	1.000

The transition probability matrix over the period of 1 year, 2 years, 3 years, 4 years and 5 years after fitting the continuous time multi-state markov model has been shown in Table 4.4. The probability of transiting to state 5 from state 1 over the years 1, 2, 3, 4 and 5 is estimated to be 0.395, 0.808, 0.959, 0.993 and 0.999, respectively and increasing over time (Table 4.4). Patients in state 3 have a probabilities of 0.121, 0.015, 0.002, 0.000 and 0.000 of staying in the same states over the period of 1, 2, 3, 4 and 5 years respectively, and decreases as the time interval increases. The probability of state 3 moving to state 5 is higher than the probability of moving to state 4, and it increases over time.

Additionally, the probability of state 4 transiting to ESRD (state 5) was higher than all other transitions to state 5 and it is increasing over the continuous time. Similar interpretations can be derived for the other elements of the transition probability matrix over the course of the given years. Generally, as time t increases, the probabilities of the patients transiting to a most worst state is increasing while the probabilities to remain in the same state is decreasing. The probabilities of a CKD patients moving from a higher state to a lower state is zero for all time periods.

Table 4.4: Estimated Transition Probability over different time t (in year).

Transition	Year 1	Year 2	Year 3	Year 4	Year 5
$1 \rightarrow 1$	0.063	0.004	0.001	0.000	0.000
$1 \rightarrow 2$	0.138	0.018	0.002	0.001	0.000
$1 \rightarrow 3$	0.221	0.072	0.012	0.002	0.000
$1 \rightarrow 4$	0.183	0.097	0.025	0.005	0.001
$1 \rightarrow 5$	0.395	0.808	0.959	0.993	0.999
$2 \rightarrow 2$	0.068	0.005	0.001	0.000	0.000
$2 \rightarrow 3$	0.231	0.043	0.006	0.001	0.000
$2 \rightarrow 4$	0.223	0.075	0.015	0.002	0.000
$2 \rightarrow 5$	0.479	0.877	0.979	0.997	0.999
$3 \rightarrow 3$	0.121	0.015	0.002	0.000	0.000
$3 \rightarrow 4$	0.191	0.036	0.005	0.001	0.000
$3 \rightarrow 5$	0.689	0.95	0.993	0.999	1.000
$4 \rightarrow 4$	0.07	0.005	0.001	0.000	0.000
$4 \rightarrow 5$	0.93	0.995	0.999	1.000	1.000
$5 \rightarrow 5$	1.000	1.000	1.000	1.000	1.000

Note: The transition probability corresponding to recovery are zero.
And the value zero (0.000) probability in the table is only approximately.

4.2.2 Transition intensity matrix and Sojourn time

4.2.2.1 Transition intensity matrix

The likelihood estimates of transition intensities between various states of CKD patients in this study are shown in Table 4.5. The transition intensity of moving from state 2 to state 2, i.e. retaining the state 2 is -0.231 whereas the transition intensity of transiting from state 2 to state 3 is 0.204, to state 4 is 0.004 and to state 5 is 0.018. Additionally the transition intensity of transitions to the worst state is increasing, as eGFR level of the patient is decreasing. The elements in each row of the transition intensity matrix sum to zero and off diagonal elements non-negative and the elements in diagonal are negative for all i equal to j. This implies that subjects in those states remain in their respective states

while the off diagonals are rates at which subjects move to other states. The transition intensities are zero for backward movement i.e. transition from a higher state to a lower state is not possible. This proves that the progression of CKD is irreversible.

Table 4.5: Estimated transition intensities.

	State 1	State 2	State 3	State 4	State 5
State 1	-0.231	0.179	0.004	0.039	0.008
State 2	0.000	-0.225	0.204	0.004	0.018
State 3	0.000	0.000	-0.173	0.168	0.005
State 4	0.000	0.000	0.000	-0.219	0.219
State 5	0.000	0.000	0.000	0.000	0.000

4.2.2.2 Mean sojourn time (MST)

Table 4.6 presents the estimated mean sojourn time, and the standard error alongside the 95% confidence interval for the mean sojourn time of each state. Mean sojourn times represent the average duration of a single stay in a particular state, calculated as $-1/\lambda_{kk}$, where λ_{kk} is the k^{th} diagonal element of the estimated transition intensity matrix. From the result, if an individual is in state 2, patient spends 4.44, $[-1/-0.225 = 4.44]$ months in that state before making a forward transition to other states.

Table 4.6: Mean sojourn time in transient states

States	**Estimate (in Month)**	**Standard Error**	**95% CI**
State 1	4.33	0.537	(3.39, 5.52)
State 2	4.44	0.385	(3.75, 5.26)
State 3	5.79	0.456	(4.96, 6.76)
State 4	4.57	0.530	(3.64, 5.74)

The interpretation of the mean sojourn times in states 1, 2, 3 and 4 are that an average stay in the eGFR states are 4.33, 4.44, 5.79 and 4.57 months respectively (Table 4.6). Thus, it is seen that a typical individual just entering a level of eGFR between 30 and 59 (moderate level), can expect to spend around twenty three weeks (5.79 months) at that level before moving to higher sever state

indicated by lower level of eGFR and the value is large compared to the estimates of other states. This proves that there is slow movement of progression of CKD in moderate level (state 3). The standard error of this estimate is 0.456, and the 95% confidence interval is (4.96, 6.76), meaning that we are 95% confident that the true mean sojourn time for state 3 is between 4.96 and 6.76 months. The mean sojourn time for state 1 which is 4.33 months is quite low in comparison to other states. The mean sojourn time for state 4 is 4.57 months which suggests the kidney failure within half year.

4.2.2.3 Forecast of the total length of stay in each state

The researcher need to forecast the total time spent in the good states and the bad states by individuals who are on CKD treatment follow up before ESRD during the study, in-order to understand the life expectancy of the patients. What makes it different from mean sojourn time is that it focuses on the expected duration an individual will spend in a particular state before reaching an endpoint or absorbing state, accounting for subsequent transitions. The estimates of the forecasted total length of stay represents the anticipated time a subject will spend in each transient state during the study period; and is estimated as on Table 4.7.

The results show that each individual is forecasted to spend approximately 4.33 months in state 1, 3.44 months in state 2, 4.18 months in state 3 and finally 4.03 months in state 4 before reaching end point. These results show that CKD patient individuals on treatment are expected to spend more time in kidney damage with normal (state 1) compared to the time spent in the rest states. Again, patient's total length of stay (life expectancies) in states 3 and 4 of the system are approximately 4.2 month and 4.0 months respectively, meaning that, the deterioration in health of a CKD patient is faster in more severe states than in lower states.

Table 4.7: Total length of stay in each state before transiting to absorbing state

States:	State 1	State 2	State 3	State 4	State 5
Total Length of Stay:	4.327	3.446	4.177	4.031	Inf

The graphical representation below illustrates the survival rates of patients in various states of chronic kidney disease (CKD) over time. It is evident that the progression to end-stage renal disease (ESRD) is slower for patients in states 1 and 2, while there is a rapid decline in survival probability around the 3-month mark for patients in state 4, indicating a rapid deterioration in kidney function. Additionally, the graph shows that patients with severe CKD have a 12-month survival probability of around 0.1, contrasting with 0.4 for moderate CKD, 0.6 for mild CKD, and 0.7 for normal CKD.

Furthermore, for individuals with severe CKD, the likelihood of surviving beyond the initial five months of the disease significantly decreases to approximately 0.3.

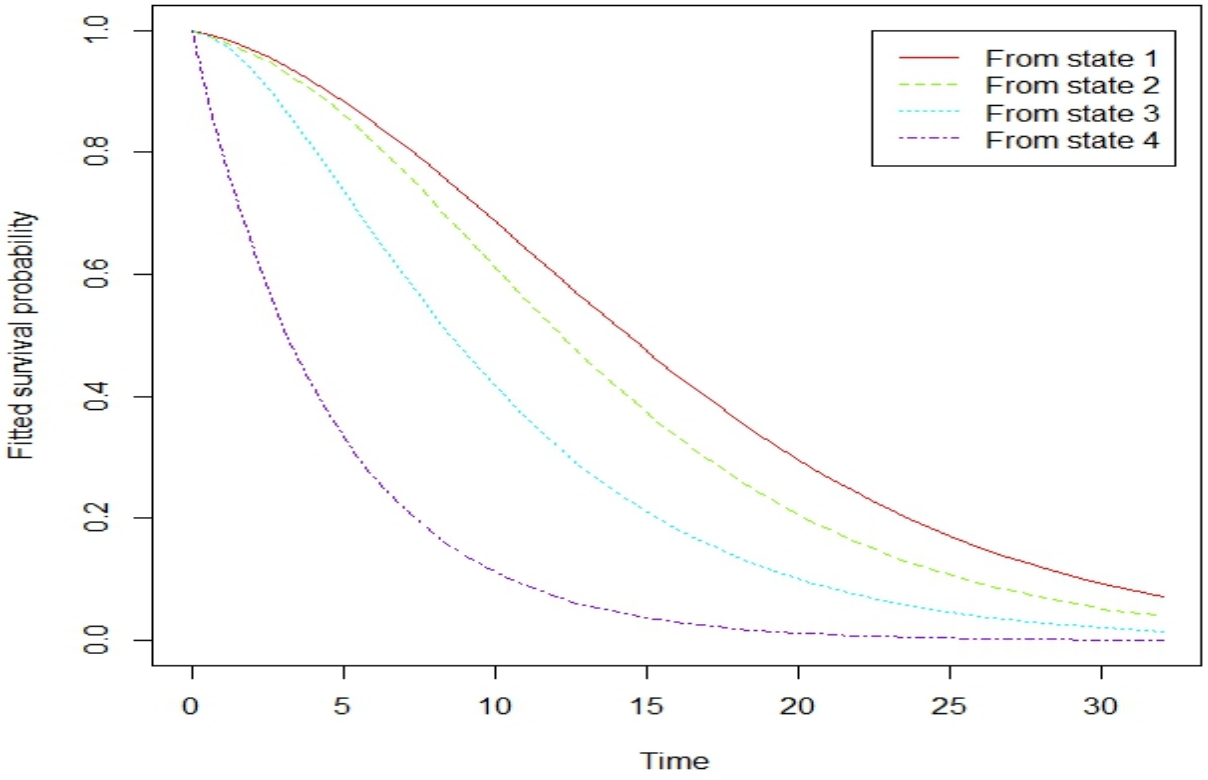

Figure 3: Plot of fitted survival probability for each state

4.2.2.4 Estimation of P-next Probabilities

The result on Table 4.8 shows the matrix of probabilities corresponding to each state k and l. Each probability denotes the probability of transition from state k to the next state l which can be calculated by $\frac{-\lambda_{kl}}{\lambda_{kk}}$ and called P-next probabilities. These probabilities are different from transition probabilities. The P-next probability indicated in the result shows that, the probability from state 1 to next higher state (state 2) is 0.897, 0.905 from state 2 to next state 3, 0.972 from state 3 to next state 4 and exact probability from state 4 to state 5. The probability from state 5 is seen to be zero, because of there is no next higher state above state 5 in this study and state 5 is absorbing states. The result of this analysis indicates that the mean sojourn times together with P-next probabilities provide more perceptible parametric information of multi-state continuous time model based on Markov processes than crude transition intensities λ_{kl}.

Table 4.8: Probability of the transition to next higher state from each state.

	State 1	State 2	State 3	State 4	State 5
State 1	0	0.897	0.019	0.168	0.037
State 2	0	0	0.905	0.017	0.079
State 3	0	0	0	0.972	0.028
State 4	0	0	0	0	1
State 5	0	0	0	0	0

4.2.3 Continuous Time Markov model with Effect of Covariates on Transition Intensities

Multi-state markov analysis in this study was carried out using the msm package in order to learn how different factors influenced the strengths of the transitions. In this section the hazard ratios for each of the covariates; socio-demographic, clinical and co-morbidity variables are estimated and computed.

4.2.3.1 Hazard ratios for socio-demographic variables

The results in table 4.9 shows that the hazard ratios and 95% confidence intervals for socio-demographic variables for different transitions between states of chronic kidney disease (CKD). The states range from 1 to 5, with higher states 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. A hazard ratio less than 1 means that the event is less likely to happen in the first group.

Male patients in state 1 have 8.2% reduced risk of experiencing state 2 compared to females but male patients in state 1 have a higher risk of illness progression to state 3, however the result was not significant at 95% CI as the confidence interval includes 1. For the transition from state 2 to state 5, the hazard ratio for gender is 1.057, which means that male patients are 5.7% more likely to progress to state 5 than female patients, with a 95% confidence interval of [0.047, 23.69] (Table 4.9). This means that we are 95% confident that the true hazard ratio for gender is between 0.532 and 1.585. However, this confidence interval includes 1, which means that the difference between male and female patients is not statistically significant at the 5% level.

For the transition from state 2 to state 5, the hazard ratio for age is 1.06, which means that for every one-year increase in age, the risk of progressing to state 5 from state 2 increases by 6%, with a 95%

confidence interval of (1.001, 1.1375). This confidence interval does not include 1, which means that the effect of age in this transition is statistically significant at 5% level. For the transition from state 3 to state 5, the hazard ratio for residence is 0.543, which means that rural resident patients are 54.3% more likely to progress to state 5 than urban resident patients, with a 95% confidence interval of (0.020, 14.87). However, this confidence interval includes 1, which means that the difference between rural and urban patients is not statistically significant at 5% level of significance.

Table 4.9: Hazard ratios and 95% confidence intervals for socio-demographic variables.

States	Gender	Age	Residence
$1 \rightarrow 1$			
$1 \rightarrow 2$	0.918 (0.532, 1.585)	0.992 (0.977, 1.107)	1.177 (0.623, 2.221)
$1 \rightarrow 3$	1.089 (0.631, 1.880)	0.641 (0.001, 2.113)	0.459 (0.001, 21.97)
$1 \rightarrow 4$	0.952 (0.213, 4.266)	0.986 (0.946, 1.027)	0.849 (0.134, 5.376)
$1 \rightarrow 5$	0.323 (0.006, 1.634)	0.998 (0.944, 1.055)	0.457 (0.017, 12.58)
$2 \rightarrow 2$			
$2 \rightarrow 3$	1.099 (0.775, 1.559)	1.004 (0.992, 1.016)	1.214 (0.837, 1.760)
$2 \rightarrow 4$	1.074 (0.765, 1.509)	0.586 (0.002, 11.01)	0.549 (0.005, 6.643)
$2 \rightarrow 5$	1.057 (0.047, 23.69)	**1.060 (1.001, 1.357)**	0.559 (0.017, 19.06)
$3 \rightarrow 3$			
$3 \rightarrow 4$	0.931 (0.663, 1.307)	1.083 (0.689, 1.703)	1.079 (0.775, 1.502)
$3 \rightarrow 5$	0.053 (0.001, 2.810)	0.529 (0.005, 1.929)	0.543 (0.020, 14.87)
$4 \rightarrow 4$			
$4 \rightarrow 5$	0.897 (0.638, 1.262)	1.066 (0.675, 1.684)	0.958 (0.718, 1.279)

Source: Jimma University Medical Center, Ethiopia, from February 2019 to February 2024.
Bold-phase indicates significance.

4.2.3.2 Hazard ratios for clinical variables

The results in table 4.10 shows the hazard ratios and 95% confidence intervals for the effect of five clinical laboratory variables on the transition of chronic kidney disease (CKD) states for patients at JUMC during their follow up period.

For the transition from state 1 to state 5 , the hazard ratio for serum creatnine is 1.30 with a 95% confidence interval of [0.203, 8.25] (Table 4.10). This means that for every unit increase in creatnine, the risk of progressing from state 1 to state 5 increases by 30%, but this effect is not statistically significant because the confidence interval includes 1. The hazard ratio for potassium is 0.98 with a 95% confidence interval of [0.509, 1.89] in the same transition. This means that for every unit increase in potassium, the risk of progressing from state 1 to state 5 decreases by 2%, but the effect is not statistically significant and the same fashion for another transitions.

It is seen from the result that hemoglobin has a negative or protective effect on all transitions. This means that higher levels of hemoglobin are associated with lower risks of progressing to more advanced stages of CKD in continuous time. In the opposite hemoglobin, variables like serum creatnine, urea protein and sodium in blood have a positive or harmful effect on all transitions. This means that higher levels of serum creatnine, urea protein and sodium in blood are associated with higher risks of progressing to more advanced stages of CKD in continuous time.

Table 4.10: Hazard ratios and 95% confidence intervals for clinical variables.

States	Serum Creatnine	Urea protein	Potassium	Hemoglobin	Sodium
$1 \rightarrow 1$					
$1 \rightarrow 2$	1.16 (0.556, 2.43)	1.02 (0.994, 1.01)	1.04 (0.992, 1.08)	0.93 (0.791, 1.10)	1.01 (0.957, 1.07)
$1 \rightarrow 3$	1.57 (0.314, 7.86)	1.02 (0.987, 1.05)	0.80 (0.244, 2.57)	0.79 (0.443, 1.41)	**1.12 (1.030, 1.21)**
$1 \rightarrow 4$	1.42 (0.301, 6.72)	1.01 (0.984, 1.04)	0.76 (0.310, 1.85)	0.86 (0.580, 1.27)	**1.11 (1.001, 1.23)**
$1 \rightarrow 5$	1.30 (0.203, 8.25)	1.04 (0.932, 1.17)	0.98 (0.509, 1.89)	0.71 (0.460, 1.09)	1.13 (0.991, 1.29)
$2 \rightarrow 2$					
$2 \rightarrow 3$	1.35 (0.814, 2.23)	**1.01 (1.001, 1.02)**	0.92 (0.721, 1.18)	0.92 (0.788, 1.07)	**1.04 (1.007, 1.07)**
$2 \rightarrow 4$	2.12 (0.703, 6.40)	**1.03 (1.004, 1.05)**	0.88 (0.270, 2.84)	0.68 (0.396, 1.18)	1.05 (0.990, 1.11)
$2 \rightarrow 5$	1.86 (0.322,10.74)	1.05 (0.990, 1.12)	0.97 (0.389, 2.44)	0.71 (0.359, 1.41)	1.06 (0.973, 1.16)
$3 \rightarrow 3$					
$3 \rightarrow 4$	**2.21 (1.515, 3.22)**	**1.01 (1.002, 1.01)**	1.01 (0.773, 1.31)	**0.57 (0.475, 0.68)**	**1.02 (1.010, 1.04)**
$3 \rightarrow 5$	1.94 (0.619, 6.09)	1.01 (0.996, 1.03)	1.26 (0.375, 4.21)	0.64 (0.301, 1.35)	**1.05 (1.001, 1.09**)
$4 \rightarrow 4$					
$4 \rightarrow 5$	**2.65 (1.905, 3.70)**	1.00 (0.999,1.02)	0.95 (0.811,1.11)	**0.60 (0.533,0.68)**	**1.04 (1.014,1.06)**

Source: Jimma University Medical Center, Ethiopia, from February 2019 to February 2024.
Bold-phase indicates significance.

Some general patterns can be observed from results that potassium has a negative or protective effect on most transitions, except for the transition from state 1 to state 2 and state 3 to other more severe states (state 4 & 5), where it has a positive or harmful effect. This tells us that potassium has a positive or harmful effect on some transitions and a negative or protective effect on others, but none of them are statistically significant. This means that potassium does not have a clear or consistent effect on the progression of CKD stages.

The most significant results are for serum creatnine which show positive associations with CKD progression for sate 3 to state 4 and state 4 to state 5 CKD transitions and a clear significance was observed for sodium from state 1 to state 2. In this result serum creatnine and sodium have positive associations with CKD progression in all transition. similarly, most significant results are for hemoglobin which show negative associations with CKD progression for sate 3 to state 4 and state 4 to state 5 transitions. Even there are positive associations with CKD progression in all transitions for urea protein, the effects are seen rare (less than 5%).

4.2.3.3 Hazard ratios for co-morbidity variables

The onset of different co-morbid illness may be occurred at different follow up time, this onset of new disease has an impact on the transition from the lower state to higher states. Therefore this impact should be studied and which comorbid has significant impact on which states should have to be identified. According to the result of hazard ratios and 95% CI, five co-morbidity variables; like diabetes, hypertension, heart disease, cardiovascular disease and anemia have significant impacts on occurrence of at-least one transition among the five CKD states, while co-morbidities like HIV/AIDS, smoking status and covid-19 have no significant impacts on the transitions (Table 4.11).

For transition from state 1 to state 2, the hazard ratio for diabetes is 2.58, which means that patients with diabetes are more likely to progress to state 2 from state 1 than patients without diabetes, with a 95% confidence interval of (1.44, 4.62). Patients without diabetes fared better in terms of rates of improvement compared to those with diabetes, and those with diabetes in states 3 and 4 were at increased risk of complications with hazard ratio [HR = 1.58] (95% CI = [1.13, 2.20]) and [HR = 3.28] (95% CI [2.65, 6.49]) respectively (Table 4.11). Heart disease has a positive significant effect on the transition from state 2 to state 3, in which the hazard ratio for heart disease is 1.60, which means that 60% patients with heart disease are more likely to progress to state 3 from state 2 than patients without heart disease, with a 95% confidence interval of (1.05, 2.46).

Table 4.11: Hazard ratios and 95% confidence intervals for co-morbidity variables.

States From →To	Diabetes	HTN	Heart Disease	CVD
1→1				
1→2	**2.58 (1.44, 4.62)**	1.28 (0.70, 2.34)	1.14 (0.61, 2.13)	1.13 (0.55, 2.34)
1→3	2.09 (0.04, 114)	1.84 (0.08, 44.3)	1.28 (0.05, 33.8)	1.18 (0.04,38.35)
1→4	1.54 (0.06, 42.3)	1.81 (0.27, 12.1)	1.18 (0.15, 9.10)	1.45 (0.13, 16.8)
1→5	1.41 (0.04, 50.3)	1.70 (0.22, 12.9)	1.54 (0.23, 10.4)	1.28 (0.09, 18.6)
2→2				
2→3	1.26 (0.88, 1.81)	1.00 (0.64, 1.57)	**1.60 (1.05, 2.46)**	1.23 (0.81, 1.89)
2→4	1.79 (0.19, 17.4)	2.02 (0.10, 39.6)	1.21 (0.03, 59.6)	1.33 (0.05, 36.4)
2→5	1.59 (0.36, 6.90)	1.53 (0.19, 12.1)	1.20 (0.09, 16.7)	1.32 (0.18, 9.76)
3→3				
3→4	**1.58 (1.13, 2.20)**	**1.94 (1.21, 3.10)**	1.23 (0.78, 1.94)	**1.72 (1.28, 2.31)**
3→5	1.21 (0.16, 8.93)	1.36 (0.14, 13.5)	1.21 (0.06, 23.5)	1.61 (0.14,1 8.5)
4→4				
4→5	**3.28 (2.13, 5.1)**	**4.14 (2.65, 6.49)**	1.41 (1.00, 1.97)	**2.58 (1.81, 3.66)**

States From →To	Smoking status	Anemia	HIV/AIDS	Covid-19
1→1				
1→2	0.534 (0.239, 1.19)	1.31 (0.49, 3.52)	0.86 (0.415, 1.77)	1.05 (0.509, 2.17)
1→ 3	1.024 (0.026, 40.8)	2.13 (0.04, 104)	1.20 (0.031, 46.7)	1.06 (0.022, 51.2)
1→ 4	0.949 (0.085, 10.6)	1.44 (0.04, 46.6)	1.21 (0.114, 12.8)	1.07 (0.105, 10.9)
1→ 5	1.201 (0.122, 11.8)	1.35 (0.03, 54.0)	0.88 (0.068, 11.4)	1.10 (0.107, 11.3)
2→ 2				
2→ 3	1.55 (0.824, 2.92)	**1.5 (1.03, 2.18)**	1.29 (0.819, 2.03)	1.13 (0.582, 2.17)
2→ 4	1.35 (0.002, 891.1)	2.08 (0.23, 18.8)	1.41 (0.043, 46.7)	1.04 (0.004, 274)
2→ 5	1.41 (0.06, 33.04)	1.62 (0.35, 7.46)	1.17 (0.160, 8.52)	1.07 (0.071, 16.1)
3→ 3				
3→ 4	1.273 (0.836, 1.93)	**2.02 (1.33, 3.07)**	1.20 (0.861, 1.67)	0.99 (0.692, 1.42)
3→ 5	1.209 (0.064, 22.51)	1.34 (0.25, 7.23)	1.45 (0.152, 13.8)	1.07 (0.146, 7.92)
4→ 4				
4→ 5	1.15 (0.817, 1.61)	**3.83 (2.8, 5.26)**	1.11 (0.784, 1.58)	0.59 (0.430, 0.82)

Source: Jimma University Medical Center, Ethiopia, from February 2019 to February 2024.
HTN= hypertension, CVD= cardiovascular disease
Bold-phase indicates significance.

The co-morbidity variable hypertension (HTN) was a risk factor and its effect on the progression from state 3 to state 4 and from state 4 to state 5 was considerable with corresponding hazard ratio [HR=1.94] (95% CI=[1.21, 3.10]) and [HR=4.14] (95% CI=[2.65, 6.49]). That means Hypertension was a significant risk factor that had a noticeable impact on the shift from states 3 and 4 to their the next worst states. Based on the transition from sever state (state 4) to dialysis state all HTN, diabetes, CVD and anemia were a risk factors and their effects on the progression are founded to be [HR=3.28] (95%CI=[2.13, 5.1]), [HR=4.14] (95%CI=[2.65, 6.49]), [HR=2.58] (95%CI=[1.81, 3.66]) & [HR=3.83] (95%CI=[2.8, 5.26]) which are significant at 95%. This shows that there is a significant kidney failure risk for patients with hypertension, diabetes, cardiovascular disease and anemia.

The result on table 4.11 also shows that 59%, 53%, 20%, 32%, 41%, 62%, 17% and 7% CKD patients with comorbidity diabetes, HTN, heart disease, CVD, smoking status, anemia, HIV/AIDS and covid-19 respectively are more likely to progress to state 5 from state 2 than patients without those comorbid illness. However, none of them are statistically significant. Additionally, non of smoking status, HIV/AIDS and covid-19 have statistically significant effects in all transitions, this means that those comorbidity illness does not have a clear or consistent effect on the progression of CKD among different states in continuous time.

In the same manner, figure 6 (for socio demographic variables), figure 7 (for clinical laboratory variables) and figure 8 (for com-morbidity illness variables) on the appendix B, reflects the effects of those prognostic factor on the prograssion of chronic kidney disease (CKD) by either accelerating or decelerating the progression between different states of the underlying disease.

4.2.4 Model comparison

A continuous-time multi-state Markov model for the effects of covariates is fitted as shown in table 4.12. Identification of covariates that have a significant effect is done by entering each covariate one after the other and performing the AIC and likelihood ratio test in comparison to the model without covariates. The AIC is a measure of the relative quality of a model, with lower values indicating better fit. The 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. The p-value is the probability of obtaining a result at least as extreme as the observed one, under the null hypothesis that there is no difference between the models. Firstly, the researcher have fitted the model with the univariate covariate then the combination of three or more significant covariates and higher number of significant covariates has been considered by combining variables that exhibit similar characteristics or behavior to improve the interpretability of the model and makes logical sense in understanding the present study.

Table 4.12: Likelihood Ratio Test Statistics, AIC and P-Values of the Selected Models.

Models	AIC	Likelihood ratio test	Df	P-value
without covariate	2185	-	-	-
Gender as a covariate	2198	3.6	8	0.893
Residence as a covariate	2208	-2.3	10	1
Age as a covariate	2203	2.5	10	0.991
SCr as a covariate	2139	66.6	10	< 0.0001
Potassium as a covariate	2215	-9.26	10	1
Urea protein as a covariate	2202	43	10	< 0.0001
Hemoglobin as a covariate	2183	42.9	10	< 0.0001
sodium as a covariate	2171	34.2	10	0.0002
Diabetes as a covariate	2155	50.9	10	0.00024
HTN as a covariate	2148	57	10	< 0.0001
Heart Disease as a covariate	2203	1.99	10	0.996
Smoking as a covariate	2199	6.14	10	0.804
CVD as a covariate	2186	19.7	10	0.032
HIV/AIDS as a covariate	2205	-0.04	10	1
Anemic as a covariate	2158	47.6	10	0.00073
COVID as a covariate	2207	-2.04	10	1.00
SCr+Na+Urea+K as a covariate	2140	106.46	40	0.00221
Diabetes+HTN+CVD as a covariates	**2135**	**110.06**	**30**	**< 0.0001**
Diabetes+HTN+CVD+Anemic	2141	104.8	40	< 0.0001
All significant variables as a covariates	2158	101	80	< 0.0001
All variables as a covariates	2176	113.6	158	0.0511

Source: Jimma University Medical Center, Ethiopia, from February 2019 to February 2024.
Age=Baseline age of patients, SCr=Serum creatnine, Na=sodium, K=potassium, HTN=hypertension, CVD=cardiovascular disease
Bold phase indicates the best model.

On comparing the likelihood ratio test statistic of 66.6 corresponding to a covariate Serum creatnine with 10 degrees of freedom, it is observed that model with the covariate Serum creatnine fits the data significantly better than the base model having no covariates as p-value is less than 0.0001. On the other hand, covariates gender, residence, age, potassium, heart disease, smoking status, HIV/AIDS and covid-19, are not a significant prognostic factors as their p-values are more than 0.05. The factors serum creatinine, hemoglobin, urea protein, sodium, diabetes, hypertension, cardiovascular disease and anemia are significant prognostic factors for the progression of CKD, as the models based on these covariates have a p-values less than 0.05. Similar interpretation can be given for the other models with combinations of prognostic variables as well.

Based on the criteria of model selection, the best model is the one that has the lowest AIC, highest likelihood ratio test, and lowest p-value. Hence, the best model is the one with the combination of Diabetes, hypertension (HTN) and cardiovascular disease (CVD) as a covariate, which has an AIC of 2135, a likelihood ratio test of 110.06, a df of 30, and a p-value of $p < 0.0001$ (Table 4.12). This means that this model fits the data better than any other models, and that those co-morbidities are a significant predictors of the progression of chronic kidney disease. The hazard ratios and 95% CI for the covariates HTN, diabetes and CVD associated with the best model as a multivariate covariate have been shown in Table 5.1, see the result on appendix C.

4.2.5 Model fit assessment

4.2.5.1 Goodness of fit test

The assessment of goodness-of-fit is a necessary approach to check whether the model fits well. Using the fitted time-continuous multi-state Markov model, the percentage prevalence were plotted to compare the expected values with the observed values. Figure 4 indicates the observed and expected percentages of baseline model prevalence in each state against time (the goodness of baseline model was evaluated). It is seen that the model fits well for states 1, 2 and 3. However, there is some misfit with respect to state 4, where the expected prevalence underestimate the observed data in between time=8 up to time = 15 months. The plots further show an increasing at a high rate on state 5 percentage

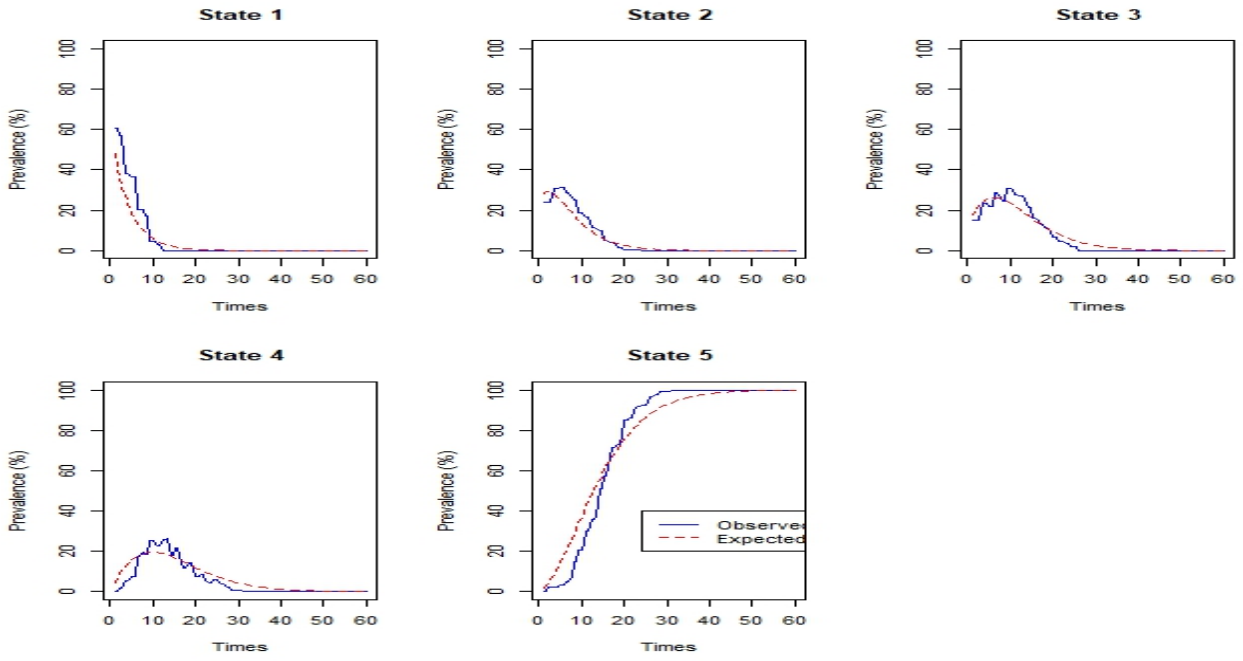

Figure 4: Plot of observed (solid lines) and expected (dashed lines) percentages of prevalence in each state against time according to the baseline model

prevalence with the fitted model underestimating the observed data after time = 15 months. There

is a sharp decrease on state 1 percentage prevalence up to time = 10 months and then after stable. The percentage prevalence for the state 2 is decreasing and after time = 20 months, the percentage prevalence is stable.

Figure 5 bellow confirms that the inclusion of covariates on the model improves the fitness of the model, since the expected prevalence is now perfectly closer to the observed prevalence for all states than the model without covariates. Additionally, the result from Kaplan Meier plot on Appendix D, reflects that the inclusion of covariates Diabetes, HTN and CVD on the model improves the fitness of our continuous time multi-state markov model.

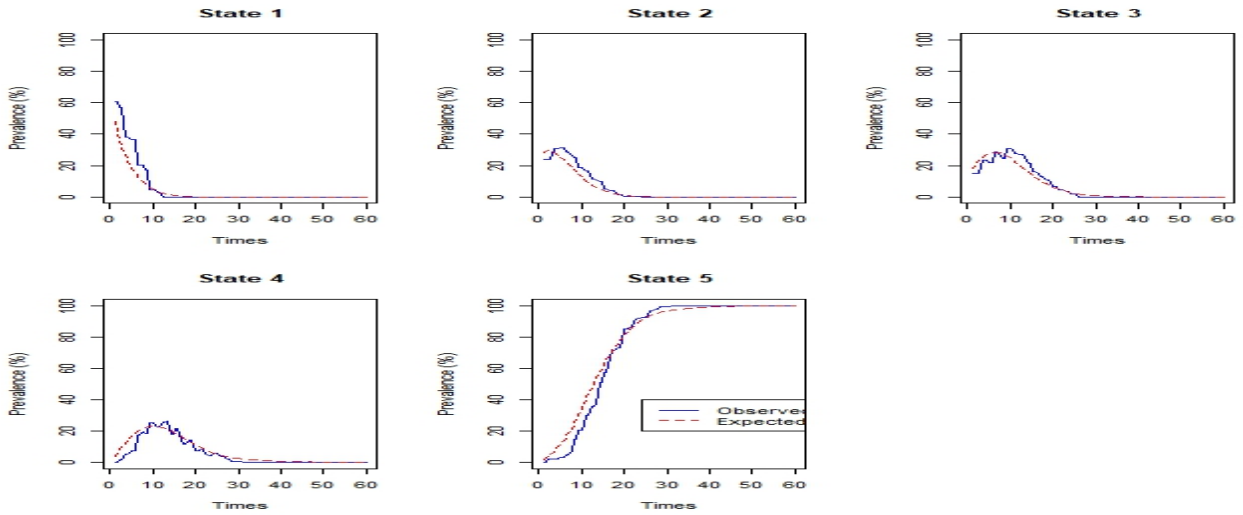

Figure 5: Plot of observed (solid lines) and expected (dashed lines) percentages of prevalence in each state against time for the model with HTN+Diabetes+CVD covariates

4.3 Discussion

In this thesis, the multi-state Markov model was used to describe the trajectory of CKD individual patients under follow up from february 2019 to february 2024 in JUMC so as to aid the understanding of CKD progression and discover factors that influence kidney deterioration. This study examined the characteristics of 194 CKD patients at JUMC. Among CKD patients admitted to JUMC between february 2019 to february 2024, the prevalence of the disease is higher in men than in women, with 106(54.6%) of the patients being male. This finding is consistent with previous studies [49, 50]. Regarding the CKD stages, CKD is asymptomatic in its early stages, with only 15.9% of patients in stage 1, while 26.2% of patients are in stage 5. The mean age of the study participants is 50.8 years, which is consistent with the fact that people in older age are at a higher risk of developing CKD [5].

The study findings also depict that an increase in serum creatinine, urea protein, potassium, and sodium in an individual's blood is associated with the development of CKD, which is consistent with existing

literature [48, 49]. However, the hemoglobin (HGB) level was found to be low in CKD patients in the dataset of our analysis. This might be explained by the fact that the patients' low level of HGB as a result of the blood removal during dialysis, which often leads to the development of anemia [72]. Hypertension and diabetes were the most prevalent co-morbidities among the patients included in the study, which aligns with the results of other studies [4, 38, 46]. Cardiovascular disease and anemia were also prevalent, although less frequent than hypertension and diabetes. These associations of CVD and anemia with development of CKD are supported with existing literature that highlights the link between these conditions and CKD [46, 47].

The MSM analysis of this study states that, the number of transitions from state i to state i in their subsequent visits are more than moving to next higher state. That means, the majority of patients remained in the same state during the study period. This suggests that the implemented treatments may be effective in stabilizing CKD progression for some patients. And it reflects that, the result highlight an essential fact that people are more concerned about their health due to the high number of transitions in the same state, which is agreed with the previous study by Grover, Sabharwal et al., (2018) [10] and Dalip et al., (2022) [73] and contradict with the result of Lintu et al (2022) [29]. This contradiction with the result of Lintu et al (2022) may be due-to this previous study uses reverse transition which is incorrect in modelling the progression of chronic kidney disease, since CKD is irreversible in nature.

Additionally, this study reveals a concerning trend of increasing transition probability to later states (particularly state 5) with worsening initial CKD state. Notably, patients in state 4 had a significantly higher transition probability (0.197) to state 5 compared to earlier states. This higher transition probability to end-stage renal disease (state 5) from state 4 of CKD patient indicates the ultimate failure of kidney function. These findings are in line with previous studies demonstrating a progressive decline in kidney function and increased risk of ESRD with advancing CKD states [10, 28, 29, 31].

As time increases the conditional probability that a patient transit from state 1 to state 2, from state 2 to state 3, from state 3 to state 4 and from state 4 to state 5 after two years (24 months) later is 0.018, 0.043, 0.036 and 0.995 respectively. The probabilities are very small indicating that as time increases the conditional probabilities of transiting to the next state is very small, except the transition from state 4 to state 5 which is a transition to most worst state (absorbing state) and indicating that the probability of transiting to most worst state is increasing with increasing time, which is supported the results of similar study in India by Grover et al, (2019) [28]. Similarly, the result of this study revealed that as time increases the probability of remaining in the same state is decreasing. This finding is in agreement with the result of previous studies by Lintu et al., (2022) and that of Grover, Sabhawal et al (2018) using hidden markov models (HMMs) [10, 29].

The diagonal entries of the transition intensity matrix are negative which indicates the negation to the instantaneous risk of moving to other state. The mean sojourn time for state 3 is 5.79 months. It means that a patient of state 3 spends on an average of 5.79 months in state 3 before moving to state 4, while it spends about 4.2 months to transit to the absorbing state (a result from total length of stay). The mean sojourn time for stage 4 is 4.57 months. This reveals the fact that the progression of disease is slow in moderate states as compared to sever states. Comparable study in India, by Lintu et al. (2022), estimated the sojourn time for state four as , 1.40 years (16.8 months) [29]. By comparing the results of this study with the study by Lintu et al, the researcher can understand that CKD rapidly progressing in JUMC (Ethiopia) than India, this rapid progress may be due to the prevalence of CKD is higher in developing countries and in Sub-Saharan Africa [4].

The researcher also provided graphical analysis for this study. The graphical results in this analysis shows that, as follow up time increases the survival rate of CKD patients were decreasing. Specifically, the graph of survival probability curve between expected survival probability and time reveals the sharp decline in the survivability of state 4 patients.

The results from Cox proportional hazards model of the effect of clinical covariates on transition intensity indicates that serum creatinine, urea protein, and sodium levels in the blood have positive or harmful effects on CKD progression, with higher levels associated with an increased risk of advancing to more severe states. Conversely, higher hemoglobin levels demonstrate a protective effect, reducing the likelihood of progression to advanced CKD stages. These findings are consistent with previous study that have shown that serum creatinine, urea protein, and sodium in blood are risk factors for CKD progression [48].

The most significant results were observed for serum creatinine, which showed positive associations with CKD progression for the transitions from state 3 to state 4 and state 4 to state 5. Sodium also showed a significant positive association with CKD progression for the transition from state 1 to state 2. Hemoglobin showed significant negative associations with CKD progression for the transitions from state 3 to state 4 and state 4 to state 5. These finding highlights the importance of monitoring and managing these factors in patients with CKD to prevent disease progression.

The finding from Cox proportional hazards model in this study also highlights the significant impact of co-morbidities on CKD progression at JUMC. According to the result of the hazards ratio, patients with diabetes were more likely to progress from moderate to sever CKD states and sever to ESRD. The study revealed a heightened risk of complications for patients with diabetes in later CKD states (state 3 and 4), suggesting an increased likelihood of progression to more severe states compared to patients without diabetes. This underscores the importance of managing diabetes alongside CKD to slow

disease progression. This finding aligns with previous study by Taha and Mohammed (2023) using data from military hospital in Iraq [31] and other studies using hidden markov modeling [10, 28, 53].

Accordingly, the results of this study suggests that hypertension emerged as a major risk factor for transitioning between later CKD states (state 3 to 4 and state 4 to 5). The hazard ratios between these transition indicates a substantial influence on disease progression. This aligns with established knowledge about the detrimental effects of uncontrolled hypertension on kidney function and effects of hypertension on CKD [10, 28, 29, 31]. The study also identified a significant association between anemia and CKD progression, particularly in later states. Future research could explore the underlying mechanisms linking these co-morbidities to CKD advancement.

Overall, the presence of hypertension, diabetes, cardiovascular disease, and anemia poses a significant risk of kidney failure (progression to end-stage renal disease) among co-morbidity variables. This study clearly indicates that patients with those conditions are at a higher risk of experiencing kidney failure, emphasizing the need for targeted interventions and management strategies for individuals with these risk factors. Considering the studies result of higher risk of experiencing kidney failure, the kidney dialysis (transplantation) leads to higher economic burdens, since patients have to undergoing hemodialysis and the cost of hemodialysis is higher in Ethiopia and patients are referred to other country for kidney transplantation.

A continuous-time multi-state markov modelling conducted in this study also employed model comparison techniques, to evaluate the effectiveness of different covariates in predicting the progression of CKD. The model is fitted with and without covariates and comparison of the models is done using the likelihood ratio test and AIC. From the results presented, it is evident that the model incorporating a combination of diabetes, hypertension, and cardiovascular disease as covariates emerges as the best-performing model based on multiple criteria. This signifies that the inclusion of these comorbidities as predictors significantly enhances the model's ability to explain the progression of CKD. In support of the selected model, various references have highlighted the role of diabetes, hypertension, and cardiovascular disease in the context of CKD progression. For instance, a study by Grover, Sabharwal et al. (2018) demonstrated the synergistic impact of Diabetes and HTN on CKD progression, underscoring the importance of considering these comorbidities in predictive models using HMMs [10].

Moreover, the inclusion of Cardiovascular Disease as a covariate aligns with findings from studies by Uddin et al. (2023) and Webster et al. (2017), emphasizing the link between CVD and CKD progression [36, 53]. These references provide support for the significance of CVD in predicting CKD outcomes and further validate its inclusion in the selected model. Finally, by integrating these key comorbidities into the predictive model, clinicians can better identify at-risk individuals and tailor

interventions to mitigate the progression of CKD effectively.

4.4 Limitation of the Study

Some limitations were addressed regarding this study. First, the patient records were inadequately kept, lacking essential follow-up data on clinical parameters. Due to this deficiency, certain patients had to be excluded from the current study. Secondly, state 3A and 3B of kidney disease in the present study were merged as state 3 due to the smaller sample size in our analysis. Additionally there is shortage of literature on application of multi-state markov modelling on the progression of chronic kidney disease, specially on the progression of CKD specific to Ethiopian context using application of the same model.

5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study concludes that the conditional probabilities of CKD patients from any good state to worst state are increasing over time. CKD progresses slowly at first but quickly in later stages and most of the patients who moved to worst stage were patients at the sever stage. Similarly, the result of mean sojourn time showed that patients with eGFR level between 30 and 59 is estimated to be larger, suggesting that there is slow movement of progression of CKD in moderate level (state 3).

Chronic kidney disease in this study predominantly affects an older, male, rural demographic; and is commonly accompanied by hypertension, diabetes, anemia, and cardiovascular disease. The laboratory tests in this study reveals significant renal impairment and related complications, which are crucial for diagnosing and staging the disease. The significant prognostic and accelerating factors associated with higher risk of deterioration from stable state to worst state are lower HGB level, higher creatnine level, large amount of sodium and urea, and patient's co-infected with diabetes, HTN, CVD and anemia.

Finally, the result of this study concludes that the mean sojourn time together with p-next probabilities provide more perceptible parametric information of multi-state continuous time model based on markov processes than the crude transition intensity. Our finding reaffirms the importance of initiating therapy at an early stage of the disease, as this could lead to the development of new treatment strategies and save more lives.

5.2 Recommendation

Based on the study findings, the researcher recommends the following points to the concerned body.

- Develop a CKD screening system: Implementing effective screening system, potentially utilizing blood and urine tests, facilitate early detection and intervention to prevent disease progression.
- The probability of patients transiting to a worst state is increasing with time, so public health education and other symposiums should be organized, to increase awareness of patients about the progression of CKD.
- The model incorporating a combination of diabetes, hypertension and cardiovascular disease is a best model for for predicting a future progression of CKD. Hence, a multi-disciplinary approach that integrates management of those co-morbidities is crucial to slow the disease progression.

- Since the the applied multi-state continuous time Markov model in this study is able to handle a complex disease process of CKD progression with different intermediate events, fitting MSM model is recommended for researches to any types of disease with intermediate events.

- Future research can explore incorporating additional factors such as socio-economic status, educational level, surgery and lifestyle choices into multi-state markov model to gain a more comprehensive understanding of CKD progression patterns; and this research can be expanded further by modeling the progression of CKD from several cohort data across different hospitals in Ethiopia using multi-state markov model.

References

[1] H. Wang, M. Naghavi, C. Allen, R. M. Barber, Z. A. Bhutta, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, M. M. Coates, *et al.*, "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015," *The lancet*, vol. 388, no. 10053, pp. 1459–1544, 2016.

[2] C. P. Kovesdy, "Epidemiology of chronic kidney disease: an update 2022," *Kidney International Supplements*, vol. 12, no. 1, pp. 7–11, 2022.

[3] V. Jha, G. Garcia-Garcia, K. Iseki, Z. Li, S. Naicker, B. Plattner, R. Saran, A. Y.-M. Wang, and C.-W. Yang, "Chronic kidney disease: global dimension and perspectives," *The Lancet*, vol. 382, no. 9888, pp. 260–272, 2013.

[4] S. Naicker, "Integrated management: chronic kidney disease, diabetes mellitus, hypertension," *African Journal of Nephrology*, vol. 16, no. 1, pp. 6–13, 2013.

[5] W. M. McClellan and N. R. Powe, "Introduction to the proceedings of a centers for disease control and prevention expert panel workshop: Developing a comprehensive public health strategy for preventing the development, progression, and complications of ckd," *American Journal of Kidney Diseases*, vol. 53, no. 3, pp. S1–S3, 2009.

[6] C. Kore, A. Tadesse, B. Teshome, K. Daniel, A. Kassa, and D. Ayalew, "The magnitude of chronic kidney disease and its risk factors at zewditu memorial hospital, addis ababa, ethiopia," *J Nephrol Ther*, vol. 8, no. 3, p. 313, 2018.

[7] T. Fiseha and Z. Tamir, "Prevalence and awareness of chronic kidney disease among adult diabetic outpatients in northeast ethiopia," *BMC nephrology*, vol. 21, no. 1, pp. 1–7, 2020.

[8] H.-C. Shih, P. Chou, C.-M. Liu, and T.-H. Tung, "Estimation of progression of multi-state chronic disease using the markov model and prevalence pool concept," *BMC Medical Informatics and Decision Making*, vol. 7, no. 1, pp. 1–12, 2007.

[9] A. K. Bello, I. G. Okpechi, M. A. Osman, Y. Cho, H. Htay, V. Jha, M. Wainstein, and D. W. Johnson, "Epidemiology of haemodialysis outcomes," *Nature Reviews Nephrology*, vol. 18, no. 6, pp. 378–395, 2022.

[10] G. Grover, A. Sabharwal, S. Kumar, and A. K. Thakur, "On the estimation of misclassification probabilities of chronic kidney disease using continuous time hidden markov models," *Journal of Nephropharmacology*, vol. 8, no. 1, pp. e07–e07, 2018.

[11] C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy, and E. Couto, "Multistate markov models for disease progression with classification error," *Journal of the Royal Statistical Society Series D: The Statistician*, vol. 52, no. 2, pp. 193–209, 2003.

[12] L. A. Inker, N. D. Eneanya, J. Coresh, H. Tighiouart, D. Wang, Y. Sang, D. C. Crews, A. Doria, M. M. Estrella, M. Froissart, *et al.*, "New creatinine-and cystatin c–based equations to estimate gfr without race," *New England Journal of Medicine*, vol. 385, no. 19, pp. 1737–1749, 2021.

[13] J. Le-Rademacher and X. Wang, "Time-to-event data: An overview and analysis considerations,"

[14] M. Hussein, G. Muleta, D. Seyoum, D. Kifle, and D. Bedada, "survival analysis of patients with end stage renal disease the case of adama hospital, ethiopia," *Clin Med Res*, vol. 6, no. 6, p. 201, 2017.

[15] S. G. Workie, T. A. Zewale, G. T. Wassie, M. A. Belew, and E. D. Abeje, "Survival and predictors of mortality among chronic kidney disease patients on hemodialysis in amhara region, ethiopia, 2021," *BMC nephrology*, vol. 23, no. 1, p. 193, 2022.

[16] K. Bakunina, H. Putter, J. Versluis, E. A. Koster, B. van der Holt, M. G. Manz, D. A. Breems, B. T. Gjertsen, J. Cloos, P. J. Valk, *et al.*, "The added value of multi-state modelling in a randomized controlled trial: The hovon 102 study re-analyzed," *Cancer medicine*, vol. 11, no. 3, pp. 630–640, 2022.

[17] W. G. Jeong, H. Choi, K. J. Chae, and J. Kim, "Prognosis and recurrence patterns in patients with early stage lung cancer: a multi-state model approach," *Translational Lung Cancer Research*, vol. 11, no. 7, p. 1279, 2022.

[18] G. Grover, P. K. Swain, K. Goel, and V. Singh, "Multistate markov modelling for disease progression of breast cancer patients based on ca15-3 marker," *Thailand Statistician*, vol. 16, no. 2, pp. 129–139, 2018.

[19] G. Grover, A. K. Gadpayle, and P. K. Swain, "A multistate markov model based on cd4 cell count for hiv/aids patients on antiretroviral therapy (art)," *International Journal of Statistics in Medical Research*, vol. 2, no. 2, p. 144, 2013.

[20] D. Seyoum, S. Wondayew, K. Sisay, and M. Tadesse, "Predicting aids disease progression using longitudinal cd4 count among adult hiv/aids patients in southwest ethiopia: Application of semi-markov process,"

[21] S. K. Mengesha, G. A. Gebremedhn, T. Ferede, and C. Atsmegiorgis, "Application of multi-state semi-markov models on hiv/aids disease progression," *i-Manager's Journal on Mathematics*, vol. 7, no. 3, p. 30, 2018.

[22] Z. Matsena Zingoni, T. F. Chirwa, J. Todd, and E. Musenge, "Hiv disease progression among antiretroviral therapy patients in zimbabwe: a multistate markov model," *Frontiers in public health*, vol. 7, p. 326, 2019.

[23] M. Bozorgnezha, S. Nodehi, M. Keyvan, and S. Azizi, "Factors affecting the intensity of nursesâĂŹ lower limbsâĂŹ varicose veins in the hospitals of north khorasan, iran, in 2016," *Journal of Biostatistics and Epidemiology*, vol. 4, no. 1, pp. 36–46, 2018.

[24] J. G. Le-Rademacher, R. A. Peterson, T. M. Therneau, B. L. Sanford, R. M. Stone, and S. J. Mandrekar, "Application of multi-state models in cancer clinical trials," *Clinical Trials*, vol. 15, no. 5, pp. 489–498, 2018.

[25] P. K. Andersen, "Multistate models in survival analysis: a study of nephropathy and mortality in diabetes," *Statistics in medicine*, vol. 7, no. 6, pp. 661–670, 1988.

[26] P. K. Andersen, S. Esbjerg, and T. I. Sørensen, "Multi-state models for bleeding episodes and mortality in liver cirrhosis," *Statistics in medicine*, vol. 19, no. 4, pp. 587–599, 2000.

[27] L. D. Amorim and J. Cai, "Modelling recurrent events: a tutorial for analysis in epidemiology," *International journal of epidemiology*, vol. 44, no. 1, pp. 324–333, 2015.

[28] G. Grover, A. Sabharwal, S. Kumar, and A. K. Thakur, "A multi-state markov model for the progression of chronic kidney disease.," *Turkiye Klinikleri Journal of Biostatistics*, vol. 11, no. 1, 2019.

[29] M. Lintu, K. Shreyas, and A. Kamath, "A multi-state model for kidney disease progression," *Clinical Epidemiology and Global Health*, vol. 13, p. 100946, 2022a.

[30] A. Milik and E. Hrynkiewicz, "On translation of ld, il and sfc given according to iec-61131 for hardware synthesis of reconfigurable logic controller," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 4477–4483, 2014.

[31] A. A. Taha and M. A. Mohammad, "A multistate model for the analyzing of chronic kidney disease progression with detecting risk factors effect," *resmilitaris*, vol. 13, no. 2, pp. 3126–3136, 2023.

[32] C. G. Cassandras and S. Lafortune, *Introduction to discrete event systems*. Springer, 2008.

[33] M. A. Tonelli, C. Wanner, A. Cass, A. X. Garg, H. Holdaas, A. G. Jardine, L. Jiang, F. Kronenberg, R. S. Parekh, T. Shoji, *et al.*, "Introduction: The case for updating and context," *Kidney international supplements*, vol. 3, no. 3, pp. 266–267, 2013.

[34] J. E. Scherberich, R. Fünfstück, and K. G. Naber, "Urinary tract infections in patients with renal insufficiency and dialysis–epidemiology, pathogenesis, clinical symptoms, diagnosis and treatment," *GMS Infectious Diseases*, vol. 9, 2021.

[35] M. P. Hossain, E. C. Goyder, J. E. Rigby, and M. El Nahas, "Ckd and poverty: a growing global challenge," *American Journal of Kidney Diseases*, vol. 53, no. 1, pp. 166–174, 2009.

[36] A. C. Webster, E. V. Nagler, R. L. Morton, and P. Masson, "Chronic kidney disease," *The lancet*, vol. 389, no. 10075, pp. 1238–1252, 2017.

[37] J. W. Stanifer, A. Muiru, T. H. Jafar, and U. D. Patel, "Chronic kidney disease in low-and middle-income countries," *Nephrology Dialysis Transplantation*, vol. 31, no. 6, pp. 868–874, 2016.

[38] E. K. Sumaili, J.-M. Krzesinski, C. V. Zinga, E. P. Cohen, P. Delanaye, S. M. Munyanga, and N. M. Nseka, "Prevalence of chronic kidney disease in kinshasa: results of a pilot study from the democratic republic of congo," *Nephrology Dialysis Transplantation*, vol. 24, no. 1, pp. 117–122, 2009.

[39] Y. M. Farag and E. El-Sayed, "Global dialysis perspective: Egypt," *Kidney360*, vol. 3, no. 7, pp. 1263–1268, 2022.

[40] J. W. Stanifer, V. Maro, J. Egger, F. Karia, N. Thielman, E. L. Turner, D. Shimbi, H. Kilaweh, O. Matemu, and U. D. Patel, "The epidemiology of chronic kidney disease in northern tanzania: a population-based survey," *PloS one*, vol. 10, no. 4, p. e0124506, 2015.

[41] S. M. Seck, D. Doupa, L. Guéye, and C. A. Dia, "Epidemiology of chronic kidney disease in northern region of senegal: a communitybased study in 2012," *Pan African medical journal*, vol. 18, no. 1, 2014.

[42] J. B. Eastwood, S. M. Kerry, J. Plange-Rhule, F. B. Micah, S. Antwi, F. G. Boa, D. Banerjee, and F. P. Cappuccio, "Assessment of gfr by four methods in adults in ashanti, ghana: the need for an egfr equation for lean african populations," *Nephrology Dialysis Transplantation*, vol. 25, no. 7, pp. 2178–2187, 2010.

[43] J. W. Stanifer, B. Jing, S. Tolan, N. Helmke, R. Mukerjee, S. Naicker, and U. Patel, "The epidemiology of chronic kidney disease in sub-saharan africa: a systematic review and meta-analysis," *The Lancet Global Health*, vol. 2, no. 3, pp. e174–e181, 2014.

[44] W. H. Organization *et al.*, "Noncommunicable diseases: progress monitor 2022," 2022.

[45] Z. Animaw, G. Walle Ayehu, and H. Abdu, "Prevalence of chronic kidney disease and associated factors among patients with chronic illness in ethiopia: A systematic review and meta-analysis," *SAGE Open Medicine*, vol. 10, p. 20503121221089442, 2022.

[46] W.-C. Lee, Y.-T. Lee, L.-C. Li, H.-Y. Ng, W.-H. Kuo, P.-T. Lin, Y.-C. Liao, T. T.-Y. Chiou, and C.-T. Lee, "The number of comorbidities predicts renal outcomes in patients with stage 3–5 chronic kidney disease," *Journal of clinical medicine*, vol. 7, no. 12, p. 493, 2018.

[47] R. Z. Alicic, M. T. Rooney, and K. R. Tuttle, "Diabetic kidney disease: challenges, progress, and possibilities," *Clinical journal of the American Society of Nephrology*, vol. 12, no. 12, pp. 2032–2045, 2017.

[48] L. Malmgren, "Kidney function during ageing and its association with bone mass, fracture and mortality," 2020.

[49] N. Amin, R. T. Mahmood, M. J. Asad, M. Zafar, and A. M. Raja, "Evaluating urea and creatinine levels in chronic renal failure pre and post dialysis: a prospective study," *Journal of cardiovascular disease*, vol. 2, no. 2, pp. 1–4, 2014.

[50] T. C. Caldeira, T. M. Sousa, M. M. Soares, I. P. Veiga, L. E. Silva, and R. M. Claro, "Temporal trend of multimorbidity of noncommunicable diseases among brazilian adults, 2006–2021," *Obesities*, vol. 3, no. 1, pp. 76–85, 2023.

[51] M. Sun, Q. Wen, H. Tu, S. Li, X. Feng, S. Wang, and X. Wu, "Associations between multimorbidity patterns of 4 chronic diseases and physical activity with all-cause mortality," *Zhonghua liu Xing Bing xue za zhi= Zhonghua Liuxingbingxue Zazhi*, vol. 43, no. 12, pp. 1952–1958, 2022.

[52] P. Sun, S. Lao, D. Du, J. Peng, and X. Yang, "Construction of life-cycle simulation framework of chronic diseases and their comorbidities based on population cohort," *Algorithms*, vol. 15, no. 5, p. 167, 2022.

[53] S. Uddin, S. Wang, A. Khan, and H. Lu, "Comorbidity progression patterns of major chronic diseases: The impact of age, gender and time-window," *Chronic Illness*, vol. 19, no. 2, pp. 304–313, 2023.

[54] A. S. Levey and L. A. Inker, "Assessment of glomerular filtration rate in health and disease: a state of the art review," *Clinical Pharmacology & Therapeutics*, vol. 102, no. 3, pp. 405–419, 2017.

[55] L. A. Inker and S. Titan, "Measurement and estimation of gfr for use in clinical practice: core curriculum 2021," *American Journal of Kidney Diseases*, vol. 78, no. 5, pp. 736–749, 2021.

[56] R. A. Hoefield, P. A. Kalra, P. Baker, B. Lane, J. P. New, D. J. O'Donoghue, R. N. Foley, and R. J. Middleton, "Factors associated with kidney disease progression and mortality in a referred ckd population," *American journal of kidney diseases*, vol. 56, no. 6, pp. 1072–1081, 2010.

[57] A. S. Levey, J. Coresh, K. Bolton, B. Culleton, K. S. Harvey, T. A. Ikizler, C. A. Johnson, A. Kausz, P. L. Kimmel, J. Kusek, *et al.*, "K/doqi clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification," *American Journal of Kidney Diseases*, vol. 39, no. 2 SUPPL. 1, pp. i–ii+, 2002.

[58] S. L. Tyas, J. C. Salazar, D. A. Snowdon, M. F. Desrosiers, K. P. Riley, M. S. Mendiondo, and R. J. Kryscio, "Transitions to mild cognitive impairments, dementia, and death: findings from the nun study," *American journal of epidemiology*, vol. 165, no. 11, pp. 1231–1238, 2007.

[59] P. K. Andersen and N. Keiding, "Multi-state models for event history analysis," *Statistical methods in medical research*, vol. 11, no. 2, pp. 91–115, 2002.

[60] P. Hougaard, "Multi-state models: a review," *Lifetime data analysis*, vol. 5, pp. 239–264, 1999.

[61] H. Putter, M. Fiocco, and R. B. Geskus, "Tutorial in biostatistics: competing risks and multi-state models," *Statistics in medicine*, vol. 26, no. 11, pp. 2389–2430, 2007.

[62] T. J. Mafu, *Modelling of multi-state panel data: the importance of the model assumptions*. PhD thesis, Stellenbosch: Stellenbosch University, 2014.

[63] G. Manzini, T. J. Ettrich, M. Kremer, M. Kornmann, D. Henne-Bruns, D. Eikema, P. Schlattmann, and L. C. de Wreede, "Advantages of a multi-state approach in surgical research: how intermediate events and risk factor profile affect the prognosis of a patient with locally advanced rectal cancer," *BMC medical research methodology*, vol. 18, no. 1, pp. 1–11, 2018.

[64] C. Cong, *Statistical Analysis and Modeling of Breast Cancer and Lung Cancer*. University of South Florida, 2010.

[65] D. Gillings, "Introduction to stochastic processes," 1974.

[66] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," in *Breakthroughs in Statistics: Methodology and Distribution*, pp. 319–337, Springer, 1958.

[67] L. F. M. Machado and J. Roca-Pardiñas, "p3state. msm: Analyzing survival data from an illness-death model," *Journal of statistical software*, vol. 38, pp. 1–18, 2011.

[68] J. Kalbfleisch and J. F. Lawless, "The analysis of panel data under a markov assumption," *Journal of the american statistical association*, vol. 80, no. 392, pp. 863–871, 1985.

[69] A. Zare, M. Mahmoodi, K. Mohammad, H. Zeraati, M. Hosseini, and K. H. Naieni, "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*, vol. 15, no. 1, pp. 441–447, 2014.

[70] G. Marshall Rivera and R. H. Jones, "Multi-state models and diabetic retinopathy," 1995.

[71] C. Jackson, "Multi-state models for panel data: the msm package for r," *Journal of statistical software*, vol. 38, pp. 1–28, 2011.

[72] A. Karaboyas, H. Morgenstern, S. Waechter, N. L. Fleischer, R. Vanholder, S. H. Jacobson, M. M. Sood, D. E. Schaubel, M. Inaba, R. L. Pisoni, *et al.*, "Low hemoglobin at hemodialysis initiation: an international study of anemia management and mortality in the early dialysis period," *Clinical kidney journal*, vol. 13, no. 3, pp. 425–433, 2020.

[73] D. Singh *et al.*, "A multistate markov model describing the progression of various deteriorating stages of chronic kidney disease," *JP Journal of Biostatistics*, vol. 21, pp. 11–28, 2022.

Appendixes

Appendix A: Different Formulas

1. A formula for the estimation of glomelural filtration rate.

   GFR is measured using the 2021 CKD-EPI Creatinine equation as:-

$$\mathrm{eGFR_{cr}} = 142 \times \min\left(\frac{Scr}{k}\right)^{\alpha} \times \max\left(\frac{Scr}{k}\right)^{-1.200} \times 0.99380^{age} \times 1.012 sex, \tag{5.1}$$

   where:

   $\Rightarrow$ $\mathrm{eGFR_{cr}}$ denotes estimated glomerular filtration rates measured in ml/min/1.73 m^2

   $\Rightarrow$ SCr denotes standardized serum creatinine measured in mg/dl

   $\Rightarrow$ sex is coded as male=0 and female=1

   $\Rightarrow$ k is 0.7 for females or 0.9 for males

   $\Rightarrow$ α is -0.241 for females or -0.302 for males

   $\Rightarrow$ min$(Scr/k, 1)$ is the minimum of Scr/k or 1.0

   $\Rightarrow$ max$(Scr/k, 1)$ is the maximum of Scr/k or 1.0.

2. The matrix exponential of scaled transition intensity using Taylor's Theorem:

$$exp(C) = 1 + \frac{C^2}{2!} + \frac{C^3}{3!} + \frac{C^4}{4!} + \dots. \tag{5.2}$$

   Where $C = tQ(\lambda)$

Appendix B: Graphical plots for univariate covariates

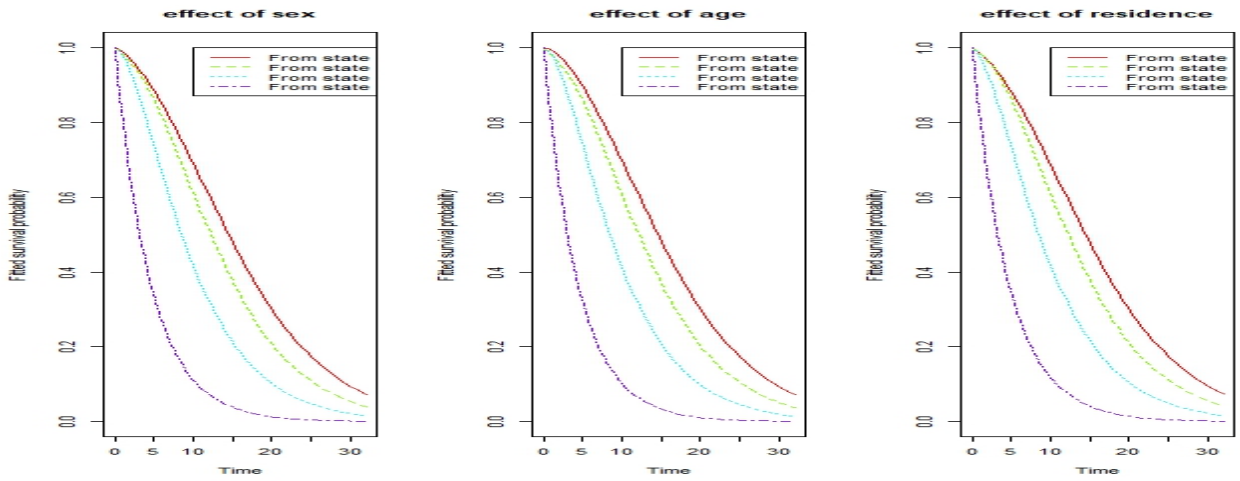

Figure 6: Plot of effect of socio-demographic variables on fitted survival probability for each state

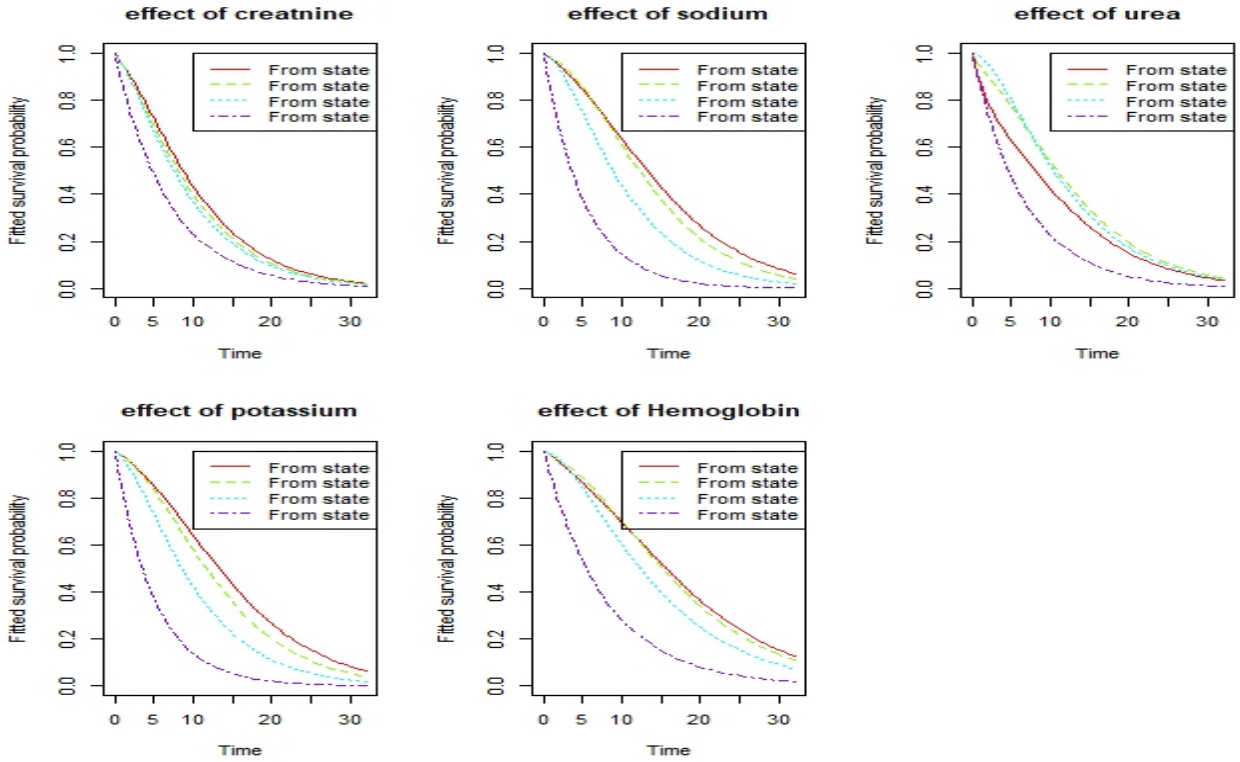

Figure 7: Plot of effect of clinical variables on fitted survival probability for each state

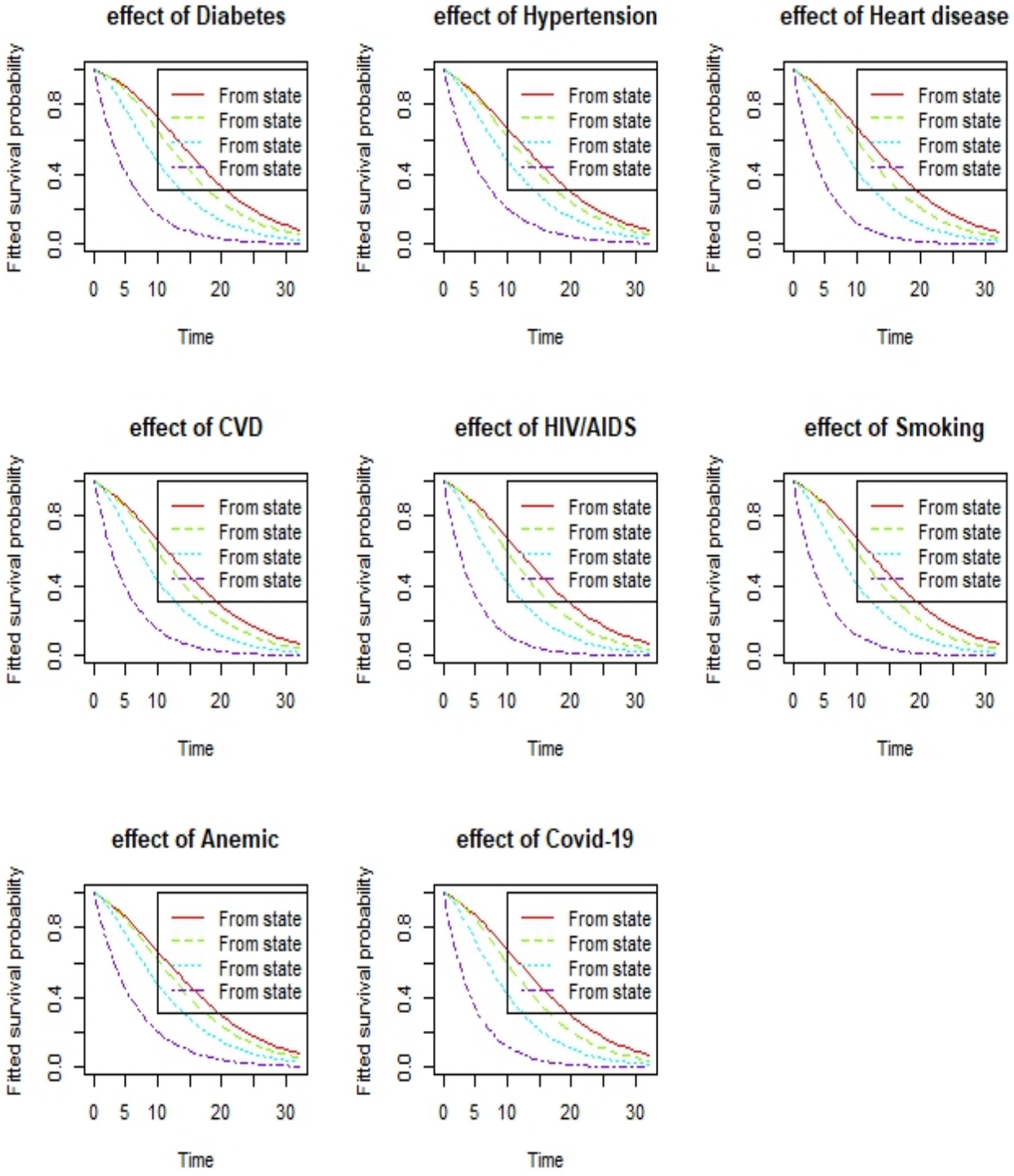

Figure 8: Plot of effect of co-morbidity variables on fitted survival probability for each state

Appendix C: Hazard ratios and 95% CI for the best Model with its plots

Table 5.1: Hazard ratios and 95% confidence intervals for the best Model .

States	Baseline	HTN	Diabetes	CVD
1 →1	-0.357 (-0.426,-0.299)			
1 → 2	0.282 (0.221, 0.359)	1.530 (0.622, 3.77)	1.731 (1.667, 3.49)	1.086 (0.599, 1.97)
1 → 3	0.025 (0.004, 0.159)	2.407 (0.002, 2561.1)	2.100 (0.001,3750.5)	1.127 (0.006,218.9)
1 → 4	0.031 (0.011, 0.091)	1.412 (0.031, 65.2)	1.411 (0.028, 72.1)	1.589 (0.256, 9.849)
1 → 5	0.019 (0.006, 0.061)	1.413 (0.020, 99.6)	1.418 (0.018, 109.8)	1.458 (0.166, 12.8)
2 → 2	-0.239 (-0.282,-0.202)			
2 → 3	0.206 (0.168, 0.252)	1.218 (0.604, 2.45)	1.165 (0.577, 2.353)	1.269 (0.867, 1.857)
2 → 4	0.017 (0.003, 0.089)	2.014 (0.006, 633.9)	1.632 (0.003, 812.9)	1.329 (0.114, 15.5)
2 → 5	0.016 (0.006, 0.041)	2.038 (0.086, 48.5)	1.758 (0.061, 50.5)	1.522 (0.307, 7.55)
3 → 3	-0.185 (-0.224,-0.153)			
3 → 4	0.173 (0.137, 0.218)	1.500 (0.830, 2.71)	0.931 (0.519, 1.669)	1.627 (0.999, 2.650)
3 → 5	0.012 (0.002, 0.068)	1.430 (0.023, 89.7)	1.144 (0.025, 52.1)	2.026 (0.061, 67.27)
4 → 4	-0.150 (-0.182,-0.125)			
4 → 5	0.150 (0.125, 0.182)	2.627 (1.797, 3.84)	1.347 (0.934, 1.939)	2.014 (1.441, 2.815)

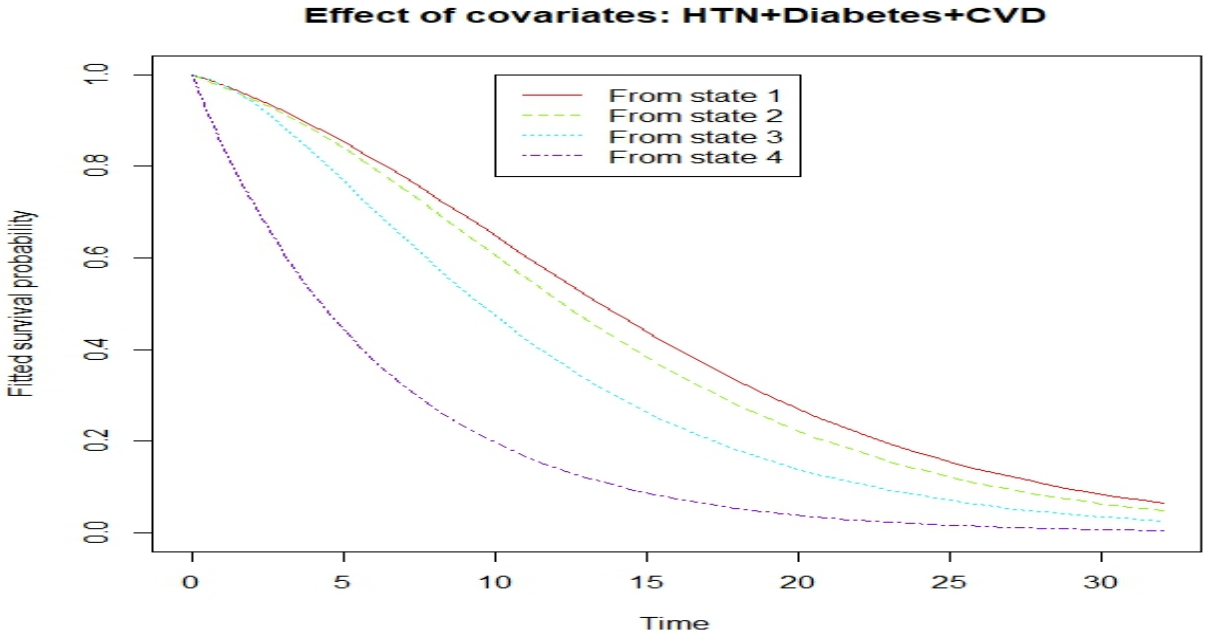

Figure 9: Effects of the final model on the survival probability of the patients

Appendix D: Graphical Plots of Kaplan Meier estimate

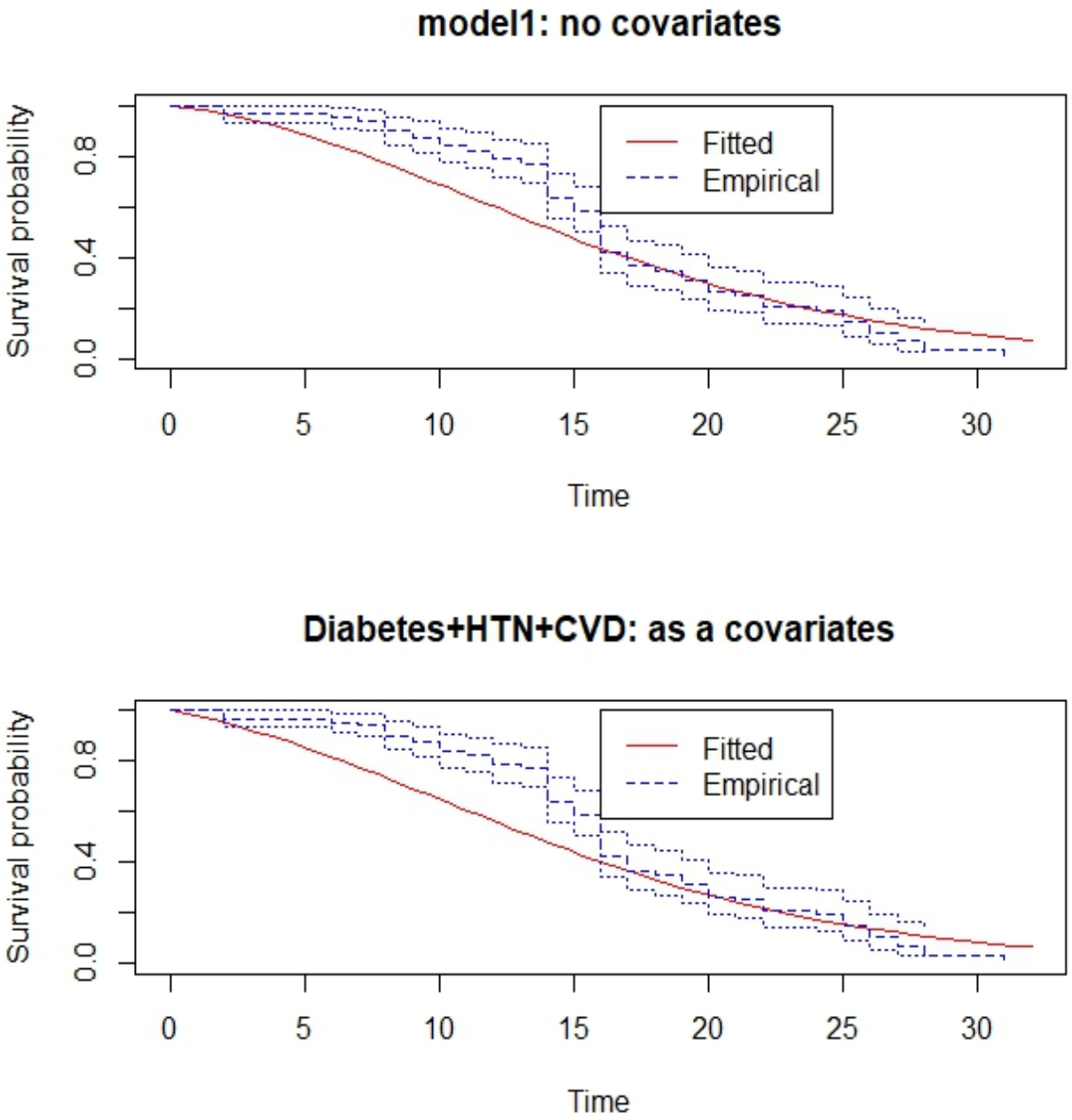

Figure 10: Plots of Kaplan Meier estimate