MEKELLE UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCINECE

DEPARTMENT OF STATISTICS

A

Thesis

On

Shared Frailty Model in Survival Analysis of Time to Discharge Dynamics for Myocardial Infarction Adult Patients at Ayder Comprehensive Specialized Referral Hospital (Jan 1, 2018 – Dec 31, 2020)

Submitted to the Department of Statistics,

College of Natural and Computational Science, Mekelle University,

In Partial Fulfillment of the Requirements for the Master of Science Degree (MSc.) in Biostatistics.

By:

Gebrewahd Tewelemedhin (CNCS/PR002/13)

Under the supervision of

Advisor: Said Mussa (Ass. Professor)

Co-Advisor: Lewtnesh Berihun (MSc.)

Nov, 2024 G.C

Mekelle, Tigray, Ethiopia

APPROVAL SHEET

MEKELLE UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF STATISTICS

Shared Frailty Model in Survival Analysis of Time to Discharge Dynamics for Myocardial Infarction Adult Patients at Ayder Comprehensive Specialized Referral Hospital (Jan 1, 2018 - Dec 31, 2020)

Submitted By:

Gebrewahd Tewelemedhin	*CNCS/PR002/13*		
Student's Name	ID.No.	Signature	Date

Approved By:

1.	*Said Mussa (Ass. professor)*		
	Advisor's Name	Signature	Date
2.	*Lewtnesh Berihun (MSc.)*		
	Department Head	Signature	Date
3.			
	Chairperson	Signature	Date
4.			
	External Examiner	Signature	Date
5.			
	Internal Examiner	Signature	Date

Declaration

I, the undersigned, declare that this research entitled, “Shared Frailty Model in Survival Analysis of Time to Discharge Dynamics for Myocardial Infarction Adult Patients at Ayder Comprehensive Specialized Referral Hospital (Jan 1, 2018 - Dec 31, 2020)” is my original work and has not been presented for any other award, and that all sources of materials used in this proposal is duly acknowledged. This research was carried out under the supervision of my principal advisor ***Said Mussa (Ass.Professor)*** Department of Statistics, College of Natural and Computational Sciences, Mekelle University in the academic year of 2024 G.C.

Name of student candidate: *Gebrewahd Tewelemedhin*

Signature with date: _____

This research has been submitted for examination with my approval as university advisor/co-advisor.

Name of the advisor: *Said Mussa (Ass.Professor)*

Signature of the Advisor with date: _____

Name of the co-advisor: *Lewtnesh Berihun(MSc.)*

Signature of the co-advisor with date: _____

Place: Mekelle University, Mekelle, Ethiopia

Date of submission: _____

Acknowledgement

First and foremost, I would like to thank the Almighty God for being the source of my capability and strength throughout my life and in accomplishing this thesis.

Secondly, I would like to express my heartfelt gratitude to my adviser, Assistant Professor *Said Mussa*, and my co-adviser, *Lewtnesh Berihun (MSc.),* for their encouragement, guidance, and constructive comments from the beginning to the end of this study. My special thanks go to all my friends and staff members. for their support throughout this journey.

Thirdly, I extend my thanks to all my friends for their support in various ways. I am also deeply grateful to Aksum University for sponsoring my master's program. Additionally, I would like to thank the Ayder Comprehensive Specialized Referral Hospital for providing all the relevant secondary data used in this study, and all individuals who contributed in one way or another to the realization of this thesis.

Finally, I convey my heartfelt appreciation to my family for their continuous encouragement, especially to my beloved father, *Tewelemedhin Tkue*, and my mother, *Haimanot Abay*.

List of Abbreviations

ACSRH	Ayder Comprehensive Specialized Referral Hospital
AFT	Accelerated Failure Time
AHA	American Heart Association
AIC	Akaike's Information Criterion
BIC	Bayesian Information Criterion
BMI	Body Mass Index
CAD	Coronary Artery Disease
CHD	Coronary Heart Disease
CVD	Cardiovascular disease
DALY	Disability Adjusted Life Years
HF	Hazard Function
ICU	Intensive Care Unit
IHD	Ischemic Heart Disease
KM	Kaplan-Meier
MI	Myocardial Infarction
MU-CHS	Mekelle University, College of Health Sciences
NSTEMI	Non-ST-elevation Myocardial Infarction
PH	Proportional Hazard
Q-Q plot	Quantile-quantile plot
STEMI	ST-elevation Myocardial Infarction
WHO	World Health Organization

Table of Contents

List of Tables

List of Figures

Abstract

Background: MI, commonly known as heart attack, happens when a blood clot obstructs the coronary arteries, resulting in decreased oxygen and nutrient supply to the heart muscle. MI continues to be a significant cause of morbidity and mortality globally, with variations in the time-to-discharge dynamics among patients. Understanding the time to discharge is crucial for optimizing patient care and resource allocation, particularly in settings with limited healthcare resources like Ayder Comprehensive Specialized Referral Hospital.

Objective: The overall objective of this study was to investigate and gain a comprehensive understanding of the time to discharge dynamics in MI patients at Ayder Comprehensive Specialized Referral Hospital, using a survival analysis with a shared frailty model.

Methods: To fulfill the study goal, secondary data from Ayder Comprehensive Specialized Referral Hospital was collected from 206 MI patients who initiated their follow-up between January 2018 and December 2020. K-M curves used to compare the survival curve for categorical variables and the univariable analysed used Cox regression model to select variable which were included in the multivariable analysis. The Cox PH model with a parametric shared frailty distribution was utilized, with the follow-up site where treatment was administered serving as a clustering effect in the models. The study employed gamma and inverse Gaussian shared frailty distributions alongside Exponential, Weibull, and log-logistic baseline models to analyze the risk factors associated with survival time until discharge, considering socio-economic and demographic factors. All fitted models were compared using the AIC and BIC values derived from the actual dataset.

Results: Of the 206 patients were seen at Ayder Comprehensive Specialized Referral Hospital during the study period. The results revealed that approximately 54% experienced the event, while 46% did not experience it by the end of the follow-up period. The AIC values for the three baseline distributions (Exponential, Weibull, and Gompertz) of the PH model were found to be 370.8967, 90.3539, and 96.2921, respectively. The corresponding BIC values for those baseline distributions were 424.1427, 146.9278, and 152.8661, respectively. The AIC values for the three baseline distributions for the Gamma shared frailty model were found to be 375.4022, 89.2839, and 98.5131, with the BIC values for the same model found to be 431.9761, 148.9278, and

155.4568, respectively. Based on the AIC and BIC values from the simulation experiment and graphical evidence, the Gamma shared frailty model with the Weibull baseline was preferred when compared to other models. The clustering effect (follow-up site) was found to be significant for modeling the risk factors of time-to-discharge patients with MI. The estimated value of theta (θ), which measures the contribution of a frailty component to the model, was 1.1056. A chi-square value of 0.00372 with one degree of freedom resulted in a p-value of 0.0031. Based on the results of the Gamma shared frailty model with the Weibull baseline, the follow-up site at the medical ward, obesity (BMI > 30), age (in years), weight, diabetes mellitus, family history of MI, uncontrolled blood pressure, high cholesterol levels, and male gender were identified as the most significant risk factors for the outcome variable, survival time to discharge. The hazard ratio and 95% confidence interval for patient age and weight were 0.9844 (CI [1.0105, 1.5116]) and 1.0101 (CI [0.0193, 0.9879]) with p-values of 0.002 and 0.019, respectively. The covariates, including follow-up site at the Medical ward, BMI with obesity (BMI>30), diabetes mellitus, family history of MI, uncontrolled blood pressure, high cholesterol levels, and male gender, exhibited hazard ratios of 2.4868 (CI [0.4281, 1.4369]), 2.3445 (CI [0.3901, 2.1253]), 2.7563 (CI [0.582, 1.5858]), 3.7139 (CI [0.0152, 1.3031]), 1.0726 (CI [0.3823, 1.8024]), 1.7318 (CI [0.3385, 1.3835]), and 4.1012 (CI [0.0110, 1.1967]), respectively, with associated p-values of 0.018, 0.001, 0.003, 0.021, 0.026, 0.013, and 0.001, indicating their respective impacts on the study outcomes.

Conclusions & Recommendation: The model suggested that there is a strong evidence of heterogeneity among follow up sites where the MI patients were treated. From the candidate models, Weibull-gamma shared frailty model was an appropriate model for the MI dataset. There was a frailty effect on the survival of the MI patients that arises due to differences in the distribution of follow up sites. The risk factors follow-up site at the medical ward, being obese, age(in years), weight, diabetes mellitus, family history of MI, uncontrolled blood pressure, high cholesterol levels, and male gender were statistically significant for the survival of MI patients whereas the other risk factors were not statistically significant. Health care providers must focus on high-risk MI patients, considering factors like age, obesity, uncontrolled blood pressure, diabetes, high cholesterol level, gender, and family history of MI.

Keywords: *Myocardial Infarction, Time to Discharge, Shared Frailty, Coronary Artery Disease, Heart Attack.*

1. Introduction

1.1 Background

MI, commonly known as a heart attack occurs when there is a blockage in the blood vessels that supply oxygen and nutrients to the heart muscle. This blockage is typically caused by the formation of a blood clot in a coronary artery, which can result in the restricted blood flow to the heart. According to the Global Burden of Disease Study 2019, MI was the leading cause of years of life lost due to premature mortality worldwide, accounting for over 16 million years lost in 2019. MI was also the second leading cause of disability-adjusted life years (DALYs) globally, accounting for over 25 million DALYs in 2019. In India MI, the leading cause of death, accounting for over a quarter of all deaths in the country. According to a study published in the Lancet Global Health, the age-standardized incidence rate of MI in India was estimated to be 104 per 100,000 person-years in 2016, with a mortality rate of 45 per 100,000 person-years. The study also found that the burden of MI in India is increasing, particularly in urban areas and among younger populations. The burden of MI in India is further compounded by factors such as low awareness and access to healthcare, inadequate healthcare infrastructure, and a high prevalence of MI risk factors such as tobacco use, unhealthy diet, and physical inactivity ("Global, Regional, and National Age-Sex Specific Mortality for 264 Causes of Death, 1980-2016: A Systematic Analysis for the Global Burden of Disease Study 2016.," 2017).

According to the WHO, MI is responsible for over 7 million deaths globally each year. Based on WHO prediction, MI will continue to be the leading cause of mortality globally up to 2030 (Mathers et al., 2009). Based on the report of Heart Disease and Stroke Statistics 2018; the global prevalence of IHD was estimated about 110.6 million, where males were more commonly affected than females and with case fatality rate of 8.9 million (Benjamin et al., 2018). MI is a leading cause of mortality and morbidity, accounts for 50% of all CVD deaths and more than 2.5 million hospitalizations worldwide each year (Guo et al., 2016).

A prospective study was conducted in Africa, Sub-Saharan countries with total of 111 (5.1%) MI patients comprising of 56% with STEMI and 44% of them NSTEMI. The study claimed that in-hospital mortality was about 6%–10% in the setting (Shavadia et al., 2012). Moreover, another

prospective study in Sub-Saharan Africa population recruited 425 patients, of which 13.5% was the prevalence of MI. About 71.5% of the total had final diagnosis of STEMI type and 28.5% of them were NSTEMI. The in-hospital mortality was reported to be 10% (N'Guetta et al., 2016).

In Ethiopia, MI continues to be a significant health problem (Shashu, 2021). Currently, MI is becoming highly prevalent and poor prognosis CVD in Ethiopia. A cross-sectional study conducted by Yedeta and his colleagues' on spectrum of MI in six main referral hospitals of Ethiopia reported that out of 6275 MI patients, 995 of them were in ACSRH, Mekelle. From the six referral hospitals, IHD accounted 9.6% of MI (Yadeta et al., 2017).

The treatment outcome of MI was observed to be high in Ethiopia, in which patient death and in-hospital complications are increasing. A retrospective cross-sectional study carried out in Tikur Anbessa Specialized Hospital, Ethiopia, from 1981 to 1986 revealed that out of 23 patients with MI in-hospital mortality was 29.4% (Fikreyesus & Bahta, 1989). Here are some epidemiological factors associated with MI are Age, gender, residence, family history of MI, cholesterol level, Diabetes mellitus, smoking use, Blood pressure, and Alcohol use etc.

1.2 Statement of the Problem

We know the fact that MI is one of the leading causes of mortality and morbidity in the world, especially in low and middle income country. Studies have been conducted to identify covariates of MI patients in Ethiopia using logistic regression and Semi-parametric proportional hazard models. But Logistic regression does not account the censoring observations, that is, it does not hold for time-to-event data, and in demographic applications, nonparametric and semi-parametric models are often used to model transition data. In such applications, it is assumed that all heterogeneity is captured by theoretically relevant covariates (Ben-Assuli et al., 2023). In many situations, however, there are ample reasons to suspect omitted or unmeasured factors. That is, while some individuals are more at risk of experiencing the event, it is unlikely that the underlying reasons for this variability are fully captured by the observed covariates. If there is unmeasured frailty, the hazard will not only be a function of the covariates but also of the frailty. To assess the true effects of the observed covariates under this circumstance, some have stressed the need to explicitly account for unobserved heterogeneity (Liu, 2013). Indeed, results from

several empirical and simulation studies have shown that accounting for unobserved heterogeneity significantly improves overall model fitness (Liu, 2013).

In this study, we argue that clustering (frailty) has an effect on modeling the understanding of time to discharge dynamics in MI patients. This effect may be attributed to the heterogeneity observed in the follow up center of patients. As a result, the shared frailty model approach proves to be relatively more effective in identifying covariates associated with the time to discharge of MI patients, thereby assisting the relevant body at ACSRH. Here is the main aim is that in my knowledge no study was conducted before on understanding time to discharge dynamics in myocardial infarction patients at ACSRH utilizing survival analysis with shared frailty model in particular.

Therefore, this study aims to answer the following scientific research questions:

- How do survival patterns vary among MI patients, considering shared frailty within the patient population?
- What are the risk factors of prolonged hospitalization for MI patients at ACSRH?
- How can the findings inform strategies for optimizing patient care and healthcare resource allocation for MI patients in this setting?
- What are the implications of the study findings for similar resource-limited healthcare settings globally?

1.3 Objective of the Study

1.3.1 General Objective of the Study

The overall objective of this study using a shared frailty model in survival analysis is to explore and comprehend the factors that impact the duration of discharge dynamics for MI patients at Ayder Comprehensive Specialized Referral Hospital.

1.3.2 Specific Objectives of the Study

- To provide insights into the factors contributing to discharge delays and propose strategies for improving the efficiency of care delivery for MI patients.
- To determine the patterns, trends, or characteristics of the time to discharge in MI patients using survival analysis with shared frailty models.

- To explore the role of shared frailty in explaining variations in the time to discharge among MI patients.
- To identify the risk factors associated with prolonged hospital stay and readmission in MI patients.

1.4 Significance of the Study

The significance of this study on discharge times for MI patients can likely be seen from different angles:

The study introduces a sophisticated statistical approach by utilizing survival analysis with a shared frailty model. This methodology allows for a more comprehensive understanding of the discharge dynamics for MI patients. It takes into account unmeasured factors and potential clustering within the hospital setting, leading to a more robust analysis. The study focuses on the critical aspect of patient care, namely the time it takes for MI patients to be discharged. By identifying factors that influence discharge times, the study can provide insights to optimize the discharge process. This optimization has the potential to improve patient outcomes, reduce hospital stays, and enhance resource management.

The study specifically examines MI patients at ACSRH. This context-specific approach enhances the significance of the study by addressing the unique challenges and characteristics of this particular healthcare setting. The findings can directly inform and guide healthcare practices at ACSRH, leading to tailored interventions and improved patient care strategies.

Although the study focuses on a specific hospital in Ethiopia, its findings may have broader implication for similar healthcare facilities globally. Understanding the dynamics of discharge times in MI patients contributes to the existing body of knowledge in the field of cardiology. The insights gained from this study can inform healthcare practices in other settings as well, expanding the knowledge base in the field.

1.5 Scope of the Study

The study is mainly focused on ACSRH, with the objective of assessing the risk factors influencing the time to discharge dynamics of MI patients receiving follow-up care at ACSRH.

1.6 Operational Definition

- **Time to Discharge:** Time to discharge refers to the duration it takes for a patient to be discharged from the hospital or healthcare facility after receiving treatment. Discharge from the hospital does not necessarily mean that the patient has fully recovered. In some cases, patients may still require ongoing medical care or rehabilitation after discharge.
- **Myocardial Infarction (MI)**: commonly known as a heart attack, occurs when blood flow decreases or stops in one of the coronary arteries of the heart, causing infarction (tissue death) to the heart muscle (Guelbert, 2022).
- **Frailty:** is defined as the unobserved or immeasurable factor that affects the response of interest (Wienke, 2010a).
- **Coronary Artery Disease (CAD):** is a condition in which the arteries that supply blood to the heart muscle become narrowed or blocked due to the buildup of plaque. This restricts blood flow to the heart and can lead to various cardiovascular problems, including heart attacks (Shashu, 2021).

2. Literature Review

2.1 NCDs as a Global Health Problem

Non-communicable diseases, including myocardial infarction, cardiovascular disease, diabetes and malignancies, can be prevented significantly by proper screening of populations for their risk factors, and by educating patients to implement healthy lifestyles, in particular by eating a healthy diet, increasing exercise and avoiding tobacco use (WHO, 2005; WHO, 2008). In addition to screening for tobacco, alcohol and other substance use, and smoking and dietary practices, other preventive medical tests, including blood pressure measurement, fasting blood glucose and fasting lipid profile, as well as measurement of waist circumference, body mass index (BMI), constitute basic parameters that can be useful for detecting early risk factors for MI, cardiovascular disease, and diabetes. Hypertension and diabetes can be treated relatively easily and inexpensively in most countries, including Ethiopia, and screening patients for these preventive parameters is useful for both treating them and educating them on lifestyle such as healthy diets and exercise.

The term 'non-communicable diseases' is a misleading term, because it includes some diseases notably, cancers of the liver, stomach, and cervix that are at least partly caused by infectious organisms, and it usually excludes mental illnesses, despite their large contribution to long-term disability. However, the four major behavioral risk factors for NCDs are: tobacco use, excessive alcohol consumption, poor diet, and lack of physical activity. Those factors are associated with disease clusters grouped under NCDs, namely MI, cardiovascular diseases, cancers, chronic pulmonary diseases, and diabetes that account for about 80% of deaths from NCD diseases (Lozano et al., 2012).

According to WHO estimates, NCDs contributed to 36 million deaths globally in 2010, and accounted for 63% of 57 million total deaths (WHO, 2010). About 80% of deaths related to NCDs occur in low- and middle-income countries, which also have a high proportion of deaths in middle age; such countries account for 90% of the 9 million NCD related deaths that occur before 60 years of age (Ndubuisi, 2021).

2.2 Non-Communicable Diseases in Ethiopia

Despite escalating rates of different NCDs in Ethiopia, they are neglected and priority of health care focuses on communicable diseases which have been considered more urgently important than NCDs. Although NCDs are believed by many to be a problem of rich societies, in reality they represent an under reported and neglected burden on health in the developing countries and one that is ever increasing (Shiferaw et al., 2018). For example one study in Ethiopia reported a prevalence of NCDs of 8.9%, with the specific observed prevalence of 0.5% for diabetes mellitus (DM), 2.6% for hypertension, and 3.0% for cardiovascular diseases (Muluneh et al., 2012).

The prevalence of hypertension was reported 10.1% in urban and 9.7% in rural areas of Sidamo zone (Giday, 2011). Status report on hypertension in Africa Consultative review for the 6*th* Session of the African Union Conference of Ministers of Health on NCDs reported a prevalence in Ethiopia of hypertension of 33% in males and 30% in females (van de Vijver et al., 2013). Another study in Gondar, Northwest Ethiopia in 2012 reported 28.3% prevalence of hypertensive of whom more than a third (37.0%) did not know they had hypertension; which agrees with other studies showing the occurrence of a ”silent epidemic” of high blood pressure in developing countries, particularly in Ethiopia (Awoke et al., 2012).

A study in Ethiopia among adults in Addis Ababa also reported a prevalence of hypertension, an important MI risk factor, of 31.5% among males and 28.9% among females (Tesfay et al., 2022). Prevalence of NCDs is increasing at a rate much higher than that of infectious diseases, maternal and prenatal diseases and nutritional deficiency diseases combined (Misganaw Dr. et al., 2014); (Muluneh et al., 2012).

2.3 Risk Factors of Myocardial Infraction

There are many risk factors associated with coronary heart disease and stroke. Some risk factors such as family history, ethnicity and age, cannot be changed. Other risk factors that can be treated or changed include tobacco exposure, high blood pressure (hypertension), high cholesterol, obesity, and physical inactivity, and diabetes, family history of MI, unhealthy diets, and harmful use of alcohol.

Hypertension (High Blood Pressure)

Blood pressure is measured as two numbers, written one over the other and recorded in millimeters of mercury for example, 120/78mmHg. The top (higher) number is the systolic pressure the pressure in the arteries as the heart is contracting and the bottom (lower) number is the diastolic pressure the pressure in the arteries when the heart is relaxed between beats. High blood pressure is defined as a repeatedly elevated systolic pressure of 140 or higher OR a diastolic pressure of 90 or higher.

Tobacco Use

Smoking is estimated to cause nearly 10% of all CVD. The risk of developing CVD is higher in female smokers, young men, and heavy smokers. There are currently about 1 billion smokers in the world today. Within two years of quitting, the risk of coronary heart disease is substantially reduced, and within 15 years the risk of CVD returns to that of a non-smoker.

Harmful use of alcohol

The harmful use of alcohol is a major risk factor for premature deaths and disabilities in the world. Hazardous and harmful drinking was responsible for 2.3 million deaths worldwide in 2004. That amounts to 3.8% of all deaths in the world. More than half of these deaths occurred as a result of NCDs, including cancers, cardiovascular disease and liver cirrhosis.

Raised Blood Glucose (diabetes)

Diabetes is defined as having a fasting plasma glucose value of 7.0mmol/l(126mg/dL) or higher. In 2010, diabetes was responsible for 1.3 million deaths globally and the global prevalence of diabetes was estimated to be 10%. CVD accounts for about 60% of all mortality in people with diabetes. The risk of cardiovascular events is from two to three times higher in people with type 1 or type 2 diabetes and the risk is disproportionately higher in women. In some age groups, people with diabetes have a two-fold increase in the risk of stroke. Patients with diabetes also have a poorer prognosis after cardiovascular events compared to people without diabetes. Cardiovascular risk increases with raised glucose values. Lack of early detection and care for diabetes results in severe complications, including heart attacks, strokes, renal failure, amputations, and blindness.

Physical Inactivity

Insufficient physical activity can be defined as less than five times 30 minutes of moderate activity per week, or less than three times 20 minutes of vigorous activity per week, or equivalent. It is the fourth leading risk factor for mortality. Approximately 3.2 million deaths and 32.1 million DALYs-representing about 2.1% of global DALYs - each year are attributable to insufficient physical activity. People who are insufficiently physically active have a 20 to 30% increased risk of all-cause mortality compared to those who engage in at least 30 minutes of moderate intensity physical activity most days of the week. The prevalence of insufficient physical activity is higher in high-income countries compared to low-income countries due to increased automation of work and use of vehicles for transport in high-income countries.

Unhealthy Diet

High dietary intakes of saturated fat, trans-fats and salt and low intake of fruits, vegetables and fish are linked to cardiovascular risk. Approximately 16 million (1.0%) DALYs and 1.7 million (2.8%) of deaths worldwide are attributable to low fruit and vegetable consumption. The amount of dietary salt consumed is an important determinant of blood pressure levels and overall cardiovascular risk and the WHO recommends a population salt intake of less than 5grams/person/day to help the prevention of CVD. Frequent consumption of high-energy foods, such as processed foods that are high in fats and sugars, promotes obesity compared to low-energy foods. High consumption of saturated fats and trans-fatty acids is linked to heart disease; elimination of trans-fat and replacement of saturated with polyunsaturated vegetable oils lowers coronary heart disease risk. Adequate consumption of fruit and vegetables reduces the risk of CVD. A healthy diet can contribute to a healthy body weight, a desirable lipid profile and a desirable blood pressure.

Cholesterol/Lipids

Raised blood cholesterol increases the risk of heart disease and stroke. Globally, one third of ischemic heart disease is attributable to high cholesterol. Overall, raised cholesterol is estimated to cause 2.6 million deaths (4.5% of total) and 29.7 million DALYS, or 2% of total DALYS globally. Lowering raised blood cholesterol reduces the risk of heart disease. In 2008, the global prevalence of raised total cholesterol among adults was 39% (37% for males and 40% for

females). The prevalence of raised total cholesterol noticeably increases according to the income level of the country. In low-income countries, around 25% of adults have raised total cholesterol, while in high-income countries; over 50% of adults have raised total cholesterol.

Overweight and Obesity

Obesity is strongly related to major cardiovascular risk factors such as raised blood pressure, glucose intolerance, type2 diabetes and dyslipidemia. Worldwide, at least 2.8 million people die each year as a result of being overweight or obese, and an estimated 35.8 million (2.3%) of global DALYs are caused by overweight or obesity. To achieve optimal health, the median BMI for adult populations should be in the range of $21{-}23 kg/m^2$, while the goal for individuals should be to maintain a BMI in the range $18.5{-}24.9\ kg/m^2$. The prevalence of raised BMI (body mass index) increases with income level of countries, up to upper-middle-income levels. The prevalence of overweight in high-income and upper-middle-income countries was more than double that of low- and lower-middle-income countries.

3. Data and Methodology

3.1 Study Area and Setting

Mekelle city, located 780 km north of Addis Ababa at an altitude of 1965 to 2220 m, is the capital city of the Tigray Regional State. It features four governmental hospitals, including ACSRH, and two private hospitals (Tilahun, 2014).

Established in 2008 and affiliated with Mekelle University, ACSRH serves approximately 10 million people across Tigray, north Afar, and parts of the Amhara region. The hospital has 500 inpatient beds and provides a range of specialty and subspecialty services through its Department of Internal Medicine, which includes 116 inpatient beds (ICU and ward) and various outpatient services. ACSRH is equipped with advanced medical technology, such as bronchoscopes, spirometers, and a cardiac catheterization laboratory (Hailu et al., 2023).

Additionally, it plays a vital role in research and the education of undergraduate and postgraduate medical and allied health professionals.

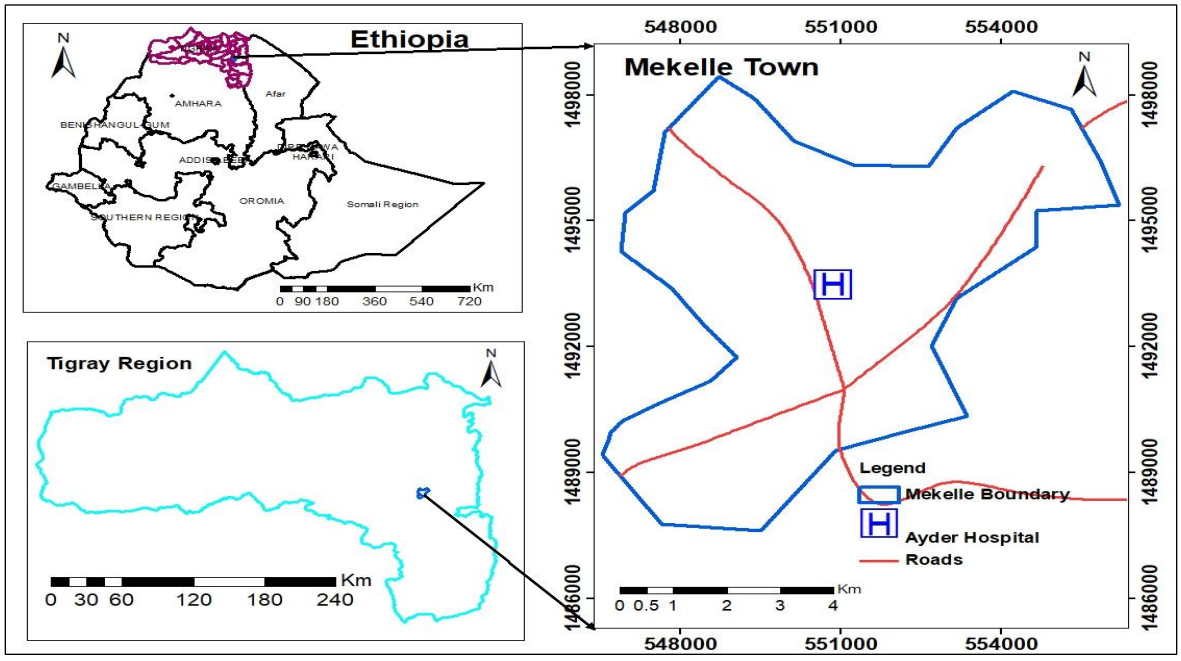


Figure 1: Maps of Mekelle city & Ayder Comprehensive Specialized Referral Hospital.

3.2 Study Population

The target population for this study is MI patients under follow-up at ACSRH, from January 01, 2018, to December 31, 2020.

3.3 Source of Data and Study Design

This study involved a retrospective review of medical records from adult MI patients registered at the MI follow-up center of ACSRH. Secondary data for this study were extracted retrospectively from patient charts at the same follow-up center. The study utilized a retrospective cohort design, involving MI patients admitted to ACSRH during a specific period.

Inclusion and Exclusion Criteria

Inclusion Criteria: In the study, patients above 18 years old who are admitted to ACSRH from January 01, 2018, to December 31, 2020, and patients who are discharged or censored within this interval, as well as patients with incomplete (missing) follow-up data, are being included.

Exclusion Criteria: In the study, patients with a history of previous repair for MI outside of ACSRH MI center, as well as patients presenting with successfully closed MI at the time of diagnosis, are being excluded. All adult MI patients admitted and discharged before January 1, 2018, and those discharged after December 31, 2020, are also being excluded

3.4 Sampling Technique and Sample Size Determination

Sampling technique is a system of taking small ratio of observation from a large population with the aim of getting information of those large populations from the sampled observation by using some statistical techniques. In this study, the simple random sampling technique was employed to select a representative sample of patients using a lottery method. An appropriate sample size is a means of gaining optimal precision of estimation with minimum cost.

The appropriate formula for determination of sample size using simple random sampling was adopted from (Cochran, 1997) as:

$$n = \frac{\frac{\left(z\alpha_{/2}\right)^2 p(1-p)}{d^2}}{1 + \frac{1}{N}\left(\frac{\left(z\alpha_{/2}\right)^2 p(1-p)}{d^2} - 1\right)}$$

Where,

- ✓ $n =$ is the required sample size
- ✓ $Z_{\alpha/2} =$ Is the critical value for the desired level of confidence (often 1.96 for a 95% confidence level).
- ✓ $P =$ is the estimated proportion of the population with a particular characteristic or outcome, $d =$ is the desired margin of error or precision, and $N =$ is the population size

3.5 Study Variables

3.5.1 Outcome Variable

The outcome variable is the survival time to discharge of MI patients, measured as the length of time from admission to the hospital until the patient is deemed medically stable and safe to leave the hospital environment, expressed in days. MI patients who survived during the study period, as well as those who were lost to follow-up, dropped out of the study, or died from causes other than MI, are considered as censored. This means that the survival data is subject to random right censoring.

3.5.2 Risk Factor Variables

In this study, the explanatory variables expected to be risk factors for MI were both categorical and continuous. While MI patients may have several risk factors, only certain variables were included in this study due to limitations in data availability. The variables included in the study are as follows:

Table 1: Risk factors of this study with variable coding

S.No.	Variables	Description	Categories
1	Age	Age in complete year of patients at start of treatment	Measured in years
2	Gender	Gender of the patients	0 = Female 1 = Male
3	Marital Status	Marital status of the patients	0=Single 1=Married 2= Widowed 3= Divorced
4	Smoking Status	Smoking status of the patients	0=Never 1=Ex-smoker 2=Current smoker
5	Educational Level	Educational level of the patients	0=No education 1=Primary 2=Secondary and above
6	Residence	Place of residence	0=Rural 1=Urban
7	Family History	Patient family history of MI	0=No 1=Yes
8	Weight	Base line weight	Measured in kg
9	BMI	Body mass index	0=Normal/underweight(BMI<25 Kg/m^2) 1=Over weight (BMI 25-29.9kg/m^2) 2=Obese(BMI>30kg/m^2
10	Diabetes mellitus	Diabetes mellitus Status	0=No 1=Yes
11	Stress	Patients stress	0= Acute stress 1= Chronic stress
12	Follow up site	Follow up site of patients	0= ICU 1=Medical ward
13	Alcohol use	Alcohol use status of patients	0= No 1=Yes
14	Cholesterol level	cholesterol level of patients	0=Normal(<200mg/dL) 1=High(>200mg/dL)
15	Blood pressure	Blood pressure of patients at start of treatment	0=Normal(Below 120/80 mmHg) 1=High(Between 120/80-179/109 mmHg) 2=Uncontrollable(Above 189/110 mmHg)

3.6 Statistical Model for MI Patients Studies: Survival Analysis

Survival analysis, also known as time-to-event analysis or event history analysis is a statistical method used to analyze the time until an event of interest, discharge of MI patients from the ACSRH, occurs. Survival analysis is one of the valuable tools for studying the time-to-event data related to time-to-discharge dynamics of MI adult patients from the hospital. Survival analysis in the context of discharge dynamics of MI patients can provide insights into the factors affecting the management and outcomes of this condition. In this study context, survival is the length of time in days until discharge of the MI patients from the hospital.

The use of survival analysis, as opposed to the use of other statistical methods, is most important when some subjects are lost to follow-up or when the period of observation is finite certain patients may not experience the event of interest over the study period. In this latter case, one cannot have complete information for such individuals. These incomplete observations are referred to as being censored. Most survival analyses consider a key analytical problem of censoring. In essence, censoring occurs when we have some information about individual survival time, but we do not know the survival time exactly. In reality, such an event can occur due to the following reasons:

- ✓ A person does not experience the event before the study end
- ✓ A person is lost to follow-up during the study period and
- ✓ A person withdraws from the study for unknown/known reasons

The most common categories of censoring are:

i) Right censoring

Survival time is said to be right censored when it is recorded from its beginning to a defined time before its end time. This type of censoring is commonly recognized survival analysis and also considered in this study. Let C denote the censoring time, that is, the time beyond which the study subject cannot be observed. The observed survival time is also referred to as follow up time. It starts at time 0 and continues until the event T or a censoring time C, whichever comes first. Let $c_1, c_2, \dots \dots . c_n$ be a sample of censoring times. And $T_1, T_2, \dots \dots T_n$ is event times. We observe a sample of couples, $(y_1, \sigma_1), (y_2, \sigma_2, \dots \dots . (\, y_n, \sigma_n)$, where for $i = 1,2,3 \dots \dots \dots n$. (Oakes, 1984).

$$y_i = \min(T_i, C_i) = \begin{cases} T_i, \text{if } T_i \le C_i \\ C_i, \text{ if } T_i \ge C_i \end{cases}, \qquad \sigma_i = I(T_i \le C_i) = \begin{cases} T_i, \text{if } T_i \le C_i \\ C_i, \text{ if } T_i \ge C_i \end{cases}$$

ii) Left censoring: Survival time is said to be left censored if an individual develops an event of interest prior to the beginning of the study.

iii) Interval censoring: Interval censoring occurs when the exact time of an event is unknown but is known to fall within a specific time interval. This complicates research design and statistical analysis, as standard methods are insufficient for handling such data. New statistical theories and computing advancements have improved the analysis of censored survival data. A key assumption in survival analysis is that censored individuals are at the same risk of subsequent failure as uncensored individuals, indicating non-informative censoring. In this study, we assume non-informative right censoring, where the response variable survival time refers to the length of stay in a center, measured in days (Oakes, 1984).

3.6.1 Survival Functions

The survivor function is defined to be the probability that the survival time of a randomly selected subject is greater than or equal to some specified time. Therefore, it gives the probability that an individual surviving beyond a specified time. Let T be a continuous random variable associated with the survival times; t be the specified value of the random variable T and $f\ (t)$ be the underlying probability density function of the 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, is given by (Oakes, 1984);

$$F(t) = P(T < t) = \int_0^t f(u)du \text{ , where; } t \ge 0 \ldots\ldots\ldots\ldots (3.1)$$

The survival function $S(t)$, is given by;

$$S(t) = P(T \ge t) = 1 - F(t), \text{ Where; } t \ge 0 \ldots\ldots\ldots\ldots\ldots (3.2)$$

From equations (3.1) and (3.2) the relationship between $f\ (t)\ and\ S\ (t)$ can be derived as

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}\big(1 - s(t)\big) = -\frac{d}{dt}s(t) \ge 0 \ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

Theoretically, as t ranges from 0 to infinity, the survivor function can be graphed as a smooth curve. Survivor functions have the characteristics that:

i) They are non-increasing function.

ii) At time $t = 0, S(0) = 1$; that is, at the start of the study, since no one has experienced the event yet, the probability of surviving past time 0 is one

iii) At time $t \to \infty, S(\infty) \to 0$; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually converge to zero.

3.6.2 Hazard Functions

The hazard function $h(t)$ gives the instantaneous potential for failing at time t, given the individual has survived up to time t. This is the conditional probability of experiencing the event of interest with in a very small time interval of size Δt having survived up to time t. It is a measure of the probability of failure during a very small interval, assuming that the individual has survived at the beginning of the interval. In addition, it is not a probability as it does not lie between $0\ and\ 1$. The hazard function, $h(t) \geq 0$ is given as (Oakes, 1984);

$$h(t) = \lim_{\Delta t \to 0} \frac{P(\text{an ind'l fails in the time interval } (t, t+\Delta t) \text{given survives until time } t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t / T \geq t)}{\Delta t}$$

By applying the theory of conditional probability, the hazard function can be expressed in terms of the underlying probability density function and the survivor function becomes;

$$h(t) = \frac{f(t)}{s(t)} = -\frac{d}{dt} \ln S(t) \ldots\ldots\ldots (3.4)$$

The corresponding cumulative hazard function $H(t)$ is defined as:

$$H(t) = \int_0^t h(u)du = -\ln S(t) \ldots\ldots\ldots\ldots\ldots (3.5)$$

And then

$$S(t) = \exp\big(-H(t)\big) \text{ and } f(t) = h(t)S(t) \ldots\ldots\ldots\ldots\ldots (3.6)$$

The survival function is most useful for comparing the survival progress of two or more groups while the hazard function gives a more useful description of the risk of failure at any time point.

3.6.3 Non Parametric Survival Methods

Nonparametric methods are often very easy and simple to understand as compared to parametric methods. Furthermore, nonparametric analyses are more widely used in situations where there is doubt about the exact form of distribution. Survival data are conveniently summarized through estimates of the survival function and hazard function. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time. These methods are said to be non-parametric methods since they require no assumptions about the distribution of survival time. Preliminary analysis of the data using non-parametric methods provides insight into the shape of the survival function for each group and get an idea of whether or not the groups are proportional, i.e., if the estimated survival functions for two groups are approximately parallel (do not cross). In order to compare the survival distribution of two or more groups, log-rank tests can be used (Oakes, 1984).

3.6.3.1 The Kaplan-Meier Estimator of Survival Functions

The Kaplan-Meier estimator is the standard non parametric estimator of the survival function, S(t), proposed by Kaplan and Meier (1958) which is not based on the actual observed event and censoring times, but rather on the ordered in which events occur. It is also called the Product-Limit estimator. KM estimator incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t. The technique becomes a little more complicated but still manageable when censored times are included. Let ordered survival times are given by $0 \leq t_1 \leq t_2 \leq \cdots \leq t_j \leq \infty$, then (Kaplan & Meier, 1958).

$$\hat{S}(t) = \begin{cases} 1 & , \quad \text{if } t < t_1 \\ \prod_{j:t_j \leq t} \left[1 - \frac{d_j}{r_j}\right], & \text{if } t \geq t_1 \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.7)$$

Where;

- d_j is the observed number of events at time t_j and
- r_j is the number of individuals at risk at time t_j

The Kaplan-Meier estimator ($\hat{S}(t)$) is a step function with jumps at the observed event times. The size of the jump at a certain event time t_j depends on the number of events observed at t_j, as well as on the pattern of the censored event times before The variance of the Product-Limit estimator is estimated by Greenwood's formula (Sawyer, 2003), and is given by;

$$Var\left(\hat{S}(t)\right) = [\hat{S}(t)]^2 \sum_{j:t_j \le t} \frac{d_j}{r_j(r_j - d_j)} \dots\dots\dots\dots (3.8)$$

Since the distribution of survival time tends to be positively skewed, the median is preferred for a summary measure. The median survival time is the time beyond which 50% of the individual sunder study are expected to survive, i.e., the value of t_{50} at $\hat{S}(t_{50}) = 0.5$

The estimated median survival time is given by $t_{50} = \min\left\{\frac{t_i}{\hat{S}(t)} < 0.5\right\}$, Where t_i is the observed survival time for the i^{th} individual, $i = 1,2, \dots n$. in general, the estimate of the p^{th} percentile is:

$$\hat{t}(p) = min\left(\frac{t_i}{\hat{S}(t)} < 1 - \frac{p}{100}\right) \dots\dots\dots\dots\dots\dots (3.9)$$

3.6.3.2 Non Parametric Comparison of Survival Functions

The simplest way of comparing the survival times obtained from two or more groups is to plot the Kaplan-Meier curves for those groups on the same graph. However, this graph does not allow us to say, with any confidence, whether or not there is a real difference between the groups. The observed difference may be a true difference, but equally, it could also be due merely to chance variation. Assessing whether or not there is a real difference between groups can only be done, with any degree of confidence, by using statistical tests. Among the various non-parametric tests one can find in the statistical literature, the Mantel-Hazel test, currently called the "log-rank" is the one commonly used non-parametric tests for comparison of two or more survival distributions. The log-rank test statistic for comparing two groups is given by (Cox & Oakes, 1984); assume that; I had grouped in to two place of delivery of obstetric fistula as health center(as group 0) and home(as group 1), now $k = 2$.

$$Q = \frac{[\sum_i^m w_i(d_{1i} - \hat{e}_{1i})]^2}{\sum_i^m w_i^2 \hat{v}_{1i}} \sim x^2_{k-1} \dots\dots\dots\dots\dots\dots (3.10)$$

Where: $\hat{e}_{1i} = \frac{n_{1i}d_i}{n_i}$ And $\hat{v}_{1i} = \frac{n_{1i}n_{0i}(n_{1i} - d_i)}{n_i^2(n_i - 1)}$

- n_{0i} is the number at risk at observed survival time $t_{(i)}$ in group 0
- n_{1i} is the number at risk at observed survival time $t_{(i)}$ in group 1
- n_i is the total number of individuals or risk before time $t_{(i)}$
- d_{1i} is the number of observed event in group 1
- d_i is the total number of event at time $t_{(i)}$
- k is the number of groups in each category

3.6.4 Cox PH Regression Model

The Cox PH regression model (introduced in a seminal paper by Cox, 1972), a broadly applicable and the most widely used method of survival analysis. Survival models are used to quantify the effect of one or more explanatory variables on failure time. This involves specification of a linear-like model for the log hazard. A parametric model based on the exponential distribution may be parameterized as follows:

$$log h_i(t|x) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

Equivalently:

$$h_i(t|x) = exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}) = \exp(\alpha)\exp(\beta' x)$$

In this case the constant α represents the log-baseline hazard since $log h_i(t) = \alpha$ when all the x's are zero. The Cox PH model is a semi-parametric model where the baseline hazard $\alpha(t)$ is allowed to vary with time.

$$\log h_i(t|x) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

$$h_i(t|x) = h_0(t)\exp(\beta_1 x_{i1} + \beta_{21} x_{i2} + \cdots + \beta_k x_{ik})$$

$$h_i(t|x) = h_0(t)\exp(x_i^T \beta) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.11)$$

Where:

- ✓ $h_0(t)$ is the baseline hazard function
- ✓ X_i is the covariates and
- ✓ β is vector of parameters for fixed effects.

The corresponding survival function for Cox-PH model is given by:

$$S(t, X) = [S_0(t)]^{\exp(\sum_{i=1}^{p} \beta_i X_i)}$$

Where, $S_0(t)$ is the baseline survival function.

With the Cox proportional hazards model the outcome is described in terms of the hazard ratio. The measure of effect is called hazard ratio. The hazard ratio of two individuals with different covariates X and X^* is given by:

$$\widehat{\mathrm{HR}} = \frac{\mathrm{h_0(t)exp}(\widehat{\beta}'\,\mathrm{X})}{\mathrm{h_0(t)exp}(\widehat{\beta}'\,\mathrm{X^*})} = \exp\left\{\sum \widehat{\beta}'\,(\mathrm{X}-\mathrm{X^*})\right\}$$

This hazard ratio is time independent, which is why this is called the proportional hazards model. The parameter of the Cox proportional hazard model refers to the hazard ratio of one group in comparison to the other groups for categorical covariates and change in hazard ratio with a unit change of the covariates for the continuous variables when other covariates are fixed.

The change in hazard ratio for the continuous covariate is given by:

$$\frac{\mathrm{h_i(t,x_k+1\,)}}{\mathrm{h_k(t,x_k)}} = \exp(\beta_\mathrm{k})$$

Which represent change in the hazard when there is a unit change in the covariate while other covariates keep constant?

For categorical explanatory variable X with a levels, the model contains $(\mathrm{a}-1)$ dummy variables defined as $D_i = 1$ if $x = i$ 0 otherwise for $i = 1,2,\ldots.,\mathrm{a}-1$. Let β_1 , β_2,…., $\beta_{\mathrm{a}-1}$ denote the coefficient of the levels of dummy variables. The ratio of the hazard of two subjects, one with X at level j and other with k $(j,k = 1,2,\ldots,\mathrm{a}-1)$, provides that the value of all other explanatory variables for this subject are the same, the hazard ratio between these two categories is given by:

$$\frac{h(t/D_j)}{h(t/D_k)} = \frac{\exp(\beta_j)}{\exp(\beta_k)} = \exp\big(\beta_j - \beta_k\big)$$

The quantity $\exp\big(\beta_j - \beta_k\big)100\%$ signifies the ratio of hazard function for subject at level j and k of covariates given the effect of other covariate keeps fixed.

3.6.4.1 Parital Likelihood Estimation for Cox PH Model

In order to fit the Cox proportional hazard model, we will estimate $h_0(t)$ and β. A more popular approach is proposed by Cox (1972) in which a partial likelihood function that does not depend on $h_0(t)$ is obtained for β. Partial likelihood is a technique which developed to make inference about the regression parameters in the presence of nuisance parameters $(h_0(t))$ in the Cox PH

model. In this part, we construct the partial likelihood function based on the proportional hazards model.

The data in survival analysis based on the sample size n are denoted by the triple (T_i, δ_i, X_i), $i = 1,2, \dots \dots, n$ where T_i is the time at which the i^{th} individual experience the event (in this research context, death), $\delta_i = 1$ if the event has occurred $\delta_i = 0$ if censored, X_i is the vector of covariate or risk factors for the i^{th} individual. we assume;

- ✓ Given X_i the life time and the censoring times are independent (non-informative censoring).
- ✓ $\tau_1 < \tau_2 < \dots \dots \tau_D$ be the D ordered distinct event at time τ_j
- ✓ We assume the there are no tied event times.

Let us define by;

- ✓ I_j is the identity of the individual who died at time τ_j
- ✓ v_j the time of the j^{th} failure at time τ_j and all information about censoring in $[\tau_{j-1}, \tau_j\ ]$

The observable data (T_i, δ_i, X_i) is represented be $\{I_j\}$ and $\{v_j\}$. Therefore;

$$P(Data) = P(\{I_1V_1, \dots \dots I_DV_D\})$$

$$= P(\{I_1V_1\}) \times P(\{I_2V_2\}|\{I_1V_1\}) \times \dots \dots . . \times P(\{I_DV_D\}|\{I_1V_1 \dots \dots I_{D-1}V_{D-1}\})$$

$$= \prod_{j=1}^{D} P(I_j|I_1, V_1, \dots, I_{j-1}, V_{j-1}, V_j) \times P\left(V_j\middle|\{I_1, V_1, \dots, I_{j-1}, V_{j-1}, V_j\}\right)$$

Due to non-informative censoring, the second term does not add much information about the parameters β.

Therefore, we define the partial likelihood as;

$$L^{partial}(\beta) = \prod_{j=1}^{D} P(I_j|\{I_1, V_1, \dots, I_{j-1}, V_{j-1}, V_j\}) = \prod_{j=1}^{D} P(I_j|H_j) \dots \dots \dots \dots \dots \dots (3.12)$$

Where:

H_j is the history of the data, up to j^{th} failure and including the failure time, but not the identity of the failing.

At each failure, we will note that the quantity $P(I_j|H_j)$ is the conditional probability that a specific individual fails at time τ_j given all the individuals that had not fail before τ_j.

We denote by $R(t)$ the set of all the individuals under study just prior to time t.

$$P(I_j|H_j) = P(individualls\ I_j\ fails|one\ individual\ fails\ in\ R(\tau_j))$$

$$= \frac{P(individuals\ I_j\ fails\ |at\ risk\ at\ \tau_j)}{\sum_{I\in R(\tau_j)} P(individuals\ I\ fails\ |at\ risk\ at\ \tau_j)}$$

$$= \frac{\lambda(\tau_j|X_j)d\tau_j}{\sum_{I\in R(\tau_j)} \lambda(\tau_j|X_j)d\tau_j} = \frac{\lambda_0(\tau_j)\exp(\beta^T X_j)}{\sum_{I\in R(\tau_j)} \lambda_0(\tau_j)\exp(\beta^T X_j)} = \frac{\exp(\beta^T X_j)}{\sum_{I\in R(\tau_j)} \exp(\beta^T X_j)}$$

And hence we get the partial likelihood;

$$L^{partial}(\beta) = \prod_{j=1}^{D} \frac{\exp(\beta^T X_j)}{\sum_{I\in R(\tau_j)} \exp(\beta^T X_j)} \dots\dots\dots\dots\dots\dots\dots\dots (3.13)$$

This is the partial likelihood defined by Cox. Note that, it does not depend on the underlying hazard function $h_0(.)$. Cox recommends treating this is as an ordinary likelihood for making inferences about β in the presence of the nuisance parameter $h_0(.)$.

3.6.5 Accelerated Failure Time Model

Although parametric models are very applicable to analyze survival data, there are relatively few probability distributions for the survival time that can be used with these models. In these situations, the accelerated failure time model (AFT) is an alternative to the PH model for the analysis of survival time data. Under AFT models we will measure the direct effect of the explanatory variables on the survival time instead of hazard. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean survival time.

The AFT model states that the survival function of an individual which covariates X at time t is the same as the survival function of an individual with a baseline survival function at a time

$t \times \exp(\alpha' X)$ Where $\alpha' = (\alpha_1, \alpha_2, \dots\dots, \alpha_p)$ is a vector of regression coefficients.

In other words, the accelerated failure-time model is defined by the relationship (Kelvin L. Moeschberger, 2013):

$$S(t|X) = S_0\{t \times \exp(\alpha' X)\}, for\ aall\ X \dots\dots\dots\dots\dots (3.14)$$

Hereby we can consider on a log-scale of the AFT model with respect to time is given analogous to the classical linear regression approach; the natural logarithm of the survival time $Y = \log(T)$ is modeled. This is the natural transformation made in linear models to convert positive variable to observations on the entire real line. A linear model is assumed for Y;

$$Y = \log(T) = \mu + \alpha' x + \delta\varepsilon$$

Where:

$\alpha' = (\alpha_1, \alpha_2, \dots\dots, \alpha_p)$ is a vector of regression coefficients.

$\mu =$ Intercept

$\delta =$ is scale parameter and

$\varepsilon =$ is the error distribution assumed to have a particular parametric distribution.

When we denote by S_0 the survival function when $X = 0$ then we find that

$$\begin{aligned} P(T > t|X) &= P(Y > \log(t)|X) \\ &= P\{\mu + \sigma\varepsilon > \log(t) - \alpha' X|X\} \\ &= P\{\exp(\mu + \sigma\varepsilon) > t \times \exp(-\alpha' X)|X\} \\ &= S_0\{t \times \exp(-\alpha' X)\} \end{aligned}$$

The effect of the covariates on the survival function is that the tie scale is changed by a factor $\exp(-\alpha' X)$, and we call this as an acceleration factor.

We can note that when;

- ✓ $\exp(-\alpha' X) > 1 \rightarrow$ The survival process accelerates.
- ✓ $\exp(-\alpha' X) < 1 \rightarrow$ The survival process decelerates.

If X is an indicator variable, this is equivalent to;

- ✓ $\alpha > 1 \rightarrow$ Time shrinks.
- ✓ $\alpha < 1 \rightarrow$ Time accelerates.

For each distribution of ε there is a corresponding distribution for T. The members of the AFT model considered in this study are the Weibull AFT, log-logistic AFT, and log-normal AFT model. The AFT models are named for the distribution of T rather than the distribution of log T. The survival function of T_i can be expressed by (Kelin and Moeschberger, 2003)

$S_i(t) = P(T_i \geq t)$

$= (\log(T_i) \geq \log(t)) = P(Y_i \geq \log(t)) = P(\mu + \alpha' x + \sigma\varepsilon \geq \log(t))$

$$= P\left(\varepsilon_i \geq \frac{\log(t) - \mu - \alpha' x}{\sigma}\right) = S_{\varepsilon_i}\left(\frac{\log(t) - (\mu + \alpha' x)}{\sigma}\right) \dots\dots\dots\dots\dots\dots\dots (3.15)$$

3.6.5.1 Weibull Accelerated Failure Time Model

The Weibull distribution (including the exponential distribution as a special case) can also be parameterized as an AFT model, and they are only family of distributions to have this property. The result of fitting a Weibull model can therefore be interpreted in either framework (Kelvin L. Moeschberger, 2013). Then the Weibull distribution is very flexible model for time-to-event data. It has a hazard rate which is monotone increasing, decreasing, or constant.

From equation (3.14), the accelerated failure time representation of the survival and hazard function of the Weibull model is given by;

First, Weibull density function can be expressed as

$$f(t, \mu, \alpha) = \frac{\alpha}{\mu}\left(\frac{\alpha}{\mu}\right)^{\alpha-1} * \exp\left(\frac{-t}{\mu}\right)^{\alpha} \dots\dots\dots\dots\dots\dots (3.16)$$

Where $\mu > 0$ and $\alpha > 0$

$$S_{\varepsilon i}(t) = \exp\left(-\exp\left(\frac{logt - (\mu + \alpha' x)}{\sigma}\right)\right) = exp\left(-\exp\left(\frac{-(\mu + \alpha' x)}{\sigma}\ t^{\frac{1}{\sigma}}\right)\right)$$

$$h_i(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} exp\left(\frac{-\mu - \alpha' x}{\sigma}\right) \dots\dots\dots\dots\dots\dots (3.17)$$

3.6.5.2 Log-logistic Accelerated Failure Time Model

The log logistic distribution has a fairly flexible function form, it is one of the parametric survival (Cox D. R., 1984) time models in which the hazard rate may be decreasing, increasing, as well as hump-shaped that is it initially increase and then decrease. In case where one comes across to censored data, using log logistic distribution is mathematically more advantageous than

other distributions. According to the study of Gupta et al. (1999), the log-logistic distribution is proved to be suitable in analyzing survival data conducted by (Cox D. R., 1972), (Cox, D. R., & Oakes, D., 1984).

The cumulative distribution function can be written in closed form is particularly useful for analysis of survival data with censoring (Bennett, 2020). The log-logistic distribution is very similar in shape to lognormal distribution, but is more suitable for use in the analysis of survival data. The log-logistic model has two parameters λ and ρ, where λ the scale parameter is and ρ is the shape parameter.

Its pdf is given by (Bennett, 2020);

$$f(t) = \frac{\lambda \rho t^{\rho-1}}{(1+\lambda t^{\rho})^2} \dots\dots\dots\dots\dots\dots (3.18)$$

The corresponding survival and hazard functions of are given by;

$$S(t) = \frac{1}{1+\lambda t^{\rho}} \dots\dots\dots\dots\dots\dots (3.19)$$

$$h(t) = \frac{\lambda \rho t^{\rho-1}}{1+\lambda t^{\rho}} \dots\dots\dots\dots\dots\dots\dots (3.20)$$

Where; $\lambda \epsilon R$, $\rho > 0$

When $\rho \leq 1$, the hazard rate decreases monotonically and when $\rho > 1$, it increases from zero to its maximum point and then decreases to zero. Suppose that the survival times have log-logistic distribution with parameter λ and ρ, under the AFT model, the hazard function for the i^{th} individual is;

$$h_i(t/x) = h_0\left(texp(-\alpha' x_i)\right) \exp(-\alpha' x_i) = \frac{\rho \exp(\lambda)\, texp(-\alpha' x_i)}{1+\exp(\lambda)\, texp(-\alpha' x_i)} \dots\dots\dots\dots\dots (3.21)$$

The log-logistic AFT model with a covariate x may be written as;

$Y = logT = \mu + \alpha' x_i + \sigma\varepsilon$, Where; $\alpha' = (\alpha_{1,}\alpha_2, \alpha_3, \dots\dots \alpha_p)$; ε has standard logistic distribution. The survival with covariate x is given as follows;

$$S_T(t|x) = \frac{1}{1+\lambda \exp(\beta' x)\, t^{\rho}} = \frac{1}{1+\exp(log\lambda + \beta' x)} \dots\dots\dots\dots\dots (3.22)$$

$$h_T(t|x) = \frac{\rho t^{\rho-1} \lambda \exp(\alpha' x)}{1+\lambda \exp(\alpha' x)\, t^{\rho}} = \frac{\rho t^{\rho-1} \lambda \exp(\alpha' x)}{1+\lambda \exp(log\lambda + \alpha' x)} \dots\dots\dots\dots\dots (3.23)$$

To interpret the factor $\exp(\beta' x)$ for log logistic model has the proportional odds (PO) property. So this model is also a proportional odds model, in which the odds of an individual surviving beyond time t are expressed as

$$\frac{S_T(t)}{1-S_T(t)} = \exp(-\alpha' x)\frac{S_0(t)}{1-S_0(t)} \dots\dots\dots\dots\dots\dots\dots (3.24)$$

The factor $\exp(-\alpha' x)$ is an estimate of how much the baseline odds of survival at any time changes when individual has covariate x. And $\exp(\alpha' x)$ is the relative odd of experiencing the event for an individual with covariate x relative to an individual with the baseline characteristics. As this representation of log-logistic regression is as accelerated failure time model with a log-logistic baseline survival function, then the log-logistic model is the only parameter model with both a proportional odds and an accelerated failure-time representation. If T_i has a log-logistic distribution, then ε_i has a logistic distribution. The survival function of logistic distribution is given by (Collett, 2003).

$$S_{\varepsilon_i}(\varepsilon) = \frac{1}{1+\exp(\varepsilon)} \dots\dots\dots\dots\dots\dots\dots (3.25)$$

Then, the AFT representation of log-logistic survival function is given by

$$S_{\varepsilon_i}(\varepsilon) = \left[1 + t^{\frac{1}{\sigma}} \exp\left(\frac{-\mu - \alpha' x}{\sigma}\right)\right]^{-1} \dots\dots\dots\dots\dots\dots (3.26)$$

And also the associated hazard function for the i^{th} individual is given by

$$h_t(t) = \frac{1}{\sigma t}\left[1 + t^{\frac{1}{\sigma}} \exp\left(\frac{-\mu - \alpha' x}{\sigma}\right)\right]^{-1} \dots\dots\dots\dots\dots (3.27)$$

If the plot of $\log\left[\frac{1-S(t)}{S(t)}\right]$ against $\log(t)$ is linear, the log-logistic distribution is appropriate for the given data set.

3.6.5.3 Log-normal Accelerated Failure Time Model

If the survival times are assumed to have a log-normal distribution, the baseline survival function and hazard function are given by (Collett, 2003):

$$S_0(t) = 1 - \phi\left(\frac{logt - \mu}{\sigma}\right), \qquad h_0(t) = \frac{\phi\left(\frac{logt}{\sigma}\right)}{\left[1 - \phi\left(\frac{logt}{\sigma}\right)\right]\sigma t'} \dots\dots\dots\dots (3.28)$$

Where μ and σ are parameters, $\phi(t)$ is the probability density function and the integral of $\phi(t)$ is the cumulative density function of the standard distribution. The survival function for the i^{th} individual is

$$S_i(t) = S_0(t|\eta_i) = S_0(t \times \exp(\mu + \alpha' x_i)) = 1 - \phi\left(\frac{logt - \alpha' x_i - \mu}{\sigma}\right) \dots\dots\dots (3.29)$$

Where $\eta_i = \exp(\alpha_1 x_1 + \alpha_2 x_2 + I + \alpha_p x_p)$. Therefore the log survival time for the i^{th} individual has normal $(\mu + \alpha' x_i, \sigma)$. The log normal distribution has the AFT property. In a group study we can easily get

$\phi^{-1}[1 - S(t)] = \frac{1}{\phi}(logt - \alpha' x_i - \mu)$, Where x_i is the value of a categorical variable which takes the value 0 in one group and 1 in the other group. This implies that the plot $\phi^{-1}[1 - S(t)]$ against $\log(t)$ will be linear if the log-normal distribution is appropriate for the given data set.

3.6.5.4 Parameter Estimation

Parameters of AFT models can be estimated by maximum likelihood method. The likelihood of n observed survival times, $t_1, t_2, t_3, \dots . t_n$, the likelihood function for right censored data is given by:

$$L(\alpha, \mu, \sigma) = \prod\prod_{j=1}^{n} f_j(t_j)^{\sigma_i} \times S_i(t_i)^{1-j} \dots\dots\dots (3.30)$$

Where:

- $f_j(t_j)$ is the density function of the i^{th} individual at time t_i,
- $S_i(t_i)$ is the survival function of the i^{th} individual at time t_i,
- σ_i is indicator variable.

The logarithm of the above equation yields;

$$\log L(\alpha, \mu, \sigma) = \sum_{j=1}^{n} \{-\delta_i \log(\delta t_i + \delta_i log f_i(x_i) + (1 - \delta_i) log S_i(W_i))\} \dots\dots\dots (3.31)$$

Where $W_i = \{logt_i - \frac{\mu + \alpha_{1i} x_i + \dots + \alpha_{pi} x_{pi}}{\delta}\}$, $Z = \{z_{ji}\}$ is vector of covariates for the j^{th} subject. The maximum likelihood parameters estimates will found by using Newton-Raphson procedure which can be done by software.

3.6.6 Shared Frailty Model

Many statistical models and methods proposed to model failure time data assume that the observations are statistically independent of each other. However, this does not hold in many applications. The concept of frailty provides a suitable way to introduce random effects in the model to account for association and unobserved heterogeneity. In its simplest form, a frailty is an unobserved random factor that modifies the hazard function of an individual or a group of individuals.

An individual is said to be if he/she is much more susceptible (exposed or infected) to adverse events than others. (Vaupel, 1979) introduces the term frailty to indicate that different individuals are at risk even though on the surface they appear to be quite similar with respect to the measurable such age, gender, weight, etc. They used the term frailty to represent an observable random effect shared by subjects with similar (unmeasured) risks in the analysis of mortality rates. A random effect describes excess risk or frailty for distinct categories, such as individual or families, over and above any measured covariates. Thus random effect or frailty models have been introduced into the statistical literature in an attempt to account for the existence of unmeasured attributes such as genotype that do introduce heterogeneity into a study population. It is recognized that individuals in the same group (cluster) are more similar than individuals in different cluster because they share similar genes, environment, custom, and culture, etc. Thus, frailty or random effect model try to account for correlations within groups (Prentice, R. L., Williams, B. J., & Peterson, A. V. et.al, 1981)

The assumption of a shared frailty model is that all individuals in cluster share the same frailty Zi, and this is why the model is called the shared frailty model. It was introduced by (Clayton, D., & Cuzick, J., 1985) and extensively studied in (Hougaard, 1986), Therneau and Grambsch (2000), and Duchateau et al. (2007). Shared-frailty models are appropriate when we wish to model the frailties as being specific to groups of subjects, such as subjects within families, kebeles, regions, etc. Here a shared frailty model may be used to model the degree of correlation within groups; i.e., the subjects within a group are correlated because they share the same common frailty.

Conditional on the frailty, the survival times in cluster i ($1 \leq i \leq$ n) are assumed to be independent. And the proportional hazard frailty model assumes (Wienke, 2010b):

$$h_{ij}(t|x_{ij}, U_i) = \exp(\beta' x_{ij} + u_i)\, h_0(t) = Z_i h_0(t) \exp(\beta' x_{ij}) \ldots\ldots\ldots (3.32)$$

An alternative when the proportional hazard assumption fails the accelerated failure time frailty model is used and given as:

$$\begin{aligned} h_{ij}(t|x_{ij}, U_i) &= \exp(\beta' x_{ij} + u_i)\, h_0(\exp(\beta' x_{ij} + u_i)t) \\ &= Z_i h_0(z_i \exp(\beta' x_{ij})\, t) \exp(\beta' x_{ij}) \ldots\ldots\ldots (3.33) \end{aligned}$$

Where, $h_0(t)$ is the baseline hazard function, $Z_i = exp(u_i)$, β is a vector of parameters to be estimated, x is a vector of observed covariates. The frailties Z_i are assumed to be identically and independently distributed random variables with common density function, $f(z, \theta)$, where θ is the parameter of the frailty distribution. The variability of Z_i determines the degree of heterogeneity among the groups. In empirical applications, the observed survival data are used to estimate the parameters of the distribution of frailty $f(z, \theta)$, and to actually predict the individual frailties. Since Z multiplies the hazard function, it has to be non-negative. Another constraint is further needed for identifiability reasons, more specifically; the mean of Z is typically restricted to unity in order to separate the baseline hazard from the overall level of the random frailties.

3.6.6.1 The Gamma Frailty Distribution

The gamma distribution has been widely applied as a mixture distribution for example (Hougaard, 1986). From a computational point of view, it fit very well into survival models, because it is easy to derive the formulas for any number of events. This is due to the simplicity of the derivatives of the Laplace transform. The gamma frailty distribution has been widely used in parametric modeling of intra-cluster dependency because of its simple interpretation, flexibility and mathematical tractability (Vaupel et al., 1979; Clayton, 1978; Oakes, 1999). To make the model identifiable, we 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 the one parameter gamma distribution. Under the restriction, the corresponding density function and Laplace transformation of gamma distribution is given by (Gutierrez, 2002);

$$f_z(z) = \frac{z_i^{\frac{1}{\theta}-1}}{\theta^{\frac{1}{\theta}}\,\Gamma(\frac{1}{\theta})} \exp\left(\frac{-z_i}{\theta}\right), \qquad \theta > 0 \ldots\ldots\ldots (3.34)$$

Where Γ(.) is the gamma function, it corresponds to a Gamma distribution Gam(μ, θ) with μ fixed to 1 for identifiability and its variance is θ. The associated Laplace transform is:

$$L(u) = (1 + \frac{u}{\theta})^{-\theta} . \theta > 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.35)$$

Note that if θ > 0, there is heterogeneity. So the large values of θ reflect a greater degree of heterogeneity among groups and a stronger association within groups. The conditional survival and hazard function of the gamma frailty distribution is given by (Gutierrez, 2002):

$$S_\theta(t) = [1 - \theta \ln(S(t))]^{\frac{-1}{\theta}} \ldots\ldots\ldots\ldots\ldots\ldots (3.36)$$

$$h_\theta(t) = h(t)\left[1 - \theta \ln\left(S(t)\right)\right]^{\frac{-1}{\theta}} \ldots\ldots\ldots\ldots\ldots\ldots (3.37)$$

Where $S(t)$ and $h(t)$ are the survival and the hazard functions of the baseline distributions. For the Gamma distribution, the Kend'll's Tau (Hougaard, 2000), which measures the association between any two event times from the same cluster in the multivariate case. It is an overall measure of dependence and independent of transformations on the time scale and the frailty model used. The associations within group members are measured by Kend'll's, which is given by:

$$\tau = \frac{\theta}{\theta + 2} \; \epsilon(0,1) \; \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.38)$$

In the parametric shared gamma frailty model with observed covariates, the unobserved frailty $z_i (i = 1,2, \ldots . . n)$ in each cluster can be estimated (Nielsen G.et'al, 1992) by the expression.

$$\widehat{\boldsymbol{z_i}} = \frac{1/\widehat{\delta^2} + \sum_{i=1}^{ni} \sigma_{ij}}{1/\widehat{\delta^2} + \sum_{i=1}^{ni} H_0(t_{ij};\hat{\theta}) exp(\hat{\beta}' x_{ij})} \ldots\ldots\ldots (3.39)$$

Here $\widehat{\delta^2}$ is the estimate of the frailty variance, $\hat{\theta}$ the vector of the estimated parameters of the cumulative baseline hazard function, $\hat{\beta}$ the vector of estimated regression coefficients ,and x_{ij} are observed covariates in my study.

3.6.6.2 Parameter Estimation

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 (Munda et al., 2012). In the second case, the baseline hazard is left unspecified and more complex techniques are available to approach that situation (Abrahantes & Duchateau, 2007). Even 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*., 201).

For right-censored clustered survival data, the observation for subject $j \in Ji = \{1, \dots, ni\}$ from cluster $i\ I = \{1, \dots, s\}$ is the couple $(yij\ \delta ij)$, where $yij = min(tij, cij)$ is the minimum between the survival time tij and the censoring time cij, and where $\delta ij = I(tij \leq cij)$ is the event indicator. When covariate information are been collected the observation will be $(yij, \delta ij, Xij)$, where Xij denote the vector of covariates for the j^{th} observation in the i^{th}cluster. In the parametric setting, estimation is based on the marginal likelihood in which the frailty have been integrated out by averaging the conditional likelihood with respect to the frailty distribution. Under assumptions of non-informative right-censoring and of independence between the censoring time and the survival time random variables, given the covariate information, the marginal log-likelihood of the observed data can be written as (Gutierrez, 2002)

$$l_{merg}(\psi, \beta, Z, X)$$

$$= \prod_{i=1}^{S} [(\prod_{j=1}^{n_j} (h_0(y_{ij}) exp(x_{ij}^T \beta))^{\delta_{ij}}))$$

$$\times \int_0^{\infty} Z_i^{di} \exp(-Z_i \sum_{j=1}^{n_i} h_0\,(y_{ij} \exp(x_{ij}^T \beta))) f(Z_i) dZ_i]$$

$$= \prod_{i=1}^{S} [(\prod_{j=1}^{n_j} (h_0(y_{ij}) exp(x_{ij}^T \beta))^{\delta_{ij}})) \times (-1)^{di} L^{di} (\sum_{j=1}^{ni} H_0(y_{ij} \exp(x_{ij}^T \beta)))]$$

Taking the logarithm, the marginal likelihood is:

$$L_{marg}(\psi, \beta, \theta, z, x) = \sum_{i=1}^{s} \left[\sum_{j=1}^{n_i} \delta_{ij} (\log(\psi)) + {x_{ij}}^T \beta \right] + \log[(-1)^{di} L^{(d)} \left[\sum_{j=1}^{ni} \psi \exp({x_{ij}}^T \beta) \right] \dots \dots (3.40)$$

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

$$L(s) = E[\exp(-Zs)] = \int_0^{\infty} (\exp(-Z_i s)\, f(Z_i) dZ_i, \qquad s \geq 0 \ldots\ldots\ldots\ldots (3.41)$$

Where ψ represents a vector of parameters of the baseline hazard function, β the vector of regression coefficients and θ the variance of the random effect. Estimates of ψ, β, θ are obtained by maximizing the marginal log-likelihood; this can be done if one is able to compute higher order derivatives of the Laplace transform up to $q = max\{d_1, \ldots, d_s\}$.

3.6.7 Model Development

The methods of selecting a subset of covariates in a PHs regression model are essentially similar to those used in any other regression models. The most common methods are purposeful selection, step-wise (forward selection and backward elimination) and best sub-set selections. Survival analysis using Cox regression method begins with a thorough univariate analysis of the association between survival time and all important covariates (Hosmer, D., and Lemeshow, S., 1998).

Recommendable procedure in selecting variables in the study

According to Hosmer and Lemeshow (1998), it is recommended to follow the steps given below.

(a) Include all variables that are significant in the univariate analysis at relaxed level and also any other variables which are presumed to be clinically important to fit the initial multivariable model.

(b) The variables that appear to be important from step one are then fitted together in a model. In the presence of certain variables others may cease to be important. As a result, backward elimination is used to omit non-significant variables from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.

(c) Variables, that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to the model from step 2, with forward selection method. This

process may result in terms in the model determined at step two ceasing to be significant.

3.6.8 Model Selection

For comparing models that are not nested, the Akaike‟s information criterion (AIC) is used which is defined as:

$$AIC = -2logL + 2(k + c + 1) \dots\dots\dots\dots (3.42)$$

Where k, is the number of covariates and c the number of model specific distributional parameters. This thesis used the AIC & BIC to compare various candidates of non- nested parametric models. The preferred model is the one with the lowest value of the AIC & BIC.

3.6.9 Model Diagnosis

3.6.9.1 Checking the Adequacy of Parametric Baseline Distributions

The graphical methods can be used to check if a parametric distribution fits the observed data. Model with the Weibull baseline has a property that the $log\ (-log(S(t))$ is linear with the log of time, where $S(t) = exp(-\lambda t^{\rho})$. Hence, $log(-log(S(t))) = log(\lambda) + \rho log(t)$. This property allows graphical evaluation of the appropriateness of a Weibull model by plotting $\log(-\log(\hat{S}(t)))$ versus $log(t)$ where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Datwyler, C., and Stucki,, T., 2011).The log-failure odd versus log time of the log-logistic model is linear. Where the failure odds of log-logistic survival model can be computed as:

$$\frac{1 - s(t)}{s(t)} = \frac{\lambda t^{\rho}/1 + \lambda t^{\rho}}{1/1 + \lambda t^{\rho}} = \lambda t^{\rho} \dots\dots\dots\dots\dots\dots\dots (3.43)$$

Therefore, the log-failure odds can be written as:

$$\log\left(\frac{1 - s(t)}{s(t)}\right) = \log(\lambda t^{\rho}) = \log(\lambda) + \rho \log(t) \dots\dots\dots\dots\dots\dots (3.44)$$

Therefore, the appropriateness of model with the log-logistic baseline can graphically be evaluated by plotting $\log\left(\frac{1-s(t)}{s(t)}\right)$ versus $log(time)$ where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Datwyler and Stucki, 2009). If the plot is straight line, log-logistic distribution fitted

the given dataset well. If the plot $\phi^{-1}[1-s(t)]$ against $log(t)$ is linear, the log-normal distribution is appropriate for the given data set.

3.6.9.2 The Quantile-Quantile Plot

A quantile-quantile or q-q plot is made to check if the accelerated failure time model provides an adequate fit to the data. The plot is based on the fact that, for the accelerated failure time model,

$$S_1(t) = S_0(\phi t) \dots\dots\dots\dots\dots\dots\dots (3.45)$$

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

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

Using the relation $S_1(t) = S_0(\phi t)$, we must have $S_0(t_{0p}) = 1-p = S_{1p}(t_{1p}) = S_0(\phi t_{1p})$ for all t. If accelerated failure time model holds, $t_{0p} = \phi t_{1p}$.To check this assumption we compute the Kaplan-Meier estimators of the two groups and estimate the percentiles t_{1p}, t_{0p}, for various values p. If we plot the estimated percentiles in group 0 versus the estimated percentiles 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, if the accelerated failure time model holds. If the curve is linear, a crude estimate of the acceleration factor ϕ is given by the slope of the line (Kelvin L. Moeschberger, 2013).

3.6.9.3 Using Residual Plots

The Cox-Snell residuals method can be applied to any parametric model and the residual plots can be used to check the goodness of fit of the model. For the parametric regression problem, analogs of the semi-parametric residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates (Kelvin L. Moeschberger, 2003).

The Cox-Snell residual for the j^{th} individual with observed survival time t_j is given by:

$$r_j = \hat{H}(T_j|X_j) = -log\hat{S}(T_j|X_j) \dots\dots\dots (3.46)$$

Where $\hat{H}$ and $\hat{S}$ are the estimated values of the cumulative hazard and survivor function of the j^{th} subject at time t_j respectively. If the model fits the data, then the $r_j's$ should have a 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 a straight line with slope unity and zero intercept. If yes, the fitted model is adequate. In general, Cox-Snell residual that provides a check of the overall fits of the model (Snell, 1968).

Table 2: The Cox–Snell residuals for the common baseline hazard functions that are considered in this study.

Base line Hazard Function	r_j
Exponential	$\hat{\lambda} t_j \exp(\hat{\beta}' X_j)$
Weibull	$\hat{\lambda} t_j^{\hat{\gamma}} \exp(\hat{\beta}' X_j)$

4. Results and Discussion

4.1 Descriptive Survival Analysis of MI patients

In this study, we considered a total of 206 MI patients who underwent follow-up at Ayder Comprehensive Specialized Referral Hospital between Jan 01, 2018, and Dec 31, 2020. The primary outcome of interest was the time to discharge from MI. Descriptive statistics of the covariates are presented in (Table 3). Out of the total 206 events, 112 (54.37%) events (discharged) were recorded, while the remaining 94 (45.63%) were censored.

Among the total of 206 patients, 142 (68.93%) were males, of whom 65 (45.77%) were censored, and 77 (54.23%) experienced the event. The remaining 64 (31.07%) patients were females, with 29 (45.31%) being censored and 35 (54.69%) experiencing the event.

The variable 'marital status' was categorized as single, married, widowed, and divorced. Among the total respondents, 18 (8.74%) were single, with 8 (44.44%) being censored and 10 (55.56%) experiencing the event. Out of the 102 (49.51%) married respondents, 49 (48.04%) were censored, and 53 (51.96%) experienced the event. Among the 54 (26.21%) respondents who were widowed, 27 (50%) were censored, and 27 (50%) experienced the event. Lastly, out of the 32 (15.54%) respondents who were divorced, 10 (31.25%) were censored, and 22 (68.75%) experienced the event.

The variable 'Diabetes mellitus status' was categorized as 'Normal', 'Pre-diabetes', and 'Diabetes', with reported frequencies of 55 (26.70%), 65 (31.55%), and 86 (41.75%), respectively. The variable 'Body mass index' was categorized as 'Normal/underweight (BMI<25)', 'Overweight (BMI 25-29.9)', and 'Obese (BMI>30)'. The frequencies for these categories were 8 (3.88%), 88 (42.72%), and 110 (53.40%), respectively.

Regarding the follow-up site, 87 (42.23%) patients were in the Intensive Care Unit (ICU), while 119 (57.77%) were in the medical ward. According to Table 3, the cholesterol level was categorized as 'Normal (<200mg/dL)' or 'High (>200mg/dL)'. Out of the total 206 patients, 15 (7.28%) had a normal cholesterol level, of which 7 (46.67%) were censored and 8 (53.33%) were

discharged. The majority, 191 (92.72%), had a high cholesterol level, with 87 (45.55%) being censored and 104 (54.45%) being discharged.

Table 3: Result of important characteristics of MI patients

Variables with category	Outcome		Total
	Event	Censored	
Gender			
Female	35 (54.69%)	29 (45.31%)	64(31.07%)
Male	77(54.23%)	65(45.77%)	142(68.93%)
Residence			
Rural	24(52.94%)	27(47.06%)	51(24.76%)
Urban	88(56.77%)	67(43.23%)	155(75.24%)
Marital status			
Single	10(55.56%)	8(44.44)	18(8.74%)
Married	53(51.96%)	49(48.04%)	102(49.51)
Widowed	27(50%)	27(50%)	54(26.21%)
Divorced	22(68.75%)	10(31.25%)	32(15.54%)
Educational level			
No education	47(46.08%)	47(53.92%)	32(15.53%)
Primary	39(63.93)	22(36.07%)	61(29.60%)
Secondary and above	18(41.86%)	25(58.14%)	43(20.87%)
Smoking status			
Never	51(53.12%)	45(46.88%)	96(46.61%)
EX-smoker	35(58.33%)	25(41.67%)	60(29.12%)
Current smoker	26(52%)	24(48%)	50(24.27%)
Family history of MI			
No	61(57.01%)	46(42.99%)	107(51.94%)
Yes	51(51.52%)	48(48.48%)	99(48.06%)
Diabetes mellitus			
Normal	31(56.36%)	24(43.64%)	55(26.70%)
Pre-diabetes	39(60%)	26(40%)	65(31.55%)
Diabetes	42(48.84%)	44(51.16%)	86(41.75%)
BMI			
Normal/underweight(BMI<25)	4(50%)	4(50%)	8(3.89%)
Over weight (BMI 25-29.9)	48(54.55%)	40(45.45%)	88(42.71%)
Obese(BMI>30)	60(54.55%)	50(45.45%)	110(53.40%)
Alcohol Use			
No	44(56.41%)	34(43.59%)	78(37.86%)
Yes	68(53.12%)	60(46.88%)	128(62.14%)
Cholesterol level			
Normal(<200mg/dL)	8(53.33%)	7(46.67%)	15(7.28%)

High(>200mg/dL)	104(54.45%)	87(45.55%)	191(92.72%)
Blood pressure			
Normal(Below 120/80 mmHg)	19(57.58%)	14(42.42%)	33(16.02%)
High(Between120/80-179/109 mmHg)	62(55.86%)	49(44.14%)	111(53.88%)
Uncontrollable(Above 189/110 mmHg)	31(50%)	31(50%)	62(30.10%)
Stress			
Acute stress	62(59.05%)	43(40.95%)	105(50.97%)
Chronic stress	50(49.50%)	51(50.50%)	101(49.03%)
Follow up site			
ICU	32(36.78%)	55(63.22%)	87(42.23%)
Medical ward	80(67.23%)	39(32.77)	119(57.77%)

From (Table 4), the study involved a diverse group of patients with a wide age range. At baseline, the mean age of the patients was 53 years, with the youngest patient being 18 years old and the oldest being 90 years old. This indicates that the study included individuals spanning different age groups. The standard error of 1.096 suggests that there was relatively little variation in age within the sample. Additionally, the 95% confidence interval of 50.595 to 54.918 provides a range within which we can be 95% confident that the true mean age of the population falls.

In terms of weight, the study included patients with varying weight profiles. The average baseline weight was 73.12 Kg, with the minimum weight being 51 Kg and the maximum weight being 92 Kg. This suggests a range of weights within the sample population. The standard error of 0.603 indicates a relatively small variation in weight within the sample. The 95% confidence interval of 71.930 to 74.312 provides a range within which we can be 95% confident that the true mean weight of the population falls.

Table 4: Summary Statistics for Continuous Variable

Variables	**Obs**	**Min**	**Max**	**Range**	**Mean**	**Std. Err**	**[95% Conf. Interval]**
Age	206	18	90	72	53	1.096	[50.595 54.918]
Weight	206	51	92	41	73.12	0.603	[71.930 74.312]
Time	206	5	25	20	15	0.315	[14.018 15.263]

Regarding the length of hospital stay (time), the study showcased variability in the time to discharge for the patients. On average, the patients had a length of stay of 15 days, with the

minimum being 5 days and the maximum being 25 days. This indicates that there was a range in the duration of hospitalization for the patients in the study. The standard error of 0.315 suggests relatively little variation in discharge time within the sample. The 95% confidence interval of 14.018 to 15.263 provides a range within which we can be 95% confident that the true mean discharge time of the population falls.

4.2 Non-parametric Survival Analysis

4.2.1 The Kaplan-Meier Estimate of Time to Discharge from MI

Non parametric survival analysis is a very important tool to visualize the survival of time to event under different covariates of the study. It gives more information on the shape of the survival hazard functions of the data set. Survival time distribution of time-to-discharge for different groups is estimated using Kaplan-Meier method and log-rank test has been employed in order to compare the survival curves of two or more groups.

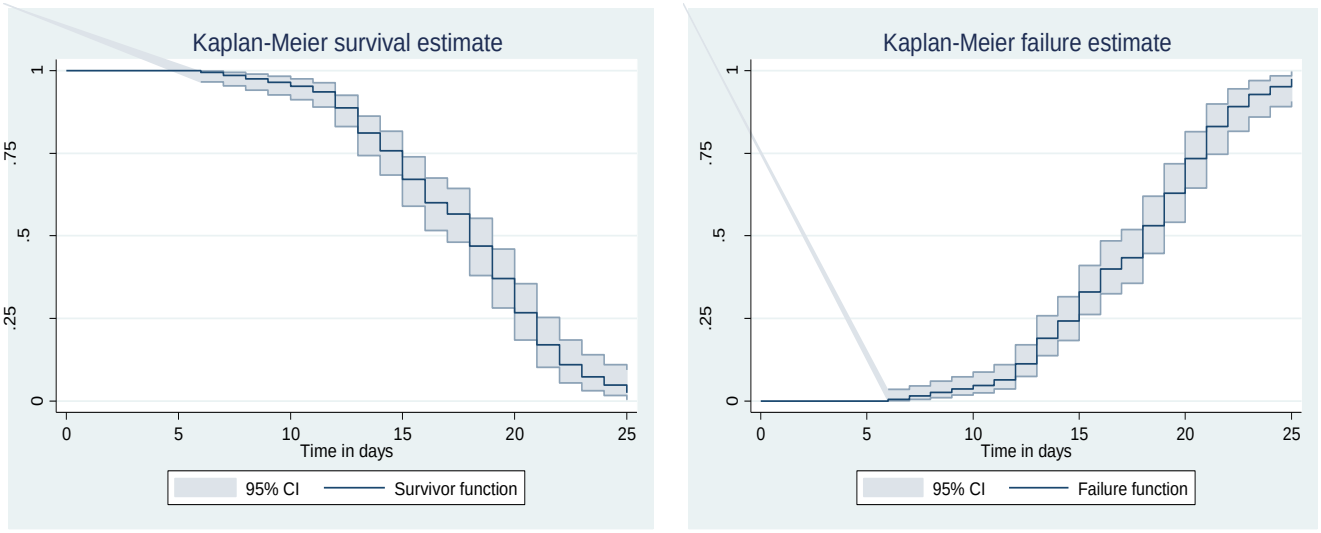


Figure 2: K-M survival and K-M failure (hazard) function of time to discharge from MI

(Figure 2) presents the Kaplan-Meier survival and failure (hazard) functions for the discharge from MI. The estimated survival function, with a 95% confidence interval, offers an important descriptive measure of the overall survival time pattern. The survivorship function demonstrates a relatively constant survival probability from the beginning until approximately 6 days. After this point, the survival time gradually declines during continued follow-up and then sharply

decreases until around 25 days. It's worth noting that the minimum value of the survivor function is not zero due to the presence of censored observations, where the largest observed time represents individuals who were still being followed up, but their exact survival time beyond the study period is unknown.

On the other hand, the hazard estimate graph for the discharge case shows a relatively constant hazard rate from the beginning until around 6 days. This suggests a consistent risk of discharge during this initial period. Subsequently, the hazard increases, indicating the occurrence of discharged subjects (events) during this time frame. The presence of subjects who have not experienced the desired outcome (discharge) beyond 7 days implies that some individuals were still under observation or follow-up at the end of the study.

4.2.2 Survival of Time to Discharge from MI for Different Group of Patients

4.2.2.1 Survival of Time to Discharge by Body Mass Index

From (Figure 3) illustrates a Kaplan-Meier survival function graph depicting the survival probabilities over time for different BMI categories. Three lines represent three distinct BMI categories: normal/underweight (BMI < 25), overweight (BMI 25-29.9), and obese (BMI > 30). The x-axis represents time in days (ranging from 0 to 25), while the y-axis represents the survival probability (ranging from 0 to 1.0).

The blue line, representing the normal/underweight BMI category, remains relatively stable and close to 1.0 for the first 10 days. It then shows a gradual decline, followed by a sharp drop after 15 days. The red line, representing the overweight BMI category, follows a similar pattern but with a slightly lower survival probability. The green line, representing the obese BMI category, exhibits a more pronounced decline in survival probability, starting at a higher probability and dropping more rapidly after 15 days.

This graph provides insights into the survival probabilities of heart attack patients based on their BMI categories. It demonstrates that individuals with a normal/underweight BMI have higher survival probabilities over time compared to those with overweight and obese BMIs. But the result of the log rank test (Table 1.2 in appendix) assured the idea ($p=0.4176$) means that this difference is not statistically significance.

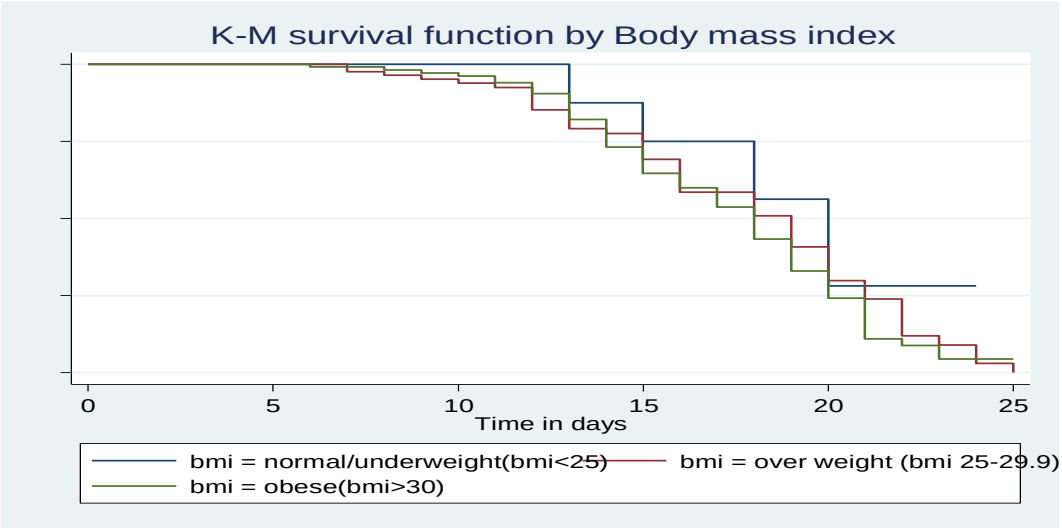


Figure 3: K-M estimates of survival for the variable Body mass index

4.2.2.2 Survival of Time to Discharge by Follow up site

Figure 4, illustrates a Kaplan-Meier survival function graph comparing the survival probabilities over time for two follow-up sites ICU and the medical ward. The x-axis represents time in days, ranging from 0 to 25, while the y-axis represents survival probability, ranging from 0 to 1.0.

The graph consists of two lines representing the different follow-up sites. The blue line corresponds to the ICU follow-up site, while the red line corresponds to the medical ward follow-up site. The blue line starts with a high survival probability close to 1.0 and gradually declines over time. Similarly, the red line also begins with a relatively high survival probability but follows a slightly lower trajectory compared to the blue line.

The graph suggests that patients with follow-up in the ICU have a slightly higher survival probability over time compared to those with follow-up in the medical ward. This observation may indicate that the ICU setting provides more effective or specialized care leading to improved survival time to discharge of the event. The result of the log rank test (Table 1.1 in appendix) assured the same idea (p=0.000) means that this difference is statistically significance.

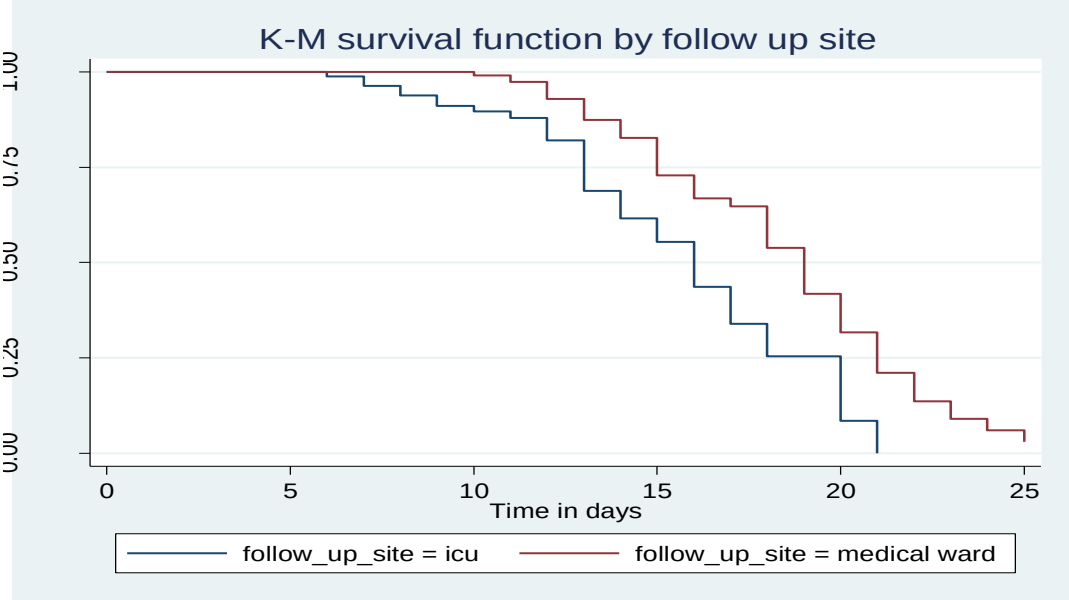


Figure 4: K-M estimates of survival for the variable follow up site.

4.2.2.3 Survival Time to Discharge by Blood Pressure

Figure 5, shows the survival probabilities for each category over time, with the lines diverging or intersecting at various points. The blue line starts with a high survival probability and remains relatively stable over time. The red line shows a decline in survival probability, indicating a lower chance of survival for patients with high blood pressure. The green line depicts a steeper decline compared to the blue line, suggesting an even lower survival probability for patients with uncontrollable blood pressure.

Based on the survival curves, it can be inferred that patients with normal blood pressure have a higher survival probability over time compared to those with high and uncontrollable blood pressure. The figure effectively illustrates the impact of blood pressure on the survival outcomes of patients and highlights the differences in survival patterns between the categories. But the result of the log rank test (Table 1.8 in appendix) assured the idea (p=0.3673) means that this difference is not statistically significance.

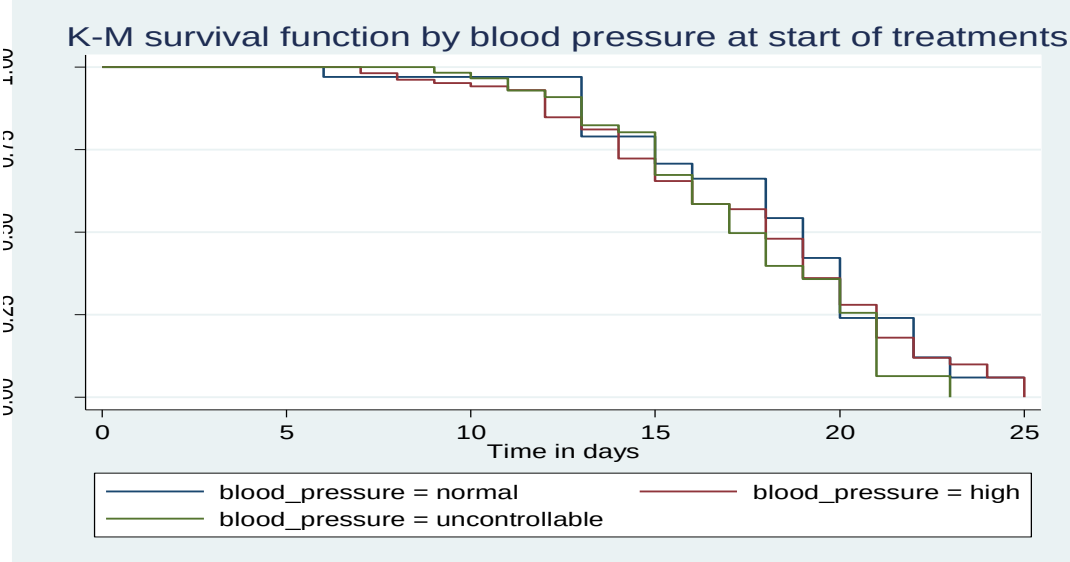


Figure 5: Survival curves of MI patients by blood pressure

4.2.2.4 Survival of Time to Discharge by Diabetes Mellitus

The figure includes three lines, each representing a different diabetes mellitus category: normal (blue), pre-diabetes (red), and diabetes (green). The blue line indicates patients with normal diabetes mellitus status, showing a high survival probability that remains relatively stable over time. The red line represents patients with pre-diabetes, displaying a gradual decline in survival probability, indicating a lower chance of survival compared to the normal category. The green line represents patients with diabetes, showing a steeper decline in survival probability, suggesting an even lower chance of survival compared to both the normal and pre-diabetes categories.

Based on the survival curves, it can be inferred that patients with normal diabetes mellitus status have higher survival probabilities over time compared to those with pre-diabetes and diabetes. The plot of survival time-to- discharge of the MI patients of the rest covariates with their long rank test is attached in the Appendix (Figures and Tables).

4.3. Parametric Survival analysis

4.3.1. Univariable Cox PH Regression analysis for the Risk factors of MI

The Cox regression for the Univariable risk factors of MI analysis, which is significant at p-value less than the cut-off point 0.25, can be candidate covariate for multivariable model building and statistical analysis. The survival experience for event of clinical and related variables of the risk factors of MI patients showed (Table 5) that gender, age, smoking status-ex-smoker, weight, BMI- Obese (BMI>30), diabetes mellitus-diabetes, alcohol use, blood pressure- uncontrollable(Above 189/110 mmHg), cholesterol level, follow up site; all had a p-value of less than the cut-off point 0.25 which means all the risk factors of MI patients were statistically associated with the survival time of time to discharge.

But, the other variables such as marital status, residence, educational level, BMI- Over weight (BMI 25-29.9), diabetes mellitus-prediabetes, smoking status- current smoker, blood pressure-high(between120/80-179/109 mmHg), and stress of the risk factors of MI patients were also not statistically associated with survival time of time to discharge.

Table 5: Univariable Cox-Regression Analysis of Risk factors

Risk factors	HR	SE	95% CI for HR	P-value
Gender				
Female**	1	1	1	1
Male	1.004	0.205	[0.672, 1.498]	0.086
Age (Continuous)	1.008	0.005	[0.997, 1.020]	0.114
Residence				
Rural**	1	1	1	1
Urban	1.41	0.326	[0.896, 2.212]	0.337
Marital_status				
Single**	1	1	1	1
Married	1.052	0.366	[0.532, 2.081]	0.884
Widowed	0.961	0.357	[0.464, 1.992]	0.916
Divorced	1.600	0.615	[0.753, 3.400]	0.321
Educational_ level				

No education**	1	1	1	1
Primary	0.872	0.184	[0.576, 1.320]	0.517
Secondary and above	0.812	0.222	[0.475, 1.388]	0.446
Smoking_status				
Never**	1	1	1	1
Ex-smoker	1.293	0.287	[0.837, 1.997]	0.247
Current smoker	0.854	0.206	[0.532, 1.370]	0.513
Weight (Continuous)	1.004	0.012	[0.980, 1.029]	0.021
BMI				
Normal/underweight(BMI < 25)**	1	1	1	1
Over weight (BMI 25-29.9)	1.639	0.845	[0.590, 4.553]	0.343
Obese(BMI>30)	1.801	0.932	[0.653, 4.969]	0.056
Diabetes_mellitus				
Normal**	1	1	1	1
Pre-diabetes	1.042	0.251	[0.649, 1.671]	0.865
Diabetes	1.341	0.321	[0.839, 2.143]	0.220
Alcohol_Use				
No**	1	1	1	1
Yes	1.328	0.260	[0.905, 1.949]	0.147
Family_history_MI				
No**	1	1	1	1
Yes	0.759	0.146	[0.521, 1.107]	0.153
Cholesterol_level				
Normal(< 200mg/dL)**	1	1	1	1
High(>200mg/dL)	1.203	0.443	[0.585, 2.477]	0.015
Blood_pressure				
Normal(Below 120/80 mmHg)**	1	1	1	1
High(Between120/80-179/109 mmHg)	1.16.	0.305	[0.693, 1.943]	0.572
Uncontrollable(Above 189/110 mmHg)	1.273	0.375	[0.714, 2.268]	0.213
Stress				
Acute stress**	1	1	1	1
Chronic stress	0.795	0.151	[0.548, 1.155]	0.329

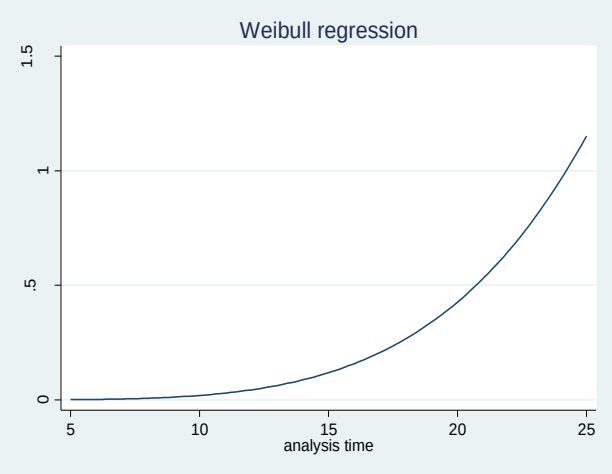
Follow_up_site				
ICU**	1	1	1	1
Medical ward	0.454	0.103	[0.291, 0.708]	0.00

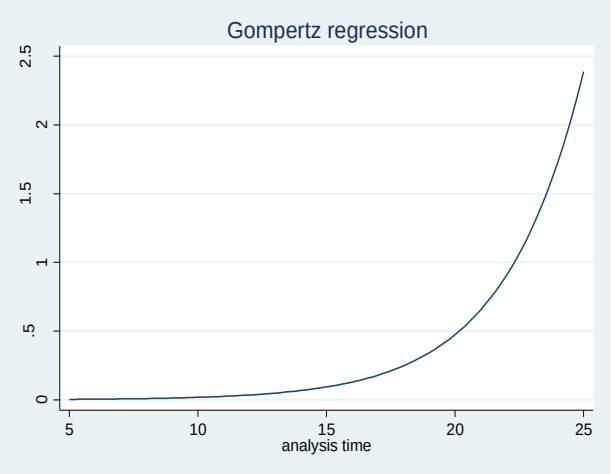
NB. ** Reference group

4.3.2 Assessing the PH Assumption

There are several options for checking whether the assumption of proportional hazard is satisfied with the data or not. In this study whether the proportional hazard assumptions were fit is checked by using formal statistical test and graphical methods. Proportional hazards mean that the hazard function of one individual is proportional to the hazard function of another individual.

Method 1; Testing proportional hazards assumption by checking the hazard ratio constant over time and constant slope over time, as shown in figure below (Figure 6), all the given curves increased at the same level and pattern each other simultaneously this indicated that they had constant hazard ratio and slope.





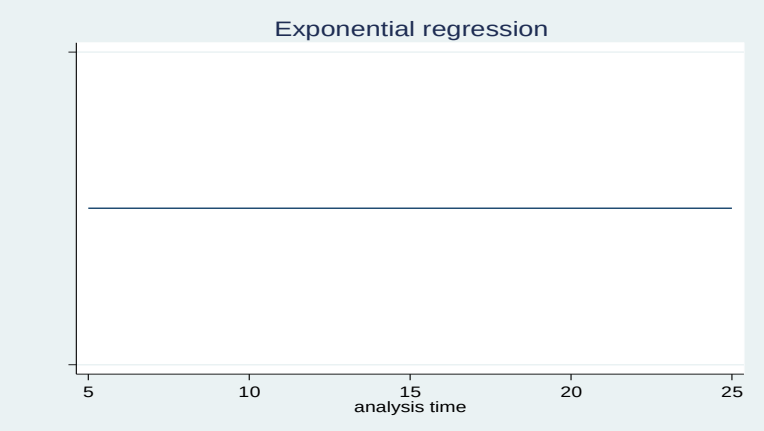


Figure 6: Testing Proportional Assumption.

Method 2: Testing the proportional hazards assumption using scaled Schoenfeld Residuals in the plots of scaled Schoenfeld residuals against time covariates body mass index, cholesterol level, age, weight, blood pressure, and follow up site of MI showed a randomness pattern and the smooth curve is a horizontal straight line and the slope is zero. Hence, these covariates satisfied the PH assumption.

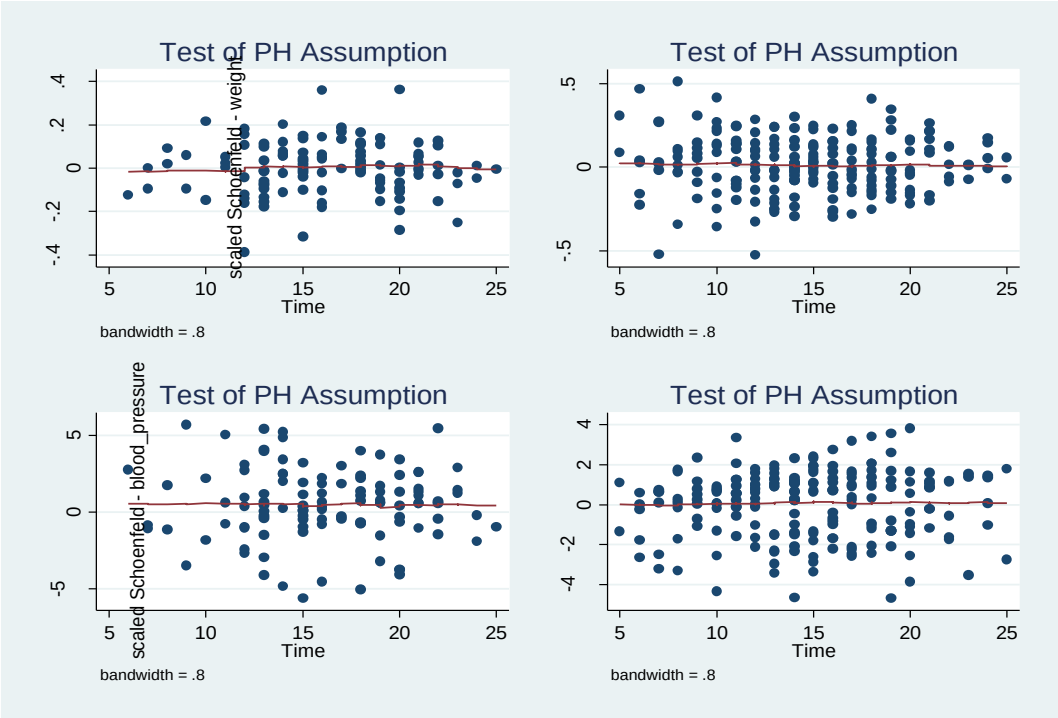


Figure 7: Testing of the PH assumption using Scaled Schoefeld residuals.

Method 3: Test of the PH assumption using global test for risk factors of MI

The graphical test is not enough to be checking the proportionality assumption of the model because graphical method of assessing the proportionality assumption is more or less subjective. So, it is better to use formal statistical tests of global test as supportive argument for proportionality test. Rho is a relation between time and residuals. The test of correlation (rho) is insignificant that indicates proportional hazards assumption is fulfilled.

As shown in the table below (Table 6), the null hypothesis states that there is no evidence to suggest a violation of the PH assumption. All covariates are insignificant at 5% level of significance, and indicating that all the covariates satisfy the proportionality assumption. The global test (p- value = 0.6854) is greater than 0.05 which indicates the proportional hazard assumptions satisfied by the covariate in the model,

Table 6: Checking PH Assumption: Global Test

Risk factors	**Categories**	**Rh_O**	**Chi-square**	**P-value**
Gender	Female**	1	1	1
	Male	0.02266	0.07	0.7948
Age	Continuous	0.03713	0.21	0.6498
Residence	Rural**	1	1	1
	Urban	0.04222	0.21	0.6446
Marital_status	Single**	1	1	1
	Married	0.04775	0.29	0.5885
	Widowed	-0.03045	0.14	0.7129
	Divorced	-0.06765	0.59	0.4408
Educational_level	No education**	1	1	1
	Primary	-0.04288	0.21	0.6434
	Secondary and above	-0.14528	2.74	0.0977
Smoking_status	Never**	1	1	1
	Ex-smoker	0.00288	0.00	0.9729
	Current smoker	0.15219	3.19	0.0741
Weight	continuous	0.01120	0.02	0.8948
BMI	Normal/underweight(BMI < 25)**	1	1	1
	Over weight (BMI 25-29.9)	-0.04520-	0.25	0.6160
	Obese(BMI>30)	-0.00817	0.01	0.9315
Diabetes_mellitus	Normal**	1	1	1
	Pre-diabetes	0.03618	0.19	0.6662

	Diabetes	0.07022	0.65	0.4196
Alcohol_use	No**	1	1	1
	Yes	0.01533	0.03	0.8539
Family_history_MI	No**	1	1	1
	Yes	-0.00330	0.00	0.9690
Cholesterol_level	Normal(< 200mg/dL)**	1	1	1
	High(>200mg/dL)	0.08771	1.12	0.2893
Blood_pressure	Normal(Below 120/80 mmHg)**	1	1	1
	High(Between120/80-179/109 mmHg)	-0.10584	1.31	0.2518
	Uncontrollable(Above 189/110 mmHg)	-0.03378	0.13	0.7163
Stress	Acute stress**	1	1	1
	Chronic stress	0.08566	0.96	0.3262
Follow_up_site	ICU**	1	1	1
	Medical ward	0.06651	0.55	0.4600
	Global Test		18.34	0.6854

NB ** Reference group

4.3.3. Model selection

As it showed in the below table (Table 7) Weibull PH model has the minimum AIC & BIC value and maximum log-likelihood value that considered it as the best fitted model for predicting risk factors of MI.

Table 7: Model Comparison: AIC, BIC & Log-likelihood Values

Model Type	Df	AIC	BIC	log-likelihood
Exponential PH	16	370.8967	424.1427	-169.4483
Weibull PH	17	90.3539	146.9278	-28.1769
Gompertz PH	17	96.2921	152.8661	-31.1460

4.3.4. Multivariable Analysis of Weibull PH Regression Model for the Risk Factors of MI

After selecting the fitted Weibull survival regression model that used to analyze the risk factors of MI patients. Multivariable analysis was done using the fitted Weibull PH regression model as shown in (Table 8). Those variables, whose p-value had less than the cut-off point at 25% level of significance in the Univariable analysis of the Cox regression model were included in the

multivariable analysis. In the multivariable analysis of the Weibull PH regression model, variables that had p-value less than 0.05 were considered as statistically significant variables and used for interpretation.

Table 8: Multivariate Analysis of the Weibull PH regression model

Risk factors with categories	**β**	**SE**	**Z**	**Exp(β)**	**P-value**	**[95% CI]**
Gender						
Female**	1	1	1	1	1	1
Male	1.2102	0.2545	0.83	3.3541	0.004	[0.7109, 1.3866]
Smoking_status						
Never**	1	1	1	1	1	1
Ex-smoker	0.4115	0.2733	1.51	1.5091	0.132	[-0.1240, 0.9471]
Current smoker	-0.2172	0.3483	-0.62	0.8047	0.533	[-0.8998, 0.4654]
Family_history_MI						
No**	1	1	1	1	1	1
Yes	1.3968	0.2229	1.78	4.0422	0.015	[0.0338, 1 .401]
Diabetes_mellitus						
Normal**	1	1	1	1	1	1
Pre-diabetes	-0.2006	0.2640	-0.76	0.8182	0.057	[-0.71814, 0.3167]
Diabetes	0.0139	0.3010	0.05	1.0139	0.013	[0.5760, 1.6040]
BMI						
Normal**	1	1	1	1	1	1
Over weight	0.6115	0.5916	1.03	1.8432	0.301	[-0.5479, 1.7711]
Obese(BMI>30)	0.8356	0.5998	1.39	2.3061	0.004	[0.3401, 2.0113]
Alcohol_use						
No**	1	1	1	1	1	1
Yes	0.5155	0.2744	1.88	1.6744	0.236	[0.0224, 1.0534]
Cholesterol_level						
Normal($< 200mg/dL$)**	1	1	1	1	1	1
High(>200mg/dL)	0.5427	0.4570	1.19	1.7206	0.035	[0.4385, 1.3529]
Blood_pressure						
Normal**	1	1	1	1	1	1
High	0.2817	0.2745	1.03	1.3254	0.305	[-0.2562, 0 .8197]
Uncontrollable	0.0682	0.3315	0.21	1.0705	0.037	[-0.5816, 0 .7181]
Follow_up_site						
ICU**	1	1	1	1	1	1
Medical ward	0.9091	0.2500	3.64	2.4820	0.010	[0.3991, 1.4191]
Age(continuous)	-0.0155	0.0082	0.68	0.9846	0.001	[0.0201, 1.0317]

Weight(continuous)	-0.0393	0.0146	0.64	0.9614	0.025	[0.0131, 1.1026]
_cons	-17.135	1.7845	-9.60	3.6e-08	0.00	[-20.633, -13.637]
Log likelihood	-30.458	LR chi2(15) = 31.09		Prob > chi2 = 0.0085		

NB ** Reference group

Depending on the result of multivariate analysis that are printed out in the table above (Table 8): gender, family history of MI, diabetes mellitus(diabetes), BMI(Obese(BMI>30)), cholesterol level, blood pressure(uncontrollable), follow up site, age, and weight are statistically significant. When the effect of the other factors keep fixed, the estimated hazard ratio of gender that were male is increased by 3.5% with 95% confidence interval [0.7109, 1.3866] compared to the reference group which is "female". Moreover, the slope of the Weibull model is 1.2102 which indicates that the increasing rate of the hazard or risk of discharge. The same for family history MI, the estimated hazard ratio of family history of MI was Yes is increased by 4.2% with 95% confidence interval [0.0338, 1 .401] compared to the reference group which is "No" while the effect the other factors hold constant. Moreover, the slope of the Weibull model is 1.3968 which indicates that the increasing rate of the hazard or risk of discharge.

When the effect of the other factors keep fixed, the estimated hazard ratio of cholesterol level that were "high (>200mg/dL)" is increased by 7.2% with 95% confidence interval [0.4385, 1.3529] compared to the reference group which is "normal (<200mg/dL)". Moreover, the slope of the Weibull model is 0.5427 which indicates that the increasing rate of the hazard or risk of discharge. Similarly, one unit increase in patients age (years) and all other variables hold constant, the risk of discharged decreased by 1.5% with 95% confidence interval [0.0201, 1.0317]. Moreover, the slope of the Weibull model is -0.0155 which indicates that the decreasing rate of the hazard or risk of discharge. When the effect of the other factors keep fixed, the estimated hazard ratio of BMI that were "Obese (BMI>30)" is increased by 3.06% with 95% confidence interval [0.3401, 2.0113] compared to the reference group which is "Normal/underweight (BMI<25)". Moreover, the slope of the Weibull model is 0.8356 which indicates that the increasing rate of the hazard or risk of discharge.

The weight of patients is increased by one kg, holding the other variables constant the risk of discharged decreased by 9.6% with 95% confidence interval [0.0131, 1.1026]. Moreover, the slope of the Weibull model is -0.0393 which indicates that decreasing rate of the hazard or risk of discharge. Assuming all the other variables hold constant, the estimated hazard ratio of follow up site that are treated in with medical ward is increased by 4.8% with 95% confidence interval [0.3991, 1.4191] compared to the reference group which is "ICU". The slope of the Weibull model is 0.9091 also indicated the increased hazard rate or risk of discharge.

4.4. Parametric Shared Frailty Model Results

4.4.1 Model Selection

In the previous section of this paper, the three PH models were fitted and compared to analyze the survival of time to discharge to identify baseline distribution and risk factors and then the Weibull model was selected based on AIC & BIC value. As it stated in the previous sections the main focus of this paper is to identify the risk factors that associated with the survival of time-to-discharge of MI patients using the evidence of parametric shared frailty model. For the data on MI the three parametric baseline distributions (Exponential, Weibull, and Gompertz) with Gamma frailty distribution were fitted by using follow up site as frailty term. The effect of frailty (random component) was significance for all (Exponential, Weibull, and Gompertz) gamma shared frailty models. According to the AIC and BIC values summarized in the table below (Table 9) the Weibull -gamma shared frailty model had the smallest AIC (89.2839) and BIC (148.9278) values than exponential-gamma and Gompertz-gamma shared frailty models which indicates that the Weibull-gamma shared frailty model is more efficient model to describe time-to-discharge dynamics of MI.

Table 9: Comparison of Gamma shared frailty model with different baseline distributions

Baseline distribution	Frailty Distribution	AIC	BIC
Exponential	Gamma	375.4022	431.9761
Weibull	Gamma	89.2839	148.9278
Gompertz	Gamma	98.5131	155.4568

4.4.2. Weibull Gamma Shared Frailty Model Analysis for the Risk Factors of MI

The age (in complete years) of MI patients was found to be statistically significant in determining the risk factors of MI. The hazard ratio and its 95% confidence interval were 0.9844 and (1.0105, 1.5116) respectively, resulting in a significant p-value of 0.002. Notably, the confidence interval excluded one, indicating that age was a crucial factor affecting the survival of MI patients. Therefore, as the age of MI patients increased, the survival rate decreased, with survival increasing by 0.9844 as the age decreased by one year. Weight was also identified as a statistically significant factor in determining the risk factors of MI. The hazard ratio and its 95% confidence interval were 0.9899 and (0.0193, 0.9879) respectively, yielding a significant p-value of 0.019. Similarly, the confidence interval did not encompass one, signifying that weight was a crucial determinant for the survival of MI patients. Consequently, as the weight of MI patients increased, their survival rate decreased, with survival decreasing by 0.9899 as weight increased by one kg.

The covariate "follow-up site at the medical ward" was statistically significant in determining the risk factors of MI. The hazard ratio and its 95% confidence interval were reported as 2.4868 and (0.4281, 1.4369) respectively, with a significant p-value of 0.018. This suggests that patients with MI who were followed at the medical ward faced different risk factors compared to those followed at the ICU, with a higher hazard ratio indicating increased risk.

Moreover, gender was identified as a significant factor in determining the risk factors of MI. The hazard ratio and its 95% confidence interval for males were 4.1012 and (0.0110, 1.1967) respectively, with females as the reference group. The associated p-value was 0.001, indicating that gender played a significant role in determining the risk factors of MI. The hazard ratio for males was 4.1012 times greater than for females, highlighting a notable disparity in risk levels between the genders.

Diabetes mellitus was revealed to be a significant factor in determining the risk factors of MI. The hazard ratio and its 95% confidence interval for individuals with diabetes were 2.7563 and (0.582, 1.5858) respectively, with diabetes mellitus (normal) as the reference group. The p-value of 0.003 indicated the importance of diabetes mellitus in assessing the risk factors of MI, with the hazard ratio for diabetes being 2.7563 times higher than for individuals with normal diabetes

mellitus. In addition, blood pressure was identified as a significant factor in determining the risk factors of MI. The hazard ratio and its 95% confidence interval for individuals with "Uncontrollable" blood pressure were 1.0726 and (0.3823, 1.8024) respectively, with "Normal" blood pressure as the reference group. The associated p-value of 0.026 indicated the significance of blood pressure in assessing the risk factors of MI. Individuals with uncontrollable blood pressure had a hazard ratio of 1.0726, signifying a higher risk compared to those with normal blood pressure.

Similarly, cholesterol level was found to be a significant factor in determining the risk factors of MI. The hazard ratio and its 95% confidence interval for individuals with "High (>200mg/dL)" cholesterol were 1.7318 and (0.3385, 1.3835) respectively, with "Normal (>200mg/dL)" cholesterol as the reference group. The p-value of 0.013 indicated the importance of cholesterol level in assessing the risk factors of MI. The hazard ratio for individuals with high cholesterol was 1.7318, suggesting an increased risk compared to those with normal cholesterol levels.

Moreover, family history of MI was revealed to be a significant factor in determining the risk factors of MI. The hazard ratio and its 95% confidence interval for individuals with a family history of MI labeled as "Yes" were 3.7139 and (0.0152, 1.3031) respectively, with "No" family history of MI as the reference group. The associated p-value of 0.021 highlighted the significance of family history in assessing the risk factors of MI. The hazard ratio for individuals with a family history of MI was 3.7139, indicating a substantially higher risk compared to those without a family history of MI.

Lastly, BMI was identified as a significant factor in determining the risk factors of MI. The hazard ratio and its 95% confidence interval for individuals classified as "Obese (BMI > 30)" were 2.3445 and (0.3901, 2.1253) respectively, with "Normal" BMI as the reference group. The p-value of 0.001 underlined the importance of BMI in assessing the risk factors of MI. The hazard ratio for individuals with an obese BMI was 2.3445, indicating a significantly higher risk compared to those with a normal BMI.

Table 10: Result of Weibull-Gamma shared frailty model

Risk factors with categories	B	SE	Z	Exp(β)	P-value	[95% CI]
Gender						
Female**	1	1	1	1	1	1
Male	1.4113	0.4245	0.91	4.1012	0.001	[0.0110, 1.1967]
Family_history_MI						
No**	1	1	1	1	1	1
Yes	1.3121	0.2229	1.78	3.7139	0.021	[0.0152, 1.3031]
Diabetes_mellitus						
Normal**	1	1	1	1	1	1
Pre-diabetes	-0.2006	0.2640	-0.76	0.8182	0.057	[-0.782, 0.3011]
Diabetes	1.0139	0.2010	0.15	2.7563	0.003	[0.582, 1.5858]
BMI						
Normal**	1	1	1	1	1	1
Over weight	0.6115	0.5916	1.03	1.8535	0.301	[-0.597, 1.7411]
Obese(BMI>30)	0.8521	0.6164	1.39	2.3445	0.001	[0.3901, 2.1253]
Cholesterol_level						
Normal(>200mg/dL)*	1	1	1	1	1	1
High(>200mg/dL)	0.5492	0.5020	1.19	1.7318	0.013	[0.3385, 1.3835]
Blood_pressure						
Normal**	1	1	1	1	1	1
High	0.2817	0.2745	1.03	1.3254	0.305	[-0.156, 0 .9301]
Uncontrollable	0.0701	0.2605	0.35	1.0726	0.026	[0.3823, 1.8024]
Follow_up_site						
ICU**	1	1	1	1	1	1
Medical ward	0.9110	0.3102	2.64	2.4868	0.018	[0.4281, 1.4369]
Age(continuous)	-0.0157	0.0082	0.68	0.9844	0.002	[1.0105, 1.5116]
Weight(continuous)	-0.0101	0.0146	0.64	0.9899	0.019	[0.0193, 0.9879]
_cons	-16.955	1.7845	-9.60	4.3e-08	0.001	[-20.633, -13.637]
theta	2.46e-08 = 1.1056					
Log likelihood	-27.6214	LR chi2(12) = 29.19 AIC= 89.2839			LR test of theta=0: chi-square = 1.5e-06 = 0.003722 p-value = 0.0031	

NB ** Reference group

4.4.3. Comparison of Weibull PH and Weibull Gamma Shared Frailty Models

Depending on (Table 11), we can observe that the results of Weibull PH and Weibull Gamma shared frailty models are quite similar but not identical. In order to compare the efficiency of the models the AIC was used. We can see from (Table 10) that the Weibull gamma shared frailty model has a minimum AIC (89.2839) than the Weibull PH model (AIC = 90.3539), indicating that Weibull gamma shared frailty model had fitted the survival of time-to-discharge of MI better than the Weibull PH model which did not take in to account the latent effect of groups. When we also look at the estimated value of coefficients of the covariates, they are altered with the inclusion of the frailty component and confidence interval for the risk factor is a little beat narrower for Weibull gamma shared frailty model. Furthermore, the variance of random effect (frailty) was significance at 5% level of significance which indicates that the parametric shared frailty model fitted the given data set better than PH model. Generally, the Weibull gamma shared frailty model is preferable over Weibull PH model for modeling of risk factors of MI patient's dataset.

Table 11: Comparison of Weibull PH and Weibull-gamma shared frailty models.

Risk factors with categories	**Weibull PH**			**Weibull-gamma Shared Frailty**		
	β	**Exp(β)**	**[95% CI]**	**β**	**Exp(β)**	**[95% CI]**
Gender						
Female**	1	1	1	1	1	1
Male	1.2102	3.3541	[0.7109, 1.3866]	1.4113	4.1012	[0.0110, 1.1967]
Family_history_MI						
No**	1	1	1	1	1	1
Yes	1.3968	4.0422	[0.0338, 1 .401]	1.3121	3.7139	[0.0152, 1 .3031]
Diabetes_mellitus						
Normal**	1	1	1	1	1	1
Pre-diabetes	-0.2006	0.8182	[-0.71814, 0.3167]	-0.2006	0.8182	[-0.7824, 0.3011]
Diabetes	0.0139	1.0139	[0.5760, 1.6040]	1.0139	2.7563	[0.5821, 1.5858]
BMI						
Normal**	1	1	1	1	1	1
Over weight	0.6115	1.8432	[-0.5479, 1.7711]	0.6115	1.8535	[-0.5976, 1.7416]
Obese(BMI>30)	0.8356	2.3061	[0.3401, 2.0113]	0.8521	2.3445	[0.3901, 2.1253]
Cholesterol_level						

Normal(< 200)**	1	1	1	1	1	1
High(>200mg/dL)	0.5427	1.7206	[0.4385, 1.3529]	0.5492	1.7318	[0.3385, 1.3835]
Blood_pressure						
Normal**	1	1	1	1	1	1
High	0.2817	1.3254	[-0.2562, 0 .8197]	0.2817	1.3254	[-0.1562, 0 .9301]
Uncontrollable	0.0682	1.0705	[-0.5816, 0 .7181]	0.0701	1.0726	[0.3823, 1.8024]
Follow_up_site						
ICU**	1	1	1	1	1	1
Medical ward	0.9091	2.4820	[0.3991, 1.4191]	0.9110	2.4868	[0.4281, 1.4369]
Age(continuous)	-0.0155	0.9846	[0.0201, 1.0317]	-0.0157	0.9844	[1.0105, 1.5116]
Weight(continuous)	-0.0393	0.9614	[0.0131, 1.1026]	0.0101	0.9899	[0.0193, 0.9879]
	AIC= 90.3539			AIC= 89.2839		
	Log likelihood = -28.1769			Log likelihood = -27.6214		

4.5. Model Diagnostics

4.5.1. Cox - Snell Residuals Plot

Based on the Cox-Snell residuals plot, statistics were applied to assess the adequacy of the model fitted to the dataset. If the model is adequate, the hazard function should closely follow the 45-degree line at the origin, except for very large values of time, where some deviations may occur. In our case, the plot of the Cox-Snell residuals for the Weibull regression model provided a straight line through the origin, in contrast to the other time-to-discharge regression models. This indicates that, when comparing the zigzag line with the reference line, the Weibull proportional hazards model fitted the data well compared to the other baseline distributions (exponential and Gompertz).

Overall, the Weibull model appears to provide the best fit to the data based on the Cox-Snell residuals plot, as its line is the closest to the 45-degree reference line. The exponential model seems to have the poorest fit, while the Gompertz model falls in between the Weibull and exponential models in terms of goodness of fit.

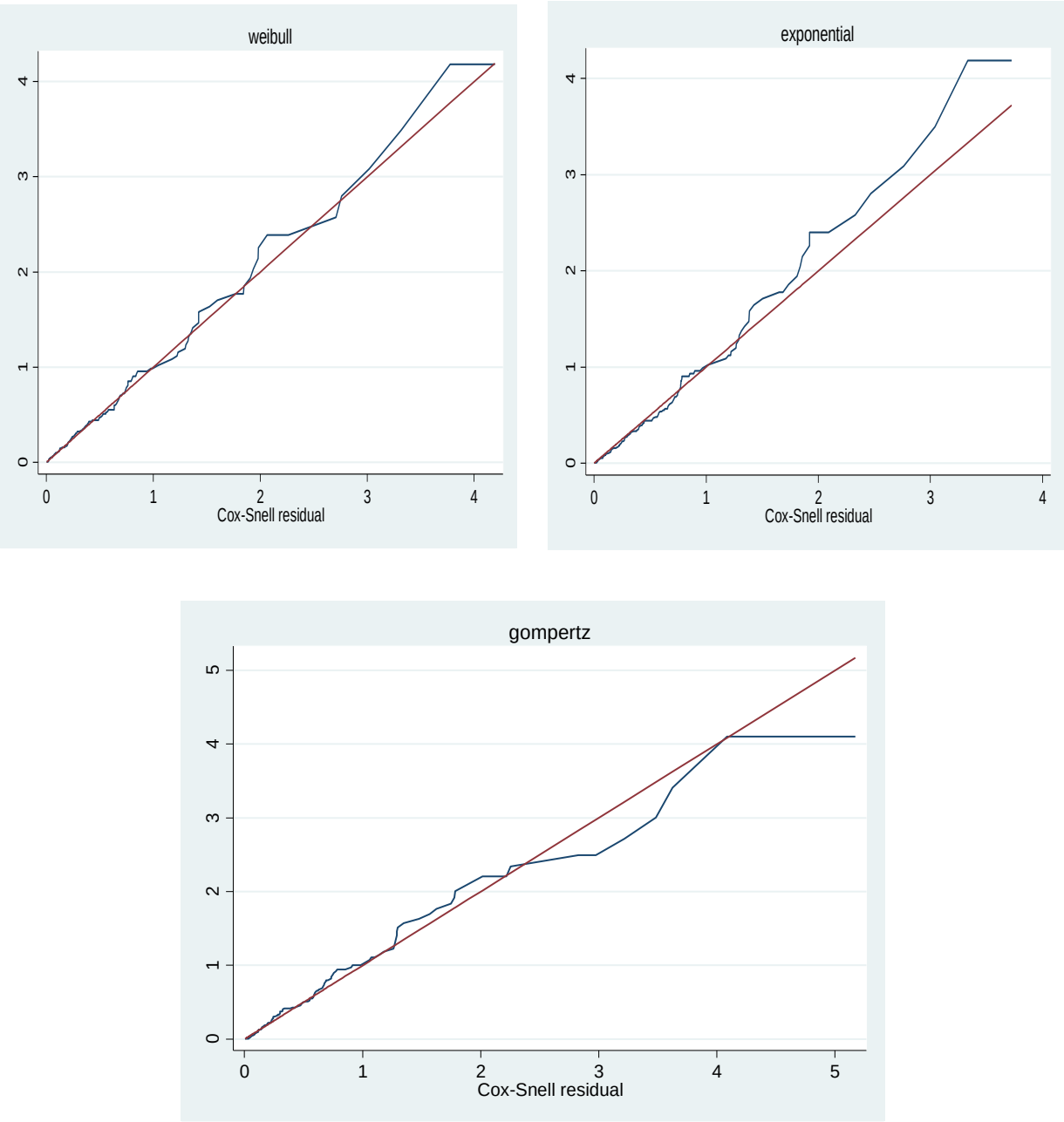


Figure 8: Cox-Snell residuals plot for diagnosing the three baseline distributions of the model.

4.5.2. Quantile – Quantile (Q-Q) plot

A quantile-quantile (Q-Q) plot is created to check whether the Cox PH model provides an adequate fit to the data across different groups of respondents. We will graphically assess the model's adequacy by comparing the significantly different factors related to MI by follow-up

site, gender, BMI, cholesterol level, diabetes mellitus, blood pressure, family history of MI, age, and weight. As shown in Figures 9 and 1.4 in the appendix, the plots appear to be approximately linear for the covariates of follow-up site, gender, BMI, cholesterol level, diabetes mellitus, blood pressure, family history of MI, age, and weight. Therefore, the Cox PH model appears to adequately describe the risk factors associated with MI.

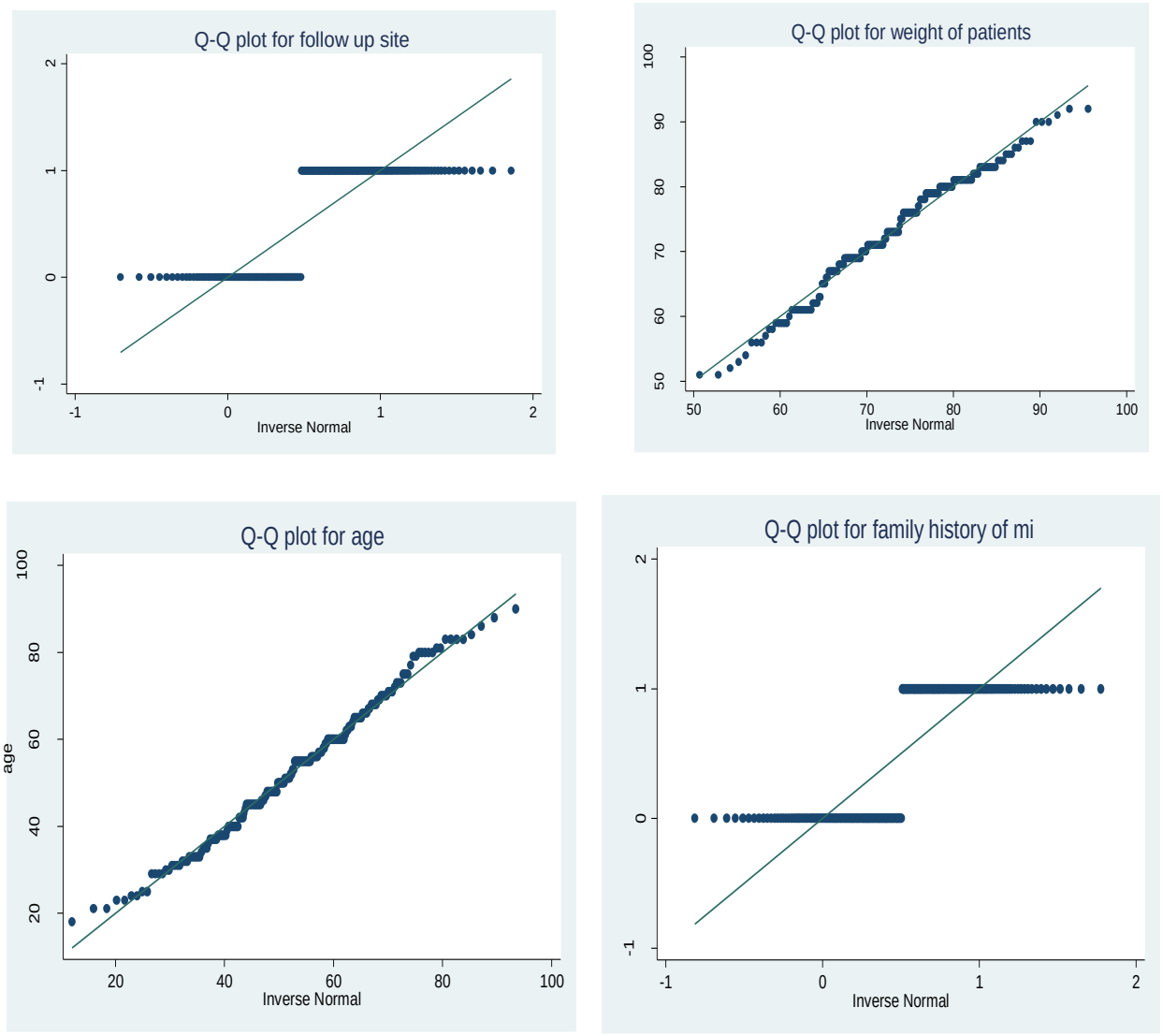


Figure 9: Quantile – Quantile (Q-Q) plot for checking model adequacy

4.6. Discussion

The main objective of this study was shared frailty model in survival analysis of MI patients is to investigate and gain a comprehensive understanding of the factors influencing the time it takes for MI patients to be discharged from Ayder. The covariates which were included in this study were gender, residence, marital status, educational level, smoking status, family history of MI, diabetes mellitus, BMI, alcohol use, cholesterol level, blood pressure, stress, follow up site, age(in complete year), and weight. The outcome variable interest was also the survival of time-to-discharge of patients with MI.

The univariate analysis related to gender, smoking status – ex-smoker, weight, age, BMI - Obese (BMI>30), diabetes mellitus - diabetes, alcohol use, blood pressure - uncontrollable(Above 189/110 mmHg), cholesterol level, follow-up site; were statistically associated with the survival time of the MI at 5% level of significant. All the significant covariates in univariate analysis were included in all multivariate analysis of the Cox PH model and comparison was done within the models using AIC where the model having minimum AIC value is selected to be the best. Weibull Cox PH was found to be the best over Exponential and Gompertz models based on AIC values. The aim of frailty model is to account the heterogeneity of subjects among different follow up centers and the clustering (group) effect was significant (p-value = 0.0031) in Weibull-gamma shared frailty model. There was heterogeneity among the follow-up centers regarding the survival of MI patients in the study area.

This study on MI patients gives an insight into survival and its risk factors at ACSRH. The study found gender, BMI(Obese(BMI>30)), cholesterol level(High(>200mg/dL)), follow up center, blood pressure(uncontrollable), diabetes mellitus(Diabetes), family history of MI, age, and weight as statistically significant and strong predictors of the MI survival.

In this study, out of the total MI patients, 142 (68.93%) were males, with 65 (45.77%) being censored and 77 (54.23%) experiencing an event. Additionally, out of 64 (31.07%) females, 29 (45.31%) were censored, and 35 (54.69%) experienced an event. The survivorship function estimates the continued constantly survival time from the beginning for around 7 days of follow up and then gradual declining for 7 days (until 28) days follow up. The minimum value of the survivor function was not zero since the largest observed time was a censored observation.

The results from the fitted Cox PH shared frailty model revealed that increasing age significantly affects the time-to-discharge for MI patients. Specifically, older age is associated with decreased survival time to discharge, indicating an increased risk for these patients. Even though the studies conducted with logistic regression model analysis, this finding is similar with Fikreyesus & Bahta.

Additionally, the study identified follow-up center as a significant predictive factor for time-to-discharge. MI patients followed in the medical ward exhibited higher risks compared to those followed in the ICU. Another important finding of this study was the association between BMI and MI risk. Patients categorized as obese had a higher risk compared to those with a normal and overweight BMI. Furthermore, the analysis indicated that increasing weight is correlated with decreased survival time to discharge.

The study also highlighted the statistical significance of blood pressure as a variable. Patients with uncontrollable blood pressure faced greater risks than those with normal blood pressure levels. Diabetes mellitus was identified as another significant factor affecting MI outcomes. Patients with diabetes had a higher risk compared to those without the condition. Similarly, high cholesterol levels (greater than 200 mg/dL) were associated with increased risk among MI patients compared to those with normal cholesterol levels.

Gender was found to be statistically significant as well; with male patients exhibiting a higher risk than female patients, this finding is similar with Benjamin.

Finally, a family history of MI was identified as a significant factor; patients with a family history of MI had a higher risk compared to those without such a history.

5. Conclusion and Recommendations

5.1. Conclusion

The study utilized a dataset on risk factors for MI in the ACSRH to conduct a shared frailty model in survival analysis for MI patients. The aim was to investigate and gain a comprehensive understanding of the factors influencing the time taken for MI patients to be discharged from Ayder. Out of the 206 patients, 54% experienced the event, while 46% did not experience it by the end of the follow-up period.

To model the risk factors for MI, both the Cox PH model and shared frailty model with different baseline distributions were employed. Among the various baseline distributions tested, the Weibull-gamma shared frailty model exhibited a better fit to the study's data compared to other parametric, shared frailty, and Cox PH models, as indicated by lower AIC values. The selected model revealed a frailty (group) effect on the time to discharge across the follow-up centers, highlighting the presence of heterogeneity and underscoring the necessity of the frailty model.

The results of the Weibull Cox PH and Weibull-Gamma shared frailty models indicated that follow-up site at the medical ward, being obese(BMI > 30), age(in years), weight, diabetes mellitus, family history of MI, uncontrolled blood pressure, high cholesterol levels, and male gender were statistically significant predictors. However, other predictor variables were not found to be statistically significant.

After identifying the significant and insignificant predictors, the study attempted to observe the results of both the Weibull PH and Weibull-Gamma shared frailty models. Although these models were quite similar, they were not identical. To compare the efficiency of the models, the AIC was utilized. From Table 11, it is evident that the Weibull-Gamma shared frailty model had the lowest AIC value (89.2839) compared to the Weibull PH model (AIC = 90.3539). This indicates that the Weibull-Gamma shared frailty model provided a better fit to the survival data of time-to-discharge for MI patients compared to the Weibull PH model, which did not account for the latent group effects.

Furthermore, when examining the estimated coefficients of the covariates, it was observed that they were altered with the inclusion of the frailty component. The confidence intervals for the risk factors were slightly narrower for the Weibull-Gamma shared frailty model. Additionally, the variance of the random effect (frailty) was found to be significant at a 5% level of significance, indicating that the parametric shared frailty model better fitted the given dataset compared to the PH model. In conclusion, the study determined that the Weibull-Gamma shared frailty model is preferable over the Weibull PH model for analyzing the risk factors in the MI dataset. How the Goodness of fit of the baseline distributions are well were also checked by means of graphical methods such Cox-Snell residual plots and Q-Q plot showed in figure 8 and 9 told us Weibull baseline distribution was better to explain the MI dataset.

5.2. Recommendations

Based on the study's findings, the following recommendations are made for the concerned bodies.

- Implement tailored interventions and monitoring protocols for high-risk MI patients, considering factors like age, obesity, diabetes, high cholesterol level, and family history of MI.
- Enhance care coordination and resource allocation for MI patients, with closer monitoring and support for those treated in medical wards versus intensive care.
- Utilize the Weibull-Gamma shared frailty model for analyzing MI risk factors, as it better accounts for heterogeneity across follow-up centers compared to the Cox PH model.
- Advocate for policy changes that integrate the identified risk factors into clinical guidelines for comprehensive MI patient management.
- Encourage multidisciplinary collaboration among healthcare providers to deliver holistic, patient-centered care addressing both medical and psychosocial needs of MI patients.
- Utilizing survival analysis to identify patients at high risk of prolonged hospitalization and implement targeted interventions to accelerate their recovery and discharge.

Reference

Anderson, J. L. (2013). 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: A report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *Circulation*, *127*(4), 362–425. https://doi.org/10.1161/CIR.0b013e3182742cf6

Awoke, A., Awoke, T., Alemu, S., & Megabiaw, B. (2012). Prevalence and associated factors of hypertension among adults in Gondar, Northwest Ethiopia: A community based cross-sectional study. *BMC Cardiovascular Disorders*, *12*(1), 1. https://doi.org/10.1186/1471-2261-12-113

Ben-Assuli, O., Ramon-Gonen, R., Heart, T., Jacobi, A., & Klempfner, R. (2023). Utilizing shared frailty with the Cox proportional hazards regression: Post discharge survival analysis of CHF patients. *Journal of Biomedical Informatics*, *140*(February), 104340. https://doi.org/10.1016/j.jbi.2023.104340

Benjamin, E. J., Virani, S. S., Callaway, C. W., Chamberlain, A. M., Chang, A. R., Cheng, S., Chiuve, S. E., Cushman, M., Delling, F. N., Deo, R., De Ferranti, S. D., Ferguson, J. F., Fornage, M., Gillespie, C., Isasi, C. R., Jiménez, M. C., Jordan, L. C., Judd, S. E., Lackland, D., … Muntner, P. (2018). Heart disease and stroke statistics - 2018 update: A report from the American Heart Association. In *Circulation* (Vol. 137, Issue 12). https://doi.org/10.1161/CIR.0000000000000558

Bennett. (2020). *sim @ doi.org*. https://doi.org/10.1002/sim.4780020223

Cochran, W. G. (1997). *Cochran_1977_Sampling_Techniques_Third_E.pdf* (pp. 76–78).

https://www.academia.edu/29684662/Cochran_1977_Sampling_Techniques_Third_Edition

Cox D. R. (1972). *Collett_1_3.PDF*.

Fikreyesus, Y., & Bahta, Y. (1989). Myocardial infarction in the Tikur Anbessa Teaching Hospital. A five-year review. *Ethiopian Medical Journal*, *27*(2), 55–61. http://europepmc.org/abstract/MED/2714260

Giday, S. A. (2011). *Strengthening Family Planning with Community- Based Nutrition Interventions in Ethiopia :* (Issue June 2011).

Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016. (2017). *Lancet (London, England)*, *390*(10100), 1151–1210. https://doi.org/10.1016/S0140-6736(17)32152-9

Guelbert, C. (2022). Myocardial Infarction. *Cardiac Vascular Nurse Certification Review*, 119–126. https://doi.org/10.1891/9780826173249.0015

Guo, J. J., Chen, Y., Du, W., Peng, H., Wang, R., Xia, Y., Xin, P., Wigle, P. R., & Papadimitropoulos, E. A. (2016). Antithrombotic Therapy and Direct Medical Costs in Patients with Acute Coronary Syndrome in Shanghai, China. *Value in Health Regional Issues*, *9*, 93–98. https://doi.org/10.1016/j.vhri.2016.01.001

Gutierrez, R. G. (2002). Parametric Frailty and Shared Frailty Survival Models. *The Stata Journal: Promoting Communications on Statistics and Stata*, *2*(1), 22–44. https://doi.org/10.1177/1536867x0200200102

Hailu, A., Gidey, K., Ebrahim, M. M., Berhane, Y., Gebrehawaria, T., Hailemariam, T., Negash, A., Mesele, H., Desta, T., Tsegay, H., Alemayohu, M. A., & Bayray, A. (2023). Patterns of Medical Admissions and Predictors of Mortality in Ayder Comprehensive Specialized Hospital, Northern Ethiopia: A

Prospective Observational Study. *International Journal of General Medicine*, *16*(January), 243–257. https://doi.org/10.2147/IJGM.S385578

Hollenberg, S. M., & Nathan, S. (2005). Myocardial infarction. *Surgical Critical Care, Second Edition*, *6*(4), 367–384. https://doi.org/10.29309/tpmj/2017.24.09.814

Hosmer, D., Lemeshow, S., & May, S. (2008). Applied Survival Analysis: Regression Modeling of Time to Event Data. *Journal of the American Statistical Association*, *95*. https://doi.org/10.2307/2669422

Kelvin L. Moeschberger. (2003). *Techniques for Censored and Second Edition*.

Kelvin L. Moeschberger, 2003. (2013). *An evaluation of the Cox-Snell residuals*. 23. https://uu.diva-portal.org/smash/get/diva2:826234/FULLTEXT01.pdf

Liu, X. (2013). Survival Models on Unobserved Heterogeneity and their Applications in Analyzing Large-scale Survey Data. *Journal of Biometrics & Biostatistics*, *05*(02). https://doi.org/10.4172/2155-6180.1000191

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S. Y., Alvarado, M., Anderson, H. R., Anderson, L. M., Andrews, K. G., Atkinson, C., Baddour, L. M., Barker-Collo, S., Bartels, D. H., Bell, M. L., … Memish, Z. A. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet (London, England)*, *380*(9859), 2095–2128. https://doi.org/10.1016/S0140-6736(12)61728-0

Mathers, C. D., Boerma, T., & Ma Fat, D. (2009). Global and regional causes of death. *British Medical Bulletin*, *92*(1), 7–32. https://doi.org/10.1093/bmb/ldp028

Misganaw Dr., A., Mariam, D. H., Ali, A., & Araya, T. (2014). Epidemiology of major non-communicable diseases in Ethiopia: A systematic review. *Journal of Health, Population and Nutrition*, *32*(1), 1–13.

Muluneh, A. T., Haileamlak, A., Tessema, F., Alemseged, F., Woldemichael, K., Asefa, M., Mamo, Y., Tamiru, S., Abebe, G., Deribew, A., & Abebe, M. (2012). Population based survey of chronic non-communicable diseases at gilgel gibe field research center, southwest ethiopia. *Ethiopian Journal of Health Sciences*, *22*(S), 7–18. http://www.ncbi.nlm.nih.gov/pubmed/23319836%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3542738

N'Guetta, R., Yao, H., Ekou, A., N'Cho-Mottoh, M. P., Angoran, I., Tano, M., Konin, C., Coulibaly, I., Anzouan-Kacou, J. B., Seka, R., & Adoh, A. M. (2016). [Prevalence and characteristics of acute coronary syndromes in a sub-Saharan Africa population]. *Annales de cardiologie et d'angeiologie*, *65*(2), 59–63. https://doi.org/10.1016/j.ancard.2016.01.001

Ndubuisi, N. E. (2021). *Noncommunicable Diseases Prevention In Low- and Middle-Income Countries : An Overview of Health in All Policies (HiAP)*. https://doi.org/10.1177/0046958020927885

Oakes, C. &. (1984). *Analysis of survival data*. *January 2002*.

Sawyer, S. (2003). The Greenwood and Exponential Greenwood Confidence Intervals in Survival Analysis. *Health (San Francisco)*, *i*, 2–5.

Schoenfeld, D. (1982a). Partial residuals for the proportional hazards regression model. *Biometrika*, *69*(1), 239–241. https://doi.org/10.1093/biomet/69.1.239

Schoenfeld, D. (1982b). Partial residuals for the proportional hazards regression model. *Biometrika*, *69*(1), 239–241. https://doi.org/10.1093/biomet/69.1.239

Shashu, B. A. (2021). The Management of Coronary Artery Disease in Ethiopia: Emphasis on Revascularization. *Ethiopian Journal of Health Sciences*, *31*(2), 439–454. https://doi.org/10.4314/ejhs.v31i2.27

Shavadia, J., Yonga, G., & Otieno, H. (2012). A prospective review of acute coronary syndromes in an urban hospital in sub-Saharan Africa. *Cardiovascular Journal of Africa*, *23*(6), 318–321. https://doi.org/10.5830/CVJA-2012-002

Shiferaw, F., Letebo, M., Misganaw, A., Feleke, Y., Gelibo, T., Getachew, T., Defar, A., Assefa, A., Bekele, A., Amenu, K., Teklie, H., Tadele, T., Taye, G., Getnet, M., Gonfa, G., Bekele, A., Kebede, T., Yadeta, D., GebreMichael, M., … Tadesse, Y. (2018). Non-communicable Diseases in Ethiopia: Disease burden, gaps in health care delivery and strategic directions. *Ethiopian Journal of Health Development*, *32*(3).

Snell, D. R. C. and E. J. (1968). *A General Definition of Residuals Author (s): D . R . Cox and E . J . Snell Published by : Wiley for the Royal Statistical Society Stable URL : https://www.jstor.org/stable/2984505*. *30*(2), 248–275.

Tesfay, F. H., Zorbas, C., Alston, L., Backholer, K., Bowe, S. J., & Bennett, C. M. (2022). Prevalence of chronic non-communicable diseases in Ethiopia: A systematic review and meta-analysis of evidence. *Frontiers in Public Health*, *10*. https://doi.org/10.3389/fpubh.2022.936482

Therneau, T. M., & Grambsch, P. M. (2013). *Modeling Survival Data: Extending the Cox Model*. Springer New York. https://books.google.com.et/books?id=oj0mBQAAQBAJ

Tilahun, F. (2014). Magnitude and Associated Factors of Cutaneous Leishmaniasis; in Mekelle City, Ayder Referral Hospital, Tigray, Northern

Ethiopia, 2014. *Clinical Medicine Research*, *3*(6), 189. https://doi.org/10.11648/j.cmr.20140306.16

van de Vijver, S., Akinyi, H., Oti, S., Olajide, A., Agyemang, C., Aboderin, I., & Kyobutungi, C. (2013). Status report on hypertension in Africa - Consultative review for the 6th Session of the African Union Conference of Ministers of Health on NCD's. *Pan African Medical Journal*, *16*, 1–17. https://doi.org/10.11604/pamj.2013.16.38.3100

Vaupel et al., 1979; Clayton, 1978; Oakes, 1982. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika*, *86*(2), 381–393. https://doi.org/10.1093/biomet/86.2.381

Vaupel, J. W. (1979). *The-impact-of-heterogeneity-in-individual-frailty @ doi.org*. https://doi.org/10.2307/2061224

WHO. (2010). Global status report on noncommunicable diseases. *World Health Organization*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Wienke, A. (2010a). Frailty models in survival analysis. *Frailty Models in Survival Analysis*, *July 2010*, 1–299. https://doi.org/10.1201/9781420073911

Wienke, A. (2010b). Frailty models in survival analysis. *Frailty Models in Survival Analysis*, *March*, 1–299. https://doi.org/10.1201/9781420073911

Yadeta, D., Guteta, S., Alemayehu, B., Mekonnen, D., Gedlu, E., Benti, H., Tesfaye, H., Berhane, S., Hailu, A., Luel, A., Hailu, T., Daniel, W., Haileamlak, A., Gudina, E. K., Negeri, G., Mekonnen, D., Woubeshet, K., Egeno, T., Lemma, K., … Tefera, E. (2017). Spectrum of cardiovascular diseases in six main referral hospitals of Ethiopia. *Heart Asia*, *9*(2), 1–5. https://doi.org/10.1136/heartasia-2016-010829

Appendix

Table 1.1: The long rank test for survival curves of time to discharge for MI by follow up site.

Covariates	Observed events	Expected events
Follow up site		
ICU	87	42.96
Medical ward	119	163.04
chi2(1) = 70.80	Pr>chi2 = 0.0000	

Table 1.2: The long rank test for survival curves of time to discharge for MI by BMI

Covariates	Observed events	Expected events
BMI		
normal/underweight(bmi<25)	8	11.76
over weight (bmi 25-29.9)	88	88.95
obese(bmi>30)	110	105.29
chi2(2) = 1.75	Pr>chi2 = 0.4176	

Table 1.3: The long rank test for survival curves of time to discharge for MI by Alcohol use

Covariates	Observed events	Expected events
Alcohol use		
No	78	90.94
Yes	128	115.06
chi2(1) = 4.04	Pr>chi2 = 0.0443	

Table 1.4: The long rank test for survival curves of time to discharge for MI by Diabetes mellitus

Covariates	Observed events	Expected events
Diabetes mellitus		
Normal	8	11.76
Prediabetes	88	88.95
Diabetes	110	105.29
chi2(2) = 7.08	Pr>chi2 = 0.0291	

Table 1.5: The long rank test for survival curves of time to discharge for MI by smoking status

Covariates	Observed events	Expected events
Smoking status		
Never	96	97.35
Ex-smoker	60	51.89
Current smoker	50	56.76
chi2(2) = 2.52	Pr>chi2 = 0.2832	

Table 1.6: The long rank test for survival curves of time to discharge for MI by gender

Covariates	Observed events	Expected events
Gender		
Male	64	63.70
Female	142	142.30
chi2(1) = 0.00	Pr>chi2 = 0.9604	

Table 1.7: The long rank test for survival curves of time to discharge for MI by stress

Covariates	Observed events	Expected events
stress		
acute stress	105	101.93
chronic stress	101	104.07
chi2(1) = 0.22	Pr>chi2 = 0.6394	

Table 1.8: The long rank test for survival curves of time to discharge for MI by blood pressure

Covariates	Observed events	Expected events
Blood pressure		
Normal	33	38.86
High	111	111.66
Uncontrollable	62	55.48
chi2(2) = 2.00	Pr>chi2 = 0.3673	

Table 1.9: The long rank test for survival curves of time to discharge for MI by marital status

Covariates	Observed events	Expected events
Marital status		
Single	18	18.27
Married	102	103.78
Widowed	54	55.66
Divorced	32	28.29
chi2(3) = 0.70	Pr>chi2 = 0.8739	

Table 1.10: The long rank test for survival curves of time to discharge for MI by place of residence.

Covariates	Observed events	Expected events
Residence		
Rural	51	55.78
Urban	155	150.22
chi2(1) = 0.69	Pr>chi2 = 0.4075	

Table 1.11: The long rank test for survival curves of time to discharge for MI by education level

Covariates	Observed events	Expected events
Education level		
No education	102	94.96
Primary	61	72.43
Secondary and above	43	38.61
chi2(2) = 3.48	Pr>chi2 = 0.1758	

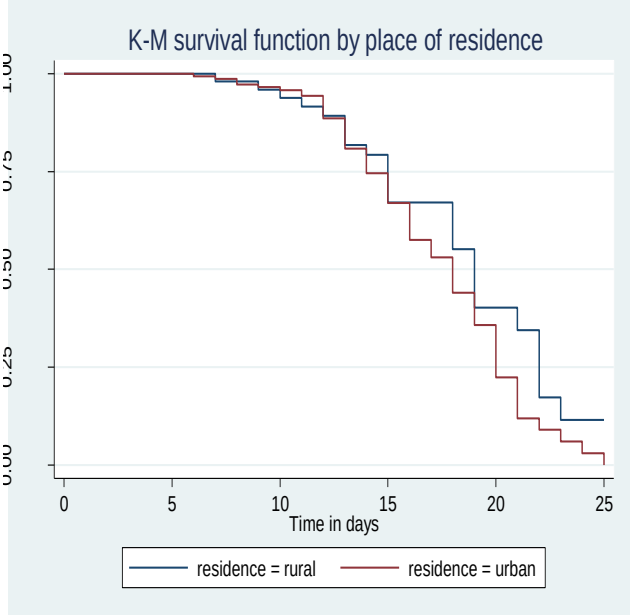
Table 1.12: The long rank test for survival curves of time to discharge for MI by cholesterol level

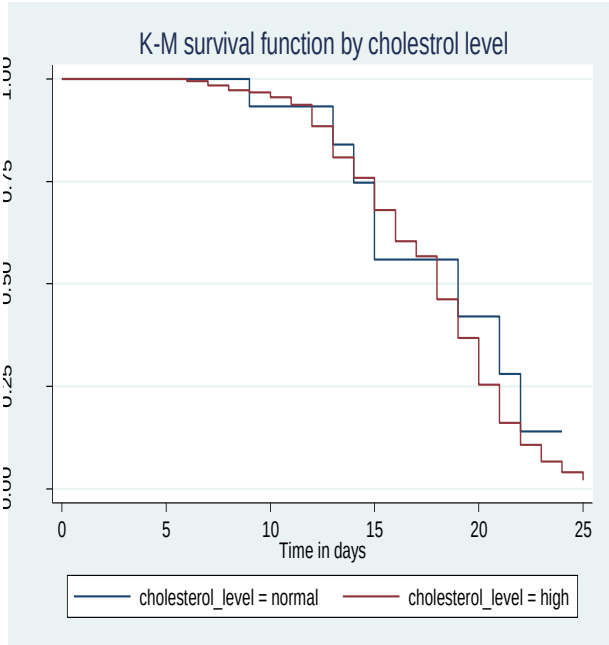
Covariates	Observed events	Expected events
Cholesterol level		
Normal	15	16.55
High	191	189.45
chi2(1) = 0.19	Pr>chi2 = 0.6598	

Table 1.13: The long rank test for survival curves of time to discharge for MI by family history of MI.

Covariates	Observed events	Expected events
Family history of MI		
No	107	102.34
Yes	99	103.66

chi2(1) = 0.51 Pr>chi2 = 0.4734





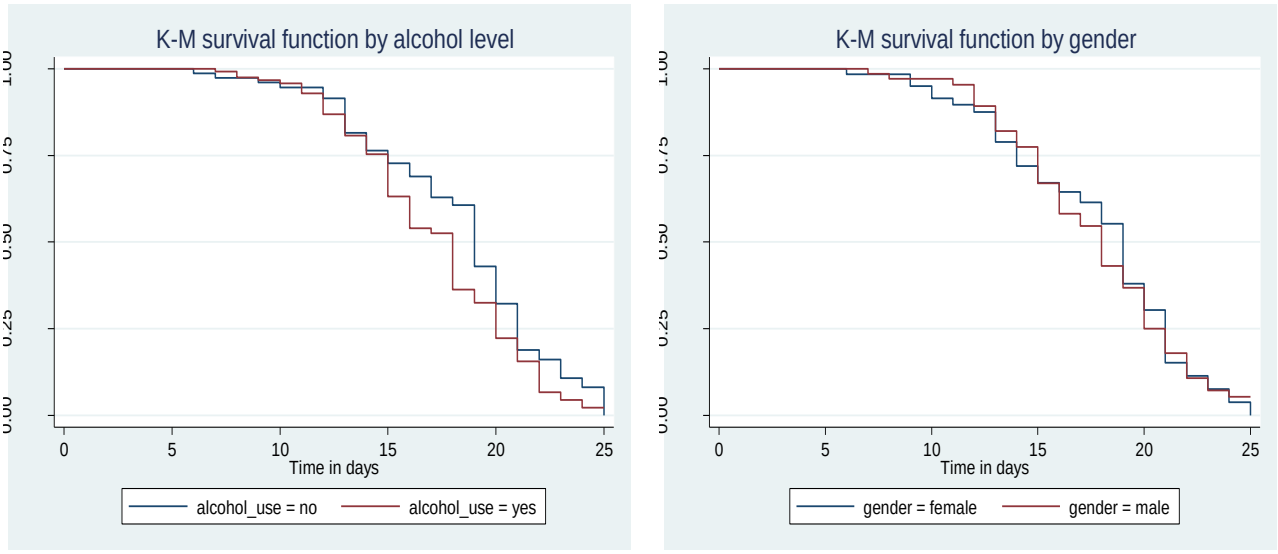
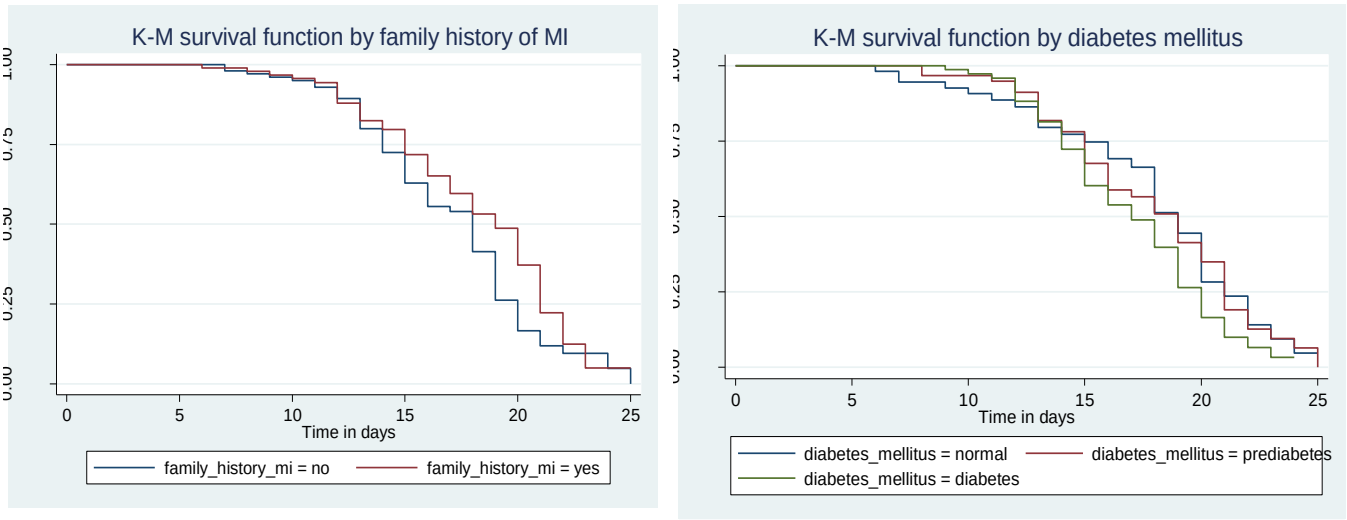


Figure 1.1: Kaplan Meier plots of time-to-discharge of MI patients by place of residence, cholesterol level, alcohol use, and gender



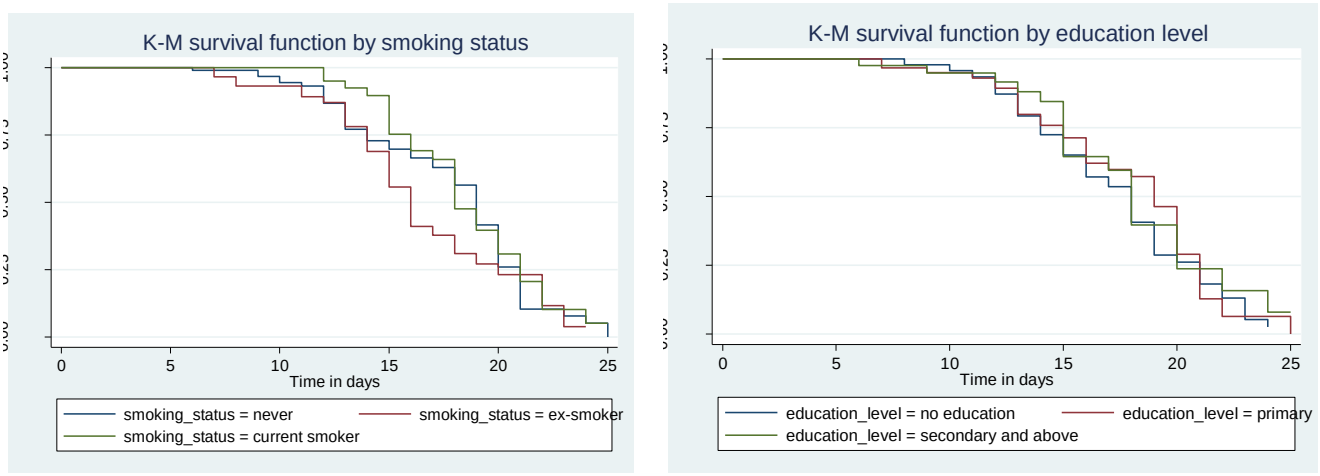


Figure 1.2: Kaplan Meier plots of time-to-discharge of MI patients by family history of MI, diabetes mellitus status, smoking status, and education level.

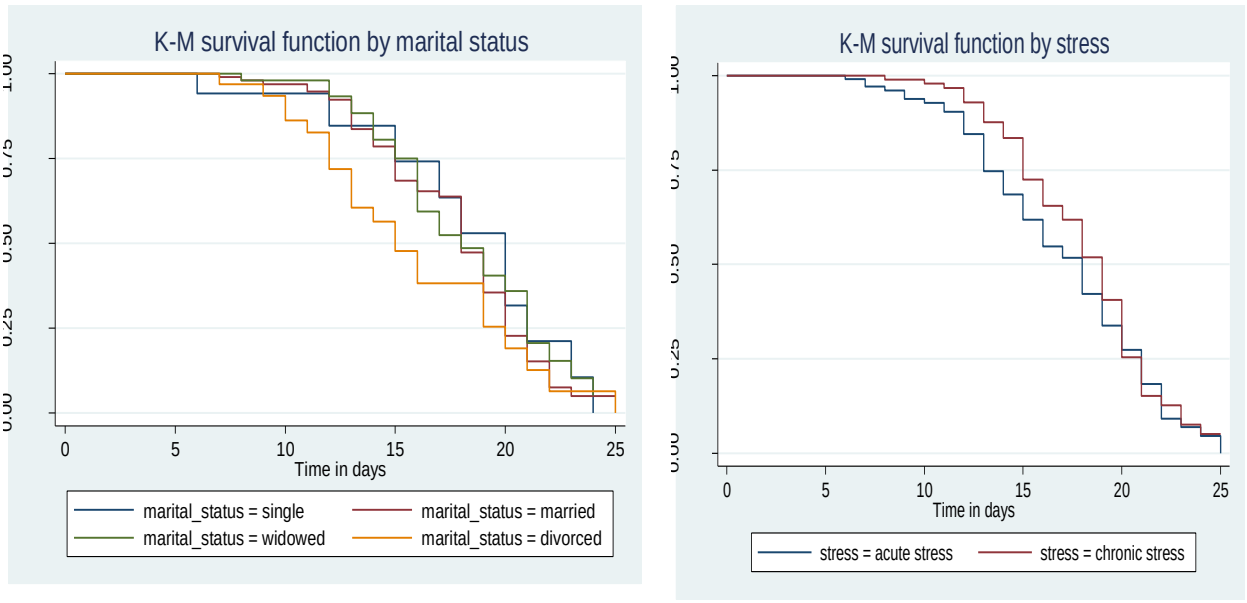
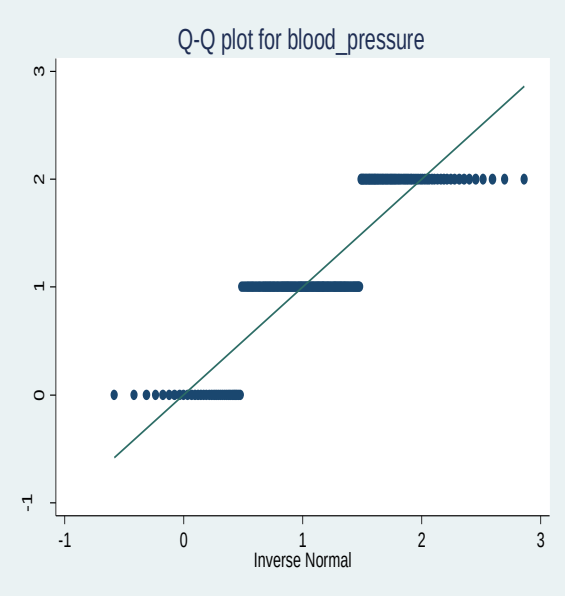
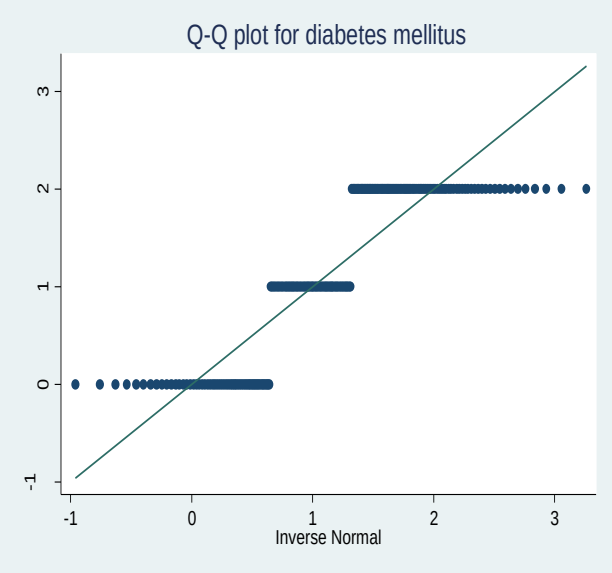
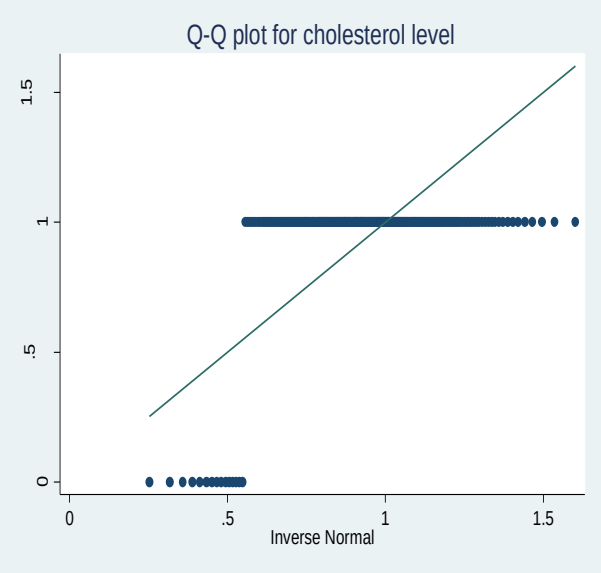


Figure 1.3: Kaplan Meier plots of time-to-discharge of MI patients by marital status, and stress


Q-Q plot for blood_pressure
-1
0
1
2
3
Inverse Normal


Q-Q plot for gender
-.5
0
.5
1
1.5
2
Inverse Normal


Q-Q plot for diabetes mellitus
-1
0
1
2
3
Inverse Normal


Q-Q plot for cholesterol level
0
.5
1
1.5
Inverse Normal

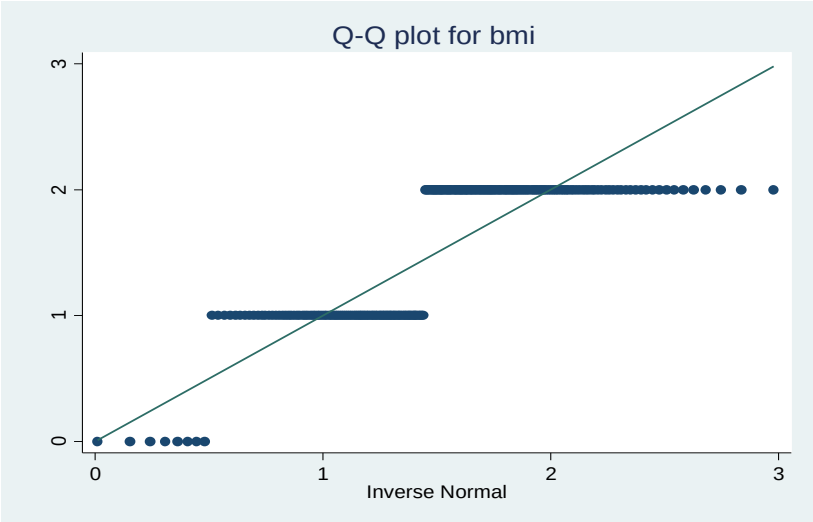


Figure 1.4: Quantile Quantile plot for checking model adequacy