College of Natural Sciences

Department of Statistics

Modeling Time to Death of Patients with Ischemic Heart Diseases at Jimma University Medical Center: A Comparison of Various Parametric Shared Frailty Models

---

By: Bekele Haile

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

December 13, 2022

Jimma, Ethiopia

Modeling Time to Death of Patients with Ischemic Heart Diseases at Jimma University Medical Center: A Comparison of Parametric Shared Frailty Models

MSc Thesis

By: Bekele Haile

Advisor: Tadele Akeba Diriba (PhD)
Co-Advisor: Demeke Kifle (MSc)

December 13, 2022
Jimma, Ethiopia

SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

MSc THESIS APPROVAL SHEET

This is to certify that the thesis titled "**Modeling Time to Death of Patients with Ischemic Heart Diseases at Jimma University Medical Center: A Comparison of Various Parametric Shared Frailty Models**" has been submitted in partial fulfillment of the requirement for the degree of Master of Science in Biostatistics to the college of natural science at Jimma University, and is a record of original research carried out by Bekele Haile Gemechu, under our supervisions, and no part of the thesis has been submitted for another degree or diploma. The assistance and the help received during the course of this investigation have been duly acknowledged. Therefore, we recommend that it be accepted as fulfilling the thesis requirement.

Tadele Akeba(PhD) Name of Major Advisor	Signature	30/12/2022 Date
Demeke Kifle(M.Sc) Name of Co-advisor Advisor	Signature	30/12/2022 Date

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

Dr. Shibru Temesgen Name of the External Examiner	Signature	29/12/2022 Date
Kenenisa Tadesse(M.Sc) Name of the Internal Examiner	Signature	30/12/2022 Date
Kibrealem Sisay (MSc) Name of the Chairperson	Signature	09/01/23 Date

Statement of Author

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate. Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Bekele Haile Gemechu

Date: _____

Signature: _____

Jimma University, Jimma, Ethiopia

Dedication

This thesis is dedicated to my family especially my mother, Megartu Soboka, who were with me at the time of my Happiness and Terrible throughout my study!!!

Acknowledgments

First of all, I would like to express my deepest gratitude to my advisor, Dr. Tadele Akeba Diriba, for his kind assistance, counsel, direction, and constructive criticism, which began with the proposal-writing process and continued through the completion of this thesis.

I also want to express my deep gratitude to my co-advisor, Mr. Demeke Kifle (MSc), for his ongoing assistance and counsel.

Additionally, I want to thank the Cardiac Center Patient Clinic at Jimma University Medical Center for their tremendous assistance in a variety of ways throughout data collection.

I praise God above all for his grace, wisdom, favor, fidelity, and safety.

Abstract

Background: Ischemic heart disease (IHD) is a disorder of cardiac function caused by insufficient blood flow to the muscle tissue of the heart. The decreased blood flow is in most cases due to coronary arteriosclerosis or to obstruction by a thrombus of the coronary arteries. Acute myocardial infarction, unstable angina, and angina pectoris are manifestations of ischemic heart disease.

Objective: The general objective of this study is to model the time to death of patients with ischemic heart disease using various parametric shared frailty models.

Methods: Different parametric frailty models were compared using exponential, weibull, and log-logistic as baseline hazard functions and the gamma as well as the inverse Gaussian for the frailty distributions, with the goal of developing an appropriate survival model that adequately describes the ischemic heart disease dataset. All models were then compared using the AIC and BIC criteria.

Results: The median time to death of the ischemic heart disease patients was about five days, with a maximum death time of thirty days, of which about 35.37% died. The clustering effect is significant in modeling the time to death of ischemic heart disease. The log-logistic model with an inverse Gaussian frailty distribution has the minimum AIC value among the models compared. According to the output of the model (log-logistic with inverse Gaussian frailty), diabetes mellitus, hypertension, obesity, smoking status, cholesterol, and other diseases were the main determinant factors of IHD.

Conclusions: Compared to other distributions employed in this study, the log-logistic with inverse Gaussian frailty model provided a superior description of the ischemic heart disease dataset. The time to death of ischemic heart disease patients vary between woredas, indicating that frailty models must be used to take into account this clustering feature.

Key words: ***Ischemic heart disease, Frailty, Heterogeneity, Parametric shared frailty Model, Time to death***

Acronyms

AFT	Acceleration Failure Time
AIC	Akaike Information Criterion
AOR	Adjusted Odds Ratio
BMI	Body Mass Index
CHD	Coronary Heart Disease
CVD	Cardio Vascular Disease
HF	Heart Failure
HDLC	High Density Lipoprotein Cholesterol
IHD	Ischemic Heart Disease
JUMC	Jimma University Medical Center
LDL	Low-Density Lipoprotein
MI	Myocardial Infarction
MLE	Maximum Likelihood Method
MUFA	Monounsaturated Fatty Acids
NHID	National Health Insurance Database
NPHS	Northwick Park Heart Study
PH	Proportional Hazards
PL	Partial Likelihood
QQ	Quantile - Quantile
USA	United State of America
WHO	World Health Organization

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background of the Study

The term ischemic heart disease (IHD) describes a group of clinical syndromes characterized by myocardial ischemia, an imbalance between myocardial blood supply and demand (Steenbergen & Frangogiannis, 2012). A disorder of cardiac function is caused by insufficient blood flow to the muscle tissue of the heart. The decreased blood flow is in most cases due to coronary arteriosclerosis or to obstruction by a thrombus of the coronary arteries (Grauss et al., 2007). The heart muscle receives less blood and oxygen when arteries narrow, this condition is also known as coronary heart disease or coronary artery disease. Acute myocardial infarction, unstable angina, and angina pectoris are realizations of IHD (Nilsson, 2008).

As long as the fundamental pathophysiologic defect in the ischemic myocardium is inadequate perfusion, ischemia is associated not only with insufficient oxygen supply but also with reduced availability of nutrients and inadequate removal of metabolic end products. The manifestations of IHD depend on the duration, severity, and acuity of the ischemic episodes. The main determinant factors or causes that clear the way for manifestations of IHD are high cholesterol, high blood pressure, smoking, diabetes mellitus, obesity, and consumption of saturated fats (Wilson, 1994). Some people can develop myocardial ischemia without experiencing any symptoms or warning indications (silent ischemia). Chest pressure or pain, typically on the left side of the body, is the most common symptom when it does occur (angina pectoris). Other warning signs and symptoms that may manifest more frequently include a rapid heartbeat, shortness of breath, nausea and vomiting, perspiration, tiredness, and discomfort in the neck, jaw, shoulder, or arm (Paul et al., 2004).

Cardiovascular diseases (CVD) cause approximately one-third of deaths worldwide (Mozaffarian et al., 2015). Among cardiovascular illnesses, IHD ranks as the most prevalent (Roth et al., 2017). Indeed, IHD is acknowledged as the main threat to sustainable development in the 21 century (Prabhakaran et al., 2018). An increasing

number of individuals with non-fatal IHD live with chronic disabilities and impaired quality of life (Moran et al., 1990). The primary pathological process that leads to IHD is atherosclerosis, an inflammatory disease of the arteries associated with lipid deposition and metabolic alterations due to multiple risk factors. More than 70% of at-risk individuals have multiple risk factors for IHD, and only 2%-7% of the general population have no risk factors (Khan et al., 2020).

CVD are the number one cause of death globally; more people die annually from CVD than from any other cause. An estimated 17.7 million people died of CVD, representing 31% of all global deaths. Of these deaths, an estimated 7.4 million were due to IHD. Of the 17 million premature deaths (under the age of 70) due to non-communicable diseases, 82% are in low-and middle-income countries, and 37% are caused by CVD (WHO, 2019)

The total rate of IHD-related mortality in Europe has decreased, particularly in the United Kingdom, the Netherlands, and Ireland. In contrast, epidemiological studies have reported that the prevalent cases and deaths related to IHD have increased rapidly in China, the Arab States, India, and Latin America. Some studies also showed that IHD is a preventable and eradicable disease if the risk factors are effectively controlled (Wang et al., 2021).

IHD, which was once thought to be uncommon in Sub-Saharan Africa, is now the eighth greatest cause of mortality in both men and women in the region. Furthermore, as a result of negative behavioral and lifestyle changes linked with urbanization and the epidemiological transition, the prevalence of IHD and concomitant morbidity are rising (Mensah, 2008). Ethiopia is one of the Sub-Saharan African countries making significant progress in combating CVD as a public health priority disease (Ali et al., 2021).

According to the latest data published by (WHO, 2020), IHD deaths in Ethiopia reached 36,530, or 6.48% of total deaths. The age-adjusted death rate is 78.39 per 100,000 populations. Furthermore, the global burden of IHD is rapidly rising as a result of the combined effects of social-demographic disadvantage, limited access to

health care, and poor health-care system performance (Vogel et al., 2021).

Therefore, it is vital to identify modifiable risk factors in addition to non-modifiable risk factors for IHD for effective health care planning and prevention. In this setting, a survival modeling framework is suitable in order to identify the factors for time to death of patients with IHD. Data that measures the time to a certain event of interest is referred to as survival data. The term survival data is applied to a wide range of occurrences, all of which can be viewed as a transition from one state to another (Bradbury et al., 2010).

Rather than focusing solely on frequency, survival analysis includes a time dimension to an event (Gerr et al., 2002). It also includes censoring, where data about the event of interest is unclear due to the patient's removal from the study. The life table created by Berkson & Gage (1952) for researching cancer survival is one of the oldest and most straightforward non-parametric tools for assessing survival data. Kaplan & Meier (1958) made a significant contribution to non-parametric approaches. Also, Cox (1972) established the proportional hazards model to quantify the inference of different factors on the timeframes of system failures.

Moreover, Clayton (1978) coined the term "frailty" to describe how diverse people are at risk, even if they appear to be relatively similar on the surface in terms of quantitative characteristics like age, gender, and weight, among others. In the examination of death rates, he adopted the term "frailty" to indicate an unobservable random effect shared among people with similar (unmeasured) risks. Over and beyond any observed covariates, a random effect represents excess risk or frailty for unique groupings, such as individuals or families.

Frailty model modifications to the Cox proportional hazard model, such as random effects and unobserved heterogeneity. Up to some observed covariates, survival analysis implicitly assumes a homogenous population, which means that all individuals in the study are, in principle, at the same risk. However, individuals differ significantly in terms of baseline traits, which has a significant impact when measuring the explanatory variables. In this case, the study population must be regarded as a diverse sample. On

the other hand, it is not always possible to include all the relevant covariates due to data limitations, and sometimes the importance of some covariates is still unknown.

As a result, to measure the heterogeneity caused by unobserved covariates, it is necessary to include a random effect term or frailty into the model (Wienke, 2010). In this thesis, shared frailty models are employed, assuming that patients in the same cluster (woreda) have similar risk factors and allowing the frailty term to be addressed at the woreda level. This is a conditional independence model in which the frailty is shared by all individuals in a cluster and is hence responsible for the emergence of dependencies between event times.

1.2 Statements of the Problem

Cardiovascular diseases, principally IHD, are the most important cause of death and disability in the majority of low and lower middle income countries (Gupta & Yusuf, 2019). In these countries more than 80% of the mortality rates due to IHD occur on individuals with low socioeconomic status (Nowbar et al., 2019). Various studies have been conducted in Ethiopia to identify covariates of IHD mortality by using logistic regression (Gebremedhin & Gebrekirstos, 2021), and Semi-parametric proportional hazard models (Abdissa, 2020).

However, the Logistic regression does not account the censoring observations, that is, it does not hold for time to event data. In demographic applications, non-parametric and semi-parametric models are often used to model transition data. In such applications, it is assumed that all heterogeneities are captured only by using theoretically relevant covariates (Trussell & Richards, 1985).

One of the most popular models for survival analysis is the Cox proportional hazard model (Cox, 1972), and in this model the hazard function may depend on unknown risk factors. But, in reality it is impossible to include the unknown risk factors in the model. This will lead to an increase in the variability of responses, which implies biased and misleading estimates might be obtained for the parameters of the Cox model (Van Dijk et al., 2008).

Nevertheless, there are numerous grounds to suspect omitted or unmeasured components in many circumstances. That is, while some people are at a higher chance of experiencing the event, the underlying causes of this variability are unlikely to be adequately reflected by the observed factors. If unmeasured frailty exists, the hazard will be a function of both the variables and the frailty (Vaupel et al., 1979). To assess the true effects of the observed covariates under this circumstance, it is crucial to explicitly account for unobserved heterogeneity.

This study generally addressed the following research questions:

- What are the main factors that have an impact on how long IHD patients live?
- Which fundamental tenet of distributional theory among exponential, weibull, loglogistic, frailty, gamma, and inverse Gaussian distributions best fits the IHD dataset?

1.3 Objectives of the Study

1.3.1 General Objective:

The general objective of this study is to model the time to death of patients with IHD using various parametric shared frailty models.

1.3.2 Specific Objectives:

The specific objectives of this study are

- To identify the main determinant (risk) factors associated with IHD
- To determine the parametric baseline hazard, which is appropriate in modeling the determinants of time to death of IHD

1.4 Significance of the Study

This study has the purpose of identifying the major contributing (risk) factors for IHD. The result of this study also provides information to the government and other concerned bodies in setting policies, strategies, and further investigation for the reduction

of IHD mortality. Moreover, this study can provide base line data for further studies in the future.

1.5 Limitation of the study

One of the study's limitations is that even though risk factors for ischemic heart disease are many, the research is limited only to the eight covariates since the records on the patient cards of secondary data do not hold some of the fundamental factors, such as other sociodemographic variables, which are clinically significant variables for the risk factors of IHD. These factors include family history (heredity) of the diseases, physical activity, stress, heavy alcohol consumption, nutrition, marital status, education status, and occupation.

1.6 Organization of the study

This research paper was organized into five chapters. The first chapter discusses the background of the study; a statement of the problems; the objectives of the study; the significance of the study; and limitations of the study. The second chapter contains a literature review as well as the concepts used in the paper. The third chapter focuses on the data and methodology for research design, data source, and population, study variables, inclusion and exclusion criteria, and survival models. The fourth chapter contains results and discussion, while the fifth chapter provides conclusions and recommendations. Finally, it presents references and different appendices.

2 Literature Review

2.1 Overview of Ischemic Heart Disease

Ischemia is defined as inadequate blood supply (circulation) to a local area due to blockage of the blood vessels supplying the area. Ischemic means that an organ (the heart) is not getting enough blood and oxygen. The discomfort felt when the heart muscle is deprived of sufficient oxygen is termed as angina pectoris (Almdal et al., 2004). IHD is a problem caused by narrowed heart (coronary) arteries that supply blood to the heart muscle due to the buildup of plaque called atherosclerosis (Verdouw et al., 1998). When the blood flow to the heart muscle is completely blocked, the heart muscle cells die, which is termed a heart attack or myocardial infarction (MI) (Badimon et al., 2012).

The global burden of disease study showed that IHD is the leading cause of death, disability, and human suffering worldwide. IHD affects about 126 million people worldwide (1,655 per 100,000), or about 1.72% of the total population and causes nine million deaths worldwide. The prevalence of IHD is still increasing. By 2030, the prevalence rate, which is currently 1,655 per 100,000 people, is anticipated to surpass 1,845. The highest incidence is now being sustained in Eastern European nations (Khan et al., 2020).

IHD was the leading cause of death in five representative nations (the United Kingdom, the United States, Brazil, Kazakhstan, and Ukraine) according to WHO data from 2005 to 2015; nevertheless, mortality from IHD has steadily declined from 2005 to 2015. With little variation in other causes of death, the age-standardized mortality rates for IHD were significantly higher in Kazakhstan (97) and Ukraine (324), compared to the United States (60), Brazil (54), and the United Kingdom (46). All 5 countries showed a progressive decline in IHD mortality, with a decline in smoking and hypertension and, in all cases, a rise in obesity and type II diabetes mellitus (Nowbar et al., 2019)

According to Wang et al. (2021), metabolic dysfunction is increasingly serving as a

primary risk factor for the emergence of a variety of comorbidities due to IHD. And it continues to be the CVD with the highest global illness burden. The global death cases from IHD climbed significantly during that time period, although the mortality rate gradually decreased, according to the methodology framework of the global burden of disease study. Notably, metabolic risk factors are now the main causes of IHD, which is partly responsible for the shift in the distribution of IHD-related mortality from industrialized to developing nations. The results point to the urgent need for effective metabolic risk factor control strategies to be put in place to stop the rise in IHD-related mortality.

Shashu (2021) conducted a facility-based study at the first private cardiac specialized hospital in Ethiopia. The frequency distribution of the outcome variables, risk factors, clinical presentations, and treatment categories were presented. The result showed that the two most frequent risk factors of cardiovascular disease are hypertension and dyslipidemia, present in 1040 (33.61%) and 831 (26.86%) patients, respectively, followed by diabetes in 508 (16.42%). Hypertension with hypertensive heart disease accounts for 1040 (49.31%), followed by IHD in 219 (10.38%) and cardiomyopathy in 133 (6.31%) patients.

By Moyehodie et al. (2022), a multicenter retrospective cohort analysis of 285 patients aged 15 or older who were being monitored from January 1, 2015, to December 31, 2019, was done. The log-rank test and the Kaplan-Meier survival curve were used to condense descriptive analyses. Following their admission to the heart failure department, the Cox-proportional hazard regression model was used to assess the risk of mortality up to five years later. The study found that the major risk factors for death were diabetes mellitus and hypertension.

According to Špinar (2012), hypertension is one of the major risk factors for ischemic heart disease and appropriate control of blood pressure is the cornerstone of both primary and secondary ischemic heart disease prevention. Effective blood pressure (BP) control is recommended in primary prevention, meaning that maintaining blood pressure <140/90 mmHg, while in secondary prevention values <130/85 mmHg used to be recommended. According to epidemiologic data, cardiovascular mortality increases

with blood pressure, starting as low as the 110/70 mmHg level. Czech, European, and American guidelines from the early 21 century recommend that blood pressure in patients with ischemic heart disease (IHD) be maintained below 130/80 mmHg.

According to Mensah (2008) IHD, which was previously considered rare in sub-Saharan Africa, now ranks eighth among the leading causes of death in men and women in the region. Furthermore, the prevalence of IHD and related morbidity may be increasing as a result of adverse behavioral and lifestyle changes associated with urbanization and the epidemiological transition. The major risk factors for IHD in sub-Saharan Africa include hypertension, smoking, diabetes,and abdominal obesity. In the Interheart Africa study, these risk factors contributed to a population-attributable risk of nearly 90% for acute myocardial infarction. In conclusion, an aggressive approach that combines environmental, policy, and legislative interventions for health promotion and primary prevention, coupled with improved access to evaluation, treatment, and control of hypertension and other major risk factors, provides the best strategy for averting an epidemic of IHD in sub-Saharan Africa.

Ostadal & Ostadal (2014) investigated epidemiological studies on sex-based differences in cardiac ischaemic injury. Studies have demonstrated that premenopausal women have a reduced risk of IHD compared with their male counterparts. The incidence of IHD in women increases after menopause, suggesting that IHD is related to declining estrogen levels. Experimental observations have confirmed the results of epidemiological studies investigating sex-specific differences in cardiac tolerance to ischaemia. Female sex also appears to favorably influence cardiac remodeling after ischaemia or reperfusion injury. Furthermore, sex-related differences in ischaemic tolerance of the adult myocardium can be influenced by interventions during the early phases of ontogenetic development.

2.2 Empirical Literature

Numerous studies have demonstrated that a variety of socioeconomic and demographic variables have an impact on the time to death of IHD patients. For instance, Abdissa (2020) performed a retrospective cohort study to evaluate all patients who received an

IHD diagnosis throughout time using data from Black Lion Specialized and Tertiary Referral Hospital in Addis Ababa, Ethiopia. Investigating the association between dependent and predictor variables was done using the Cox-regression model. According to the findings, patients with IHD have a greater chance of developing incident HF. In these individuals, increasing age, LDL cholesterol, diabetes mellitus, decreased hemoglobin, and a dilated left atrium were the primary predictors of incident HF. Such patients require frequent monitoring and more thorough care.

According to Abdissa et al. (2021), IHD patients were recruited and followed retrospectively over a 24-month period. The Cox regression model was employed to identify risk factors for incident heart failure with reduced or preserved ejection fraction. The result showed that of the 153 patients with new onset HF, 60.1% (92/153) were those with reduced ejection fraction, while 39.9% (61/153) were those with preserved ejection fraction. Besides diabetes (HR 2.07 [95% CI: 1.33-3.22], P = 0.001) and left atrium dimension (HR 1.03 [95% CI: 1.001-1.065], P = 0.04), age 46-55 (HR 0.4 [95% CI: 0.17-0.94], P = 0.036), age 66 and above (HR 0.36 [95% CI: 0.13-0.98], P = 0.047); and increasing left ventricular diastolic (diastolic LVD) (HR 1.06 [95% CI: 1.03-1.09], P < 0.001) were associated with heart failure with reduced ejection fraction. The finding suggests that age, diabetes, and diastolic LVD are predictors of heart failure with reduced or preserved ejection fraction in patients with IHD.

Tromp et al. (2021) investigate global differences in prevalence, association with outcome, and treatment of IHD in patients with acute HF in the international registry to assess medical practice. With a HF treatment registry and longitudinal observation a total of 18,539 patients with acute HF were prospectively enrolled from 44 countries and 365 centers in the report of the HF registry. Patients with a history of IHD, an ischemic event causing admission for acute HF, or coronary revascularization were classified as IHD. Clinical characteristics, treatment, and outcomes of patients with and without IHD were explored. The results compared with 8,766 (47%) patients without IHD, 9,773 (53%) patients with IHD were older, more likely to have a left ventricular ejection fraction <40% HF with reduced ejection fraction, and reported more comorbidities. IHD was more common in lower income countries than in high-income

countries (61% vs. 48%). Patients with IHD from countries with low health care expenditure per capita or without health insurance were less likely to undergo coronary revascularization or use anticoagulants at discharge. IHD was independently associated with worse cardiovascular death (hazard ratio: 1.21; 95% CI: 1.09 to 1.35). The association between IHD and cardiovascular death was stronger in HF with reduced ejection fraction compared with HF with preserved ejection fraction (P <0.001).

Tate et al. (1998) performed a prospective investigation of CVD as it develops in a cohort of 3983 young men, 1094 study members (27%) developed clinical evidence of IHD. Blood pressure, body weight, smoking, and the presence of diabetes mellitus have been recorded at regular intervals throughout the follow-up period. Using measurements from examinations every five years between ages 40 and 75 years, age-specific Cox proportional hazard models were fitted to relate these risk factors to IHD. The result revealed that the adjusted relative risk of IHD for systolic blood pressure, diastolic blood pressure, and smoking was found to significantly ($p < 0.001$) decline with advancing age. The adjusted relative risk for body mass index and the presence of diabetes mellitus for ischemic heart disease did not vary with age ($p > 0.05$). After age 65 years, these risk factors were of little value for the prediction of IHD. The relative risk and statistical significance of blood pressure and smoking, as risk factors for IHD, decline with age.

Barbiero et al. (2009) conducted cross-sectional, population-based study, with a stratified probabilistic sample of secondary schools in Porto Alegre, comprising a total of 511 schoolchildren. Data on family risk factors, anthropometry, and eating habits were collected. The result revealed that the prevalence of excess weight was 27.6% among the schoolchildren, with 17.8% being overweight (BMI $\geq 85^{th}$ and $< 95\%$) and 9.8 % obese (BMI $\geq 95\%$). Overweight was more prevalent in females (19.9 %) and obesity in males (11.8%). As a conclusion, obesity had a highly significant impact on the survival time of patients with IHD.

Sakboonyarat & Rangsin (2018) carry out a cross-sectional study to assess national outcomes among patients with diabetes who visited public hospitals in Thailand to evaluate the status of care among patients with diabetes aged at least 18 years who

received medical treatment in the target hospital for the last 12 months. The results showed that a total of 25,902 patients with diabetes were included in this study. IHD was detected among 918 patients (3.54%; 95% CI: 3.32-3.77). Multivariate analysis was conducted to determine which factors were most associated with IHD, and the results showed age (AOR 1.05; 95% CI: 1.04-1.05), being male (AOR 1.78; 95% CI: 1.53-2.07), hypertensive comorbidity (AOR 2.10; 95% CI: 1.68-2.62), being in Health Region (AOR 1.93; 95% CI: 1.54-2.35), presenting hyperglycemic crisis (AOR 1.53; 95% CI: 1.14-2.06) and insulin therapy (AOR 1.40; 95% CI: 1.17-1.66) were the highest associated factors for IHD in this population. and concluded that IHD was a problem among patients with diabetes.

Rashid et al. (2019) worked on a case control study among 142 newly diagnosed IHD female patients registered in government hospitals in Terengganu, Malaysia and their 1:1 frequency matched population controls. Data on sociodemographic and socioeconomic profiles, co-morbidities, lifestyle factors related to physical activities, dietary fat intake, stress, passive smoking history, anthropometric measurements, and biochemical markers were obtained. Middle-aged women were recruited with women diagnosed with diabetes (AOR=1.92, 95% CI: 1.11-3.31), having low HDLC (AOR=3.30, 95% CI: 1.28-8.27), those with positive family history of IHD (AOR=1.92, 95% CI: 1.13-3.26) and passive smokers (AOR=2.99, 95% CI: 1.81-4.94) were at higher odds of IHD.

Mahendra et al. (2015) recruited patients with type 2 diabetes mellitus based on the inclusion and exclusion criteria. History of IHD evidence of ischemia was obtained. Retinopathy was diagnosed by direct opthalmoscopy. Fasting glucose levels, lipid profiles, and plasma fibrinogen levels were all measured. Statistical analysis was carried out by the Chi-square test and the Student's test. The results revealed that the prevalence of metabolic syndrome in Type 2 diabetes mellitus is 58% and plasma fibrinogen is significantly higher in these patients. Macro and micro vascular complications are frequent in long-standing patients with type 2 diabetes mellitus. Thus, the presence of metabolic syndrome and hyperfibrinogenemia may contribute to the early development of macro (IHD) and micro (retinopathy) vascular complications.

Gebremedhin & Gebrekirstos (2021) investigated a facility-based unmatched case-

control study from November 16 to March 20, 2020, among patients with IHD and those patients who visited the three hospitals in the Wolaita Zone. A total of 557 study participants (140 cases and 417 controls) were included in a ratio of 1:3, The results of logistic regression showed that,the adjusted odds ratio for having no formal education (AOR = 3.18; 95% CI: 1.59-6.34), previous history of hypertension (AOR = 2.84; 95% CI: 1.73-4.66), physical inactivity (AOR = 2.23; 95% CI: 1.32-3.76), inadequate intake of fruit and vegetable consumption (AOR = 2.43; 95% CI: 1.40-4.22), palm oil use for food preparation (AOR = 2.12; 95% CI: 1.23, 3.63) and obesity (AOR = 5.68; 95% CI: 2.63-12.23) increased the occurrence of the disease. As a conclusion, although ischemic heart disease is preventable using relatively simple and inexpensive lifestyle changes, it is projected to cause a preventable loss of life.

Altaseb (2020) investigated a retrospective cohort study on CHD patients whose age is $\geq$ 18 years. The total deaths of people with CHD in Ethiopia reached 47712 in 2018, where 7.81% of deaths in Ethiopia were due to CHD. In the survival analysis, the Cox PH model, was used to identify potential associations between the survival time and the study variables. The result of Cox PH regression analysis showed that the age of CHD patients, diabetes mellitus, hypertension, and atherosclerotic heart diseases were significantly associated with the survival of CHD patients.

Gona et al. (2021) conducted an epidemiological analysis of the prevalence of IHD for 16 southern African development communities using global burden of diseases study data. An ensemble model and spatiotemporal Gaussian regression were used to estimate mortality due to IHD. The result revealed that obesity in adult females increased 1.54-fold from 12.0% (uncertainty interval: 11.5-12.4) to 18.5% (17.9-19.0), where as in adult males, obesity nearly doubled from 4.5 (4.3-4.8) to 8.8 (8.5-9.2). In children, obesity more than doubled in both sexes, and overweight increased by 27.4% in girls and by 37.4% in boys. Mean BMI increased by 0.7 in adult males from 22.4 (21.16-23.1) to 23.1 (22.23-24.0) and by 1.0 in adult females from 23.8 (22.9-24.7) to 24.8 (23.8-25.8). In 2019, the prevalence of obesity was highest in South Africa 44.7 (42.5-46.8), Swaziland 33.9 (31.7-36.0), and Lesotho 31.6 (29.8-33.5). The corresponding prevalence in males for the three countries was 19.1 (17.5-20.7), 19.3 (17.7-20.8), and 9.2 (8.4-10.1),

respectively. Congo and Madagascar had the lowest prevalence of adult obesity, with 5.6 (4.8-6.4) and 7.0 (6.1-7.9), respectively, in females in 2019, and in males from 4.9 (4.3-5.4) in Congo and 3.9 (3.4-4.4) in Madagascar. The finding suggests that a high body mass index is associated with IHD.

As stated by Crump et al. (2017), obesity is known risk factors for IHD. A national cohort study of all 1,547,407 military conscripts was conducted in Sweden. Body mass index (BMI) measurements were examined in relation to IHD. The results revealed that there were 38,142 men diagnosed with IHD in 39.7 million person-years of follow-up. High BMI or low aerobic fitness was associated with a higher risk of IHD, adjusting for family history and socioeconomic factors. The combination of high BMI (overweight/obese versus normal) was associated with the highest IHD risk (incidence rate ratio, 3.11; 95% CI: 2.91-3.31; $P<0.001$). In this large cohort study, high BMI at age 18 was associated with a higher risk of IHD in adulthood. Low aerobic fitness appeared to account for a similar number of IHD cases among those with normal versus high BMI. This finding suggests that interventions to prevent IHD should begin early in life and include not only weight control but aerobic fitness, even among people of normal weight.

3 Data and Methodology

3.1 Description of Study Area

This study has been conducted at Jimma University Medical Center. JUMC is one of the oldest public hospitals in the country. It was established in 1930 E.C by Italian invaders for the service of their soldiers. Geographically, it is located in Jimma City, 357 km southwest of Addis Ababa, the capital city of Ethiopia.

3.1.1 Study design, Population and Period

A retrospective study was conducted to obtain data on IHD patients who were recorded at JUMC. All IHD patients who had registered at JUMC between January 1, 2016, and June 1, 2022 comprised the study's population. The information from the patient's registration card and registration log book has been carefully scrutinized; any insufficient information found has been verified in the file and, if found to be such, has been excluded from analysis. Accordingly, the information was gathered from patient follow-up records using the study's factors.

3.1.2 Inclusion and Exclusion Criteria

Inclusion criteria: All patients who supplied complete information, including the variables of interest, in the registration book or on the cards, were considered eligible for the study. To be included in the study, the patients had to have received at least one hospital therapy.

Exclusion criteria: If the registration book or card didn't provide enough details regarding the research factors, patients weren't eligible. As a result, these patients were excluded.

3.1.3 Data collection methods

Ethical permission has been obtained from the Jimma University Medical Center, Jimma, Ethiopia. Then secondary data was taken based on data existing in the hospital by a trained enumerator and the principal investigator using a checklist (data

extraction form).

Starting time: The starting time of the interval (in days) is the time origin or the beginning of the study. The entry of the survival data would be considered from the day that the ischemic heart disease patients started their diagnosis, when the patient first received the treatment.

Ending time: The time (in days) when the ischemic heart disease patients died or were lost to follow-up at the beginning of June 2022 (the end of the study) means that the type of survival data is right-censored.

3.2 Variable Description

3.2.1 Dependent Variable

The response variable in this study is the time to death of IHD, measured (in days) from the day that the patient starts treatment after admission to discharge. The survival status of an IHD patient at discharge time was one of the two: event (death = 1) or censored = 0.

3.2.2 Independent Variable

Several explanatory variables are used in this study to investigate the risk factors for the time to death of patients with IHD. A patient's residential place (woreda) is considered as a clustering variable (effect) in all frailty models. The candidate predictor variables included in the study and the codes of categories were described in Table 3.1.

Table 3.1: List of explanatory variables with description and categories

Variables	Definition and Categories
Sex	Sex of patients (0 = Female; 1 = Male)
Age	Age group of patients (1 = 0-14; 2 = 15-47; 3 = 48-63; 4 = $\geq$ 64)
Diabete Mellitus	patients having diabetes (0 = No; 1 = Yes)
Hypertension	patients having high blood pressure (0 = No; 1 = Yes)
Obesity	patients having overweight/obesity (0 = No; 1 = Yes)
Smoking status	Smoking status of patients (0 = No; 1 = Yes)
Cholesterol	patients having high cholesterol (0 = No; 1 = Yes)
Other diseases	patients having other co-infection (0 = No; 1 = Yes)
Residential place (Woreda)	Agaro = 1; Chora Botor = 2; Dedo = 3; Gatira = 4; Gera = 5; Gomma = 6; Guma = 7; Jimma = 8; Kersa = 9; Limmu Ganat = 10; Limmu Kosa = 11; Limmu Sakka = 12; Mana = 13; Nono = 14; Omo Nada = 15; Sakka = 16; Shabe Sonbo = 17; Setema = 18; Sigimo = 19; Sokoru = 20; Yabu = 21; Other = 22

3.3 Survival Data Analysis

3.3.1 The Survival Model

The time until an event occurs is the outcome variable of interest in survival analysis, which is a collection of statistical processes for data analysis. Years, months, weeks, or days from the start of an individual's follow-up until an event occurs are referred to as time; conversely, time can refer to an individual's age when an event occurs. Death, disease occurrence, relapse from remission, recovery, or any other identified experience of interest that may occur to an individual are all examples of events. The time variable in a survival analysis is commonly referred to as "survival time" since it indicates how long an individual has survived over a given period of time. More frequently, mortality,

disease incidence, or some other bad individual experiences are considered as events of interest in survival analysis, and these events can also be referred to as failures (Kleinbaum et al., 2012).

Let T be a random variable associated with the survival times, t be the realization of the random variable T and $f(t)$ be probability density function of survival time t. The cumulative distribution function $F(t)$, which represents the probability that a subject selected at random will have a survival time less than some stated value t, for t ≥ 0 is given by

$$F(t) = p(T \leq t) = \int_0^t f(u)du. \tag{1}$$

The survival function is given by

$$S(t) = p(T > t) = \int_t^\infty f(u)du. \tag{2}$$

The hazard function becomes

$$h(t) = f(t)/S(t) = -\frac{d}{dt} lnS(t), \tag{3}$$

Survival model is usually expressed in terms of hazard function.

$$S(t) = exp(-\int_0^t h(u)du) = exp(-H(t)). \tag{4}$$

The cumulative hazard function is defined as

$$H(t) = \int_0^t h(u)du. \tag{5}$$

Under the parametric approach, the baseline hazard is a parametric function and the vector of its parameters are estimated together with the regression coefficients and the frailty parameters.

3.3.2 Non-parametric methods

Non-parametric survival analyses are more commonly utilized in cases when the exact form of the distribution is unknown. Estimates of the survival function and hazard function are used to summarize the data in survival analysis. The survival distribution estimation method yields descriptive statistics such as the median survival time. These

methods are referred to as non-parametric since they do not make any assumptions about the survival time distribution. The Kaplan-Meier, Nelson-Aalen, and Life Tables are the most extensively used survival and hazard functions estimation methods (Gatabazi, 2016).

The Kaplan-Meier (KM) estimator is a non-parametric survival function estimator that may be used to estimate survival probabilities from both censored and uncensored survival times (Kaplan & Meier, 1958). Suppose that r individuals have failures in a group of individuals, and $0 \leq t_{(1)} \leq t_{(2)} \ ... \ < t_{(r)} < \infty$ be the observed ordered death times. $r_{(j)}$ be size of the risk at $t_{(j)}$, where risk set encompasses individuals alive and uncensored before $t_{(j)}$. Let $d_{(j)}$ be the number of observed events at $t_{(j)}$, j = 1,2,... r. Then the Kaplan-Meier estimator of the survival probability of developing disease at any time is given by

$$\hat{S}(t) = \prod_{t_{(j)} \leq t} [\frac{r_{(j)} - d_{(j)}}{r_{(j)}}]. \tag{6}$$

The Cumulative hazard function of the Kaplan-Meier estimator can be estimated as:

$$\hat{H}(t) = -ln(\hat{S}(t)). \tag{7}$$

The log-rank test, first proposed by Breslow (1975), allows for comparison of the survival curves for two or more groups. It gives information on the significance of the difference in the survival of two groups of patients.

3.4 Modeling Frailty

Ordinary survival models deal with the simplest case of independent and identically distributed data. This is based on the assumption that the study population is homogeneous. But it is a basic observation of medical statistics that individuals differ greatly. So do the effects of a drug or the influence of various explanatory variables. Variability is a term used to describe this heterogeneity, and it is widely recognized as one of the most important sources of variability in medical and biological applications (Grover & Seth, 2014).

Frailty models are variations on the proportional hazards model, more commonly known as the Cox model. This is the most commonly used survival analysis model. In most clinical applications, survival analysis implicitly assumes that the population being investigated is homogeneous. This means that all participants in the study are, in theory, exposed to the same risks (risk of death and risk of disease recurrence). In many cases, the study population cannot be assumed to be homogeneous and must instead be viewed as a heterogeneous sample, implies, a collection of people with varying risks (Sargent, 1998).

The random effects (frailty) element is added into standard models of analysis to account for unmeasured variables or linked survival data. The frailty model, which was first proposed in the biostatistical literature by Vaupel et al. (1979) and is studied in depth by Hougaard (2000), Janssen & Duchateau (2011) and Wienke (2010), accounts for this baseline heterogeneity. It's a variation on the proportional hazards model in which the hazard function is determined by an unobservable random variable called frailty, which operates multiplicatively on the hazard function. Because observations within a subgroup share unmeasured risk factors that push them to quit earlier than other subgroups, models designed in terms of group-level frailties are commonly referred to as "shared frailty models." Individual level frailty-based models

In certain circumstances, it is impossible to measure all relevant factors at once, for instance. The value of various variables is still unknown in relation to the disease of interest, sometimes for economic reasons. The frailty method is a statistical modeling technique aimed at accounting for variation induced by unmeasured covariates. In statistical terms, a frailty model is a random effect model for time-to-event data where the random effect (the frailty) has a multiplicative effect on the baseline hazard function (Wienke et al., 2003).

3.4.1 Shared Frailty Model

Cox proportional hazard model (1972) can be used to control or evaluate the impact of a variety of explanatory variables or covariates on people's survival periods. The Cox model's popularity stems from the fact because the regression coefficients may be

estimated even though the model's baseline hazard is unknown. Thus, the Cox model is well-suited if one is more interested in the parameter estimates than the shape of the hazard. The Cox model, on the other hand, implicitly assumes that populations are homogeneous, implying that all people are equally vulnerable to failure (Klein & Moeschberger, 2003). However, including all important risk factors is unfeasible. As a result, it's critical to think of the population as heterogeneous, with individuals who face varying risks. The Cox model is extended to include frailty models. A frailty model is a time-to-event random effect model in which the random effect has a multiplicative inference on the baseline hazard (Kleinbaum et al., 2012).

Many statistical methods for modeling failure time data are based on the assumption that the observations are statistically independent (Kalbfleisch & Prentice, 2011). However, in many cases, this is not the case. The shared frailty model is a conditional model in which all participants in a cluster share frailty. The shared frailty concept is what causes event times to be dependent on one another. Because the frailties in each cluster are believed to be random, it's also known as a mixture model. Also, it is assumed that all event times in a cluster are independent for the given frailty. Frailty is a convenient technique to incorporate random effect, association, and unobserved heterogeneity into survival data models. Frailty is an unobserved random proportionality factor that alters the hazard function of an individual or related persons in its most basic form (Wienke et al., 2003).

Based on the population hazard function acquired from life tables, Vaupel et al. (1979) used the frailty approach to estimate the individual hazard function. Given the frailties, the shared frailty technique posits that all failure times in a cluster are conditionally independent. The frailty term's value is constant throughout time and is shared by all individuals in the cluster, therefore it's what causes the cluster's event times to be dependent. In shared frailty models, this reliance is always positive. There are two types of frailty models: those that use univariate time as an endpoint and those that use multivariate survival endpoints. Even though frailty may differ from group to group, shared frailty models assume that similar observations share frailty. In effect, observations within the same group are associated due to shared frailty. The shared frailty

model assumes that members of a subgroup or pair have the same frailty denoted by u, but that frailty varies from group to group (Hougaard, 1995).

Conditional on the random effect, called the frailty denoted by u_i, the survival times in cluster i ($1 \leq$ i $\leq$ n) are assumed to be independent and the proportional hazard frailty model denoted by

$$h_{ij}(t/x_{ij}, u_i) = h_0(t)exp(\beta' x_{ij} + u_i). \tag{8}$$

If the proportional hazards assumption is not fulfilled the alternative is the accelerated failure time frailty model which denoted by

$$h_{ij}(t/x_{ij}, u_i) = h_0(exp(\beta' x_{ij} + u_i)t)exp(\beta' x_{ij} + u_i), \tag{9}$$

where i indicates the i^{th} cluster and j indicates the j^{th} individual for the i^{th} cluster, $h_0(.)$ is the baseline hazard function, u_i is the random term of all the subjects in cluster i, x_{ij} is the vector of covariates for subject j in cluster i and β the vector of regression coefficients. If the number of subjects n_i is one for all groups, the univariate frailty model is obtained. Unless otherwise the model is called the shared frailty model as result of all the subjects in the same cluster share the same frailty value. Suppose that Z = $exp(u_i)$ and let Z is distributed as gamma or inverse Gaussian distribution, in addition on an unobservable random variable Z, which acts multiplicatively on the baseline hazard function (Hanagal, 2019). The assumption of shared frailty model is that subjects in the same cluster share the same frailty value Z_i, where i = 1,2,...n. Fore this reason it is called shared frailty model. But frailty differ from group to group (Huang & Wolfe, 2002).

3.4.2 Test of unobserved heterogeneity

For frailty models variance of random terms (θ) is estimated to obtain ideas in the outcome in the clusters. If θ is large as well as significantly different from zero, it indicates heterogeneity among the cluster and strong association among individuals in the same cluster. In other way, when θ is equal to zero, the frailties are identically equal to one which implies that the cluster effects are not present and events are independent within and across cluster (Glidden & Vittinghoff, 2004).

3.5 Parameterization

In order for the hazard ratio to remain constant throughout time, the term proportionate hazards (PH) denotes that the hazard function of one group is proportional to the hazard function of the other group (Klein, 1992). Therefore, the hazard ratio is given by HR = $\exp(\beta' X_{ij})$, where β' = 1,2,..., p is a vector of regression coefficients and X_{ij} is the vector of covariates for subject j in cluster i. The accelerated failure-time (AFT) model explains how the survival time might increase or decrease depending on the predictor variables. This is the acceleration factor indicated by ϕ is $\exp(\alpha' X_{ij})$, where α' = 1,2,..., p is, in the case of the AFT model, a vector of regression coefficients. The relationship between exponential, weibull, and log logistic survival models and β is given by

(a) For exponential β_j = -α_j, the exponential PH and AFT are in fact the same model, except that the parameterization is different, and hence HR = $\exp(-\alpha_j)$ is the hazard ratio of the j^{th} group with the reference groups.
(b) For weibull β_j = -$\alpha_j\rho$, where ρ is the shape parameter and hence, HR = $\exp(-\alpha_j\rho)$ is the hazard ratio of the j^{th} group with reference groups.
(c) For loglogistic, β_j = -$\alpha_j\rho$, where ρ is the shape parameter and OR = $\exp(-\alpha_j\rho)$ indicates the failure odds ratio of the j^{th} group with reference groups. The log-logistic model, which has a constant OR for two groups, is a proportional odds (PO) model.

3.5.1 Baseline Survivor and Hazard Functions

Baseline hazard can be assumed in frailty models like proportional hazards model, parametric and non-parametric forms. If non-parametric form is assumed for baseline hazard ($h_0(t)$) , then semi parametric proportional hazards model is considered and the estimates are usually obtained by using Expectation-Maximization algorithm. When parametric form for $h_0(t)$ is assumed, the baseline hazard function is a parametric function and the vector of its parameters estimated together with the regression coefficients and the frailty parameter(s). Then maximum likelihood estimates can be obtained by maximizing the likelihood function. If one of the four functions (density function, survival function, hazard function, and cumulative hazard function) is speci-

fied, it specifies the other three functions of the above baselines as stated under section 3.3.1. Parameter λ is reparameterized in terms of predictor variables and the regression parameters. In parametric models, the shape parameter ρ is fixed.

Table 3.2: Baseline distribution for Survival and Hazard functions

Distribution	h(t)	S(t)	f(t)	Parameter space
Exponential	λ	$\exp(-\lambda t)$	$\lambda \exp(-\lambda t)$	$\lambda > 0$
Weibull	$\rho\lambda t^{\rho-1}$	$\exp(-\lambda t^{p})$	$\rho\lambda t^{\rho-1} \exp(-\lambda t^{p})$	$\lambda,\ \rho > 0$
Log-logistic	$\frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}}$	$\{1+\lambda t^{\rho}\}^{-1}$	$\frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}}\ \{1+\lambda t^{\rho}\}^{-1}$	$\lambda \in :\ \mathbf{R},\ \rho > 0$

3.6 Frailty Distribution

The structure, properties, and applicability of survival models to practical problems depend on the nature of time-to-event data. The main premise behind the frailty model is that people have various levels of frailty and that the most frail will die sooner than the least frail. As a result, a systematic selection of strong individuals occurs, distorting what is observed. The distribution of frailty among individuals determines the precise nature of the link between individual and population aging. For the unobserved covariates, several distributions can be used, resulting in qualitative and quantitative changes. In particular, the variance of the frailty distribution determines the degree of heterogeneity in the study population.

The classical and widely used frailty model is based on a proportional hazards model with a random effect (frailty). The frailty z_i is an unobservable realization of a random variable Z with probability density function f(.) which is the frailty distribution. Since z_i multiplies the hazard function, Z has to be non-negative. Here, Z is considered as a random mixture variable, varying across the population. Note that a scale factor common to all subjects in the population maybe absorbed into the baseline hazard function, so that frailty distributions are standardized to mean of Z is one. The variance parameter σ^2 is variance of Z is interpretable as a measure of heterogeneity across the population in baseline risk. When σ^2 is small, then the values of Z are closely

concentrated around one. If σ^2 is large, then values of Z are more dispersed, inducing greater heterogeneity in the individual hazards. Frailty increases the individual risk and is sometimes called liability or susceptibility in other settings. All individuals, apart from an individual constant Z are assumed to follow the same mortality pattern.

What may be observed in a population is not the individual hazard but the net result for a number of individuals with different values of the random variable Z. One important problem in the area of frailty models is the choice of the frailty distribution. The Laplace transform for frailty receives special attention because it can easily express unconditional survival and hazard functions. Hence, the likelihood function can also be expressed by means of the Laplace transform. This is the reason why frailty distributions with easy Laplace transforms are so popular and allow maximum likelihood methods in parameter estimation. Different distributions have been proposed for the frailty term. In this study, gamma and inverse Gaussian frailty distributions were applied. In both cases, the degree of independence is represented by a single heterogeneity parameter denoted by θ.

3.6.1 Gamma Frailty Distribution

Choosing the frailty distribution is often exercised by the problem at hand in terms of the model implications. As a mixing distribution, the gamma distribution has been widely used by Janssen & Duchateau (2011), Kong & Kaddoum (2019) and Oakes (1982). It fits very well as a mixture distribution to failure data from a computational and analytical standpoint (Wienke, 2010). Hougaard (1986) claimed that the gamma and inverse Gaussian distributions are important and mathematically tractable as frailty distributions for a heterogeneous population between groups.

The gamma distribution is a versatile distribution that can take many different forms. The gamma distribution is most commonly used for its mathematical simplicity. This is due to the Laplace transform's derivative being simple, allowing maximum likelihood approaches to be employed for parameter estimation (Wienke et al., 2003). Another rationale for choosing the gamma distribution as the frailty distribution is its flexible shape (Manda, 2020). Although the gamma distribution is the most often utilized

frailty distribution for mathematical reasons, Hougaard (1995) pointed out that there are no biological reasons for using it.

To make the model identifiable, restrict that expectation of the frailty equals one and variance be finite. So that only one parameter needs to be estimated. Thus, the distribution of frailty Z is one parameter gamma distribution. The density of a gamma-distributed random variable with parameter $\theta > 0$ is given by:-

$$f_z(z_i) = \frac{z_i^{(1/\theta)-1} exp(-z_i/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}} \tag{10}$$

Where $\Gamma(.)$ is gamma function. It corresponds to a Gamma distribution (μ, θ) with μ fixed to one for identifiability and its variance becomes θ. With corresponding Laplace transform

$$\mathrm{L(s)} = (1 + \tfrac{s}{\theta})^{-\theta} \text{ , where } \theta > 0$$

The conditional survival function of the gamma frailty distribution is given by

$$S_\theta(\mathrm{t}) = (1 - \theta ln(S(t)))^{-1/\theta}$$

And also the conditional hazard function is given by

$$h_\theta(\mathrm{t}) = \mathrm{h(t)}(1 - \theta ln(S(t)))^{-1}$$

Where S(t) and h(t) are the survival function and the hazard functions of the baseline distributions. in case of Gamma distribution it measures the association between any two event times from the same cluster in the multivariate case. and can be determined as

$$\tau = \tfrac{\theta}{\theta+2} \text{ , where } \tau \in (0,1)$$

3.6.2 Inverse Gaussian Frailty Distribution

The inverse Gaussian distribution was introduced as a frailty distribution alternative to the gamma distribution by Hougaard (1984) and has been used for example by Manton et al. (1986), Price & Manatunga (2001) and Kheiri et al. (2007). The unconditional survival and hazard functions have straightforward closed-form formulas, similar to the gamma frailty model, which makes the concept appealing. The correlation between

related event times is of particular importance in the multivariate scenario. Distinct frailty distributions result in different dependent structures. inverse Gaussian frailties produce dependency in the middle time (Hougaard, 1995). A fit assessment should accompany the selection of frailty distributions. It's only reasonable to base the mean of the frailty variable on the observed filtration, which should be around one (Pipper & Martinussen, 2004).

The probability density function of an inverse normal distributed random variable with mean one and variance $\sigma^2 = \theta$ is:

$$f(z) = \frac{1}{\sqrt{2\pi\theta z^3}} exp(-\frac{1}{2\theta z}(z-1)^2) \tag{11}$$

where as $\theta > 0$, z > 0. Consequently, the Laplace transform of the inverse normal distribution is given by:

$$\text{L(s)} = \exp(\tfrac{1}{\theta}(1 - \sqrt{1+2\theta s})) \text{ , } \theta \text{ and s} > 0$$

Hence, the conditional survival and hazard function take the forms

$$S_\theta(t) = \exp(\tfrac{1}{\theta}(1 - \sqrt{1-2\theta lnS(t)})) \text{ , } \theta > 0$$

$$h_\theta(t) = \text{h(t)}\ (1-2\theta lnS(t))^{-1/2} \text{ , } \theta > 0$$

An Inverse Gaussian distributed frailty yields a Kendall's Tau given by

$$\tau = \tfrac{1}{2} - \tfrac{1}{\theta} + 2\ \tfrac{exp(2/\theta)}{\theta^2}\ \int_{2/\theta}^{\infty} \tfrac{exp(-u)}{u} du \text{ , where } \tau \in (0,\ \tfrac{1}{2})$$

3.7 Parameter Estimation Method

Estimation of the frailty model can be parametric or semi-parametric. In the former case, a parametric density is assumed for the event times, resulting in a parametric baseline hazard function. Estimation is then conducted by maximizing the marginal log-likelihood. In the second case, the baseline hazard is left unspecified and more complex techniques are available to approach that situation (Abrahantes et al., 2007). Though semi parametric estimation offers more flexibility, the parametric estimation will be more powerful if the form of the baseline hazard is somehow known in advance (Munda et al., 2012).

The Frailty models account for the clustering present in grouped event time data. For a right-censored clustered survival data, the observation for subject j $\in J_i$ = (1,...,n_i) from cluster $i \in I$ = (1,...,s) is the couple (y_{ij}, δ_{ij}), where y_{ij} = min(t_{ij}, c_{ij}) is the minimum between the survival time t_{ij} and the censoring time c_{ij}, and indicator δ_{ij} = $I(t_{ij} \leq c_{ij})$ is one for a subject where the event has taken place, while $\delta_{ij} = 0$ for censored observation. When covariate information's been collected the observation will be (y_{ij}, δ_{ij}, x_{ij}), where x_{ij} denote the vector of covariates for the ij^{th} observation. In case of the parametric setting, estimation is based on the marginal likelihood in which the frailties have been integrated out by averaging the conditional likelihood with respect to the frailty distribution.

Based on the assumption of right censoring of independence between the censoring time and the survival time of random variables, given the covariate information, the marginal log-likelihood of the observed data can be given by

$$l_{marg}(\varphi,\beta,\theta, \text{z, X}) = \sum_{i=1}^{s} \left(\sum_{j}^{n_i} \delta_{ij}(\log(h_0(y_{ij})) + X_{ij}^T \beta)\right) + \log((-1)^{d_i} L^d \left(\sum_{j}^{m} H_0(y_{ij})\exp(X_{ij}^T \beta)\right)).$$

Where $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ is the number of events in the i^{th} clusters and $L^q(.)$ is the q^{th} derivative of the Laplace transform of the frailty distribution Z is defined as

$$L_{(s)} = \text{E(exp(-Zs))} = \int_0^\infty \exp(Z_i s)\ f(Z_i)dz_i, \text{ where s} > 0.$$

$$L^q(\text{S}) = -1^{(q)} \int_0^\infty Z^q \exp(\text{-Zs})\text{f(z)dz , q} > 0$$

Where φ represents a vector of parameters of the baseline hazard function, β the vector of regression coefficients and θ the variance of the random effect. The estimates of φ, β, θ are obtained by maximizing the marginal log-likelihood stated above. This can be done if one is able to compute higher order derivatives $L^q(.)$ of the Laplace transform up to q = max($d_1, ..., d_s$).

3.8 Comparison of Models

In statistics, model selection is a process researchers use to compare the relative values of different statistical models and determine which one is the best fit for the observed

data. The most commonly used model selection criteria are the Akaike information criterion (AIC) and Bayesian information criterion (BIC). A data-driven model selection method such as an adapted version of Akaike's information criterion AIC is used to find the truncation point of the series (Akaike, 1974). AIC and BIC are given by the expression

$$AIC = -2\log(L) + 2(k + c + 1) \tag{12}$$

$$BIC = -2\log(L) + 2kln(N) \tag{13}$$

where L is the maximized likelihood value, k is the number of covariates, c the number of model specific distributional parameter, and N is the total sample size. In this study, AIC and BIC criterion were used to compare various candidates for parametric frailty models. The R software was used for comparison manipulation.

3.9 Model Diagnostics

Checking the adequacy of a model in describing a dataset is an essential part of any statistical analysis. It is generally recommended to assess the adequacy before using a model for decision-making purposes. Ideally, we would like our model to be flexible and parsimonious, with the ability to fit a wide range of data satisfactorily. Thus, an assessment of the quality of a fit and adherence to model assumptions are as important as model development in any statistical analysis. Many of the model diagnostic procedures are based on graphical assessment. The graphical methods can be used to check if a parametric distribution fits the observed data or not (Dätwyler & Stucki, 2011).

3.9.1 Evaluation of the Parametric Baselines

The model with the Weibull baseline has a property that the log(-log($\hat{S}$(t))) is linear with log time, where $\hat{S}$(t) = exp(- λ t^{ρ}). Hence, log(-log($\hat{S}$(t))) = log(λ) + ρlog(t). The intercept and slope of the line will be rough estimate of log(λ) and ρ respectively. This property allows a graphical evaluation of the appropriateness of a Weibull model by plotting log(-log($\hat{S}$(t))) versus log(t) where $\hat{S}$(t) is the Kaplan-Meier survival estimate (Goel et al., 2010). The appropriateness of the model with the exponential baseline can graphically be evaluated by plotting -log($\hat{S}$(t)) versus time. The plot should be

linear and pass through the origin (Costella, 2010). The appropriateness of the model with the log logistic baseline can graphically be evaluated by plotting log(1 - $\hat{s}$(t)/$\hat{s}$(t)) versus log(t). The log-failure odd versus log time of the log-logistic model is linear with slope ρ then the survival time follows a log-logistic distribution. Where the failure odds of log-logistic survival model can be computed as follows

$$1 - s(t)/s(t) = \frac{\lambda t^{\rho}}{1+\lambda t^{\rho}} \Big/ \frac{1}{1+\lambda t^{\rho}} = \lambda t^{\rho}.$$

Then, the log-failure odds becomes

$$\log(1 - s(t)/s(t)) = \log(\lambda t^{\rho}) = \log(\lambda) + \rho \log(t),$$

which is the liner function of log(t) (Dätwyler & Stucki, 2011).

3.9.2 The Cox-Snell Residuals

Cox-Snell residuals are a type of standardized residuals used in reliability analysis. A residual is the difference between an observed data point and a predicted or fitted value. A Cox-Snell residual considers the distribution and estimated parameters from the lifetime regression model. The Cox-Snell residuals are equal to the negative of the natural log of the survival probability for each observation. Cox-Snell residual for the j^{th} individual with observed survival time t_j is given by $r_j = \hat{H}(T_j/X_J)$ -log$\hat{S}(T_j/X_j)$, where $\hat{H}$ and $\hat{S}$ are the estimated values of the cumulative hazard and survivor function of the j^{th} subjects at time t_j respectively. If the model fits the data, then the $r_j's$ should have standard ($\lambda = 1$) exponential distribution, so that a hazard plot of r_j versus the Nelson-Aalen estimator of the cumulative hazard of the $r_j's$ should be straight line with slope unity and intercept zero. If yes, the fitted model is appropriate. In general, residual check the overall fit of the model (Cox & Snell, 1968).

Table 3.3: The three Cox Snell residuals baseline hazard functions

Model	r_j
Exponential	$\hat{\lambda} t_j \exp(\hat{\beta}' X_j)$
Weibull	$\hat{\lambda}\ t_j^{\hat{\rho}}\ \exp(\hat{\beta}' X_j)$
Log-logistic	$\ln\left[\frac{1}{1+\hat{\lambda} t_j^{\hat{\rho}} exp(\hat{\beta}' X_j)}\right]$

3.9.3 Quantile-Quantile plot

The accelerated failure-time models are an alternative to the proportional hazards model when comparing two groups. To see if this provides a sufficient fit to the data, a quantile-quantile or q-q plot is used. The graph is based on the notion that in accelerated failure-time models,

$$S_1(t) = S_0(\theta t),$$

where S_0 and S_1 are the survival functions in the two groups and θ is the acceleration factor. Let t_{0p} and t_{1p} be the p^{th} percentiles of groups 0 and 1, respectively, that is

$$t_{kp} = S_k^{-1}(1-p), k = 0, 1.$$

By the relationship $S_1(t) = S_0(\theta t)$, we must have $S_0(t_{0p}) = 1 - p = S_1(t_{1p}) = S_0(\theta t_{1p})$ for all t. If the accelerated failure time model holds, $t_{0p} = \theta t_{1p}$. To see this assumption we compute the KaplanMeier estimators of the two groups and estimate the percentiles T_{1P}, T_{0P}, for various values of p. If we plot the estimated percentile in group 0 versus the estimated percentile in group 1 (i.e., plot the points t_{1p}, t_{0p} for various values of p), the graph should be a straight line through the origin, when the accelerated failure time model holds. If the curve is linear, the slope of the line provides a rough estimate of the acceleration factor q.

4 Result and Discussion

4.1 Descriptive Statistics

To learn more about the distribution of the variables, descriptive statistics are employed and descriptive summaries are given in Table 4.1. The shortest and longest event times recorded during the follow-up of IHD patients were 1 and 30 days, respectively. About 64.63 percent of those IHD patients were censored (right censored), while the remaining 35.37 percent have since passed away. The median event time of IHD was 5 days. Regarding Sex, 26.69%, of the IHD patients were female and the remaining were male during the follow-up study. Based on residential place, Jimma had the highest death rate of IHD 29 (26.36%), followed by Mana 9 (8.17%) among the twenty-two woredas. And also, Dedo, Sakka, and others have the same death experience as 8 (7.27%). But Chora Botor and Limmu Ganat did not experience death in general.

By observing the diabetes mellitus of IHD patients, most IHD patients were 65.27% non-diabetic and the death proportion seemed highest for those IHD patients who were diabetic, which was 54.55% compared to non-diabetic, which was 45.45%. Looking at the cholesterol level of IHD patients, about 47.27% and 52.73% were IHD patients with cholesterol and those without cholesterol, respectively, in which IHD patients with cholesterol seemed to have a lower survival time because the death proportion for patients with cholesterol (68.18%) is higher than for IHD patients without cholesterol (31.82%). Most IHD patients have no obesity, 72.35%, and the remaining have obesity. And the death rate appears to be highest in IHD patients who were obese, at 50.91%, compared to 49.09% in non-obese patients. Moreover, about 52.73% of IHD patients had other diseases, and the rest did not.

Table 4.1: Descriptive summary of covariates of IHD mortality in JUMC

Patient's status				
Covariates	Categories	Number of censored(%)	Number of death(%)	Total
Sex	Female	52 (25.87)	31 (28.18)	83 (26.69)
	Male	149 (74.13)	79 (71.82)	228 (73.31)
Age	0-14	4 (1.99)	1 (0.91)	5 (1.61)
	15-47	75 (37.31)	43 (39.09)	118 (37.94)
	48-63	79 (39.31)	38 (34.55)	117 (37.62)
	≥ 64	43 (21.39)	28 (25.45)	71 (22.83)
Diabetes mellitus	No	153 (76.12)	50 (45.45)	203 (65.27)
	Yes	48 (23.88)	60 (54.55)	108 (34.73)
Hypertension	No	154 (76.62)	54 (49.09)	208 (66.88)
	Yes	47 (23.38)	56 (50.91)	103 (33.12)
Obesity	No	171 (85.07)	54 (49.09)	225 (72.35)
	Yes	30 (14.93)	56 (50.91)	86 (27.65)
Smoking status	No	165 (82.09)	101 (91.82)	266 (85.53)
	Yes	36 (17.91)	9 (8.18)	45 (14.47)
Cholesterol	No	129 (64.18)	35 (31.82)	164 (52.73)
	Yes	72 (35.82)	75 (68.18)	147 (47.27)
Other disease	No	118 (58.71)	29 (26.36)	147 (47.27)
	Yes	83 (41.29)	81 (73.64)	164 (52.73)

Residential place	Agaro	14 (6.97)	5 (4.55)	19 (6.11)
(woreda)	Chora Botor	9 (4.48)	0 (0.00)	9 (2.89)
	Dedo	8 (3.98)	8 (7.27)	16 (5.14)
	Gatira	5 (2.49)	3 (2.73)	8 (2.57)
	Gera	3 (1.49)	1 (0.91)	4 (1.29)
	Gomma	6 (2.99)	6 (5.45)	12 (3.86)
	Guma	7 (3.48)	1 (0.91)	8 (2.57)
	Jimma	48 (23.88)	29 (26.36)	77 (24.26)
	Kersa	7 (3.48)	7 (6.36)	14 (4.50)
	Limmu Ganat	5 (2.49)	0 (0.00)	5 (1.61)
	Limmu Kosa	8 (3.98)	3 (2.73)	11 (3.54)
	Limmu Sakka	1 (0.49)	5 (4.55)	6 (1.93)
	Mana	11 (5.47)	9 (8.17)	20 (6.43)
	Nono	5 (2.49)	2 (1.82)	7 (2.25)
	Omo Nada	13 (6.46)	5 (4.55)	18 (5.79)
	Sakka	5 (2.49)	8 (7.27)	13 (4.18)
	Shabe Sonbo	6 (2.99)	3 (2.73)	9 (2.89)
	Setema	7 (3.48)	2 (1.82)	9 (2.89)
	Sigimo	5 (2.49)	2 (1.82)	7 (2.25)
	Sokoru	7 (3.48)	1 (0.91)	8 (2.57)
	Yabu	6 (2.99)	2 (1.82)	8 (2.57)
	Other	15 (7.46)	8 (7.27)	23 (7.41)
		median(days) = 5		

4.2 Survival of Significantly Different Groups

The Log-rank test and Kaplan-Meier survival estimates are used to look into the significance of the difference in survival experience among different factors. Therefore, the log-rank test results presented in Table 4.2 show that there is a statistically significant difference in experiencing death events among the groups of diabetes mellitus, hypertension, obesity, smoking status, cholesterol, and other diseases at 5% level of significance.

Table 4.2: Log rank test for equality of Survival function of different groups

Covariate	Chi-square value	df	Pr > Chi-Square
Sex	0.1	1	0.8
Age	1.5	3	0.7
Dibetes mellitus	28	1	<0.001
Hypertension	27.9	1	<0.001
Obesity	50.3	1	<0.001
Smoking status	6.4	1	0.01
Cholesterol	32.2	1	<0.001
Other disease	30.7	1	<0.001

The Kaplan-Meier estimator survival curve can be used to estimate survival function among different covariates so that one can make a comparison. Separate graphs of the estimates of the Kaplan-Meier survivor functions are constructed for different categorical covariates. In general, the survivorship pattern of one is lying above another, which means the group defined by the upper curve has a better survival than the group defined by the lower curve (Figure 4.1).

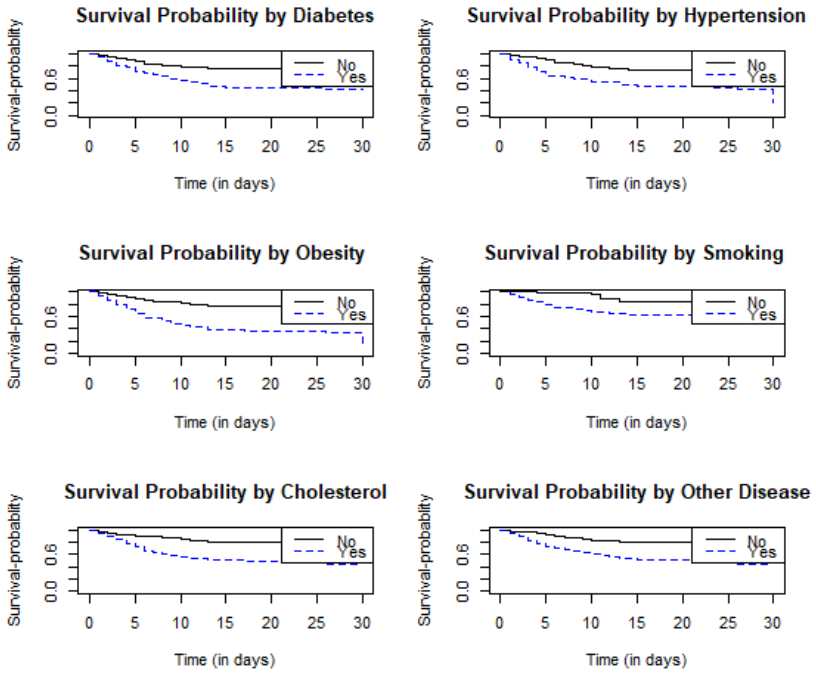

Figure 4.1: Survival curves for significantly different groups

4.3 Test of unobserved heterogeneity

Likelihood ratio is used for comparing the models with and without frailty. It is used for testing the null hypothesis H_0: $\theta = 0$ versus the alternative hypothesis H_1: $\theta > 0$. Heterogeneity among parameter θ from the frailty models was estimated using the mariginal Likelihood techniques. Multivariable analysis was done by assuming the exponential, Weibull and loglogistic baseline hazard functions for Gamma and Inverse Gaussian shared frailty distributions. The results given in Table 4.3 shows that the likelihood ratio tests of variance of random term (θ) for exponential gamma, exponential inverse gaussian, weibull gamma, Weibull inverse gaussian, loglogistic gamma and loglogistic inverse Gaussian shared frailty models were 149, 159, 167, 153, 168 and 154 with p values of <0.001 for all shared frailty model. Thus from this results we can conclude that unobservable heterogeneity is significant in all models at 5% level of significance.

The heterogeneity parameter (variance of random effect) is highest for loglogistic in-

verse Gaussian shared frailty model (θ =0.948) followed by loglogistic Gamma shared frailty model (θ = 0.569) and the least (θ=0.005) for exponential gamma shared frailty model with exponential baseline hazard. The Kendall's tau (τ) is used to measure the dependence within the clusters (woredas) and it is higher for the higher variance of random effect (θ) values. From the results of this study the values of Kendall's tau (τ) for exponential gamma, exponential inverse gaussian , weibull gamma, Weibull inverse gaussian, loglogistic gamma and loglogistic inverse Gaussian shared frailty models were 0.022, 0.019, 0.02, 0.017, 0.018.and 0.051,respectively. This evidence shows that, on average, there is a positive correlation between times to deaths within the clusters (woredas).

Table 4.3: Test of unobserved heterogeneity by using likelihood ratio test

Shared frailty model	LRT	θ	τ	P value
Exponential Gamma	149	0.005	0.022	<0.001
Exponential Inverse Gaussian	159	0.075	0.019	<0.001
Weibull Gamma	167	0.561	0.020	<0.001
Weibull Inverse Gaussian	153	0.081	0.017	<0.001
Loglogistic Gamma	168	0.569	0.018	<0.001
Loglogistic Inverse Gaussian	154	0.948	0.051	<0.001

LRT = likelihood ratio test of θ = 0 with 0 and 1 degrees of freedoms, θ = variance of random terms, τ = kedall"s tau

4.4 Univariable Analysis

A univariate analysis was performed to select variables to be included in the multivariable analysis. The outputs in univariable analysis (Annex I) show that the covariates of IHD, diabetes mellitus, hypertension, obesity, smoking status, cholesterol, and other diseases are statistically significant in the entire models used. The confidence interval of the acceleration factors for those covariates does not include one in all models. This shows that those variables are determinant factors for the time to death of ischemic

heart disease. However, sex and age of patients are statistically not significant according to all the models at a 0.25 level of significance. Therefore, based on this result, it is better to ignore those insignificant covariates and to do multivariable analysis using the significant covariates.

4.5 Model Comparison

For the data on IHD, the parametric baseline distribution with Gamma and the inverse Gaussian shared frailty distribution were fitted. In survival analysis, to compare the efficiency of parametric shared frailty models, the Akaike Information Criterion (AIC), which assesses the goodness of fit of a statistical model, was used. It is the most commonly applicable criterion for selecting the model. The multivariable analysis of loglogistic inverse Gaussian models AIC value, which is 908.6871, is the lowest of all the models, indicating that it is the most effective model among the parametric frailty models for describing the IHD dataset (Table 4.4).

Table 4.4: The Multivariable Parametric Shared Frailty Models' AIC and BIC Values

Model			
Baseline hazard function	Frailty distribution	AIC	BIC
Exponential	Gamma	916.4669	944.2260
	Inverse-Gaussian	914.5860	957.5034
Weibull	Gamma	919.3614	999.1133
	Inverse-Gaussian	916.5015	964.5191
Loglogistic	Gamma	910.8247	990.3987
	Inverse-Gaussian	**908.6871**	938.2056

AIC = Akaike's Information Criterion

BIC = Bayesian's Information Criterion

4.6 Multivariable Analysis

Similar to univariable survival analysis, multivariable survival analysis was performed by taking into account the three baseline hazard functions (exponential, Weibull, and

log-logistic) and two frailty distributions (gamma and inverse Gaussian), using the six most important covariates from univariable analysis of the models. Based on multivariable analysis output, a log-logistic inverse Gaussian frailty model is selected by using AIC. In multivariable frailty models, the covariates diabetes mellitus, hypertension, obesity, Smoking status, cholesterol, and other diseases are significant factors for time to death in ischemic heart disease.

The results of this study suggested that diabetes mellitus had a significant effect on the time to death of the IHD. Patients with diabetes mellitus had significantly different death times than the reference group. As a result, the ϕ of dying of IHD patients with diabetes mellitus was 0.471 times that of those without diabetes mellitus ($\phi = 0.471$, CI: 0.314, 0.706), indicating that the time to death of diabetic patients was reduced by 52.9% when compared to those without diabetes mellitus.

Consequently, the findings of this study suggested that hypertension had a notable impact on the IHD patient's time to death. Compared to patients with and without hypertension, those with hypertension had considerably shorter death times. The ϕ of dying in hypertensive patients was 0.521 times that of those without hypertension ($\phi = 0.521$, CI: 0.346, 0.784), indicating that the time to death in hypertensive patients was reduced by 47.9% when compared to patients who were not hypertensive.

Depending on the result, obesity had a significantly different time to death from IHD than the reference groups. And the ϕ of being died of patients with obesity was 0.279 times the factor of those without obesity ($\phi = 0.279$, CI: 0.185, 0.422), this indicates that the time to death of patients who were obese was reduced by 72.1% when compared with patients who were not obese. Therefore, IHD patients who had obesity had a shortened time of death. The acceleration factor and its 95% confidence interval for smoking is 2.724 (1.381, 5.372). In the 95% confidence interval of the acceleration factors, one is not included, which implies that smoking status determines the time to death of ischemic heart disease. The p-values also support this, which is less than 0.05 for smoking status (yes), when compared to smoking status (no) as the reference category.

Moreover, the acceleration factor, ϕ of being died of patients with cholesterol was 0.393 times the factor of those without obesity (ϕ = 0.393, CI: 0.264, 0.585), this indicates that the time to death of patients who were cholesterol was reduced by 60.7% when compared with patients who were not cholesterol. Finally, the ϕ of being died of patients with other disease was 0.382 times the factor of those without other disease (ϕ = 0.382, CI: 0.246, 0.593), this indicates that the time to death of patients who were infected by other disease was reduced by 61.8% when compared with patients who were not infected by other disease source (Table 4.5)

Table 4.5: Results of Loglogistic Inverse Gaussian multivariable shared frailty model.

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P-value
Intercept		5.231	186.887	0.291	(105.556, 330.883)	<0.001
Diabetes mellitus	No(rf)		1			
	Yes	-0.752	0.471	0.207	(0.314, 0.706)	<0.001
Hypertension	No(rf)		1			
	Yes	-0.651	0.521	0.208	(0.346, 0.784)	0.002
Obesity	No(rf)		1			
	Yes	-1.273	0.279	0.209	(0.185, 0.422)	<0.001
Smoking status	No(rf)		1			
	Yes	1.002	2.724	0.346	(1.381, 5.372)	0.004
Cholesterol	No(rf)		1			
	Yes	-0.933	0.393	0.203	(0.264, 0.585)	<0.001
Other disease	No(rf)		1			
	Yes	-0.962	0.382	0.224	(0.246, 0.593)	<0.001
θ = 0.948	λ = -5.588					
ρ = 0.998	τ = 0.051				AIC = 908.6871	

Source = JUMC, $\hat{\beta}$ = Coefficients, St.err = Standard error, 95% CI = confidence interval for acceleration factor, rf = reference, θ = Variance of random term, τ = kedall's tau, ϕ = Acceleration factor, AIC = Akaike's information criteria

4.7 Checking for overall goodness of fit

4.7.1 Diagnostic Plots of the Parametric Baselines

The overall goodness of fit was used to evaluate the model. It is desirable to make this determination in order to know if a fitted parametric model accurately describes the data or not. Of the three parametric baseline graphs, the log-logistic curve is more linear than the others. This suggests that a log-logistic baseline hazard is a better choice for the IHD dataset (Figure 4.2).

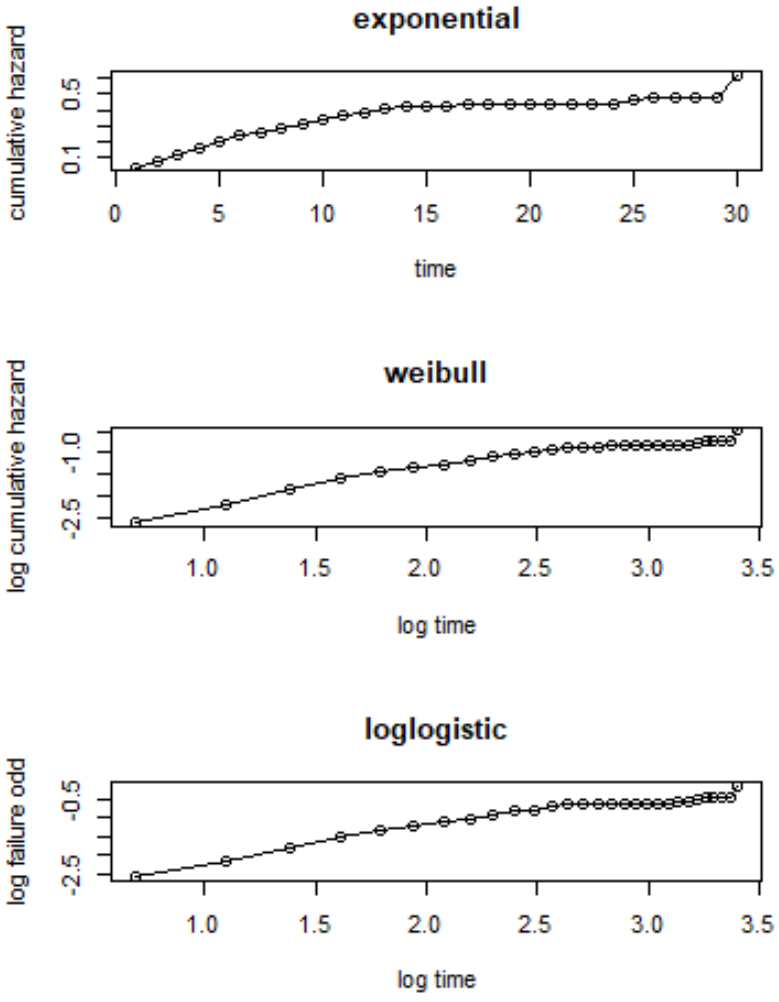

Figure 4.2: Diagnostic plot for baselines hazards

4.7.2 Cox Snell Residual plots

One method to assess how well the model fits the data is to look at the Cox-Snell residual. Relatives should resemble a censored sample from a unit exponential distribution if the model fits. In other words, departures from the norm ought to be minimal. Figure 4.3 shows the plot for the residuals for the exponential, Weibull, and logarithmic models fitted to the data using maximum likelihood estimation and cumulative hazard functions. The residuals plot for the loglogistic hazard function can be observed to be rather near to the 45-degree straight line through the origin. The graphic demonstrates that the log-logistic model that was fitted to the data is generally acceptable.

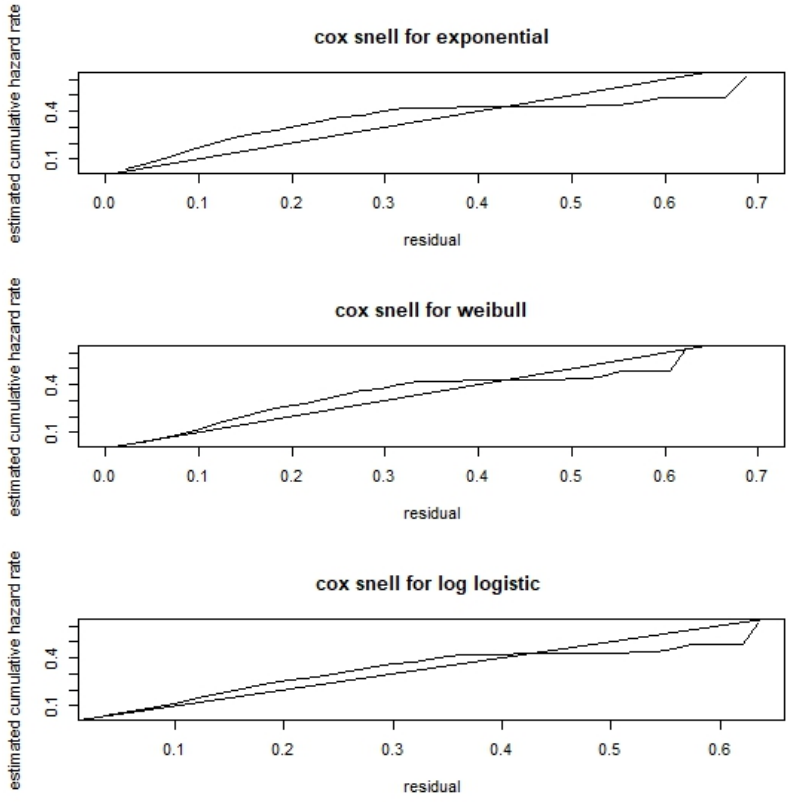

Figure 4.3: Cox Snell residual plots

4.7.3 Adequacy of Accelerated Failure Time

To determine whether the accelerated failure time provides a satisfactory fit to the data used by two different populations, a quantile-quantile or q-q plot is created. By contrasting the substantially different groups (diabetes mellitus, hypertension, obesity, smoking status, cholesterol, and other disease), graphically assess the accelerated failure-time model's suitability (Figure 4.4). For all factors, the figures appeared to be linear. As a result, an accelerated failure time model was developed, using the log-logistic as a baseline.

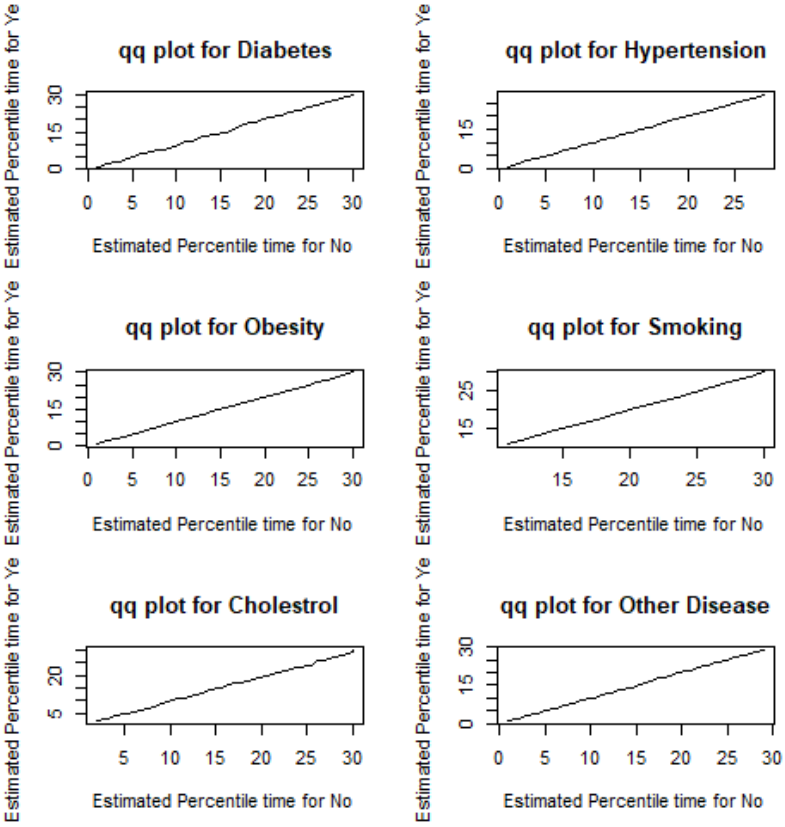

Figure 4.4: Quantile-Quantile Plots

4.8 Discussion

Ischaemic heart disease (IHD) is estimated to be the leading cause of mortality in the world and in high-income countries it is also the leading cause of premature mortality and disability (Guilbert, 2000). The primary goal of this study was to use parametric shared frailty models to model the time to death of patients with IHD. Among many parametric frailty models, Gamma and Inverse Gaussian shared frailty model were used with exponential,,Weibull and loglogistic as baseline hazard function.

Woredas were utilized as clusters in this study because populations within a certain woredas typically share certain factors, such as the environment, medical facilities, and other factors, in determining the time to death of ischemic heart disease. The likelihood ratio test was used to examine the impact of clustering (unobserved heterogeneity) between the clusters, and the results revealed that in all shared frailty models, the variance of the random effect term is significantly different from zero at the 5% level of significance.

Sex, age, diabetes mellitus, hypertension, obesity, smoking status, cholesterol, and other diseases were factors that were of concern for this study. According to the univariable analysis shown in (Appendix I), the time to death from IHD was strongly influenced by diabetes mellitus, hypertension, obesity, smoking, cholesterol, and other diseases. A similar study done at different places and times using different methods by Nowbar et al. (2019), Moyehodie et al. (2022), Mensah (2008), and Sakboonyarat & Rangsin (2018), as illustrated in the literature, agrees that the aforementioned factors are highly related to IHD. The multivariable analysis includes all relevant factors from the univariate analysis.

The AIC and BIC criteria were used to compare various parametric distributions with two shared frailty distributions of the models for various multivariable analyses. According to Munda et al. (2012), the model with the lowest AIC and BIC provides the greatest fit for the parametric shared frailty model. In this investigation, the multivariable log-logistic-inverse Gaussian shared frailty model showed a low AIC value of 908.6871, making it a suitable model for modeling IHD data. The BIC value also

supports this decision.

The results of this study showed that diabetes mellitus, hypertension, and other diseases (co-infections), which reduced time to death by a factor of $\phi = 0.471$, 0.521, and 0.382, respectively, compared to the reference groups, had a significant impact on the survival time of IHD patients at a 5 percent level of significance (Table 4.5). The results of a comparable study by Altaseb (2020), Špinar (2012), and Mahendra et al. (2015) which looked at the risk factors for ischemic heart disease, suggested that diabetes mellitus, hypertension, and other diseases were strongly connected with the survival of IHD patients. In addition, consistent studies conducted by Abdissa et al. (2021) and Moyehodie et al. (2022) as discussed in the literature also indicated that diabetes mellitus and hypertension (high blood pressure) were the known risk factors of IHD.

The outcome of this study also showed that ischemic heart disease patients' time to death is significantly influenced by blood cholesterol levels, obesity and smoking. This shows that the survival time of IHD patients is reduced by a factor of $\phi = 0.279$, 0.393, and 2.724 respectively, when compared to the reference categories as the impact of these diseases increases. According to studies done by Barbiero et al. (2009), and Gebremedhin & Gebrekirstos (2021) obesity and serum cholesterol, particularly LDL cholesterol, are all thought to be risk factors for IHD. Furthermore, according to Mensah (2008), Rashid et al. (2019), and Gona et al. (2021) studies, blood cholesterol levels, obesity, and smoking were among the major determinant factors in time to death of patients with IHD.

Even though the weibull model is the most commonly used parametric model since it supports proportional hazards and an accelerated life-time model (Hougaard, 2000), the log-logistic baseline was the best fit for the IHD data set when compared to the exponential and weibull hazard functions. Diagnostics graphs were created to evaluate the model's suitability. As a result, figure 4.2 log-logistic plot of the log failure odds with log time was more linear than the exponential (cumulative hazard versus time) and weibull (log cumulative hazard versus log time) plots, demonstrating that the log logistic baseline accurately reflected the IHD dataset.

The cumulative hazard plot for the Cox Snell residuals of the exponential, Weibull, and log-logistic models also supported this conclusion (Figure 4.3). The log-logistic model performed the best, as evidenced by the plots being closer to the line in this instance. A q-q plot was created in order to determine whether the accelerated failure time offered a good fit to the data utilized by two separate populations defined by diabetes mellitus, hypertension, obesity, smoking status, cholesterol, and other disease (Figure 4.4). For each covariate, plot seemed to follow a linear pattern. As a result, the accelerated failure time model was utilized using the log-logistic as a baseline. So, selecting a survival model to fit event periods is not strictly necessary. Comparing the baseline hazard function and the frailty distribution is necessary before choosing the best model for accurate inference.

5 Conclusions and Recommendations

5.1 Conclusions

This study's objective was to model the determinants of time to death of IHD by using different parametric baselines with different shared frailty models using a dataset of IHD patients obtained from JUMC. Out of the total 311, about 35.37% experienced an event (death) and 64.63% did not experience an event (censored).

Almost everywhere, extending survival times is the main goal. To do that, it is preferable to pinpoint the determining variables that are connected to ischemic heart disease's time to death. From the literature, a number of covariates were chosen as the predictor variables. The log-logistic-inverse Gaussian frailty model is the most suitable statistical model among many parametric frailty models that accurately predicted the time to death of ischemic heart disease patients who were identified at Jimma University Medical Center. The heterogeneity in the woredas of the patients causes a clustering effect on the time to death of IHD.

The result of the log-logistic-inverse Gaussian shared frailty model showed that the factors that determine the time to death of ischemic heart disease were diabetes mellitus, hypertension, obesity, smoking status, cholesterol, and other diseases. The findings of the study are generally consistent with data from earlier studies about relationships between these variables and the time to death of ischemic heart disease. In general, patients with the aforementioned predictors had shorter survival times.

The goodness of fit of the baseline distribution was determined using graphical methods and Cox-Snell residuals plots in figures 4.2 and 4.3, which revealed that the log-logistic distribution is superior to the exponential and Weibull distributions in explaining the time to death of the IHD dataset.

5.2 Recommendations

The following suggestions are provided for decision-makers and the general public. Based on the study's findings and keeping the limitations in mind, the study makes

the following recommendations.

- Ischemic heart disease patients with diabetes mellitus, hypertension, obesity, smoking status, cholesterol, and other diseases should receive extra care from medical personnel and medication.
- Additional socio-demographic, clinical, and other characteristics that were excluded from the study due to the source's data limitations will be incorporated into future research to provide more details about the determining factors that affect the survival times of IHD patients (like marital status, heredity of patients, physical activity, stress, alcoholic consumption, unhealthy nutrition, and soon).

References

Abdissa, S. G. (2020). Predictors of incident heart failure in a cohort of patients with ischemic heart disease. *The Pan African Medical Journal*, *35*.

Abdissa, S. G., et al. (2021). Risk factors for new-onset heart failure with reduced or preserved ejection fraction in patients with ischemic heart disease: A cohort study. *Journal of Clinical Sciences*, *18*(3), 133.

Abrahantes, J. C., Legrand, C., Burzykowski, T., Janssen, P., Ducrocq, V., & Duchateau, L. (2007). Comparison of different estimation procedures for proportional hazards model with random effects. *Computational statistics & data analysis*, *51*(8), 3913–3930.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Ali, S., Misganaw, A., Worku, A., Destaw, Z., Negash, L., Bekele, A., et al. (2021). The burden of cardiovascular diseases in ethiopia from 1990 to 2017: evidence from the global burden of disease study. *International Health*, *13*(4), 318–326.

Almdal, T., Scharling, H., Jensen, J. S., & Vestergaard, H. (2004). The independent effect of type 2 diabetes mellitus on ischemic heart disease, stroke, and death: a population-based study of 13 000 men and women with 20 years of follow-up. *Archives of internal medicine*, *164*(13), 1422–1426.

Altaseb, G. (2020). *Survival modeling of coronary heart disease patients: A case study of university of gondar referral hospital, gondar, ethiopia.* Unpublished doctoral dissertation, Getachew Altaseb.

Badimon, L., Padró, T., & Vilahur, G. (2012). Atherosclerosis, platelets and thrombosis in acute ischaemic heart disease. *European Heart Journal: Acute Cardiovascular Care*, *1*(1), 60–74.

Barbiero, S. M., Pellanda, L. C., Cesa, C. C., Campagnolo, P., Beltrami, F., &

Abrantes, C. C. (2009). Overweight, obesity and other risk factors for ihd in brazilian schoolchildren. *Public health nutrition*, *12*(5), 710–715.

Berkson, J., & Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, *47*(259), 501–515.

Bradbury, A. W., Adam, D. J., Bell, J., Forbes, J. F., Fowkes, F. G. R., Gillespie, I., et al. (2010). Bypass versus angioplasty in severe ischaemia of the leg (basil) trial: an intention-to-treat analysis of amputation-free and overall survival in patients randomized to a bypass surgery-first or a balloon angioplasty-first revascularization strategy. *Journal of vascular surgery*, *51*(5), 5S–17S.

Breslow, N. E. (1975). Analysis of survival data under the proportional hazards model. *International Statistical Review/Revue Internationale de Statistique*, 45–57.

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.

Costella, J. P. (2010). A simple alternative to kaplan–meier for survival curves. *Peter MacCallum Cancer Centre Working Paper (No*.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, *30*(2), 248–265.

Crump, C., Sundquist, J., Winkleby, M. A., & Sundquist, K. (2017). Interactive effects of obesity and physical fitness on risk of ischemic heart disease. *International journal of obesity*, *41*(2), 255–261.

Dätwyler, C., & Stucki, T. (2011). *Parametric survival models.*

Gatabazi, P. (2016). *Survival analysis and its stochastic process approach with application to diabetes data*. University of Johannesburg (South Africa).

Gebremedhin, M. H., & Gebrekirstos, L. G. (2021). Dietary and behavioral risk factors of ischemic heart disease among patients of medical outpatient departments in southern ethiopia: Unmatched case-control study. *Integrated Blood Pressure Control*, *14*, 123.

Gerr, F., Marcus, M., Ensor, C., Kleinbaum, D., Cohen, S., Edwards, A., et al. (2002). A prospective study of computer users: I. study design and incidence of musculoskeletal symptoms and disorders. *American journal of industrial medicine*, *41*(4), 221–235.

Glidden, D. V., & Vittinghoff, E. (2004). Modelling clustered survival data from multicentre clinical trials. *Statistics in medicine*, *23*(3), 369–388.

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, *1*(4), 274.

Gona, P. N., Kimokoti, R. W., Gona, C. M., Ballout, S., Rao, S. R., Mapoma, C. C., et al. (2021). Changes in body mass index, obesity, and overweight in southern africa development countries, 1990 to 2019: Findings from the global burden of disease, injuries, and risk factors study. *Obesity Science & Practice*, *7*(5), 509–524.

Grauss, R. W., Winter, E. M., Tuyn, J. van, Pijnappels, D. A., Steijn, R. V., Hogers, B., et al. (2007). Mesenchymal stem cells from ischemic heart disease patients improve left ventricular function after acute myocardial infarction. *American Journal of Physiology-Heart and Circulatory Physiology*, *293*(4), H2438–H2447.

Grover, G., & Seth, D. (2014). Application of frailty models on advance liver disease using gamma as frailty distribution. *SRL*, *3*, 42–50.

Guilbert, J. (2000). The world health report 1999making a difference. report of the director general, world health organization, geneva. *Education for Health*, *13*(1), 126–128.

Gupta, R., & Yusuf, S. (2019). Challenges in management and prevention of ischemic heart disease in low socioeconomic status people in llmics. *BMC medicine*, *17*(1), 1–11.

Hanagal, D. D. (2019). Various frailty models. In *Modeling survival data using frailty models* (pp. 85–121). Springer.

Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, *71*(1), 75–83.

Hougaard, P. (1986). A class of multivanate failure time distributions. *Biometrika*, *73*(3), 671–678.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime data analysis*, *1*(3), 255–273.

Hougaard, P. (2000). Shared frailty models. In *Analysis of multivariate survival data* (pp. 215–262). Springer.

Huang, X., & Wolfe, R. A. (2002). A frailty model for informative censoring. *Biometrics*, *58*(3), 510–520.

Janssen, P., & Duchateau, L. (2011). Frailty model. In *International encyclopedia of statistical science* (pp. 544–546). Springer.

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, *53*(282), 457–481.

Khan, M. A., Hashim, M. J., Mustafa, H., Baniyas, M. Y., Al Suwaidi, S. K. B. M., AlKatheeri, R., et al. (2020). Global epidemiology of ischemic heart disease: results from the global burden of disease study. *Cureus*, *12*(7).

Kheiri, S., Kimber, A., & Meshkani, M. R. (2007). Bayesian analysis of an inverse gaussian correlated frailty model. *Computational statistics & data analysis*, *51*(11), 5317–5326.

Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, 795–806.

Klein, J. P., & Moeschberger, M. L. (2003). Semiparametric proportional hazards regression with fixed covariates. In *Survival analysis* (pp. 243–293). Springer.

Kleinbaum, D. G., Klein, M., et al. (2012). *Survival analysis: a self-learning text* (Vol. 3). Springer.

Kong, L., & Kaddoum, G. (2019). Secrecy characteristics with assistance of mixture gamma distribution. *IEEE Wireless Communications Letters*, *8*(4), 1086–1089.

Mahendra, J., Kumar, S. D., Anuradha, T., Talikoti, P., Nagaraj, R., & Vishali, V. (2015). Plasma fibrinogen in type 2 diabetic patients with metabolic syndrome and its relation with ischemic heart disease (ihd) and retinopathy. *Journal of clinical and diagnostic research: JCDR*, *9*(1), BC18.

Manda, S. (2020). Flexible modeling of frailty effects in clustered survival data. In *Computational and methodological statistics and biostatistics* (pp. 489–509). Springer.

Manton, K. G., Stallard, E., & Vaupel, J. W. (1986). Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association*, *81*(395), 635–644.

Mensah, G. (2008). Ischaemic heart disease in africa. *Heart*, *94*(7), 836–843.

Moran, A., Forouzanfar, M., Roth, G., Mensah, G., & Ezzati, M. (1990). others. 2014a.. *The Global Burden of Ischemic Heart Disease in*, 1493–501.

Moyehodie, Y. A., Muluneh, M. W., Belay, A. T., & Fenta, S. M. (2022). Time to death and its determinant factors among patients with chronic heart failure in northwest ethiopia: A retrospective study at selected referral hospitals. *Frontiers in Cardiovascular Medicine*, *9*.

Mozaffarian, D., Benjamin, E., Go, A., Arnett, D., Blaha, M., Cushman, M., et al. (2015). Mohler 3rd. *ER, Moy, CS, Muntner, P., Mussolino, ME, Nasir, K., Neumar, RW, Nichol, G., Palaniappan, L., Pandey, DK, Reeves, MJ, Rodriguez, CJ, Sorlie, PD, Stein, J., Towfighi, A., Turan, TN, Virani, SS, Willey, JZ, Woo, D., Yeh, RW, Turner, MB*.

Munda, M., Rotolo, F., & Legrand, C. (2012). parfm: Parametric frailty models in r. *Journal of Statistical Software*, *51*, 1–20.

Nilsson, S. (2008). *Chest pain and ischemic heart disease: Diagnosis and management in primary health care*. Unpublished doctoral dissertation, Institutionen för medicin och hälsa.

Nowbar, A. N., Gitto, M., Howard, J. P., Francis, D. P., & Al-Lamee, R. (2019). Mortality from ischemic heart disease: Analysis of data from the world health organization and coronary artery disease risk factors from ncd risk factor collaboration. *Circulation: cardiovascular quality and outcomes*, *12*(6), e005375.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(3), 414–422.

Ostadal, B., & Ostadal, P. (2014). Sex-based differences in cardiac ischaemic injury and protection: therapeutic implications. *British journal of pharmacology*, *171*(3), 541–554.

Paul, S. A., Simons, J. W., & Mabjeesh, N. J. (2004). Hif at the crossroads between ischemia and carcinogenesis. *Journal of cellular physiology*, *200*(1), 20–30.

Pipper, C. B., & Martinussen, T. (2004). An estimating equation for parametric shared frailty models with marginal additive hazards. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(1), 207–220.

Prabhakaran, D., Jeemon, P., Sharma, M., Roth, G. A., Johnson, C., Harikrishnan, S., et al. (2018). The changing patterns of cardiovascular diseases and their risk factors in the states of india: the global burden of disease study 1990–2016. *The Lancet Global Health*, *6*(12), e1339–e1351.

Price, D. L., & Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in medicine*, *20*(9-10), 1515–1527.

Rashid, N. A., Nawi, A. M., & Khadijah, S. (2019). Exploratory analysis of traditional risk factors of ischemic heart disease (ihd) among predominantly malay malaysian women. *BMC public health*, *19*(4), 1–8.

Roth, G. A., Johnson, C., Abajobir, A., Abd-Allah, F., Abera, S. F., Abyu, G., et al. (2017). Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *Journal of the American college of cardiology*, *70*(1), 1–25.

Sakboonyarat, B., & Rangsin, R. (2018). Prevalence and associated factors of ischemic heart disease (ihd) among patients with diabetes mellitus: a nation-wide, cross-sectional survey. *BMC cardiovascular disorders*, *18*(1), 1–7.

Sargent, D. J. (1998). A general framework for random effects survival analysis in the cox proportional hazards setting. *Biometrics*, 1486–1497.

Shashu, B. (2021). Profile of patients presenting to addis cardiac hospital, addis ababa, ethiopia, from may 2007 to december 2011. *East African Medical Journal*, *98*(7), 3966–3974.

Špinar, J. (2012). Hypertension and ischemic heart disease. *Cor et Vasa*, *54*(6), e433–e438.

Steenbergen, C., & Frangogiannis, N. G. (2012). Ischemic heart disease. In *Muscle* (pp. 495–521). Elsevier Inc.

Tate, R. B., Manfreda, J., & Cuddy, T. E. (1998). The effect of age on risk factors for ischemic heart disease: the manitoba follow-up study, 1948–1993. *Annals of Epidemiology*, *8*(7), 415–421.

Tromp, J., Ouwerkerk, W., Cleland, J. G., Angermann, C. E., Dahlstrom, U., Tiew-Hwa Teng, K., et al. (2021). Global differences in burden and treatment of ischemic heart disease in acute heart failure: Report-hf. *Heart Failure*, *9*(5), 349–359.

Trussell, J., & Richards, T. (1985). Correcting for unmeasured heterogeneity in hazard models using the heckman-singer procedure. *Sociological methodology*, *15*, 242–276.

Van Dijk, P. C., Jager, K. J., Zwinderman, A. H., Zoccali, C., & Dekker, F. W. (2008). The analysis of survival data in nephrology: basic concepts and methods of cox regression. *Kidney international*, *74*(6), 705–709.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*(3), 439–454.

Verdouw, P. D., Doel, M. A. van den, Zeeuw, S. de, & Duncker, D. J. (1998). Animal models in the study of myocardial ischaemia and ischaemic syndromes. *Cardiovascular research*, *39*(1), 121–135.

Vogel, B., Acevedo, M., Appelman, Y., Merz, C. N. B., Chieffo, A., Figtree, G. A., et al. (2021). The lancet women and cardiovascular disease commission: reducing the global burden by 2030. *The Lancet*, *397*(10292), 2385–2438.

Wang, F., Yu, Y., Mubarik, S., Zhang, Y., Liu, X., Cheng, Y., et al. (2021). Global burden of ischemic heart disease and attributable risk factors, 1990–2017: A secondary analysis based on the global burden of disease study 2017. *Clinical Epidemiology*, *13*, 859.

WHO. (2019). *Global action plan on physical activity 2018-2030: more active people for a healthier world*. World Health Organization.

WHO. (2020). *Who guideline on use of ferritin concentrations to assess iron status in populations*. World Health Organization.

Wienke, A. (2010). *Frailty models in survival analysis*. Chapman and Hall/CRC.

Wienke, A., Arbeev, K., Locatelli, I., Yashin, A. I., et al. (2003). A simulation study of different correlated frailty models and estimation strategies. *URL http://www. demogr. mpg. de/en/projects_publications/publications_1904/mpidr_working_papers/all. htm*.

Wienke, A., et al. (2003). Frailty models. *Demogr Res*, *49*, 0–13.

Wilson, P. W. (1994). Established risk factors and coronary artery disease: the framingham study. *American journal of hypertension*, *7*(7_Pt_2), 7S–12S.

Wogi, A. A., Wakweya, S. T., & Tesfay, Y. Y. (2018). Determinants of time-to-under-five mortality in ethiopia: Comparison of parametric shared frailty models. *International Journal of Biomedical and Clinical Engineering (IJBCE)*, *7*(1), 1–24.

6 Annex

6.1 Annex I: Univariable Analysis for Frailty Models

Table 6.1: Exponential versus Gamma Univariable frailty model

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Sex	Coeff	3.8069	45.008	0.219	(29.303, 69.130)	<0.001
	Female(rf)		1			
	Male	0.0735	1.076	0.217	(0.703, 1.647)	0.74
Age	Coeff	4.497	89.739	1.01	(12.308, 654.301)	<0.001
	0-14(rf)		1			
	15-47	-0.737	0.478	1.02	(0.064, 3.533)	0.47
	48-63	-0.504	0.604	1.02	(0.081, 4.463)	0.62
	≥ 64	-0.729	0.482	1.02	(0.064, 3.593)	0.48
Diabetes mellitus	Coeff	4.21	67.552	0.160	(49.337, 92.491)	<0.001
	No(rf)		1			
	Yes	-1.02	0.362	0.194	(0.247, 0.529)	<0.001
Hypertension	Coeff	4.16	64.333	0.136	(49.272, 83.998)	<0.001
	No(rf)		1			
	Yes	-1.00	0.367	0.191	(0.252, 0.533)	<0.001
Obesity	Coeff	4.29	72.847	0.174	(51.780, 102.486)	<0.001
	No(rf)		1			
	Yes	-1.36	0.257	0.195	(0.175, 0.376)	<0.001
Smoking status	Coeff	3.657	38.736	0.100	(31.830, 47.141)	<0.001
	No()		1			
	Yes	0.933	2.541	0.348	(1.285, 5.026)	0.007
Cholesterol	Coeff	4.47	87.688	0.195	(59.835, 128.504)	<0.001
	No(rf)		1			
	Yes	-1.23	0.292	0.209	(0.193, 0.439)	<0.001
Other disease	Coeff	4.57	97.016	0.209	(64.392, 146.166)	<0.001
	No(rf)		1			
	Yes	-1.23	0.290	0.221	(0.188, 0.448)	<0.001

Table 6.2: Exponential versus Inverse Gaussian Univariable frailty model

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Sex	Coeff	3.7450	42.306	0.201	(28.508, 62.784)	<0.001
	Female(rf)		1			
	Male	0.0738	1.076	0.215	(0.705, 1.642)	0.73
Age	Coeff	4.474	87.703	1.01	(12.131, 634.025)	<0.001
	0-14(rf)		1			
	15-47	-0.777	0.459	1.02	(0.062, 3.380)	0.45
	48-63	-0.544	0.580	1.02	(0.078, 4.276)	0.59
	≥ 64	-0.758	0.468	1.02	(0.063, 3.478)	0.46
Diabetes mellitus	Coeff	4.20	66.845	0.161	(48.786, 91.588)	<0.001
	No(rf)		1			
	Yes	-1.02	0.361	0.195	(0.247, 0.529)	<0.001
Hypertension	Coeff	4.165	64.380	0.143	(48.605, 85.274)	<0.001
	No(rf)		1			
	Yes	-0.999	0.368	0.192	(0.252, 0.536)	<0.001
Obesity	Coeff	4.25	70.173	0.170	(50.277, 97.943)	<0.001
	No(rf)		1			
	Yes	-1.36	0.256	0.195	(0.175, 0.376)	<0.001
Smoking status	Coeff	3.672	39.348	0.127	(30.698, 50.435)	<0.001
	No()		1			
	Yes	0.919	2.505	0.351	(1.260, 4.983)	0.009
Cholesterol	Coeff	4.47	86.964	0.203	(58.368, 129.571)	<0.001
	No(rf)		1			
	Yes	-1.24	0.288	0.210	(0.191, 0.435)	<0.001
Other disease	Coeff	4.56	95.668	0.208	(63.623, 143.853)	<0.001
	No(rf)		1			
	Yes	-1.24	0.290	0.221	(0.188, 0.447)	<0.001

Table 6.3: Weibull versus Gamma Univariable frailty model

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Sex	Coeff	3.9268	50.742	0.243	(31.516, 81.696)	<0.001
	Female(rf)		1			
	Male	0.0921	1.096	0.264	(0.653, 1.839)	0.73
Age	Coeff	4.897	133.887	1.25	(11.494, 1559.569)	<0.001
	0-14(rf)		1			
	15-47	-1.023	0.359	1.26	(0.030, 4.229)	0.42
	48-63	-0.761	0.467	1.26	(0.039, 5.506)	0.55
	≥ 64	-1.000	0.367	1.27	(0.030, 4.393)	0.43
Diabetes mellitus	Coeff	4.42	83.351	0.229	(53.170, 130.663)	<0.001
	No(rf)		1			
	Yes	-1.15	0.317	0.241	(0.197, 0.508)	<0.001
Hypertension	Coeff	4.39	80.866	0.222	(52.357, 124.899)	<0.001
	No(rf)		1			
	Yes	-1.13	0.321	0.243	(0.199, 0.517)	<0.001
Obesity	Coeff	4.46	86.076	0.234	(54.398, 136.200)	<0.001
	No(rf)		1			
	Yes	-1.48	0.228	0.237	(0.143, 0.363)	<0.001
Smoking status	Coeff	3.83	45.947	0.161	(33.499, 63.019)	<0.001
	No()		1			
	Yes	1.07	2.926	0.426	(1.270, 6.740)	0.12
Cholesterol	Coeff	4.69	108.586	0.266	(64.506, 182.789)	<0.001
	No(rf)		1			
	Yes	-1.35	0.259	0.246	(0.159, 0.420)	<0.001
Other disease	Coeff	4.81	123.023	0.287	(70.143, 215.770)	<0.001
	No(rf)		1			
	Yes	-1.37	0.253	0.269	(0.149, 0.429)	<0.001

Table 6.4: Weibull versus Inverse Gaussian Univariable frailty model

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Sex	Coeff	3.8984	49.323	0.250	(30.246, 80.435)	<0.001
	Female(rf)		1			
	Male	0.0845	1.088	0.254	(0.661, 1.790)	0.74
Age	Coeff	4.760	116.736	1.20	(11.108, 1226.800)	<0.001
	0-14(rf)		1			
	15-47	-0.911	0.402	1.20	(0.038, 4.238)	0.45
	48-63	-0.652	0.521	1.20	(0.049, 5.498)	0.59
	≥ 64	-0.908	0.403	1.21	(0.037, 4.309)	0.45
Diabetes mellitus	Coeff	4.43	83.778	0.225	(53.915, 130.183)	<0.001
	No(rf)		1			
	Yes	-1.17	0.309	0.245	(0.191, 0.499)	<0.001
Hypertension	Coeff	4.43	84.135	0.212	(55.572, 127.377)	<0.001
	No(rf)		1			
	Yes	-1.19	0.304	0.249	(0.186, 0.495)	<0.001
Obesity	Coeff	4.40	81.515	0.222	(52.787, 125.879)	<0.001
	No(rf)		1			
	Yes	-1.49	0.225	0.241	(0.140, 0.361)	<0.001
Smoking status	Coeff	3.82	45.611	0.165	(33.017, 63.009)	<0.001
	No()		1			
	Yes	1.06	2.897	0.423	(1.264, 6.637)	0.12
Cholesterol	Coeff	4.61	100.843	0.257	(60.951, 166.844)	<0.001
	No(rf)		1			
	Yes	-1.34	0.261	0.247	(0.160, 0.423)	<0.001
Other disease	Coeff	4.78	119.291	0.279	(69.044, 206.105)	<0.001
	No(rf)		1			
	Yes	-1.38	0.250	0.271	(0.147, 0.425)	<0.001

Table 6.5: Loglogistic versus Gamma Univariable frailty model

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Sex	Coeff	3.5226	33.871	0.251	(20.728, 55.348)	<0.001
	Female(rf)		1			
	Male	0.0545	1.055	0.280	(0.609, 1.828)	0.85
Age	Coeff	4.551	94.720	1.20	(9.078, 988.273)	<0.001
	0-14(rf)		1			
	15-47	-1.129	0.323	1.21	(0.030, 3.440)	0.35
	48-63	-0.841	0.431	1.21	(0.040, 4.598)	0.49
	≥ 64	-1.079	0.339	1.22	(0.031, 3.689)	0.38
Diabetes mellitus	Coeff	4.00	54.731	0.212	(36.094, 82.991)	<0.001
	No(rf)		1			
	Yes	-1.19	0.304	0.246	(0.187, 0.492)	<0.001
Hypertension	Coeff	3.98	53.509	0.207	(35.646, 80.322)	<0.001
	No(rf)		1			
	Yes	-1.23	0.291	0.249	(0.179, 0.475)	<0.001
Obesity	Coeff	4.03	56.123	0.218	(36.577, 86.114)	<0.001
	No(rf)		1			
	Yes	-1.53	0.217	0.236	(0.137, 0.345)	<0.001
Smoking status	Coeff	3.38	29.259	0.160	(21.390, 40.022)	<0.001
	No()		1			
	Yes	1.19	3.281	0.406	(1.481, 7.267)	0.003
Cholesterol	Coeff	4.3	73.795	0.251	(45.109, 120.723)	<0.001
	No(rf)		1			
	Yes	-1.4	0.245	0.235	(0.154, 0.389)	<0.001
Other disease	Coeff	4.39	80.822	0.265	(48.071, 135.887)	<0.001
	No(rf)		1			
	Yes	-1.40	0.247	0.259	(0.149, 0.411)	<0.001

Table 6.6: Loglogistic versus Inverse Gaussian Univariable frailty model

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Sex	Coeff	3.5220	33.852	0.263	(20.224, 56.662)	<0.001
	Female(rf)		1			
	Male	0.0346	1.035	0.267	(0.613, 1.746)	0.9
Age	Coeff	4.427	83.662	1.12	(9.267, 755.283)	<0.001
	0-14(rf)		1			
	15-47	-1.014	0.362	1.13	(0.039, 3.306)	0.37
	48-63	-0.716	0.488	1.13	(0.053, 4.442)	0.52
	≥ 64	-1.007	0.365	1.14	(0.039, 3.384)	0.38
Diabetes mellitus	Coeff	4.01	54.998	0.205	(36.786, 82.227)	<0.001
	No(rf)		1			
	Yes	-1.23	0.293	0.250	(0.179, 0.478)	<0.001
Hypertension	Coeff	4.0	54.486	0.192	(37.413, 79.348)	<0.001
	No(rf)		1			
	Yes	-1.3	0.272	0.254	(0.165, 0.449)	<0.001
Obesity	Coeff	3.97	52.871	0.204	(35.468, 78.815)	<0.001
	No(rf)		1			
	Yes	-1.54	0.213	0.239	(0.133, 0.341)	<0.001
Smoking status	Coeff	3.37	29.211	0.174	(20.763, 41.098)	<0.001
	No()		1			
	Yes	1.15	3.172	0.396	(1.458, 6.899)	0.004
Cholesterol	Coeff	4.21	67.689	0.242	(42.090, 108.858)	<0.001
	No(rf)		1			
	Yes	-1.40	0.246	0.235	(0.155, 0.391)	<0.001
Other disease	Coeff	4.36	78.099	0.257	(47.220, 129.170)	<0.001
	No(rf)		1			
	Yes	-1.41	0.243	0.260	(0.146, 0.405)	<0.001

6.2 Annex II: Multivariable Analysis for Frailty Models

Table 6.7: Results of Exponential Gamma multivariable shared frailty model.

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Intercept		5.654	285.308	0.246	(176.243, 461.866)	<0.001
Diabetes mellitus	No(rf)		1			
	Yes	-0.677	0.508	0.206	(0.339, 0.760)	0.001
Hypertension	No(rf)		1			
	Yes	-0.568	0.566	0.207	(0.377, 0.849)	0.006
Obesity	No(rf)		1			
	Yes	-1.207	0.298	0.199	(0.202, 0.441)	<0.001
Smoking status	No(rf)		1			
	Yes	0.907	2.476	0.354	(1.236, 4.957)	0.01
Cholesterol	No(rf)		1			
	Yes	-0.887	0.411	0.213	(0.271, 0.625)	<0.001
Other disease	No(rf)		1			
	Yes	-1.012	0.363	0.233	(0.230, 0.574)	<0.001
$\theta = 0.005$	$\lambda = 0.003$		$\tau = 0.022$		AIC = 916.4669	

Source = JUMC, $\hat{\beta}$ = Coefficients, St.err = Standard error, 95% CI = confidence interval for acceleration factor, rf = reference, θ = Variance of random term, λ = scale, τ = kedall's tau, ϕ = Acceleration factor, AIC = Akaike's information criteria

Table 6.8: Results of Exponential Inverse Gaussian multivariable shared frailty model.

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Intercept		5.702	299.524	0.260	(179.968, 498.503)	<0.001
Diabetes mellitus	No(rf)		1			
	Yes	-0.665	0.514	0.209	(0.341, 0.774)	0.002
Hypertension	No(rf)		1			
	Yes	-0.580	0.560	0.210	(0.371, 0.844)	0.006
Obesity	No(rf)		1			
	Yes	-1.234	0.291	0.201	(0.196, 0.432)	<0.001
Smoking status	No(rf)		1			
	Yes	0.943	2.568	0.359	(1.269, 5.196)	0.009
Cholesterol	No(rf)		1			
	Yes	-0.903	0.405	0.214	(0.266, 0.617)	<0.001
Other disease	No(rf)		1			
	Yes	-1.040	0.353	0.238	(0.221, 0.563)	<0.001
$\theta = 0.075$	$\lambda = 0.003$		$\tau = 0.019$		AIC = 914.5860	

Source = JUMC, $\hat{\beta}$ = Coefficients, St.err = Standard error, 95% CI = confidence interval for acceleration factor, rf = reference, θ = Variance of random term, λ = scale, τ = kedall's tau, ϕ = Acceleration factor, AIC = Akaike's information criteria

Table 6.9: Results of weibull Gamma multivariable shared frailty model.

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Intercept		5.757	316.417	0.353	(158.539, 631.515)	<0.001
Diabetes mellitus	No(rf)		1			
	Yes	-0.569	0.565	0.208	(0.376, 0.850)	0.006
Hypertension	No(rf)		1			
	Yes	-0.559	0.571	0.209	(0.379, 0.860)	0.007
Obesity	No(rf)		1			
	Yes	-1.221	0.294	0.204	(0.197, 0.439)	<0.001
Smoking status	No(rf)		1			
	Yes	0.957	2.604	0.347	(1.318, 5.145)	0.006
Cholesterol	No(rf)		1			
	Yes	-0.917	0.399	0.210	(0.264, 0.603)	<0.001
Other disease	No(rf)		1			
	Yes	-1.060	0.346	0.240	(0.216, 0.554)	<0.001
$\theta = 0.561$	$\lambda = 0.004$					
$\rho = 0.982$	$\tau = 0.02$				AIC = 919.3614	

Source = JUMC, $\hat{\beta}$ = Coefficients, St.err = Standard error, 95% CI = confidence interval for acceleration factor, rf = reference, θ = Variance of random term, τ = kedall's tau, ϕ = Acceleration factor, AIC = Akaike's information criteria

Table 6.10: Results of weibull Inverse Gaussian multivariable shared frailty model.

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Intercept		5.654	285.293	0.317	(153.223, 531.199)	<0.001
Diabetes mellitus	No(rf)		1			
	Yes	-0.650	0.521	0.211	(0.345, 0.788)	0.002
Hypertension	No(rf)		1			
	Yes	-0.571	0.565	0.208	(0.375, 0.849)	0.006
Obesity	No(rf)		1			
	Yes	-1.216	0.296	0.210	(0.196, 0.447)	<0.001
Smoking status	No(rf)		1			
	Yes	0.933	2.541	0.355	(1.267, 5.095)	0.009
Cholesterol	No(rf)		1			
	Yes	-0.891	0.410	0.215	(0.269, 0.625)	<0.001
Other disease	No(rf)		1			
	Yes	-1.027	0.358	0.240	(0.223, 0.573)	<0.001
$\theta = 0.081$	$\lambda = 0.004$					
$\rho = 0.981$	$\tau = 0.017$				AIC = 916.5015	

Source = JUMC, $\hat{\beta}$ = Coefficients, St.err = Standard error, 95% CI = confidence interval for acceleration factor, rf = reference, θ = Variance of random term, τ = kedall's tau, ϕ = Acceleration factor, AIC = Akaike's information criteria

Table 6.11: Results of Loglogistic Gamma multivariable shared frailty model.

Covariate Variable	Category	$\hat{\beta}$	ϕ	St. err	95% CI	P value
Intercept		5.355	211.592	0.332	(110.338, 405.766)	<0.001
Diabetes mellitus	No(rf)		1			
	Yes	-0.722	0.485	0.205	(0.324, 0.726)	<0.001
Hypertension	No(rf)		1			
	Yes	-0.609	0.543	0.208	(0.362, 0.817)	0.003
Obesity	No(rf)		1			
	Yes	-1.275	0.279	0.206	(0.186, 0.418)	<0.001
Smoking status	No(rf)		1			
	Yes	1.005	2.732	0.343	(1.394, 5.354)	0.003
Cholesterol	No(rf)		1			
	Yes	-0.969	0.379	0.201	(0.256, 0.562)	<0.001
Other disease	No(rf)		1			
	Yes	-0.978	0.376	0.224	(0.242, 0.583)	<0.001
$\theta = 0.569$	$\lambda =$ -5.597					
$\rho = 1.000$	$\tau = 0.018$				AIC = 910.8247	

Source = JUMC, $\hat{\beta}$ = Coefficients, St.err = Standard error, 95% CI = confidence interval for acceleration factor, rf = reference, θ = Variance of random term, τ = kedall's tau, ϕ = Acceleration factor, AIC = Akaike's information criteria