College of Natural Sciences

Department of Statistics

Modeling Time to Kidney Failure of the Patients at Adama Hospital Medical College: Application of Copula Model

---

By: Firomsa Shewa

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

August 2021

Jimma, Ethiopia

Jimma University

College of Natural Sciences

Department of Statistics

Modeling Time to Kidney Failure of the Patients at Adama Hospital
Medical College: Application of Copula Model

By: Firomsa Shewa

Advisor: Akalu Banbeta (Ass't Prof.)
Co-Advisor: Jaleta Abdisa (MSc.)

August 2021

Jimma, Ethiopia

Approval Sheet

As thesis research advisors, we here by certify that we have read the thesis prepared by Firomsa Shewa under our guidance, which is entitled "Modeling Time to Kidney Failure of the Patients at Adama Hospital Medical College: Application of Copula Model", in its final format is found that (1) its format, citations, and bibliographical style are consistent and acceptable and ful-fill University and department style requirements; (2) its illustrative materials including tables and figures are in place, (3) the final manuscript is satisfactory to graduate committee and is ready for submission to the University library.
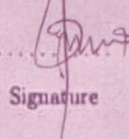
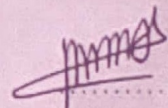|		
Akalu Banbeta (Ass't Prof.)		02|09|2021
Advisor	Signature	Date
Jaleta Abdisa(MSc.)		02/09/2021
Co-advisor	Signature	Date

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. We recommend that the thesis has been accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

Shibru Temesgen (PhD)		01|09|2021
External Examiner	Signature	Date
Geremew Muleta (Ass't Prof.)		02|09|2021
Internal Examiner	Signature	Date
Kibrealem Sisay (MSc.)		02|09|2021
Chairman, Post graduate coordinator	Signature	Date
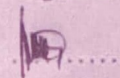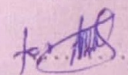
1

Declaration

I declare that, this thesis is a result of my genuine work and all sources of materials used for writing it have been duly acknowledged. I have submitted this thesis to Jimma University in partial fulfillment for the Degree of Master of Science in Biostatistics. And also, I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Firomsa Shewa   . . . .

    Name   Signature   Date

Acknowledgment

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

AHMC: Adama Hospital Medical College

AIC: Akaike Information Criterion

AKF: Acute Kidneys Failure

BIC: Bayesian Information Criterion

BMI: Body Mass Index

CI: Confidence Interval

EDTA: European Dialysis and Transplant Association

ERA: European Renal Association

ESKD: End-stage of Renal Disease

Final llk: Joint maximum log-likelihood

GFR: Glomerular Filtration Rate

HR: Hazard Ratio

NKF: National Kidney Foundation

OR: Odd Ratio

PH: Proportional Hazard

PIOS: pseudo in and out of sample

PO: Proportional Odds

RRT: Rena Replacement Therapy

SE: Standard Error

SSA: Sub-Saharan Africa

USA: United States of America

USRDS: United States Renal Data System

WHO: World Health Organization

Definition of Symbols

η: Dependence parameter

η_{ind}:   Value of η if independent

τ: Kendall's tau

λ: Scale parameter

k: Shape parameter

a: Scale parameter for Gompertz

b: Rate parameter for Gompertz

Abstract

Background: Kidney failure is an irreversible disease in which one or both kidneys are unable to adequately filter waste products from the blood. Bi-variate time to event endpoints may be correlated as they come from the same subject. However, classical survival analysis assumes that survival times of different subjects are independent. Thus, this study aimed to model time to right and left kidney failure of the patient at Adama Hospital Medical College.

Methods: The data for this study was the chronic kidney disease patients under follow up at Adama Hospital Medical College fromfrom 1st January 2015 to 30th January 2020 . The copulas are used to join the bi-variate time to event endpoints to the one dimensional marginal distribution functions. The dependence between the time to right and left kidney failure of the patient was quantified using the copula parameter, while the effect of covariates were modeled using the parametric marginal survival model. Akaike information criterion and Bayesian information criterion were used for the models comparison.

Results: Of all 431 patients, 170 (39.4%) failed at least one kidney during the follow-up period. The Log-logistic marginal distribution with Clayton copula model revealed that sex of patients, hypertension, family history of kidney disease, obesity and age of patients were the most significant factor that associated with time to kidney failure. The dependence parameter was 1.4 (p-value < 0.0001).

Conclusions: The Log-logistic marginal distribution with Clayton copula model fit the kidney failure dataset well. Being male, older adult, obese, hypertensive and having family history of kidney disease were the most risk factors that leads to kidney failure. There is the dependence between the time to right and left kidney failure of the patient.

Key words: Bi-variate Events; Copula Model; Dependence; Kidney Failure

1 Introduction

1.1 Background of the Study

Kidney failure is an irreversible disease in which one or both kidneys are unable to adequately filter toxins and waste products from the blood. Glomerular filtration rate (GFR) levels less than 60 mL/min/1.73m^2 for 3 or more months are a sign of chronic kidneys disease[1]. Glomerular filtration rate value less than 15 mL/min/1.73m^2 indicates kidney failure[2].

Kidney failure is a worldwide public health problem, with increasing incidence and prevalence, high costs and poor outcomes[3]. The incidence and prevalence of end-stage of renal risease (ESRD) have doubled in the past 10 years and are expected to continue to rise steadily[4]. The United States Renal Data System (USRDS) in 2016, reported incidence rates of ESRD in Taiwan, United States of America (USA), Jalisco region of Mexico and Thailand were 493, 378, 355 and 346 per million population per year respectively[5]. The epidemiology of ESRD in Iran showed that the prevalence and incidence of ESRD have been increasing in Iran, from 238 cases per million population to 357 per million population and 49.9 per million population to 63.8 per million population, respectively from 2000 to 2006 years[6].

The Global Kidney Health Atlas in 2019 found that, by 2030, 14.5 million people around the world will have ESRD, yet only 5.4 million will be treated due to economic, social, and political factors[7]. Additionally, more than 2 million people will die each year due to little or no access to hemodialysis or kidney transplantation[8].

Kidney disease is at least 3-4 times more frequent in Africa than in developed countries[9]. In Africa, the vast majority cases of ESRD remain undiagnosed and untreated, which leading to almost certain mortality[10]. The burden ESRD in Sub-Saharan Africa (SSA) is unknown but is probably high. Access to dialysis for ESRD is limited by insufficient infrastructure and catastrophic out-of-pocket costs[11]. In African, patients with ESRD have the lowest access to Rena Replacement Therapy (RRT), with only 9–16% being treated, in central and eastern Africa, the treatment rate is estimated to

be as low as 1–3% because of it's cost[12].

According to World Health Organization (WHO) comprehensive and integrated action of the means to prevent and control chronic diseases in developing countries, kidney diseases are growing problem in developing countries like Ethiopia[13]. The prevalence of kidney disease 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[14]. According to the WHO data published in April 2011, kidney disease deaths in Ethiopia reached 12,038 or 1.47% of total deaths. The study conducted at Adama Hospiatal Medical College in 2016 revealed that among 500 ESRD 27.40% were died[15].

Bi-variate data arise frequently in many research areas such as health, epidemiology and economics. In medical studies, it is common to record two event times for each patient. Particular examples include failure times of paired human organs, (kidneys, eyes, lungs and breasts)[16]. These type of events are correlated as they come from the same subject[17]. In analysis of such bi-variate survival data, the key element is an appropriate account for dependence between event times[18]. It is of interest to estimate and quantify the dependence between the bi-variate event times and the effects of covariates under the dependence structure.

Classical survival analysis techniques assume that survival times of different subjects are independent. Although this assumption may be valid in many situations, it may be violated in others. Indeed, survival times are frequently not independent of each other; there may be a natural association because individuals share biological and/or environmental conditions. For example, kidneys within a patient will be more alike than kidneys from different patients because of genetic influence. Such data are known as clustered or correlated survival data. In survival studies when the event times are dependent, performing of the analysis using methods based on independent assumption, leads to biased estimation. This is why the alternative framework of the bi-variate survival analysis for time to bi-variate event endpoints has been developed[19]. Bi-variate survival analysis is a branch of survival analysis, which deals with two events per subject and deals with dependence between failure times including influence of covariates

2

on failure times in the presence of dependence[20].

Copula is used for the statistical analysis of bi-variate event times[21]. A copula can be used to link two event times by specifying their dependence structure[20]. It has the tendency to models, the two marginal distributions and the between-margin dependence separately, allowing flexibility in marginal models. Moreover, the challenge from censoring can be naturally handled through the marginal distributions within the copula function. Consequently, measures of dependence, such as Kendall's tau, can be derived from a copula without influenced by the marginal distributions[22]. One unique feature of copula is that it models the two marginal distributions and the between margin dependence separately, allowing flexibility in marginal models and straightforward interpretation for covariate effects[23].

1.2 Statement of the Problem

Nowadays, bi-variate time to events endpoints are often used in clinical trials for studying bilateral diseases. However, only few studies has been conducted regarding of bivariate events time. Bi-variate survival data are correlated as they come from the same subject; analyzing such data requires model specifications on the dependence between bi-variate events time. Classical survival analysis techniques assume that survival times of different subjects are independent. But, the time to right and left kidney failure of the patient are not independent of each other because a pair of kidneys share the same biological gene in common[17].

Chen, *et al.* fitted marginal models for bi-variate censored data[24]. However, the approach deals with the marginal likelihood function directly and ignores the dependence structure between the failure times. This study addressed research problems using the copula models. Unlike the marginal approach the copula based methods directly connect the two marginal distributions through a copula function to construct the joint distribution, of which the copula parameter explore the dependence between the time to right and left kidneys failure of the patient and the parametric marginal distribution model assess the effects of covariates under dependence structure.[23].

The prevalence of kidney disease in Ethiopia is high and has increased in the last few years to be 12.2% with an increased prevalence of diabetes and hypertension and it is becoming one of the public health problems[14]. However, not much attention is given and few studies have been conducted in our country regarding kidney failure. Generally, since the researcher did not yet find a study conducted on modeling time to kidney failure of the patients using the bi-variate survival model at national level and the cases under study is found to be really a predominant issue, it happened to be a reason to conduct this study.

Thus, this study addressed the following research questions:

- Is there dependence between the time to right and left kidney failure of a patient?

- What are the factors that significantly affect the time to kidney failure of the patients?

- Which statistical model could fit the kidney failure dataset well?

1.3   Objectives of the Study

General Objective

The general objective of this study is modeling the time to kidney failure of the patients at Adama Hospital Medical College using the copula model.

Specific Objectives

The specific objectives of the study are:

- To identify whether there is a dependence between the time to right and left kidney failure of a patient.

- To determine the significant factors that affect time to kidney failure of the patients.

- To identify the copula model that best predicts time to kidney failure of the patients well.

1.4 Significance of the Study

This study identified and gave detailed information on the major risk factors which lead to kidney failure of the patients. For academicians or statisticians, it will direct to thoughts and genuine interest on the subject matter for further research specially, when two endpoints are dependent. Finally, the results obtained from this study may be used as a baseline or reference to pave the way for conducting further related studies and used as input for stakeholders (health policy makers).

2   Literature Review

2.1   Definition and Overview of the Kidney Failure

Kidney failure is classified as either acute kidney failure (AKF), which develops rapidly and may resolve and chronic kidney failure (CKF), which develops slowly and can often be irreversible[25]. Kidney disease is evaluated in terms of the overall renal function GFR, renal ultrasound or the presence of kidney damage established by either kidney biopsy or other markers of kidney damage[26]. Kidney failure is also known as the end-stage of renal disease, and it is a medical condition in which the kidneys are functioning at less than 15% of normal levels. There are five stages of kidney disease. Stage 1: kidney damage with GFR $\geq$ 90 mL/min/1.73m^2 (with normal GFR), stage 2: kidney damage with GFR 60-89 mL/min/1.73m^2 (increased risk damage), stage 3: GFR 30-59 mL/min/1.73^2 (decrease GFR), stage 4: GFR 15-29 mL/min/1.73m^2 (regardless of kidney damage) and stage 5: GFR <15 mL/min/1.73^2 (kidney failure) treated by dialysis or transplantation[27].

Currently, more than 2 million people worldwide receive treatment with dialysis or a kidney transplant to stay alive, yet this number may only represent 10% of people who actually need treatment to live[28]. It is estimated that kidney disease affects 31 million people in the United States alone, and globally 1 in 10 people have some form of kidney[29]. Similarly, number of cases of kidney failure will increase disproportionately in countries, like China and India, where the number of elderly people are increasing[30].

In Africa continents, less than 2% of the patients with ESRD have access to RRT, and the ESRD rate is increasing at 6% to 8% per year[31]. By 2030, more than 70% of patients with ESRD are estimated to be living in low-income countries, such as those in SSA[32]. According to WHO, comprehensive and integrated action in low-income countries, such as Ethiopia, kidney disease is a growing problem. The incidence of kidney disease in Ethiopia is rising because of increased risk factors such as high blood pressure and diabetes mellitus[13].

2.2 The Economic Burden of the Kidney Failure

Kidney failure has been observed to be a major threat to the world's health, and in some African countries, it is a death sentence. It affects economically productive young adults between the ages of 20–50 in SSA as against the middle age and elderly in the developed world[33]. It imposes disproportionate, incalculable human suffering and a catastrophic economic burden on the African continent in several respects. Kidney failure is the ultimate stage of CKD, represents a major issue for the national health service, due to the high increase of its incidence and prevalence as well as for the high costs for treatments. In middle-income countries, treatment with dialysis or kidney transplantation creates a huge financial burden for most of the people who need it[34].

Globally, the cost of dialysis care ranges from US$100 to $200[35]. In the US, treatment of chronic kidney disease is likely to exceed $48 billion per year[11]. In Australia, treatment cost for kidney failure is estimated to US$12 billion[36]. Of all 47 countries in SSA, only 6.3% have a functioning renal transplant program and the cost of renal transplant in SSA ranges between US$3,000 and $20,000[37].

A study conducted in Nigeria shows that less than 1% of patients can afford treatment for more than three months mainly because of financial constraint[38]. The cross-sectional study conducted in Sudan, showed the annual cost of hemodialysis two sessions per week was US$24,732[39]. In Ethiopia, cross-sectional study was conducted, among 172 ESRD patients undergoing hemodialysis treatment, the overall mean annual cost of hemodialysis treatment for end-stage renal disease patients was 121,089.27 birr (US$4466.59)[40].

2.3 Major Risk Factors of the Kidney Failure

Risk factors that leading to kidney failure of the patients includes: diabetes, high blood pressure, family history of kidney disease, older age, anemia and obesity[41].

Hypertension disease:- Hypertension is a leading risk factor for ESRD. Vaes, *et al.* were conducted a retrospective cohort study during a 10-year time interval (2002-2012), in order to explore the correlation between blood pressure and kidney function decline,

the study found that hypertension (blood pressure $\geq 140/90$ mm Hg) was a significant risk factor for the development of ESRD (HR = 1.25 [95% CI: 1.22, 1.28])[42]. Similarly, Vupputuri, *et al.* conducted cohort study among 890 hypertensive patients who attended the Hypertension Clinic at the Veterans Administration Medical Center of New Orleans between 1996 and 2002[43]. The study revealed that higher systolic and diastolic blood pressure was associated with relative risks 1.81[95% CI:1.29 to 2.55] and 1.55[95% CI: 1.08 to 2.22], respectively, for early kidney function decline. Also in a large Japanese cohort study, Tozawa, *et al.* were conducted study that includes 98759 patients[44]. The study found that higher baseline blood pressure (hypertension) was a significant risk of development of ESRD.

Diabetes mellitus:- Diabetes is the commonest cause of ESRD requiring renal replacement therapy[45]. Around 20-30% of people with diabetes develop ESRD[46]. In the baseline cohort analysis of a large Medicare American study the presence of diabetes was found to double the risk of developing ESRD compared with those without diabetes (OR = 2.04[95% CI: 2.00 to 2.09], $p < 0.0001$)[47]. In a community based longitudinal cohort study of the patients from the Framingham Offspring Study 2,585 individuals were monitored over 12 years. In multivariate analysis those with diabetes at baseline had an increased rate of development of ESRD (OR = 2.60[95% CI 1.44 to 4.70])[48].

Obesity:- One of the strongest yet modifiable risk factors for kidney failure in the twenty-one century is obesity. A large-scale epidemiological study from Sweden demonstrated the role of obesity in ESRD[49]. Obesity anytime during lifetime was linked to 3-4 to increases in ESRD risk[50]. It may contribute to the pathogenesis of kidney damage through inflammation, oxidative stress, endothelial dysfunction, prothrombotic state, hypervolemia, and adipokine derangements[51].

Family History: Kidney disease runs in families. Patients may be more likely to get kidney disease if patients have a close relative with kidney disease. Mekiya, *et al.* used Cox regression to analysis survival analysis of patients with ESRD[15]. The study found that, family history of kidney failure was a significant risk factors for ESRD (HR = 1.88[95% CI: 1.2 to 2.9]). Song, *et al.* had screened incident dialysis patients between 1

January 1995 and 31 December 2003 in the US[52]. The study found that, nearly 23% of incident dialysis patients had close relatives with ESRD. Another cross-sectional hospital based study conducted in Khartoum State showed that the prevalence of ESRD is 19.3% among relatives[53]. A team of researchers from the US and Europe looked at the Deoxyribonucleic Acid (DNA) sequences in more than 65,000 people. A team of researchers found that, Common variations in several genes were to be more common among people with kidney disease than in those with normal kidney function[54].

Sex:- The study conducted by Tangri, *et al.* showed that more women than men have kidney diseases, but men are more likely to reach ESRD sooner than women[55]. The Japanese Society for Dialysis Therapy have demonstrated that ESRD is more frequent among men[56]. In one study, a total of 107,192 subjects (51,122 men and 56,070 women) from Okinawa, Japan participated in a 10-year follow-up where odds ratio for ESRD was 1.41 among male participants[57]. A new analysis of data in the European Renal Association with European Dialysis and Transplant Association, shows that men are affected by kidney failure much more often than women[58].

Age:- Kidney function declines with age in almost everyone, and the proportion of older people with GFR readings below 60 approaches 50%[59]. Chadban, *et al.* conducted individual-level meta-analysis including 2051244 participants from 33 general population during a period of 1972 to 2011 with a mean follow-up time of 5.8 years. The study revealed that age was a significant risk factor, the older adulthood age compared to young adulthood with a odds ratio of 2.5 (95% CI: 1.8 to 3.2)[60]. In Turkey, a population-based survey of ESRD study was conducted, the 95% CI of OR for ESRD was [1.45 to 2.18] among older adulthood age subjects[61]. The Framingham Offspring study established a graded risk associated with age (OR of 2.36 per 10 year age increment; 95% CI 2.00 to 2.78)[48].

Smoking:- Smoking is associated with a greater risk of kidney failure; the risk increases with an increase in the smoking duration, number of cigarettes smoked daily, and pack-years[62]. A good quality Swedish case control study provides supportive evidence for smoking as a significant risk factor for ESRD in a community based population[49]. Odds ratios increased with increasing frequency and duration of smoking. 16-30 pack

years of smoking increased the risk of ESRD significantly [OR = 1.32] and > 30 pack years [OR = 1.52]. Orth, *et al.*[63] conducted study in order to explore the effects of smoking on renal function. The study found that, each additional five smoked cigarettes per day was associated with an increase in ESRD by 31%.

Alcohol consumption:- Regular heavy drinking more than four drinks a day has been found to double the risk of ESRD. But only a few studies have been done on drinking alcohol and the risk of ESRD, drinking even a few alcoholic beverages per week (three to four drinks) increased the risk of ESRD. Heavy drinkers who also smoke have an even higher risk of kidney problems. Smokers who are heavy drinkers have about five times the chance of developing ESRD than people who don't smoke or drink alcohol to excess[64].

2.4   Overview of the Copula Model

The first approach for analyzing bi-variate time to event endpoints is a marginal method, which was developed under the general estimation equation framework[65]. A robust sandwich estimator from the estimating equation is used to estimate the variance-covariance matrix of the regression parameter. Wei, *et al.* considered the semi-parametric Cox model and proposed to estimate the regression parameter under a working independence assumption by which observations in each cluster are treated as independent of one another[66]. By applying a sandwich estimator, it takes into account the fact that observed event times are correlated however, the strength of such correlations cannot be explicitly modeled under this marginal approach[65].

Copulas are functions that join or "couple" multivariate distribution functions to their one dimensional marginal distribution functions. Copulas are of interest to statisticians for two main reasons: Firstly, to study scale-free measures of dependence, and secondly, as a starting point for constructing families of bi-variate distributions[67].

One of the earliest distribution families for modeling correlated bi-variate measurements is the copula family, originated from Sklar's theorem, in which the joint distribution is modeled as a function of each marginal distribution together with an association

parameter[68][69]. Copula function provides a parametric assumption about the association between two correlated margins[65].

The applications of copulas in multivariate survival analysis became active after David George Clayton introduced his bi-variate survival model[69]. His work yielded one of the most important copulas for bi-variate survival analysis, known as the Clayton copula[20]. Burzykowski, develop the most successful papers on copula-based survival models with two-step method for analyzing dependence between two correlated endpoints[70]. Indeed, copula based methods are adaptive to survival data with complex dependence structure clustered survival data[20][71].

Copulas have become a popular multivariate modeling tool in many fields where multivariate dependence is of interest and the usual multivariate normality is in question. In many areas of statistics the main goal or objective is to model the data in order to explain a response variable. However, sometimes the interest goes behind this objective and the aim is to study dependencies or correlation between them.

Copulas have provided flexible survival models and unified statistical methods. In addition, copulas provide measures of dependence Kendall's tau, that are free from the model specifications of the marginal survival distributions. Cox regression model with parametric baseline hazard function is used for marginal survival distribution.

Generally this study is conducted to assess the most risk factors of the kidney failure taking into account the dependence between the time to right and left kidneys failure of the patient.

3   Data and Methodology

3.1   Description of the Study Area

The study was conducted at Adama Hospital Medical College (AHMC). AHMC was previously known by the names of Hayilemariam Mamo Memorial Hospital and Adama Referral Public Hospital at different times. It is one of the first medical hospital situated in Adama town, located in Oromia region, 100 km to Southeast of Addis Ababa, Ethiopia. The hospital was upgraded to medical college in 2003 E.C because of its location, patient load and staff capacity. The hospital is serving a catchment population of more than 6 million; from four regions and a city administration (Oromia, Amhara, Afar, Somali, and Dire-Dawa). The hospital has 232 beds capacity and serving on average 1000 patients per day at six medical case teams and different specialty clinics.

3.2   Study Population and Period

A retrospective study was conducted to assess the risk factors that leading to kidney failure of the patients based on hospital registry in AHMC. This data is secondary data recorded at the hospital from patient's registry date to the event time or censoring time. The population of this study was all patients with kidney disease who had been registered at AHMC starting from 1st January 2015 to 30th January 2020. The total number of patients considered in the study was 431. A patient may experience one of the following four cases: a) [1, 1] if both kidneys of the patient are failed, b) [1, 0] if the only right kidney of the patient is failed, c) [0, 1] if the only left kidney of the patient is failed, or d) [0, 0] if both kidneys of the patient are not failed.

3.3   Inclusion and Exclusion Criteria

Inclusion Criteria:- Patients who had GFR levels less than 60 mL/min/1.73m^2 and registered with full information including in the registration log book or in the patients identification card will consider to be eligible for the study.

Exclusion criteria:- Patients who had GFR levels greater than 60 mL/min/1.73m^2 and insufficient information about one of the vital variables either in the registration

book or in the card will not eligible. In addition, the patients who had born naturally with only one kidney or born with two kidneys, but only one of them works were excluded from the study.

3.4   Data Collection Procedure

The data set used for this study was collected from patients individual card. All the data had been carefully reviewed from the registration log book and patients registration card; if any inadequate information counters it has been checked from the file and excluded from analysis if proven to be inadequate. For the data collection, one health professional and two experienced data collectors under the supervision of the researcher were contributed.

3.5   Data Structure for Bi-variate Events

The table below illustrates data structures required for modeling time to bi-variate event endpoints (Modeling time to kidneys failure of the patients).

Table 3.1: Data Structure for Bi-variate Time to Event Endpoints

id	ind	obs_times	status	sex	hypertension	obesity	family history	smoking
1	1	461	0	male	no	no	no	non-smoker
1	2	461	0	male	no	no	no	non-smoker
2	1	872	0	female	yes	no	yes	non-smoker
2	2	378	1	female	yes	no	yes	non-smoker
3	1	240	1	male	no	no	no	smoker
3	2	840	0	male	no	no	no	smoker
4	1	450	1	female	yes	yes	no	non-smoker
4	2	660	1	female	yes	yes	no	non-smoker

id:- patient identification number

ind:- margin indicator (1 for right and 2 for left kidney)

obs_times:- times (in days) at which the event occurs or censoring.

status:- 1 event and 0 censoring.

For each kidney, the event of interest is the time from the day of the patients registered at hospital to the kidney failure in days. Event time is the time from hospital registration for kidney disease to when GFR dropped below 15 mL/min/1.73m^2 (call it "kidney failure"). In kidney failure dataset, the times to right and left kidneys failure of the patients cannot be precisely observed, leading to bi-variate censored. Bi-variate right-censored data occur when the study ends prior to the occurrence of one or both events. The right censoring could also happen if the event still does not occur at the last assessment time. Therefore, the time to kidneys failure the patients are bi-variate right censored. Censoring was caused by death, dropout, refer to other hospital or end of the study.

3.6   Study Variables

Response Variable

The response variable in this study is the time in days to kidney failure of patients starting from the day of the patients registered at hospital. Kidney failure means when GFR dropped below 15 mL/min/1.73m^2 (call it "kidney failure").

Independent Variables

The explanatory variables that expected as the predictor factors that associated with kidney failure according to different literature source in this study are described as follows:-

Table 3.2: Explanatory Variables and its Code.

Variables	Categories	Codes
Sex	Female	0
	Male	1
Residence	Rural	0
	Urban	1
Diabetes mellitus	No	0
	Yes	1
Hypertension disease	No	0
	Yes	1
Anemia disease	No	0
	Yes	1
Smoking status	Non-smoker	0
	Smoker	1
Alcohol consumption	No	0
	Yes	1
Family history	No	0
	Yes	1
Obesity	No	0
	Yes	1
Age	≤ 35	0
	36-55	1
	≥ 56	2

3.7   Statistical Methods

3.7.1   Survival Data Analysis

The uni-variate survival analysis assume that, event times are independent of each other. However, this assumption can be violated when the study units are paired such as twins, married couples or bilateral disease such as kidneys. In the presence of the dependence between the two event times, bi-variate survival analysis needs to be

considered[17]. Bi-variate survival analysis involves the study of failure times, including the influence of covariates, in the presence of dependence[72].

Copulas are functions that join multivariate distribution functions to their one dimensional marginal distribution functions. The Copula parameter is used to demonstrate the dependence structure between the time to right and left kidney failure of the patient and the parametric marginal distribution is used to investigates the effect of covariates on time to kidney failure of the patients under dependence structure. One unique feature of copula is that it models the two marginal distributions and the between margin dependence separately, allowing flexibility in marginal models and straightforward interpretation for covariate effects[23].

3.7.2   Copula Models for Bi-variate Events

Copula is used to link two event times by specifying their dependence structure. First, let we introduce the notation for bi-variate time to event data. Assume that there are n patients. Let (T_{1i}, T_{2i}) and (C_{1i}, C_{2i}), $i = 1, 2, 3, \ldots, n$ denote the bi-variate failure times and censoring times for the ith patients respectively. Then for each patient, we observe $D_i = \{(Y_{ji}, \Delta_{ji}, Z_{ji}); Y_{ji} = min(T_{ji}, C_{ji}), \Delta_{ji} = I(T_{ji} \leq C_{ji}), j = 1, 2\}$, where C_{ji} is the censoring time of T_{ji}, Δ_{ji} is the censoring indicator and Z_{ji} is the covariate vector. Let $S_j(t_j) = P(T_j > t_j)$, $j = 1, 2$ denotes marginal survival function, $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$ denote the joint survival distribution and $f(t_1, t_2) = \partial^2 S(t_1, t_2)/\partial t_1 \partial t_2$ denote density function for (T_1, T_2) respectively[65].

Copula functions provide a parametric assumption about the dependence between two correlated margins. The parameter η in copula function describes the dependence between T_1 and T_2. By Sklar's theorem, one can model the joint distribution by modeling the copula function and the marginal distributions separately[68]. The theorem is stated as: if marginal survival functions $S_1(t_1) = P(T_1 > t_1)$ and $S_2(t_2) = P(T_2 > t_2)$ for T_1 and T_2 are continuous, then there exists a unique copula function C_η such that for all $t_1, t_2 \geq 0$

$$S(t_1, t_2) = C_\eta\{S_2(t_2), S_1(t_1)\}, t_1, t_2 \geq 0. \tag{1}$$

Here, the function C_η is called a copula function and its parameter η measures the dependence between the two margins. Define the density function for C_η to be $c_\eta = \partial^2 C_\eta(u,v)/\partial u \partial v$, then the joint density function of T_1 and T_2 can be expressed as:

$$f(t_1, t_2) = c_\eta\{S_1(t_1), S_2(t_2)\}f_1(t_1)f_2(t_2), t_1, t_2 \geq 0. \tag{2}$$

The Copula function is robust in modeling various dependence structures and has nice properties. The dependence measurement Kendall's tau (τ) can be directly obtained as a function of η in some copula models.

Archimedean Copula Family

The most popular copula families for bi-variate events data is the Archimedean copula family, which is one of the most popular copula families because of its flexibility and simplicity[65]. A copula C_η belongs to an Archimedean family if it can be expressed as:

$$C_\eta(u,v) = \phi_\eta\{\phi_\eta^{-1}(u) + \phi_\eta^{-1}(v)\},$$

where u and v are two uniformly distributed margins; ϕ_η is the generator function, which is a continuous, strictly decreasing and convex function; ϕ_η^{-1} is the inverse of ϕ_η. The copula parameter η has a one-to-one correspondence with the popular dependence measure Kendall's tau. Three most frequently used Archimedean copulas in survival analysis are:

Clayton Copula

The Clayton copula is expressed as:

$$C_\eta(u,v) = (u^{-\eta} + v^{-\eta} - 1)^{-1/\eta}, \eta \in (0, \infty),$$

and its generator function is given by:

$$\phi_\eta(t) = \eta^{-1}(t^\eta - 1),$$

For a Clayton copula, the association parameter η corresponds to Kendall's tau as $\tau = \eta/(2+\eta)$. Thus, T_1 and T_2 are positively associated when $\eta > 0$ and are independent when $\eta \to 0$[69].

17

Gumbel Copula

The Gumbel copula is expressed as:

$$C_\eta(u,v) = exp\left[-\{(-logu)^\eta + (-logv)^\eta\}^{1/\eta}\right], \eta \in [1,\infty),$$

and its generator function is given by:

$$\phi_\eta(t) = (-log(t))^\eta,$$

For a Gumbel copula, $\tau = 1 - 1/\eta$, meaning T_1 and T_2 are positively associated when $\eta > 1$ and are independent when $\eta = 1$[73].

Joe Copula

The Joe copula is expressed as:

$$C_\eta(u,v) = 1 - \{(\tilde{u})^\eta + (\tilde{v})^\eta - (\tilde{u}\tilde{v})^\eta\}^{1/\eta}, \eta \in [1,\infty),$$

Where $\tilde{u} = 1 - u$ and $\tilde{v} = 1 - v$

and its generator function is given by:

$$\phi_\eta(t) = -log(1 - (1-t)^\eta),$$

For a Joe copula family, Kendall's tau τ is given by below equation, meaning T_1 and T_2 are positively associated when $\eta > 1$ and are independent when $\eta = 1$[72].

$$\tau = 1 - \sum_{k=1}^{\infty} \frac{1}{k(\eta + 2)\{\eta(k-1) + 2\}} \tag{3}$$

Kendall's tau(τ) is most frequently used in practice as a measure of dependence between time to bi-variate event endpoints.

18

Table 3.3: Archimedean Copula Families and its Measures of Dependence

Copula family	Range of η	η_{ind}	Generator:$\phi_\eta(t)$	Kendall's tau:τ_η	Range of τ
Clayton	$0 \le \eta < \infty$	0	$\eta^{-1}(t^\eta - 1)$	$\eta/\eta + 2$	$0 \le \tau \le 1$
Gumbel	$1 \le \eta < \infty$	1	$(-log(t))^\eta$	$1 - 1/\eta$	$0 \le \tau \le 1$
Joe	$1 \le \eta < \infty$	1	$-log(1 - (1-t)^\eta)$	.	$0 \le \tau \le 1$

For Joe, there is no closed form, but equation (3) is evaluated numerically.

3.7.3 Parametric Marginal Models

To assess the effect of covariates on time to kidney failure of the patients, it is necessary to choose a regression model for the margins. A parametric survival model is one in which survival time is assumed to follow a known distribution. Parametric regression models is more efficient than its corresponding non-parametric or semi-parametric models, because its estimation is based on both time and event information[74]. It is more efficient, leading to smaller standard errors and more precise estimates[75]. The supported marginal models are Proportional hazards models (Weibull and Gompertz) and Proportional odds (Log-logistic).

Generally, marginal survival model in terms of hazard function is given by:

$$\lambda_j(t_{ji}|Z_{ji}) = \lambda_{0j}(t_{ji})exp(\beta'Z_{ji}), j = 1, 2, i = 1, ..., n \qquad (4)$$

where λ_{0j} is the baseline hazard function for the jth margin, Z_{ji} are the covariates for the ith patient with j^{th} margin and β are the coefficient of covariate.

The marginal survival distribution for T_{ji} given covariate Z_{ji} can be expressed as:

$$S_j(t_{ji}|Z_{ji}) = P(T_{ji} \ge t_{ji|Z_{ji}}) = S_{0j}(t_{ji})^{exp(\beta'Z_{ji})}, j = 1, 2, i = 1, ..., n \qquad (5)$$

Where, $S_{0j}(t_{ji})$ is the baseline survival distribution and given by $exp\{-\int_0^{t_{ji}} \lambda_{0j}(s)d(s)\}$

Choice of marginal survival Distributions

The choice of the appropriate parametric marginal distribution form is the most difficult part of survival analysis. The specification of the parametric marginal distribution form should be driven by the study hypothesis, along with prior knowledge and biologic plausibility of the shape of the baseline hazard.

Weibull Marginal Distribution

The Weibull is the most widely used survival time distribution model. The Weibull distribution assumes a monotonic hazard that can either be increasing or decreasing but not both. It has two parameters. The Weibull marginal survival distribution can be written as:

$$S_j(t_j|Z_{ji}) = exp\{(-\lambda_j t_j^{k_j})e^{\beta' Z_{ji}}\}, j = 1, 2, i = 1, ..., n \qquad (6)$$

where where $t_j > 0$ is the failure time, $\lambda_j > 0$ and $k_j > 0$ are the scale and shape parameters of the baseline Weibull marginal distribution, Z_{ji} are the covariates for the ith patient with j^{th} margin and β are the coefficient of covariate.

Gompertz Marginal Distribution

The Gompertz distribution is a PH model that is equal to the log-Weibull distribution, so the log of the hazard function is linear in time. The distribution was introduced by Gompertz in 1825, as a model for human mortality. This distribution has an exponentially increasing failure rate and is often appropriate for actuarial data, as the risk of mortality also increases exponentially over time. The Gompertz marginal survival distribution can be written as:

$$S_j(t_j|Z_{ji}) = exp\{-\frac{b_j}{a_j}(e^{a_j t_j} - 1)e^{\beta' Z_{ji}}\}, j = 1, 2, i = 1, ..., n \qquad (7)$$

where where $t_j > 0$ is the survival time, $b_j > 0$ and $a_j > 0$ are the scale and shape parameters of the baseline Gompertz marginal distribution, Z_{ji} are the covariates for the ith patient with j^{th} margin and β are the coefficient of covariate.

Log-logistic Marginal Distribution

The Log-logistic distribution has a fairly flexible functional form, it is one of the parametric survival time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped that is it initially increases and then decreases. In cases where one comes across censored data, using log-logistic distribution is mathematically more advantageous than other distributions[76]. The log-logistic distribution is not a PH

model, but it is a Proportional Odds (PO) model. The distribution is the natural one to use in conjunction with the proportional odds model. In fact, it is the only distribution to share both accelerated failure time property and the proportional odds property. This means that it is subject to the PO assumption, the advantage is the coefficients can be interpreted as odds ratios[77]. If survival times for individuals are assumed to have a log-logistic distribution, the Marginal survival distribution is given by:

$$S_j(t_j|Z_{ji}) = \{1 + \lambda_j t_j^{k_j} e^{\beta' Z_{ji}}\}^{-1}, j = 1, 2, i = 1, ..., n \tag{8}$$

where where $t_j > 0$ is the failure time, $\lambda_j > 0$ and $k_j > 0$ are the scale and shape parameters of the baseline Log-logistic marginal distribution, Z_{ji} are the covariates for the ith patient with j^{th} margin and β are the coefficient of covariate.

Table 3.4:   Summary of Marginal Baseline Distributions

Distribution	$S_0(t)$	$h_o(t)$	Parameter space
Weibull	$exp\{-\lambda t^k\}$	$\lambda k t^{k-1}$	$\lambda, k > 0$
Gompertz	$exp\{-\frac{b}{a}(e^{at} - 1)$	be^{at}	a,b > 0
Log-logistic	$\{1 + \lambda t^k\}^{-1}$	$\lambda k t^{k-1}/1 + \lambda t^k$	$\lambda, k > 0$

3.7.4   The Novel Two-step Estimation

Likelihood estimation is required to fit a statistical model to data and provide estimates for the model's parameters, with the most common approach being maximum likelihood estimation. Joint maximum likelihood estimation is used to obtain estimates for parameters in the marginal distribution models and the dependence parameters. In this study, each patient experiences one of the four cases: (i) $\delta_1 = \delta_2 = 1$ if both kidneys are failed, (ii) $\delta_1 = 1$ and $\delta_2 = 0$ if only right kidney is failed, (iii) $\delta_1 = 0$ and $\delta_2 = 1$ if only left kidney is failed or (iv) $\delta_1 = \delta_2 = 0$ if both kidneys are not failed. Each case has its own likelihood. Combining the four cases, the joint likelihood for the observed data $D = \{D_i\}_{i=1}^n$ can be written as

$$L_n(\theta|D) = \prod_{i=1}^n f(y_{i1}, y_{i2}|Z_{i1}, Z_{12})^{\delta_{i1}\delta_{i2}} \times \left[-\frac{\partial S(y_{i1}, y_{i2})|Z_{i1}, Z_{i2}}{\partial y_{i1}}\right]^{\delta_{i1}(1-\delta_{i2})}$$

$$\times \left[-\frac{\partial S(y_{i1}, y_{i2})|Z_{i1}, Z_{i2}}{\partial y_{i2}} \right]^{(1-\delta_{i1})\delta_{i2}} \times S(y_{i1}, y_{i2}|Z_{i1}, Z_{i2})^{(1-\delta_{i1})(1-\delta_{i2})}$$

$$= \prod_{i=1}^{n} [C_\eta \{S_1(y_{i1}|z_{i1}), S_2(y_{i2}|z_{i2})\} f_1(y_{i1}|Z_{i1}) f_2(y_{i2}|Z_{i2})]^{\delta_{i1}\delta_{i2}}$$

$$\times \left[-\frac{\partial C_\theta \{S_1(y_{i1}|Z_{i1}), S_2(y_{i2}|Z_{i2})\}}{\partial y_{i1}} \right]^{\delta_{i1}(1-\delta_{i2})} \times \left[-\frac{\partial C_\eta \{S_1(y_{i1}|Z_{i1}), S_2(y_{i2}|Z_{i2})\}}{\partial y_{i2}} \right]^{(1-\delta_{i1})\delta_{i2}}$$

$$\times C_\eta \{S_1(y_{i1}|Z_{i1}), S_2(y_{i2}|Z_{i2})\}^{(1-\delta_{i1})(1-\delta_{i2})}$$

where $(\delta_{i1}, \delta_{i2}) \in \{(0,0),(0,1),(1,0),(1,1)\}$

The estimation procedure for the unknown parameter θ is generally applicable for any selected Archimedean copula families and marginal distribution models, where $\theta = (\beta_1', \beta_2', \eta, S_{01}, S_{02})'$. In principle, we can maximize the joint log-likelihood function based on above formula directly, written as $l_n(\theta) = logL_n(\theta|D) = \sum_{i=1}^{n} logL(\theta|D_i)$. Due to the complex structure of the log-likelihood function, a novel two-step estimation procedure is used, which is proven to be computationally more stable and efficient than the one-step procedure[23]. Essentially, the two-step procedure implements an extra step to obtain appropriate initial values for all the unknown parameters. In step 1, we first obtain initial estimates of the parameters in marginal distributions (β_j, S_{0j}) based on marginal likelihood functions. Then we maximize the pseudo joint likelihood (with the initial estimates of (β_j, S_{0j}) plugged in) to get an initial estimate of the dependence parameter η. Then in step 2, we maximize the joint likelihood with estimates from step 1 being initial values to obtain the final estimate[23][65]. The estimation procedure is described below:

1. Obtain initial estimates of θ_n:

   - $\left(\hat{\beta}_{jn}^{(1)}, \hat{S}_{0j}^{(1)} \right) = \underset{(\beta_j, S_{0j})}{argmax} \; l_{jn}(\beta_j, S_{0j})$, where l_{jn} denotes the log-likelihood for the marginal model, $j = 1, 2$,

   - $\hat{\eta}_n^{(1)} = \underset{\eta}{argmax} \, l_n\{\hat{\beta}_n^{(1)} = (\hat{\beta}_{1n}^{(1)}, \hat{\beta}_{2n}^{(1)}, \eta, \hat{S}_{01}^{(1)}, \hat{S}_{02}^{(1)})\}$, where $\hat{\beta}_{jn}^{(1)}$ and $\hat{S}_{j2}^{(1)}$ are the initial estimates and l_n is the joint log-likelihood.

2. Simultaneously maximize the joint log-likelihood to get final estimates:

$$\hat{\theta}_n = (\hat{\beta}_n, \hat{\eta}, \hat{S}_{01}^{(1)}, \hat{S}_{02}^{(1)}) = \underset{(\beta, \eta, S_{01}, S_{02})}{argmax} \, l_n(\beta, \eta, S_{01}, S_{02}) \text{ with initial values}$$

$(\hat{\beta}_n^{(1)}, \eta_n^{(1)}, \hat{S}_{01}^{(1)}, \hat{S}_{02}^{(1)})\}$, obtained from step 1.

3.7.5   Model Selection

Several model selection procedures have been proposed for copula-based time to events end points models. In 2000 a model selection procedure based on nonparametric estimation of the bivariate joint survival function within the class of Archimedean copulas was proposed[78]. For model diagnostics, Chen *et al* proposed a penalized pseudo-likelihood ratio test for copula models in noncensored data[24]. Recently, Zhang *et al* proposed a goodness of fit test for copula models using the pseudo in and out of sample (PIOS) method[79]. Then Mei extended this PIOS method to censored survival data without covariates[80].

It is essential to select appropriate copula families and marginal baseline distributions. Copulas are not nested relative to each other. Thus information criteria such as: Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) is useful to choose the best-fitting copula. To explore the best statistical model which fit the kidney failure datset among Archimedean copula families and marginal baseline distribution, the researcher used AIC and BIC which are given by

$$AIC = -2logL(D; \hat{\theta}) + 2k,$$

$$BIC = -2logL(D; \hat{\theta}) + kln(n),$$

where

- k = the number of parameters estimated by the model;

- n = number of observation;

- $L(D; \hat{\theta}$ = the joint maximized value of the likelihood function of the model, where $\hat{\theta}$ are the parameter values that maximize the likelihood function

3.7.6 Model diagnostics checking

a) Evaluation of Marginal Distribution (Graphical Checking)

Regardless of which type of model is fitted and how the variables are selected to be in the model, it is important to evaluate how well the model fits the data. Model with weibull distribution has property that the log of cummulative hazard is linear with the log of time, where $S(t) = exp(-\lambda t^k)$, log of cumulative hazard is $\log(\lambda) + k\log(t)$. This property allows a graphical evaluation of the appropriateness of weibull model by plotting log of cumulative hazard versus log of time[81]. The plot should resemble a straight line if the weibull assumptions holds.

Model with the Gompertz distribution has a property that the log of hazard is linear with the time, where $S(t) = exp\{-\frac{b}{a}(e^{at} - 1)$. Hence, log of hazard is $\log(b) + at$. This property allows a graphical evaluation of the appropriateness of a Gompertz model by plotting log of hazard versus time[77]. The plot should resemble a straight line if the assumptions holds.

The log-failure odd versus log time of the log-logistic model is linear. Where the failure odds of log-logistic survival model can be computed as:

$$\frac{1 - S(t)}{S(t)} = \lambda t^k$$

$\log(1-S(t)/S(t)) = \log(\lambda) + k\log(t)$ Therefore the appropriateness of model with the log logistic distribution can be graphically evaluated by plotting $\log(1 - S(t)/S(t))$ versus log time[82]. The plot should resemble a straight line if the assumptions holds.

b) Adequacy of Archimedean copula families

To check the adequacy of Archimedean copula families it is possible to use scatter plots of joint survival distribution or scatter plot of bi-variate event times, Emura, *et al.*[83] and Wang[78]. In this study, the researcher used scatter plots of joint survival distribution for various copula families in order to choose Archimedean copula families which fit the kidney failure dataset well.

c) Checking the Assumption of Proportional Hazard

To check the proportionality assumption we could plot $ln\left(-ln(S(t))\right)$ versus t. If the two survival curves do not intersect and are parallel, it clearly provides evidence the assumption is valid. PH assumption is required for specifications of parametric Weibull and Gompertz survival models[82].

To check the assumption of proportional odds log$\{ S_i(t/[1-S_{(t)}])\}$ are plotted against log t. If the plots shows parallel straight lines, this would indicate the log-logistic model is appropriate. Parallel curves plot also suggest that the proportional odds assumptions is valid[77].

Statistical Software Used

- R software version 4.0.5 with CopulaCenR packages was used for data analysis.

4 Results and Discussion

4.1 Exploratory Data Analysis

A total of 431 patients were considered in this study. Of all patients, 170 (39.4%) failed at least one kidney during the follow-up period. From 431 patients, 51 (11.8%), 43 (10%), 76 (17.6%) failed only right kidney, only left kidney and both kidneys, respectively, while 261 (60.6%) were not failed both kidneys. The overall median kidney failure time was 897 days, while the minimum and maximum observed event times were 270 and 1080 days, respectively.

From the total of patients, 237 (55.0%) were females and 194 (45.0%) were males. During the follow-up period, 16 (3.7%), 15 (3.5%) and 42 (9.7%) of the female patients were failed only right kidney, only left kidney and both kidneys, respectively, while 35 (8.1%), 28 (6.5%) and 34 (7.9%) of the male patients were failed only right kidney, only left kidney and both kidneys, respectively.

Regards to smoking status, smoker incorporates 144 (33.4%) of the total patients where, 20 (4.6%), 21 (4.9%) and 26 (6.0%) were failed only right kidney, only left kidney and both kidneys, respectively. Looking for patients who had diabetes mellitus, 21 (4.9%), 18 (4.2%) and 32 (7.4%) were failed only right kidney, only left kidney and both kidneys, respectively. About 258 (59.8%), patients were from urban communities, 29 (6.7%), 28 (6.5%) and 54 (12.5 %) were failed only right kidney, only left kidney and both kidneys, respectively.

Majority, about 238 (55.2%) of patients who had hypertension were participated in this study, 34 (7.9%), 30 (7.0%) and 55 (12.7%) were failed only right kidney, only left kidney and both kidneys, respectively. Similarly, patients who had family history of kidneys disease incorporates 170(39.4%) of the total patients, where 26 (6.0%), 21 (4.9%) and 26 (6.0%)were failed only right kidney, only left kidney and both kidneys respectively.

Table 4.1: Descriptive statistics on number of a pair of kidneys failure

Variables	Category	Number of a pair of kidneys (%)				Total
		(0, 0)	(1, 0)	(0, 1)	(1, 1)	
Sex	Female	164(38.1%)	16(3.7%)	15(3.5%)	42(9.7%)	237(55%)
	Male	97(22.5%)	35(8.1%)	28(6.5%)	34(7.9%)	194(45%)
Residence	Rural	114(26.5%)	22(5.1%)	15(3.5%)	22(5.1%)	173(40.2%)
	Urban	147(34.1%)	29(6.7%)	28(6.5%)	54(12.5)	258(59.8%)
Smoking	Non smoker	184(42.7%)	31(7.2%)	22(5.1%)	50(11.6%)	287(66.6%)
	Smoker	77(17.9%)	20(4.6%)	21(4.9%)	26(6.0%)	144(33.4%)
Diabetes mellitus	No	183(42.4%)	30(7%)	25(5.8%)	44(10.2%)	282(65.4%)
	Yes	78(18.1%)	21(4.9%)	18(4.2%)	32(7.4%)	149(34.6%)
Hypertension	No	142(33.0%)	17(3.9%)	13(3.0%)	21(4.9%)	193(44.8%)
	Yes	119(27.6%)	34(7.9%)	30(7.0%)	55(12.7%)	238(55.2%)
Family History	No	164(38.1%)	25(5.8%)	22(5.1%)	50(11.6%)	261(60.6%)
	Yes	97(22.5%)	26(6.0%)	21(4.9%)	26(6.0%)	170(39.4%)
Alcohol	No	162(37.6%)	25(5.8%)	15(3.5%)	45(10.4%)	247(57.3%)
	Yes	99(23.0%)	26(6.0%)	28(6.5%)	31(7.2%)	184(42.7%)
Anemia	No	173(40.2%))	30(7.0%)	24(5.6%)	40(9.3%)	267(61.9%)
	Yes	88(20.4%)	21(4.8%)	19(4.4%)	36(8.3%)	164(38.1%)

Variables	Category	Number of a pair of kidneys (%)				Total
		(0, 0)	(1, 0)	(0, 1)	(1, 1)	
Obesity	No	179(41.6%)	9(2.0%)	20(4.7%)	28(6.5%)	236(54.8%)
	Yes	82(19.0%)	42(9.7%)	23(5.3%)	48(11.1%)	195(45.2%)
Age	≤35	169(39.2%)	4(0.9%)	13(3.0%)	6(1.4%)	192(44.5%)
	36-55	59(13.7%)	14(3.2%)	9(2.1%)	30(7.0%)	112(26%)
	≥56	33(7.7%)	33(7.7%)	21(4.9%)	40(9.2%)	127(29.5%)
	Total	261(60.6%)	51(11.8%)	43(10%)	76(17.6%)	431(100%)

Source: Adama Hospital Medical College, Ethiopia, from 1st January 2015 to 30th January 2020.

4.2 Statistical Analysis

Uni-variable and multi-variable analysis were applied. In uni-variable analysis, the model which contains each covariate at a time were fitted to determine variables that have the potential for being included in the multi-variable analysis. Covariates with p-value less than 25% in the uni-variable analysis were considered for multi-variable analysis. The full multi-variable analysis was fitted including all the potential covariates that were significant at 25% level of significance in uni-variable analysis. For multi-variable analysis, variables with p-value less than 5% were selected as significant covariates.

4.2.1 Uni-variable Analysis

The prognostic factors considered in the study were the sex, residence, smoking status, diabetes mellitus, hypertension, family history of kidney disease, alcohol consumption, anemia, obesity and age of patients. Outputs from uni-variable analysis (Appendix-1), covariates like: sex, diabetes mellitus, hypertension, family history of kidney disease, anemia, obesity and age (56 years and older) were significant at 25% level of significance

in all models. This indicates that they have a power to be included in the multi-variable analysis. However, Residence, smoking status and alcohol consumption were not significantly at 25% level of significance, and they were excluded from multi-variable analysis.

4.2.2 Multi-variable Analysis and Model Comparison

Similarly, multi-variable analysis were also fitted using parametric marginal distribution like: Weibull, Gompertz and Log-logistic with Archimedean copula family like: Clayton, Gumbel and Joe with significant covariates in uni-variable analysis at 25% level of significance. The researcher used AIC, BIC and Final llk to compare various candidates of parametric marginal distribution with copula family models.

From Table 4.2, Log-logistic marginal distribution with Clayton copula model has the smallest value of AIC and BIC, which are 4260.953 and 4297.548, respectively. And also the model has the higher value of Final llk which is -2121.477. This indicates that Log-logistic marginal distribution with Clayton copula model is the best statistical model that fits the kidney failure dataset well. Since the selected baseline marginal distribution is Log-logistic, no need to discuss more about proportional hazard marginal distribution.

The measure of dependence parameter Kendaell's tau τ is highest when we assume the Log-logistic marginal distribution (0.412) followed by Weibull marginal distribution (0.404) with Clayton copula model (Table 4.2).

Table 4.2: The Comparisons of the Models

Marginal distribution	Archimedean copula	AIC	Final llk	BIC	τ
	Clayton	4275.536	-2128.768	4312.131	0.404
Weibull	Gumbel	4282.812	-2132.406	4319.407	0.301
	Joe	4289.760	-2135.880	4326.355	0.241
	Clayton	4386.068	-2686.034	4414.530	0.380
Gompertz	Gumbel	4406.376	-2696.188	4434.839	0.183
	Joe	4413.636	-2699.818	4442.099	0.126
	Clayton	4260.953	-2121.477	4297.548	0.412
Log-logistic	Gumbel	4277.563	-2129.781	4314.158	0.266
	Joe	4285.078	-2133.539	4321.673	0.205

The multi-variable analysis of the Log-logistic marginal distribution with Clayton copula model is shown in Table 4.3; and the output of the other multi-variable analysis are similarly drawn (Appendix-2).

Analysis based on the Log-logistic marginal distribution with Clayton copula model, the result shows that the copula parameter η is significant at 5% level of significance. This indicates the time to right and left kidneys failure of the patient is dependent. Kendall's tau (τ) is 0.4123. This shows that the dependence between the time to right and left kidney failure of the patient is 41.2%.

According to the selected model hypertension, family history, obesity and age (56 years and older) were significant at 5% level of significance. This indicates that they are the most risk factor that leads to kidney failure of the patients. However, according to the model anemia, age (36-55 years old) and diabetes mellitus were not significant at 5% level of significance (Table 4.3).

Table 4.3: The Log-logistic-Clayton Copula Multi-variable Analysis

Variables	Estimate	SE	P-Value	OR(95% CI)
Sex (Male)	0.3678	0.1836	0.0383	1.4445[1.0079, 2.0703]
Diabetes mellitus (Yes)	0.0957	0.1828	0.6007	1.1004[0.7690, 1.5460]
Hypertension (Yes)	0.7466	0.2227	0.0008	2.1098[1.3636, 3.2645]
Family history (Yes)	0.4221	0.1943	0.0299	1.5252[1.0422, 2.2320]
Anemia (Yes)	0.1816	0.1943	0.2044	1.1991[0.8135, 1.7549]
Obesity (Yes)	0.4051	0.2024	0.0394	1.4995[1.0084, 2.2296]
Age (36-55)	0.0358	0.2228	0.8725	1.0364[0.6697, 1.6040]
Age ($\geq$ 56)	0.6459	0.2664	0.0153	1.9077[1.1318, 3.2156]
η	1.4034	0.2653	< 0.0001	

Final llk = -2121.477		$\tau = 0.412$
	AIC = 4260.953	

Source: Adama Hospital Medical College, Ethiopia; from 1ˢᵗ January 2015 to 30ᵗʰ January 2020.

The odds ratio for the patients who had hypertension is 2.1098[95%CI: 1.3636, 3.2645] and p-value is 0.0008. This shows that the odds of time to kidney kidney failure for the patients who had hypertension is twice that for the patients who had not hypertension.

The odds ratio for the patient who had obesity is 1.4995[95%CI: 1.0084, 2.2296] and p-value is 0.0394. This shows that the odds of time to kidney failure for the patients who had obesity is 50.0% more than as compared to the patients who had not obesity. The odds ratio for the patients who had family history of kidney disease is 1.5252[95%CI: 1.0422, 2.2320] and p-value is 0.0299. This shows that the odds of time to kidney failure for the patients who had family history of kidney disease is 52.5% more than as compared to the patients who had no family history of kidney disease.

The odds ratio for male patients is 1.4445[95%CI: 1.0079, 2.0703] and p-value is 0.0383. This shows that the odds of time to kidney failure for the male patient is 44.5% more

than as compared to the female patients. The odds ratio for patients age (56 years and older) is 1.9077[95%CI: 1.1318, 3.2156] and p-value is 0.0153. This shows that the odds of time to kidneys failure for the patients age (56 years and older) is 1.9077 times of the patients age less than 35.

4.3   Model Diagnostics

a) Evaluation of Marginal Baseline Distribution

The plots of Log-logistic and Weibull are more linear than the Gompertz plot, but the Log-logistic plot is more linear than the Weibull plot, only few observations are scattered at the beginning time. The patterns suggest that the Log-logistic marginal distribution fits the kidney failure dataset well.
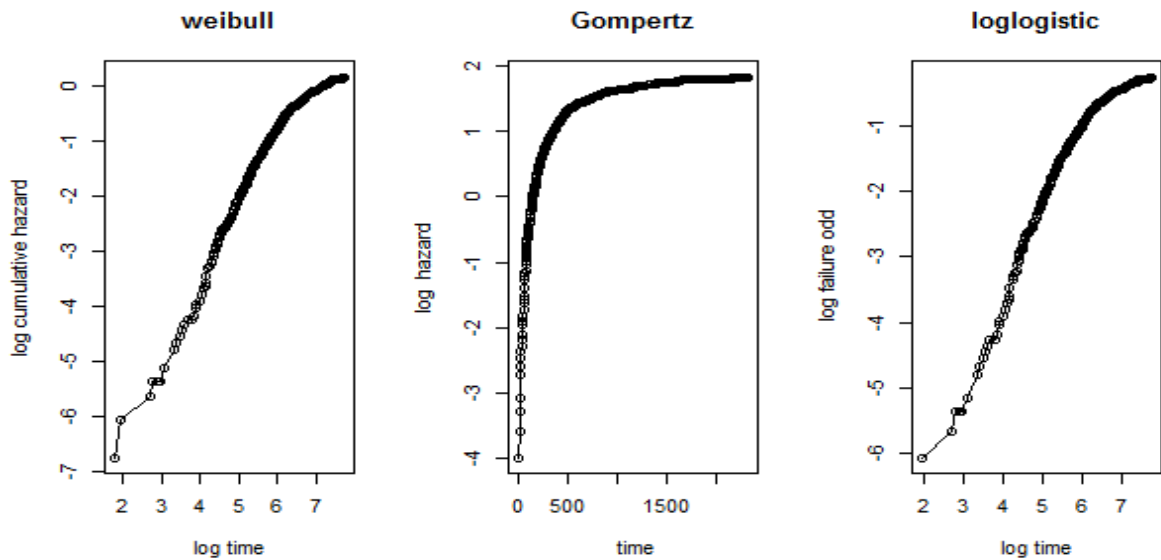
Figure 4.1: Graphical Evaluation of Marginal Distributions

b) Adequacy of Archimedean Copula Families

The scatter plot of joint survival distribution shows that the dependence between the time to right and left kidney failure of the patient is positive. Clayton scatter plot shows the time to right and left kidney failure of the patient appear to behave more

closely or condense than Gumbel and Joe. The scatter plot suggests that the Clayton copula family fit the kidney failure dataset well.
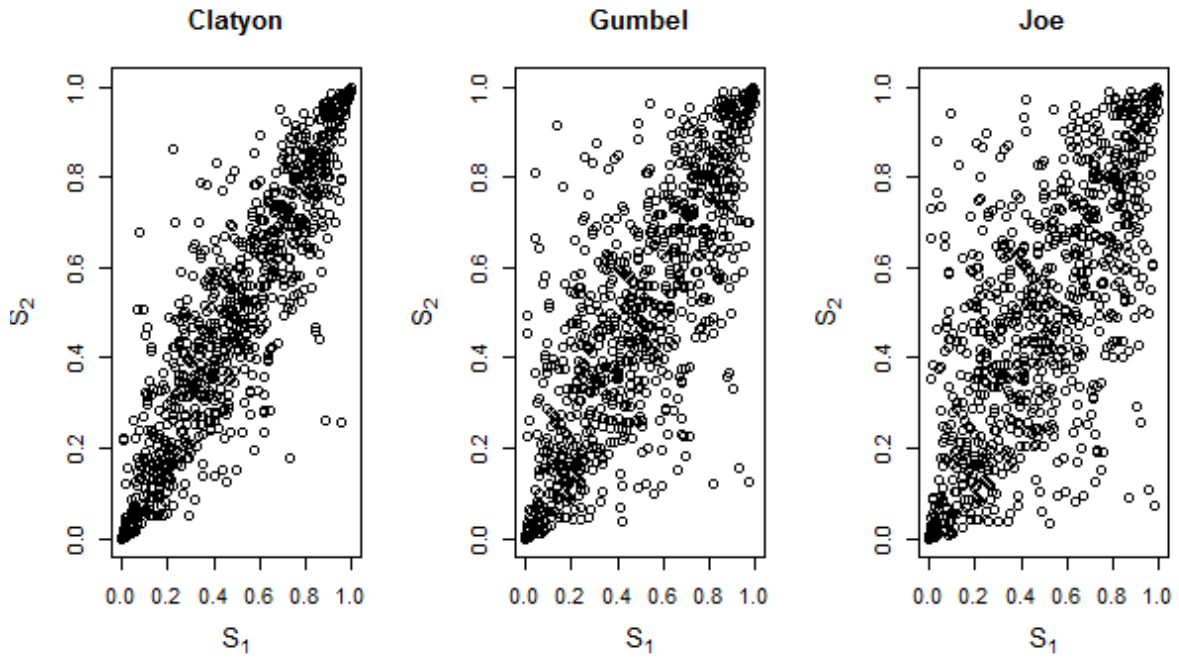
Figure 4.2: The Scatter Plots of Joint Survival Distribution for Copulas Family

c) Checking the Assumption of Proportional Odds

The plot shows a parallel curve. Parallel curves plot indicates that the proportional odds assumption is valid (Appendix-3 for other covariate).
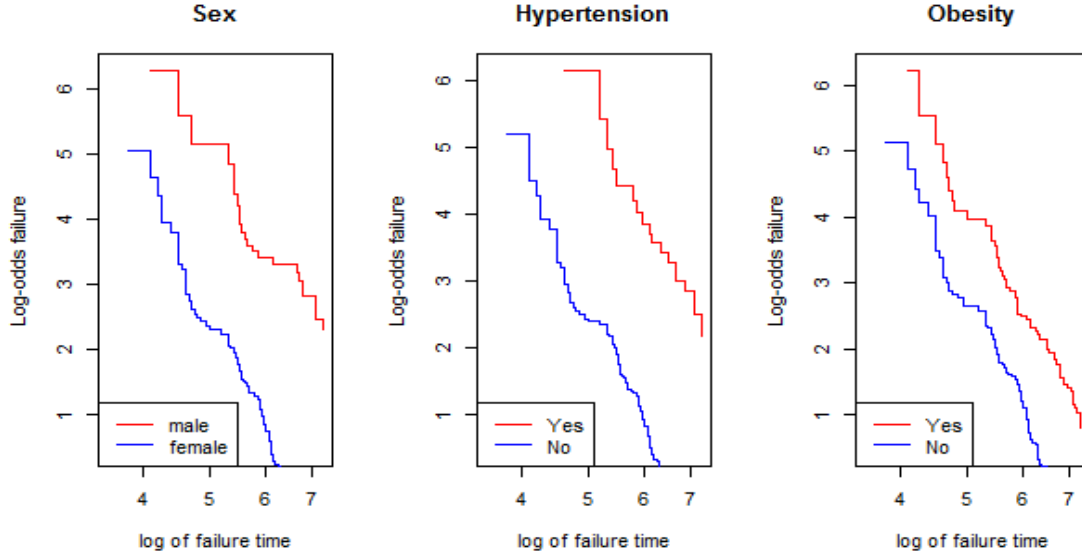
4.4   Discussion

The descriptive result shows that of all 431 patients, 170 (39.4%) failed at least one kidney during the follow-up period, while 261 (60.6%) were not failed both kidneys. The overall median kidney failure time was 897 days, while the minimum and maximum observed event times were 270 and 1080 days, respectively. The level of dependence between the time to right and left kidneys failure of the patient was 41.2%.

The findings of this study revealed that the failure of one side kidney (either right or left kidney failure) predict the failure of other side kidney of the patient ($\eta = 1.4034$ and p-value $= < 0.0001$). This might be due to the facts that a pair of kidneys in a patient is more alike than a pair of kidneys from different patients because of the genetic influences. This consolidates the idea that the failure times of paired human organs are correlated as they come from the same subject[17][18][23].

According to the result of the Log-logistic marginal distribution with Clayton copula analysis, the results suggest that hypertension is significantly associated with the time to kidney failure of the patients. The odds ratio for the patients who had hypertension

34

is 2.110, suggesting that the odds of time to kidney failure for the patients who had hypertension is twice that of the patients who had no hypertension. This is consistent with the study conducted by Vaes, et al. at the Veterans Administration Medical Center of New Orleans. The study revealed that hypertension was a significant risk factor for the development of ESRD (HR = 1.25 (95% CI: 1.22 to 1.28))[42]. This may be due to over time, uncontrolled high blood pressure can cause arteries around the kidneys to narrow, weaken or harden. These damaged arteries are not able to deliver enough blood to the kidney tissue. The findings of this study also consistent with the previous studies like[43][44]

Accordingly, the results of this study suggest that obesity is significantly associated with the time to kidney failure of the patients. The odds ratio for the patient who had obesity is 1.50, shows that the odds of time to kidney failure for the patient who had obesity is higher than for the patients who had no obesity. This might be due to extra weight forces the kidneys to work harder and filter waste above the normal level. Over time, this extra work increases the risk of kidney disease. This result is consistent with the study[50].

The findings of this study reveals that family history of kidney disease is significantly associated with the time to kidney failure of the patients. The odds ratio for the patient who had family history of kidney disease is 1.525, shows that the odds of time to kidney failure for the patient who had a family history of kidney disease is higher than for the patients who had no family history of kidney disease. This is in line with the study conducted by Mekiya, *et al.* in order to analyze survival analysis of patients with ESRD using cox regression[15]. The study suggested that family history was a significant risk factor for the ESRD (HR = 1.88 [95% CI: 1.2 to 2.9]). This study is agree with the findings in other studies like[52][53].

Similarly, the sex of patients is significantly associated with the time to kidney failure of the patients. The odds ratio for male patients is 1.445, indicating that the odds of time to kidney failure for male patients is higher than female patients. This is consistent with the study conducted by Iseki, *et al.* in order to analyze the risk factor associated with ESRD in Japan[57]. The study showed that the odds ratio for ESRD

was 1.41 among male participants. This may be due to the higher testosterone levels in men cause a loss in kidney function.

The findings of this study also shows that age is significantly associated with the time to kidney failure of the patients. The odds ratio for older adulthood patient is 1.908, showing that the odds of kidney failure for older adulthood patients is higher than for young adulthood patients. This may be due to as age increases the amount of kidney tissue decreases and kidney function diminishes. This is consistent with the study conducted by Chadban, *et al.* on individual-level meta-analysis[60]. The study revealed that age was a significant risk factor for the development of ESRD (the older adulthood compared to young adulthood with a OR = 2.5 (95% CI: 1.8 to 3.2)).

5 Conclusion and Recommendation

5.1 Conclusion

This study revealed that some variables considered in this study have a significant association with the time to kidney failure of the patients. Based on the findings in the preceding chapter, this study arrives at the following conclusions.

Based on exploratory data analysis, of all 431 patients, 170 (39.4%) failed at least one kidney during the follow-up period. The overall median kidney failure time was low. The level of dependence between the time to right and left kidney failure of the patient was strong.

The Log-logistic marginal distribution with Clayton copula model is the best statistical model that describes the kidney failure dataset well. The result of Log-logistic marginal distribution with Clayton copula model shows that being male, older adulthood, obese, hypertensive and having family history of kidney disease are the most risk factors that significantly associate with time to kidney failure of the patients.

Furthermore, the copula parameter showed that there is the dependence between the time to right and left kidney failure of the patient.

5.2 Recommendations

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

- As hypertensive and obese are the risk factor of time to kidney failure of the patients, controlling the high blood pressure and overweight might prevent the onset of kidney failure.

- Kidney disease runs in families, so having periodic screenings for kidney disease is recommended.

- The time to kidney failure risk is high for older adults, so it is better to give special care to them.

- As at least one kidney failure predicts the failures of the other one, it is better to treat the damaged (failed) one before it gets worse.

- This study is only modeling the time to kidney failure of the patients using the copula model. But patients come to Adama Hospital Medical College from different clusters (regions). Hence, future study should apply other models to account for the effect of clusters.

- Further research should apply copula models for bi-variate time to events endpoint, in order to assess the dependence between the bi-variate events time.

References

[1] Russell, C. D., Bischoff, P. G., Kontzen, F. N., Rowell, K. L., Yester, M. V., Lloyd, L. K., ... & Dubovsky, E. V. (1985). Measurement of glomerular filtration rate: single injection plasma clearance method without urine collection. Journal of nuclear medicine, 26(11), 1243-1247.

[2] Levey, A. S., Coresh, J., Balk, E., Kausz, A. T., Levin, A., Steffes, M. W., ... & Eknoyan, G. (2003). National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Annals of internal medicine, 139(2), 137-147.

[3] Thurlow, J. S., Joshi, M., Yan, G., Norris, K. C., Agodoa, L. Y., Yuan, C. M., & Nee, R. (2021). Global epidemiology of end-stage kidney disease and disparities in kidney replacement therapy. American journal of nephrology, 52(2), 98-107.

[4] Wilson, R., Krefting, L. H., Sutcliffe, P., & VanBussel, L. (2010). Incidence and prevalence of end-stage renal disease among Ontario's James Bay Cree. Canadian journal of public health = Revue canadienne de sante publique, 83(2), 143-146.

[5] Collins, A. J., Foley, R. N., Gilbertson, D. T., & Chen, S. C. (2015). United States Renal Data System public health surveillance of chronic kidney disease and end-stage renal disease. Kidney international supplements, 5(1), 2-7.

[6] Mousavi, S. S. B., Soleimani, A., & Mousavi, M. B. (2014). Epidemiology of end-stage renal disease in Iran: a review article. Saudi Journal of Kidney Diseases and Transplantation, 25(3), 697.

[7] Ethier, I., Johnson, D. W., Bello, A. K., Ye, F., Osman, M. A., Levin, A., ... & Board, S. E. A. R. (2021). International Society of Nephrology Global Kidney Health Atlas: structures, organization, and services for the management of kidney failure in Oceania and South East Asia. Kidney International Supplements, 11(2), e86-e96.

[8] Luyckx, V. A., Tonelli, M., & Stanifer, J. W. (2018). The global burden of kidney disease and the sustainable development goals. Bulletin of the World Health Organization, 96(6), 414.

[9] Naicker, S. (20013). End-stage renal disease in sub-Saharan and South Africa. Kidney International, 63, S119-S122.

[10] Ashuntantang, G., Osafo, C., Olowu, W. A., Arogundade, F., Niang, A., Porter, J., ... & Luyckx, V. A. (2017). Outcomes in adults and children with end-stage kidney disease requiring dialysis in sub-Saharan Africa: a systematic review. The Lancet Global Health, 5(4), e408-e417.

[11] Wang, V., Vilme, H., Maciejewski, M. L., & Boulware, L. E. (2016, July). The economic burden of chronic kidney disease and end-stage renal disease. In Seminars in nephrology (Vol. 36, No. 4, pp. 319-330). WB Saunders.

[12] Liyanage, T., Ninomiya, T., Jha, V., Neal, B., Patrice, H. M., Okpechi, I., ... & Perkovic, V. (2015). Worldwide access to treatment for end-stage kidney disease: a systematic review. The Lancet, 385(9981), 1975-1982.

[13] Walt, G. (2004). WHO's World Health Report 2003: shaping the future depends on strengthening health systems. British Medical Journal, 328(7430), 6-7.

[14] Kore, C. (2020). The magnitude of chronic kidney disease and its risk factors at Zewditu Memorial Hospital, Addis Ababa, Ethiopia. J Clin Nephrol Ther 2020.

[15] Hussein, M., Muleta, G., Seyoum, D., Kifle, D., & Bedada, D. (2017). Survival Analysis of Patients with End Stage Renal Disease the Case of Adama Hospital, Ethiopia. Clinical Medicine Research. Vol. 6, No. 6, 2017, pp. 201-208.

[16] Hanagal, D. D., Pandey, A., & Ganguly, A. (2017). Correlated gamma frailty models for bivariate survival data. Communications in Statistics-Simulation and Computation, 46(5), 3627-3644.

[17] Hougaard P. (2000) Dependence structures. In: Analysis of Multivariate Survival Data. Statistics for Biology and Health. Springer, New York, NY.

[18] Mahé, C., & Chevret, S. (1999). Estimating regression parameters and degree of dependence for multivariate failure time data. Biometrics, 55(4), 1078-1084.

[19] Rondeau, V., Pignon, J. P., Michiels, S., & Mach-NC Collaborative Group. (2015). A joint model for the dependence between clustered times to tumour progression and deaths: A meta-analysis of chemotherapy in head and neck cancer. Statistical methods in medical research, 24(6), 711-729.

[20] Emura, T., Matsui, S., & Rondeau, V. (2019). Survival Analysis with Correlated Endpoints: Joint Frailty-Copula Models. Springer Singapore.

[21] Geerdens, C., Claeskens, G., & Janssen, P. (2016). Copula based flexible modeling of associations between clustered event times. Lifetime data analysis, 22(3), 363-381.

[22] Nelsen, R. B. (2007). An introduction to copulas. Journal of Multivariate Analysis, 60(1), 111-122.

[23] Sun, T., & Ding, Y. (2020). CopulaCenR: Copula based Regression Models for Bivariate Censored Data in R. The R Journal, 12(1), 266-282.

[24] Chen, M. H., Tong, X., & Zhu, L. (2013). A linear transformation model for multivariate interval-censored failure time data. Canadian Journal of Statistics, 41(2), 275-290.

[25] Khan, T. M., & Khan, K. N. M. (2015). Acute Kidney Injury and Chronic Kidney Disease. Veterinary Pathology, 52(3), 441–444.

[26] Sanyaolu, A., Okorie, C., Annan, R., Turkey, H., Akhtar, N., Gray, F., & Nwaduwa, I. C. (2018). Epidemiology and management of chronic renal failure: a global public health problem. Biostatistics Epidemiol Int J, 1(1), 00005.

[27] Parmar, M. S. (2002). Chronic renal disease. Bmj, 325(7355), 85-90.

[28] Woo, K. T., Choong, H. L., Wong, K. S., Tan, H. B., & Chan, C. M. (2012). The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. Kidney international, 81(10), 1044-1045.

[29] Lewis, J. (2010). Racial differences in chronic kidney disease (CKD) and end-stage renal disease (ESRD) in the United States: a social and economic dilemma. Clinical nephrology, 74, S72-7.

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

[31] Arogundade, F. A., & Barsoum, R. S. (2008). CKD prevention in Sub-Saharan Africa: a call for governmental, nongovernmental, and community support. American Journal of Kidney Diseases, 51(3), 515-523.

[32] Durand, A. C., Jouve, E., Delarozière, J. C., Boucekine, M., Izaaryene, G., Crémades, A., ... & Gentile, S. (2018). End-stage renal disease treated in Provence-Alpes Côte d'Azur: 12-years follow-up and forecast to the year 2030. BMC nephrology, 19(1), 1-11.

[33] Ogundele, S. B. (2018). Chronic kidney disease in sub-Saharan Africa. Saudi Journal of Kidney Diseases and Transplantation, 29(5), 1188.

[34] Van Biesen, W., Jha, V., Abu-Alfa, A. K., Andreoli, S. P., Ashuntantang, G., Bernieh, B., ... & Garcia, G. G. (2020). Considerations on equity in management of end-stage kidney disease in low-and middle-income countries. Kidney international supplements, 10(1), e63-e71.

[35] Xie, Y., Bowe, B., Mokdad, A. H., Xian, H., Yan, Y., Li, T., ... & Al-Aly, Z. (2018). Analysis of the Global Burden of Disease study highlights the global, regional, and national trends of chronic kidney disease epidemiology from 1990 to 2016. Kidney international, 94(3), 567-581.

[36] Facts, G. (2015). About Kidney Disease [Internet]. The National Kidney Foundation, 2016.

[37] Kaze, A. D., Ilori, T., Jaar, B. G., & Echouffo-Tcheugui, J. B. (2018). Burden of chronic kidney disease on the African continent: a systematic review and meta-analysis. BMC nephrology, 19(1), 1-11.

[38] Okpechi, I. G. (2017). ESKD in sub-Saharan Africa: will governments now listen?. The Lancet Global Health, 5(4), e373-e374.

[39] Elsharif, M. E., GARIBALLA, E. E., & GADOUR, M. (2010). Costs of Hemodialysis and Kidney Transplantation in Sudan A Single Center Experience.. IRANIAN JOURNAL OF KIDNEY DISEASES (IJKD), 4(4), 282-284.

[40] Kassa, D. A., Mekonnen, S., Kebede, A., & Haile, T. G. (2020). Cost of Hemodialysis Treatment and Associated Factors Among End-Stage Renal Disease Patients at the Tertiary Hospitals of Addis Ababa City and Amhara Region, Ethiopia. ClinicoEconomics and Outcomes Research: CEOR, 12, 399.

[41] Crews, D. C., Bello, A. K., & Saadi, G. (2019). 2019 World Kidney Day Editorial-burden, access, and disparities in kidney disease. Brazilian Journal of Nephrology, 41(1), 1-9.

[42] Vaes, B., Beke, E., Truyers, C., Elli, S., Buntinx, F., Verbakel, J. Y., ... & Van Pottelbergh, G. (2015). The correlation between blood pressure and kidney function decline in older people: a registry-based cohort study. BMJ open, 5(6), e007571.

[43] Vupputuri, S., Batuman, V., Muntner, P., Bazzano, L. A., Lefante, J. J., Whelton, P. K., & He, J. (2003). Effect of blood pressure on early decline in kidney function among hypertensive men. Hypertension, 42(6), 1144-1149.

[44] Tozawa, M., Iseki, K., Iseki, C., Kinjo, K., Ikemiya, Y., & Takishita, S. (2003). Blood pressure predicts risk of developing end-stage renal disease in men and women. Hypertension, 41(6), 1341-1345.

[45] Collins, A. J., Foley, R. N., Gilbertson, D. T., & Chen, S. C. (2015). United States Renal Data System public health surveillance of chronic kidney disease and end-stage renal disease. Kidney international supplements, 5(1), 2-7.

[46] Gheith, O., Farouk, N., Nampoory, N., Halim, M. A., & Al-Otaibi, T. (2016). Diabetic kidney disease: world wide difference of prevalence and risk factors. Journal of nephropharmacology, 5(1), 49.

[47] Foley, R. N., Murray, A. M., Li, S., Herzog, C. A., McBean, A. M., Eggers, P. W., & Collins, A. J. (2005). Chronic kidney disease and the risk for cardiovascular disease, renal replacement, and death in the United States Medicare population, 1998 to 1999. Journal of the American Society of Nephrology, 16(2), 489-495.

[48] Fox, C. S., Larson, M. G., Leip, E. P., Culleton, B., Wilson, P. W., & Levy, D. (2004). Predictors of new-onset kidney disease in a community-based population. Jama, 291(7), 844-850.

[49] Ejerblad, E., Fored, C. M., Lindblad, P., Fryzek, J., Dickman, P. W., Elinder, C. G., ... & Nyrén, O. (2004). Association between smoking and chronic renal failure in a nationwide population-based case-control study. Journal of the American Society of Nephrology, 15(8), 2178-2185.

[50] Nguyen, S., & Hsu, C. Y. (2007). Excess weight as a risk factor for kidney failure. Current opinion in nephrology and hypertension, 16(2), 71-76.

[51] Mirrakhimov, A. E. (2012). Obstructive sleep apnea and kidney disease: is there any direct link?. Sleep and Breathing, 16(4), 1009-1016.

[52] Song, E. Y., McClellan, W. M., McClellan, A., Gadi, R., Hadley, A. C., Krisher, J., ... & Freedman, B. I. (2009). Effect of community characteristics on familial clustering of end-stage renal disease. American journal of nephrology, 30(6), 499-504.

[53] Elemam, A. M., Ismail, A. M., & Mohammed, M. I. (2019). Obese Family Members of Chronic Renal Failure Patients are at Higher Risk for Developing Kidney Diseases: In a Cross-sectional Study. Sudan Journal of Medical Sciences, 14(3).

[54] Li, G. (2015). Genetic factors for end-stage renal disease. Journal of Integrative Nephrology and Andrology, 2(2), 46.

[55] Tangri, N., Stevens, L. A., Griffith, J., Tighiouart, H., Djurdjev, O., Naimark, D., ... & Levey, A. S. (2011). A predictive model for progression of chronic kidney disease to kidney failure. Jama, 305(15), 1553-1559.

[56] Takamatsu, N., Abe, H., Tominaga, T., Nakahara, K., Ito, Y., Okumoto, Y., ... & Doi, T. (2009). Risk factors for chronic kidney disease in Japan: a community-based study. BMC nephrology, 10(1), 1-10.

[57] Iseki, K. (2005). Factors influencing the development of end-stage renal disease. Clinical and experimental nephrology, 9(1), 5-14.

[58] Kramer, A., Pippias, M., Noordzij, M., Stel, V. S., Afentakis, N., Ambühl, P. M., ... & Jager, K. J. (2018). The european renal association–european dialysis and transplant association (ERA-EDTA) registry annual report 2015: a summary. Clinical kidney journal, 11(1), 108-122.

[59] Ravani, P., Quinn, R., Fiocco, M., Liu, P., Al-Wahsh, H., Lam, N., ... & Tonelli, M. (2020). Association of Age With Risk of Kidney Failure in Adults With Stage IV Chronic Kidney Disease in Canada. JAMA network open, 3(9), e2017150-e2017150.

[60] Chadban, S. I., Matsushita, K., Sang, Y., Mahmoodi, B. K., Black, C., Ishani, A., ... & Chronic Kidney Disease Prognosis Consortium, F. T. (2012). Age and association of kidney measures with mortality and end-stage renal disease. Jama, 308(22), 2349-2360.

[61] Süleymanlar, G., Utaş, C., Arinsoy, T., Ateş, K., Altun, B., Altiparmak, M. R., ... & Serdengeçti, K. (2011). A population-based survey of Chronic Renal Disease In Turkey—the CREDIT study. Nephrology Dialysis Transplantation, 26(6), 1862-1871.

[62] Choi, H. S., Han, K. D., Oh, T. R., Kim, C. S., Bae, E. H., Ma, S. K., & Kim, S. W. (2019). Smoking and risk of incident end-stage kidney disease in general population: A Nationwide Population-based Cohort Study from Korea. Scientific reports, 9(1).

[63] Orth, S. R., Schroeder, T., Ritz, E., & Ferrari, P. (2005). Effects of smoking on renal function in patients. Nephrology Dialysis Transplantation, 20(11), 2414-2419.

[64] Reynolds, K., Gu, D., Chen, J., Tang, X., Yau, C. L., Yu, L., ... & He, J. (2008). Alcohol consumption and the risk of end-stage renal disease among Chinese population. Kidney international, 73(7), 870-876.

[65] Liu, Y. (2017). Novel Single and Gene-Based Test Procedures for Large-Scale Bivariate Time-to-Event Data, with Application to a Genetic Study of AMD Progression (Doctoral dissertation, University of Pittsburgh).

[66] Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American statistical association, 84(408), 1065-1073.

[67] Kotz, S., Johnson, H. L.,& Read, C. B. (1982). Encyclopedia of statistical sciences (No. 519.5 E5).

[68] Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris, 8, 229-231.

[69] Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65(1), 141-151.

[70] Burzykowski T, Molenberghs G, Buyse M. (Eds.) (2005). The Evaluation of Surrogate Endpoints. New York: Springer.

[71] Peng M, Xiang L, Wang S. (2018). Semiparametric regression analysis of clustered survival data with semicompeting risks. Compt Stat Data Anal, 124:53-70.

[72] Joe, Harry. Multivariate models and multivariate dependence concepts. CRC Press, 1997.

[73] Gumbel, E. J. (1960). Bivariate exponential distributions. Journal of the American Statistical Association, 55(292), 698-707.

[74] Montaseri, M., Charati, J. Y., & Espahbodi, F. (2016). Application of parametric models to a survival analysis of hemodialysis patients. Nephro-urology monthly, 8(6).

[75] Kim, G., Silvapulle, M. J., & Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. Computational Statistics & Data Analysis, 51(6), 2836-2850.

[76] Gupta, R. D., & Kundu, D. (1999). Theory & methods: Generalized exponential distributions. Australian & New Zealand Journal of Statistics, 41(2), 173-188.

[77] Collett, D. (2003). Modelling survival data in medical research (No. 610.72 C721m). Florida, US: CRC Press.

[78] Wang, A. (2010). Goodness-of-fit tests for Archimedean copula models. Statistica Sinica, 441-453.

[79] Zhang, S., Okhrin, O., Zhou, Q. M., & Song, P. X. K. (2016). Goodness-of-fit test for specification of semiparametric copula dependence models. Journal of Econometrics, 193(1), 215-233.

[80] Mei, M. (2016). A goodness-of-fit test for semi-parametric copula models of right-censored bivariate survival times. Statistics in biosciences, 2(2), 154-179.

[81] Dätwyler, C., & Stucki, T. (2011). Parametric survival models. International Journal of Statistics and Probability, 6(6), 85-91.

[82] Klein, J. P., & Moeschberger, M. L. (2003). Survival analysis: techniques for censored and truncated data (Vol. 1230). New York: Springer.

[83] Emura, T., Lin, C. W., & Wang, W. (2010). A goodness-of-fit test for Archimedean copula models in the presence of right censoring. Computational Statistics & Data Analysis, 54(12), 3033-3043.

Appendices

Reference category are: Sex (Female), Residence (Rural), Smoking (Non-smoker), Diabetes mellitus (No), Hypertension (No), Family history (No), Alcohol consumption (No), Anemia (No), Obesity (No), Age (≤ 35).

Appendix-1: Uni-variable Analysis

Table 1: Wibull-Clayton Uni-variable Analysis

Variables	Estimate	SE	P-Value
Sex (Male)	0.3665	0.1460	0.0121*
η	2.0773	0.3407	< 0.0001*
Residence (Urban)	0.0393	0.1375	0.7751
η	2.1326	0.3518	<0.0001*
Smoking (Smoker)	0.0570	0.1403	0.6843
η	2.1279	0.3510	< 0.0001*
Diabetes mellitus (Yes)	0.2752	0.1365	0.0439*
η	2.0836	0.3424	< 0.0001*
Hypertension (Yes)	0.9268	0.1552	< 0.0001*
η	2.0639	0.3392	< 0.0001*
Family history (Yes)	0.6452	0.1403	< 0.0001*
η	2.1429	0.3509	< 0.0001*
Alcohol consumption (Yes)	0.1367	0.1343	0.3086
η	2.1166	0.3484	< 0.0001*
Anemia (Yes)	0.5845	0.1366	< 0.0001*
η	2.0892	0.3439	< 0.0001*
Obesity (Yes)	0.7966	0.1450	< 0.0001*
η	2.1199	0.3471	< 0.0001*
Age (36-55)	0.0054	0.1739	0.9751
Age (≥ 56)	0.6207	0.2653	0.0240*
η	2.1870	0.3590	< 0.0001*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 2: Gompertz-Clayton Uni-variable Analysis

Variables	Estimate	SE	P-Value
Sex (Male)	0.3077	0.1207	0.0108*
η	1.9075	0.2988	< 0.0001*
Residence (Urban)	0.1270	0.1588	0.4237
η	1.9000	0.3096	< 0.0001*
Smoking (Smoker)	0.0396	0.1332	0.7661
η	1.9340	0.2974	< 0.0001*
Diabetes mellitus (Yes)	0.2656	0.1238	0.0319*
eta	1.9326	0.2905	< 0.0001*
Hypertension (Yes)	0.8096	0.1623	< 0.0001*
η	1.8794	0.2944	< 0.0001*
Family history (Yes)	0.5888	0.1181	< 0.0001*
η	1.8782	0.2942	< 0.0001*
Alcohol consumption (Yes)	0.1271	0.1293	0.3259
η	1.9224	0.2922	< 0.0001*
Anemia (Yes)	0.4886	0.1407	0.0005*
η	1.9378	0.3035	< 0.0001*
Obesity (Yes)	0.7277	0.1597	< 0.0001*
η	1.9067	0.2898	< 0.0001*
Age (36-55)	0.0046	0.1850	0.9801
Age (≥ 56)	0.6685	0.3216	0.0376*
η	1.9091	0.2902	< 0.0001*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 3: Log-logistic-Clayton Uni-variable Analysis

Variables	Estimate	SE	P-Value
Sex (Male)	0.3806	0.1844	0.0390*
η	1.9253	0.3120	< 0.0001*
Residence (Urban)	0.1997	0.1774	0.2602
η	1.8940	0.3112	< 0.0001*
Smoking (Smoker)	0.0411	0.1803	0.8198
η	1.9510	0.3161	< 0.0001*
Diabetes Mellitus (Yes)	0.2543	0.1750	0.1461*
η	1.9549	0.3160	< 0.0001*
Hypertension (Yes)	1.0534	0.1935	< 0.0001*
η	1.9451	0.3144	< 0.0001*
Family history (Yes)	0.7754	0.1789	< 0.0001*
η	1.9469	0.3147	< 0.0001*
Alcohol consumption (Yes)	0.1498	0.1736	0.3884
η	1.9400	0.3150	< 0.0001*
Anemia (Yes)	0.5902	0.1754	0.0008*
η	1.9499	0.3152	< 0.0001*
Obesity (Yes)	0.8931	0.1824	< 0.0001*
η	1.9481	0.3168	< 0.0001*
Age (36-55)	0.0225	0.2226	0.9195
Age (≥ 56)	0.6399	0.2654	0.0159*
η	1.9598	0.3168	< 0.0001*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 4: Weibull-Gumbel Uni-variable Analysis

Variable	Estimate	SE	P-Value
Sex (Male)	0.3820	0.1521	0.0120*
η	1.5779	0.0973	< 0.0001*
Residence (Urban)	0.1803	0.1548	0.2529
η	1.6070	0.0999	< 0.0001*
Smoking (Smoker)	0.1726	0.1585	0.2612
η	1.5913	0.0976	< 0.0001*
Diabetes mellitus (Yes)	0.4264	0.1423	0.0027*
η	1.5771	0.0964	< 0.0001*
Hypertension (Yes)	1.1053	0.1613	< 0.0001*
η	1.5478	0.0949	< 0.0001*
Family History (Yes)	0.8924	0.1452	< 0.0001*
η	1.5955	0.0981	< 0.0001*
Alcohol consumption (Yes)	0.1816	0.1943	0.3501
η	1.5931	0.0979	< 0.0001*
Anemia (Yes)	0.7306	0.1423	< 0.0001*
η	1.5584	0.0954	< 0.0001*
Obesity (Yes)	1.0184	0.1520	< 0.0001*
η	1.5576	0.0955	< 0.0001*
Age (36-55)	0.0107	0.1739	0.9511
Age (≥ 56)	0.6498	0.2668	0.0140*
η	1.5558	0.0925	< 0.0001*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 5: Gompertz-Gumbel Uni-variable Analysis

Variables	Estimate	SE	P-Value
Sex (Male)	0.2995	0.1402	0.03271*
η	1.2567	1.7603	0.0483*
Residence (Urban)	0.1268	0.1598	0.4024
η	1.2637	1.6903	0.0172*
Smoking (Smoker)	0.1716	0.1575	0.2757
η	1.2635	1.7306	0.03373*
Diabetes mellitus (Yes)	0.3269	0.1381	0.0179*
η	1.2614	1.7247	0.0318*
Hypertension (Yes)	0.9657	0.1573	< 0.0001*
η	1.2309	1.6802	0.0204*
Family History (Yes)	0.7971	0.1227	< 0.0001*
η	1.2564	1.7352	0.0367*
Alcohol consumption (Yes)	0.1217	0.1146	0.3388
η	1.2620	1.7571	0.0455*
Anemia (Yes)	0.6106	0.1615	0.0002*
η	1.2432	1.9201	0.0495*
Obesity (Yes)	0.9091	0.1695	< 0.0001*
η	1.2458	1.8990	0.0489*
Age (36-55)	0.0034	0.1960	0.9860
Age (≥ 56)	0.6278	0.2634	0.0156*
η	1.2333	1.5900	0.0004*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 6: Log-logistic-Gumbel Uni-variable Analysis

Variables	Estimate	SE	P-Value
Sex (Male)	0.3829	0.1882	0.0419*
η	1.4529	0.0867	< 0.0001*
Residence (Urban)	0.1961	0.1852	0.2898
η	1.4693	0.0877	< 0.0001*
Smoking (Smoker)	0.1829	0.1813	0.3542
η	1.4616	0.0867	< 0.0001*
Diabetes mellitus (Yes)	0.3157	0.1793	0.0782*
η	1.4573	0.0865	< 0.0001*
Hypertension (Yes)	1.3023	0.1986	< 0.0001*
η	1.4468	0.0863	< 0.0001*
Family History (Yes)	1.0995	0.1823	< 0.0001*
η	1.4614	0.0870	< 0.0001*
Alcohol Consumption (Yes)	0.1754	0.1772	0.2645
η	1.4601	0.0866	< 0.0001*
Anemia (Yes)	0.7466	0.1786	< 0.0001*
η	1.4614	0.0872	< 0.0001*
Obesity (Yes)	1.1825	0.1882	< 0.0001*
η	1.4570	0.0869	< 0.0001*
Age (36-55)	0.0244	0.2210	0.9122
Age (≥ 56)	0.6900	0.2321	0.0029*
η	1.4445	0.0862	< 0.0001*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 7: Weibull - Joe Univariable Analysis

Variable	Estimate	SE	P-Value
Sex (Male)	0.3865	0.1499	0.0099*
η	1.7330	0.1349	< 0.0001*
Residence (Urban)	0.1283	0.1498	0.4061
η	1.7811	0.1340	< 0.0001*
Smoking (Smoker)	0.1804	0.1487	0.2529
η	1.7576	0.1352	< 0.0001*
Diabetes mellitus (Yes)	0.1859	0.1438	0.1961*
η	1.7366	0.1325	< 0.0001*
Hypertension (Yes)	1.1339	0.1590	< 0.0001*
η	1.7215	0.1313	< 0.0001*
Family History (Yes)	0.9419	0.1430	< 0.0001*
η	1.7620	0.1357	< 0.0001*
Alcohol consumption (Yes)	0.1279	0.1381	0.3153
η	1.7601	0.1366	< 0.0001*
Anemia (Yes)	0.7640	0.1397	0.0001*
η	1.6600	0.1275	< 0.0001*
Obesity (Yes)	1.0526	0.1497	< 0.0001*
η	1.6782	0.1288	< 0.0001*
Age (36-55)	0.0087	0.1747	0.9603
Age (≥ 56)	0.6810	0.2324	0.0034*
η	1.6557	0.1264	< 0.0001*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 8: Gompertz-Joe Uni-variable Analysis

Variables	Estimate	SE	P-Value
Sex (Male)	0.2974	0.1405	0.0343*
η	1.2372	0.0675	< 0.0001*
Residence (Urban)	0.1481	0.1426	0.2989
η	1.2408	0.0677	< 0.0001*
Smoking (Smoker)	0.1754	0.1572	0.2645
η	1.2418	0.0662	< 0.0001*
Diabetes mellitus (Yes)	0.3248	0.1400	0.0204*
η	1.2487	0.0677	< 0.0001*
Hypertension (Yes)	0.9764	0.1594	< 0.0001*
η	1.2473	0.0681	< 0.0001*
Family History (Yes)	0.8109	0.1247	< 0.0001*
η	1.2412	0.0678	< 0.0001*
Alcohol consumption (Yes)	0.1193	0.1506	0.4158
η	1.2384	0.0676	< 0.0001*
Anemia (Yes)	0.6178	0.1608	0.0001*
η	1.2395	0.0680	< 0.0001*
Obesity (Yes)	0.9215	0.1715	< 0.0001*
η	1.2425	0.0679	< 0.0001*
Age (36-55)	0.0064	0.1729	0.9857
Age ($\geq$ 56)	0.5725	0.1998	0.0052*
η	1.2499	0.0682	< 0.0001*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 9: Log-logistic-Joe Uni-variable Analysis

Variable	Estimate	SE	P-Value
Sex (Male)	0.3875	0.1859	0.0371*
η	1.5375	0.1203	< 0.0001*
Residence (Urban)	0.1423	0.1763	0.2527
η	1.5692	0.1230	< 0.0001*
Smoking (Smoker)	0.1641	0.1787	0.2505
η	1.5556	0.1201	< 0.0001*
Diabetes mellitus (Yes)	0.3471	0.1762	0.0489*
η	1.5486	0.1194	< 0.0001*
Hypertension (Yes)	1.3499	0.1961	< 0.0001*
η	1.5318	0.1199	< 0.0001*
Family History (Yes)	1.1620	0.1798	< 0.0001*
η	1.5536	0.1210	< 0.0001*
Alcohol consumption (Yes)	0.1267	0.1741	0.2786
η	1.5499	0.1213	< 0.0001*
Anemia (Yes)	0.7871	0.1755	< 0.0001*
η	1.4973	0.1149	< 0.0001*
Obesity (Yes)	1.2355	0.1857	< 0.0001*
η	1.5467	0.1210	< 0.0001*
Age (36-55)	0.0198	0.22200	0.9289
Age ($\geq$ 56)	0.6809	0.2324	0.0034*
η	1.5838	0.1239	< 0.0001*

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Appendix-2: Multi-variable Analysis

Table 10: Weibull - Clayton multi-variable Analysis

Variables	Estimate	SE	P-Value	HR(95% CI)
λ	1295.4167	86.3746	< 0.0001*	
k	3.0089	0.1460	< 0.0001*	
Sex (Male)	0.3201	0.1532	0.0367*	1.377[1.020, 1.8597]
Diabetes mellitus (Yes)	0.1885	0.1386	0.1738	1.207[0.9202, 1.5844]
Hypertension (Yes)	0.6501	0.1761	0.0002*	1.916[1.3565, 2.7082]
Family History (Yes)	0.2462	0.1511	0.1032	1.279[0.9512, 1.7201]
Anemia (Yes)	0.2314	0.1487	0.1196	1.260[0.9417, 1.6869]
Obesity (Yes)	0.4110	0.1611	0.0107*	1.508[1.0999, 2.0685]
Age (36-55)	0.0232	0.1750	0.8945	1.023[0.7263, 1.4422]
Age ($\geq$56)	0.6275	0.2624	0.0168*	1.873[1.1199, 3.1324]
η	1.3530	0.2624	< 0.0001*	

Final llk = -2128.768 $\qquad\qquad\qquad\qquad\qquad$ $\tau = 0.4035$

AIC = 4275.536

Source: Adama Hospital Medical College, Ethiopia; from 1ˢᵗ January 2015 to 30ᵗʰ January 2020.

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

57

Table 11: Gompertz-Clayton multi-variable Analysis

Variables	Estimate	SE	P-Value	HR(95% CI)
a	909.3604	35.5338	<0.0001*	
b	2.8484	0.1358	<0.0001*	
Sex (Male)	0.2744	0.1216	0.0240*	1.316[1.0368, 1.6678]
Diabetes mellitus (Yes)	0.1889	0.1313	0.1503	1.208[0.9339, 1.5624]
Hypertension (Yes)	0.5585	0.1839	0.0024*	1.748[1.2191, 2.5065]
Family history (Yes)	0.2517	0.1292	0.0519	1.286[0.9985, 1.6240]
Anemia (Yes)	0.1909	0.1411	0.1760	1.210[0.9179, 1.5960]
Obesity (Yes)	0.3884	0.1656	0.0190*	1.475[1.0659, 2.040]
Age (36-55)	0.0214	0.1905	0.9107	1.022[0.7033, 1.4841]
Age ($\geq$56)	0.6749	0.3145	0.0319*	1.964[1.0603, 3.6375]
η	1.2283	0.2505	<0.0001*	

Final llk = -2686.034 $\tau = 0.3805$

AIC = 4386.068

Source: Adama Hospital Medical College, Ethiopia; from 1ˢᵗ January 2015 to 30ᵗʰ January 2020.

Table 12: Weibull-Gumbel multi-variable Analysis

Variables	Estimate	SE	P-Value	HR(95% CI)
λ	1551.70	122.93	< 0.0001*	
k	2.7445	0.1478	< 0.0001*	
Sex (Male)	0.3888	0.1594	0.0147*	1.475[1.0794, 2.0162]
Diabetes mellitus (Yes)	0.2810	0.1441	0.0512	1.325[0.9985, 1.7568]
Hypertension (Yes)	0.7053	0.1842	0.0001*	2.025[1.4110, 2.9046]
Family History (Yes)	0.3440	0.1555	0.0269*	1.411[1.0400, 1.9132]
Anemia (Yes)	0.3220	0.1687	0.0574	1.380[0.9914, 1.9207]
Obesity (Yes)	0.5491	0.1701	0.0012*	1.732[1.2407, 2.4169]
Age (35-55)	0.019	0.1744	0.9127	1.019[0.7241, 1.4345]
Age (≥ 56)	0.5608	0.1937	0.0038*	1.752[1.1986, 2.5611]
η	1.4301	0.0836	< 0.0001*	

Final llk = -2132.406 $\qquad\qquad\qquad\qquad$ τ= 0.3008

AIC = 4282.812

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 13: Gompertz-Gumbel multi-variable Analysis

Variables	Estimate	SE	P-Value	HR(95% CI)
a	985.3055	47.9576	< 0.0001*	
b	2.4513	0.1323	< 0.0001*	
Sex (Male)	0.3254	0.1415	0.0215*	1.385[1.0493, 1.8270]
Diabetes mellitus (Yes)	0.2350	0.1491	0.1149	1.265[0.9444, 1.6942]
Hypertension (Yes)	0.6236	0.1750	0.0004*	1.866[1.3239, 2.6290]
Family History (Yes)	0.3374	0.1361	0.0132*	1.401[1.0732, 1.8298]
Anemia (Yes)	0.2644	0.1647	0.1084	1.303[0.9433, 1.7989]
Obesity (Yes)	0.5037	0.1816	0.0056*	1.6548[1.1593, 2.3622]
Age (35-55)	0.0114	0.1889	0.9517	1.012[0.6985, 1.4646]
Age ($\geq$ 56)	0.5608	0.1937	0.0038	1.752[1.1985, 2.5613]
η	1.2245	0.6004	0.0414*	

Final llk = -2696.188 τ = 0.1834

AIC = 4406.376

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 14: Log-logistic-Gumbel multi-variable Analysis

Variables	Estimate	SE	P-Value	OR(95% CI)
λ	1253.7	96.495	< 0.0001*	
k	3.4155	0.1858	< 0.0001*	
Sex (Male)	0.3523	0.2039	0.0841	1.422[0.9538, 2.1210]
Diabetes mellitus (Yes)	0.1961	0.1852	0.2898	1.217[0.8463, 1.7491]
Hypertension (Yes)	0.8756	0.2317	0.0002*	2.400[1.5242, 3.780]
Family History (Yes)	0.5801	0.1958	0.0031*	1.786[1.2169, 2.6219]
Anemia (Yes)	0.2500	0.1983	0.2075	1.284[0.8705, 1.4751]
Obesity (Yes)	0.5932	0.2136	0.0055*	1.810[1.1907, 2.7508]
Age (35-55)	0.0275	0.2212	0.9012	1.028[0.6662, 1.5858]
Age ($\geq$56)	0.6945	0.2324	0.0028*	2.003[1.270, 3.1582]
η	1.3632	0.0771	< 0.0001*	

Final llk = -2129.781 $\tau = 0.2664$

AIC = 4277.563

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 15: Weibull-Joe multi-variable Analysis

Variable	Estimate	SE	P-Value	HR(95% CI)
λ	1627.5797	131.9546	< 0.0001*	
k	2.6741	0.1491	< 0.0001*	
Sex (Male)	0.4026	0.1557	0.0097*	1.496[1.1023, 2.0295]
Diabetes mellitus (Yes)	0.2904	0.1604	0.0686	1.337[0.9763, 1.8309]
Hypertension (Yes)	0.7003	0.1804	0.0001*	2.014[1.4144, 2.8688]
Family History (Yes)	0.3635	0.1526	0.0172*	1.438[1.0665, 1.9398]
Anemia (Yes)	0.3480	0.1511	0.0213*	1.416[1.0532, 1.9045]
Obesity (Yes)	0.5759	0.1665	0.0005*	1.779[1.2835, 2.4650]
Age (36-55)	0.0161	0.1751	0.9267	1.016[0.7210, 1.4323]
Age ($\geq$ 55)	0.5553	0.1932	0.0040	1.742[1.1932, 2.5447]
η	1.5680	0.1178	< 0.0001*	

Final llk = -2135.880 $\qquad\qquad\qquad\qquad\qquad$ $\tau = 0.2413$

AIC = 4289.760

Source: Adama Hospital Medical College, Ethiopia; from 1ˢᵗ January 2015 to 30ᵗʰ January 2020.

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 16: Gompertz-Joe multi-variable Analysis

Variable	Estimate	SE	P-Value	HR(95% CI)
a	1015.9411	52.0377	< 0.0001*	
b	2.3641	0.1337	< 0.0001*	
Sex(Male)	0.3359	0.1386	0.0154*	1.399[1.0663, 1.8360]
Diabetes mellitus (Yes)	0.2358	0.1523	0.1215	1.266[0.9392, 1.7063]
Hypertension (Yes)	0.6323	0.1742	0.0003*	1.882[1.3376, 2.6477]
Family History (Yes)	0.3407	0.1353	0.0118*	1.406[1.0784, 1.8329]
Anemia (Yes)	0.2776	0.1617	0.0859	1.320[0.9615, 1.8121]
Obesity (Yes)	0.5201	0.1843	0.0048*	1.682[1.1722, 2.4140]
Age (36-55)	0.0052	0.1897	0.9780	1.005[0.6931, 1.4579]
Age ($\geq$ 55)	0.5547	0.1998	0.0055*	1.741[1.1772, 2.5762]
η	1.2520	0.0554	< 0.0001*	

Final llk = -2699.818 $\qquad\qquad\qquad\qquad\qquad\qquad$ $\tau = 0.1255$

AIC =4413.636

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.

Table 17: Log-logistic-Joe multi-variable Analysis

Variables	Estimate	SE	P-Value	OR(95% CI)
λ	1309.7278	101.9087	< 0.0001*	
k	3.3544	0.1869	< 0.0001*	
Sex (Male)	0.3762	0.1991	0.0589	1.457[0.9861, 2.1520]
Diabetes mellitus (Yes)	0.2161	0.1806	0.2315	1.241[0.8712, 1.7684]
Hypertension (Yes)	0.8900	0.2278	< 0.0001*	2.435[1.5582, 3.8053]
Family History (Yes)	0.5965	0.1924	0.0019*	1.816[1.2453, 2.6475]
Anemia (Yes)	0.2781	0.1934	0.1506	1.321[0.9039, 1.9294]
Obesity (Yes)	0.6357	0.2089	0.0023*	1.888[1.2540, 2.8437]
Age (35-55)	0.0220	0.2222	0.92111	1.022[0.6613, 1.5801]
Age ($\geq$ 56)	0.6868	0.2328	0.0032*	1.987[1.2592, 3.1365]
η	1.4569	0.1049	< 0.0001*	

Final llk = -2133.539 $\hspace{4cm}$ τ = 0.2048

AIC = 4285.078

Source: Adama Hospital Medical College, Ethiopia; from 1st January 2015 to 30th January 2020.
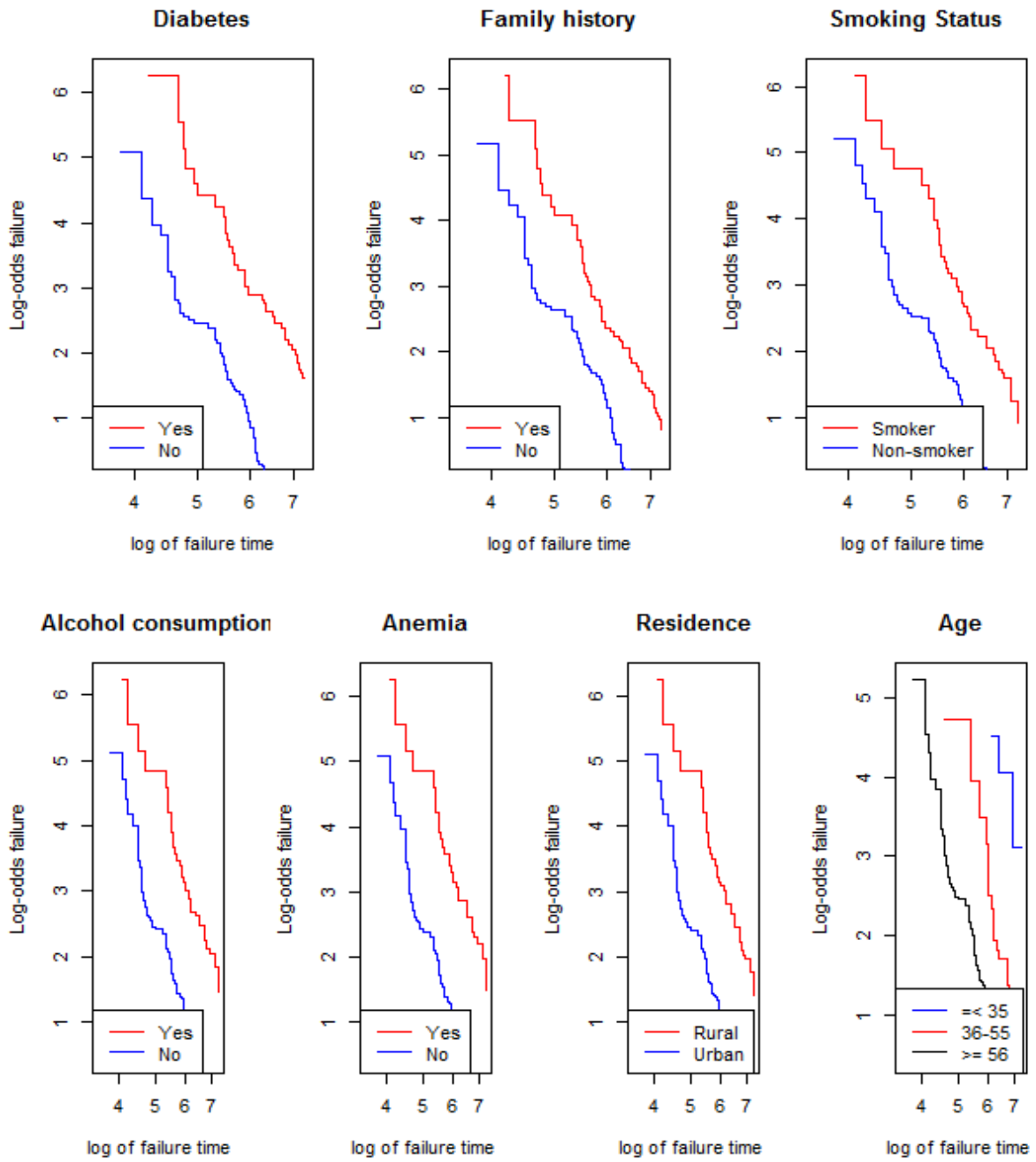
Appendix-3: Assumption Checking for Proportional Odds

Figure: 1 Graphical checking for proportional odds assumption