DETERMINANTS OF CARDIOVASCULAR DISEASE COMPLICATIONS AMONG HYPERTENSIVE PATIENTS: AT JIMMA UNIVERSITY TEACHING HOSPITAL

By: Habtamu Ayele

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

Jimma, Ethiopia

August 2020

DETERMINANTS OF CARDIOVASCULAR DISEASE COMPLICATIONS AMONG HYPERTENSIVE PATIENTS AT JIMMA UNIVERSITY TEACHING HOSPITAL: ANALYSIS BASED ON SURVIVAL MODELS

Master of Science in Biostatistics Thesis

By: Habtamu Ayele

Advisors: Akalu Banbeta (Asst. Prof)

Abiyot Negash (MSc)

Jimma, Ethiopia

August 2020

SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

MSc THESIS APPROVAL SHEET

We, the undersigned, member of the Board of Examiners of the final open defense by **Habtamu Ayele** have read and evaluated his/her thesis entitled **"Determinants of Cardiovascular Disease Complications among Hypertensive Patients: Analysis Based On Survival Models"** and examined the candidate. This is therefore to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree Master of Science in **Biostatistics** (_____).

Jaleta Abdisa (MSc)		28/08/2020
Name of the Chairperson	Signature	Date
Akalu Banbeta (Asst. Prof)		28/08/2020
Name of Major Advisor	Signature	Date
Dechasa Bedada (Asst. Prof)		28/08/2020
Name of the Internal Examiner	Signature	Date
Dr. Shibru Temesgen		28/08/2020
Name of the External Examiner	Signature	Date

ACKNOWLEDGEMENT

I wish to express my sincere appreciation to my advisor Akalu Banbeta (PhD Candidate) who has the substance of a genius: for his convincingly guided, kind support, advice, constructive comments and encouraged me to be professional and do the right thing on my thesis.

Similarly, I would like to pay my special regards to my co-advisor Mr. Abiyot Negash (MSc) for his guidance, advice and kind support from proposal construction to the final thesis.

I heartily thank Madda Walabu University for helping me the financial support for this study and Jimma University for all their facilitates, by giving me advisor, co-advisor and instructors.

I wish to acknowledge the support and great love of my family and my fiance (Kena Amansisa). They kept me going on and this work would not have been possible without their input. Also I would pray to the Lord of life and death to stay him into heaven, my father Ayele Necho, who I have lost this year due to this complications.

I would like to thank Dr. Ahmed Mohamed, the general chief clinical director of Jimma University Teaching Hospital for his permission to collect the data from the hospital. I am especially grateful to Mr. Ifa Bane (Nurse), Mr. Habtamu Abebe (MSc in Bio-statistics) and Miss. Melat Wodaje (MSc in Epidemiologist) for their guidance and helps during the data collection.

Moreover, I would like to thank God, the creator and Lord of all, for his protection, helpful and faithfulness.

ABSTRACT

Background: Cardiovascular diseases complications are a group of disorders of the heart and blood vessels. Globally an estimated 17.9 million peoples were died from cardiovascular disease. This covers 31% of all global deaths, which three quarters of are take place in developing countries. Hypertension is the major cause for increasing cardiovascular disease complications. It's influence is high with additional risk factors. The aim of this study is to determine the major risk factors of Cardiovascular disease complications among hypertensive patients at Jimma University Teaching Hospital.

Method: Retrospective cohort type study was conducted to hypertensive patients in Jimma University Teaching Hospital, 2017. Using total of 343 hypertensive patients, who fulfilled all inclusion criteria, semi-parametric and several parametric survival models were applied to analyze the determinants of cardiovascular disease complications. By assessing the overall goodness of fitted models, log-logistic accelerated failure time model which can fit the data well and had smallest akaike information criterion value were selected as the appropriate fit model.

Result: About 40.23% of hypertensive patients were experienced cardiovascular disease complications with minimum, maximum and median time of 2, 35 and 28 months respectively. According to result from the chosen log-logistic model, potential subsets of covariates; age, residence place, proteinuria, systolic, diastolic and combination of both systolic and diastolic blood pressures were significant prognostic covariates for cardiovascular disease complications of hypertensive patients.

Conclusion: Log-logistic acceleration failure time model were chosen for determinants of cardiovascular disease complications among hypertensive patients at Jimma University Teaching Hospital, 2017. More than 50% of hypertensive patients were from urban, had diabetes mellitus, proteinuria and hyperlipidemia. Of these additional risk factors, proteinuria and urban residence place had greatest impact on cardiovascular disease complications, through shortening the expected time.

Key words: Cardiovascular disease, Hypertension, Semi-parametric, Parametric

ACRONYMS

AFT:	Acceleration Failure Time
AIC:	Akaike Information Criterion
ANOVA:	Analysis of Variance
BP:	Blood Pressure
CAD:	Coronary Artery Diseases
CHD:	Coronary heart disease
CV :	Cardiovascular
CVD:	Cardiovascular Diseases
CO:	cardiac outputs
DBP:	Diastolic Blood Pressure
DM:	Diabetes Mellitus
JUTH:	Jimma University Teaching Hospital
KM :	Kaplan-Meier estimates
LMIC:	Low and middle-income countries
MLE:	Maximum Likelihood Method
mm Hg:	Millimetre of mercury
NCD:	Non-Communicable Diseases
PH :	Proportional Hazards
PL:	Partial Likelihood
PPH:	Parametric Proportional Hazard
QQ:	Quantile - Quantile
RHD:	Rheumatic Heart Disease
SBP:	Systolic Blood Pressure
WHF:	World Heart Federation
WHO:	World Health Organization

Contents

ACKNOWLEDGEMENT **i**

ABSTRACT **iii**

ACRONYMS **iv**

1 Introduction **1**

1.1 Background of the Study 1

1.2 Statements of the Problem 4

1.3 Objectives of the Study 5

1.3.1 General Objective 5

1.3.2 Specific Objectives 5

1.4 Significance of the Study 6

1.5 Limitations of the Study 6

2 Literature Review **7**

2.1 Development of Cardiovascular Disease Complications 7

2.1.1 Definitions 7

2.1.2 Cardiovascular Disease Complications for Hypertensive Patients 8

2.1.3 Impact of Cardiovascular Disease 9

2.1.4 The Major Risk Factors 10

2.2 Survival Analysis 13

3 Methodology **15**

3.1 Study Area 15

3.2 Study Design 15

3.3 Target Population 15

3.4 The Data Set 15

3.5 Inclusive and Exclusive Criteria 16

3.6 Variables in the Study 17

3.6.1 Dependent Variable . . . 17
3.6.2 Independent Variables . . . 17
3.7 Statistical Methods . . . 19
3.7.1 The Survival Models . . . 19
3.7.1.1 Descriptive Approaches of Survival Analysis . . . 19
3.7.1.2 Semi-parametric Cox-regression Model . . . 20
3.7.1.3 Parametric Survival Models . . . 21
3.7.2 Comparison of Models . . . 23
3.7.3 Model Diagnostics . . . 25
3.7.3.1 Checking for Proportional Hazard Assumption . . . 25
3.7.3.2 Parametric Baseline (Graphical Checking) . . . 26
3.7.3.3 Cox-Snell Residuals . . . 26
3.7.3.4 The Quantile-Quantile Plot . . . 28

4 Results **29**
4.1 Descriptive Statistics . . . 29
4.1.1 Survival Function for Different Groups of Categorical Covariates . . . 30
4.2 Statistical Analysis . . . 31
4.2.1 Model Choice and Variable Selection . . . 32
4.3 Model Diagnostics . . . 36
4.3.1 Cox-Snell Residual Plots . . . 36
4.3.2 Plots of Parametric Baselines for Several Parametric Models . . . 37
4.3.3 The Quantile-Quantile Plot . . . 39
4.4 Discussion . . . 40

5 Conclusions and Recommendations **43**
5.1 Conclusions . . . 43
5.2 Recommendations . . . 43

References **45**

APPENDIXES **53**

Appendix 1: Log-logistic Univariable Analysis 53

Appendix 2. Multivariable Analysis for all the Candidate Models 54

List of Tables

3.1 Description of independent variables in this study. 18

3.2 Summary table of the selected parametric survival distributions 23

3.3 Linear $\hat{H}$ of parametric models in some function of time. 26

4.1 Descriptive statistics of variables in this study. 30

4.2 AIC and ρ of parametric survival models 33

4.3 Multi-variable analysis using the log-logistic parametric survival model 34

List of Figures

0.1 ii
2.1 Types of CVD highly caused by hypertension 9
4.1 Hazard and survival function plot of log-logistic model at mean of covariates . . . 36
4.2 Cox-Snell residuals from the fitted cox-regression, exponential, weibull, log-normal and log-logistic models 37
4.3 Graphical evaluation for parametric models 38
4.4 Q-Q plot of hypertensive patients data using survival times of Residence place . . . 39

1 Introduction

1.1 Background of the Study

Cardiovascular diseases (CVD) are a group of disorders of the heart and blood vessels. It is a broad, umbrella term used to describe all conditions affecting the heart and circulatory system, including coronary heart disease, stroke, heart attack, aortic, cerebrovascular, rheumatic heart disease and congenital heart disease [1]. CVD vary throughout the world in type and distribution especially between the developed and developing countries. In developed countries, coronary artery disease is a leading cardiovascular problem. And, in developing country like sub-Saharan Africa hypertension, rheumatic heart disease (RHD), cardiomyopathy and congenital heart disease are reported to be the major cardiovascular disease conditions [2, 3].

Globally more peoples die due to CVD than any other diseases. In 2013, an estimated 17.3 million and in 2016, 17.9 million peoples are died from CVD, representing 31% of all global deaths. From these deaths, 85% are due to heart attack and stroke. Heart attacks and strokes are usually acute events and are mainly caused by a blockage that prevents blood from flowing to the heart or brain. Over three quarters of CVD deaths take place in low- and middle-income countries. For example, out of the 38 million deaths due to noncommunicable diseases (NCD) in 2012, 82% are in low- and middle-income countries, and 37% are caused by CVD [4].

Cardiovascular disease is a unique double burden challenge for the whole of Africa [5]. Traditionally, in Africa a communicable diseases like HIV-AIDS are accounted for the greatest burden of mortality [6]. This burden is now fast shifting towards chronic non-communicable diseases, which majority of is CVD. This is why being termed as a "double burden of disease" in Africa [7].

National representative surveys of NCDs and their risk factors in Ethiopia are not available. However, there are some studies which have estimates on death due to CVD . For example, from a report of national strategic action plan for prevention of NCD (2014-2016), in Ethiopia there are about 34% of annual death rate due to NCDs, of which 15% are accounted for CVD complica-

tions. Specifically, population-based 'STEPS' survey conducted in Jimma (south-west Ethiopia) from 2008 – 2009 showed a substantial burden of NCDs and their risk factors within the community. In this, the prevalence of cardiovascular diseases and hypertension in Jimma was reported as 3% and 2.6% respectively [8]. Health complications related to cardiovascular disease, especially for the urban communities in Ethiopia is becoming the major health threat. This threat is mostly manifested with obesity, high blood pressure, dyslipidemia, heart diseases and diabetes [9]. One type of cardiovascular disease complication, stroke is the one of top ten causes of death in Ethiopia and it is strongly caused by hypertension [10].

There are many risk factors that contributed to the development of CVD complications. Among all risk factors for CVD, the major cause for leading CVD complication is hypertension [11]. Hypertension, or abnormally high blood pressure is a worldwide serious public health problem. Most people with hypertension, which also known as a "silent killer" are in unaware of the problem because it may have no warning signs or symptoms. When it is not treated properly, sufferers can develop CVD complications such as strokes, heart attacks, kidney failure or others. Hypertension can independently contributed as the risk of cardiovascular events, but its impact is greatly influenced by associated risk factors. Majority of hypertensive patients have additional risk factors for cardiovascular disease (CVD) complications. For example, Framingham Heart Study displays as about 17% of women and 19% of men with hypertension had this as their only CVD risk factor, while 32% of women and 30% of men with this hypertension had 3 or more additional risk factors [12]. Various risk factors considered in this thesis were age, residence place, diabetes mellitus, protienuria, hyperlipidemia, systolic BP, diastolic BP and others. This study were identified the risk factors which highly associated to CVD complications of hypertensive patients at JUTH.

Globally, an estimated **one billion** peoples are with hypertension, which more than **nine millions** are die due to this in each year [13]. Until recently, hypertension was mainly associated with more affluent regions of the world. However, the condition is increasingly emerging in low and middle-income countries (LMICs) [14]. For example, Africa has the highest rate of high blood pressure in the world. Nearly, about 30% of African adults were estimated to have hypertension in 2014, and

in 2016 it was raised to 46% of total adults. Especially, in Sub-Saharan Africa the cause is highly increasing and about 150 million peoples will be estimated to have hypertension in 2025 [15]. In Ethiopia there is lack of representative surveys on hypertension, but some meta-analysis study estimated as the prevalence of the disease is 19.6% [16].

The data to be used in this study contains censored information and skewed time. Therefore, the usual regression model will not help to perform the analysis of the data [17]. Hereby, we have considered various techniques in survival analysis, which is the statistical tool used to analyze CVD complication of hypertensive patients data. There are a some studies conducted focusing on determinants of cardiovascular disease complications for hypertensive patients. Most of them were used Cox-regression, which is semi-parametric survival model for analyzing the data [18, 19, 20]. To decide the outperformed model fit for CVD complications of hypertensive patients, some decision process was done among semiparametric and several parametric models. The goal of this study is, analyzing the major risk factors that will lead hypertensive patients to CVD complication at Jimma University Teaching Hospital (JUTH).

1.2 Statements of the Problem

About half (50%) of burdens related to CVD are linked to the complication from raised blood pressure or hypertension [21]. For example, lowering BP reduces risk of stroke, myocardial infraction and heart failure about 40%, 20%-25% and 50% respectively [22]. However, BP management of hypertensive patients who are at high risk for cardiovascular disease complication is cost-effective [23]. Since, the risk of CVD is high (75%) in low- and middle-income countries, it is better to identify the significant risk factors which leads patients to CVD.

Non communicable diseases and their related risk factors are growing and becoming a double burden in Ethiopia [24] . In the country, representative surveys on NCDs and their risk factors are not available [8]. However, some hospital based studies in the country showed that the prevalence of death due to cardiovascular disease was high. For example, a study from Addis Ababa investigating cause of death using verbal autopsy showed that 24% deaths were due to CVD [18]. As a result, the impact of CVD and associated risk factors, specially hypertension is the current issue.

In 2017 about 52.7% of hypertensive patients in JUTH had uncontrolled hypertension [20]. This means, hypertension has been a major health problem in JUTH. Since, uncontrolled hypertension is the high leading cause of CVD complications, the current situation calls for intervention in view of cardiovascular disease complication at the specified place.

There are studies conducted on CVD complications. However, they were not focused specifically on hypertensive patients. Therefore, to fill this gap and all the stated above, the researcher conduct this study which identifies a risk factor that calls patients to CVD complications.

Researchers in medical sciences often prefer semi-parametric to the parametric models. Similarly, most of CVD related studies conducted were used semi-parametric cox-regression model. This is due to it involves minimal assumptions and requires less model checking efforts. However, parametric models may give more precise and efficient estimates of quantities of interest than semi-parametric model [25]. Using some decision criteria created by Melinda [26], the choice of

appropriate survival model from different types of semi- and parametric models was done. This can fill the gap on ambiguous choice of appropriate survival model for further studies on CVD complication of hypertensive patients.

This study were addressed on the following research questions.

- What is the estimated median and rate of time patients develop CVD?
- What are the major risk factors which leads CVD complications for hypertensive patients?
- Which survival model is appropriate or preferable for CVD among hypertensive patients data?

1.3 Objectives of the Study

1.3.1 General Objective

The general objective of the study is to determine the major risk factors of CVD complications among hypertensive patients at Jimma University Teaching Hospital, 2017.

1.3.2 Specific Objectives

The specific objectives of the study are:

- To estimate the median and rate of time at patients develop CVD.
- To make comparison between groups of patients to know their significant difference on developing CVD.
- To select the appropriate survival model that fit the data well.
- To identify the significant risk factors which leads patients CVD complication.

1.4 Significance of the Study

This thesis identified and gave discussions on the risk factors which plays critical roll on developing CVD complication at Jimma University Teaching Hospital in 2017. Also, groups of patients who are at higher risks to develop the disease were identified. This have more advantage for health professionals (physicians) in order to give good treatment for identified stakeholders on identified risk factors. Solved complexity of the disease after treatment have its own advantage for the stakeholders and family to save them selves economically and from disease complexity. This is useful for the country (government) to minimize number of deaths of adults in working age. Also, the thesis will helps academician or readers, on the selection of appropriate survival model(s) for further study on determinants of CVD among hypertensive patients data.

1.5 Limitations of the Study

This study have also its own limitations. Within variability of changes in age, systolic and diastolic blood pressure covariates through time were not addressed. Because, the study used only the baseline values. Patients with incomplete information was excluded. Due to this, the number of participants became minimized, and some necessary information may be lost.

2 Literature Review

2.1 Development of Cardiovascular Disease Complications

2.1.1 Definitions

Cardiovascular system consists of the heart, which is an anatomical pump, with its intricate conduits (arteries, veins, and capillaries) that traverse the whole human body carrying blood. Heart is muscular organ weighing between 250 - 350 grams located obliquely in the mediastinum. It functions as pump supplying blood to the body and accepting it in return for transmission to the pulmonary circuit for gas exchange [27]. The pumping action of heart usually maintains a balance between cardiac outputs (CO) and venous return. CO is amount of blood pumped out by each ventricle in one minute. The normal adult blood volume is 5 liters and it usually passes through the heart once a minute [28].

The cardiac cycle refers to events that occur during one heart beat and is split into ventricular systole (contraction) and diastole (relaxation). A normal heart rate is approximately 72 beats per minute, and cardiac cycle spreads over 0.8 seconds [29]. When heart's functions become compromised, this is known as cardiovascular disease. Cardiovascular disease (CVD) is a general term used to describe disorders that can affect the heart (cardio) and/or the body's system of blood vessels (vascular).

Most cardiovascular diseases reflects chronic conditions, conditions that develop or persist over a long period of time. Some peoples are born with conditions that predispose them to CVD complication specially, heart disease and stroke [30]. However, most people who develop CVD do so because of various risk factors such as diabetes, hypertension, dyslipidemia, central obesity, increased inflammation, and procoagulant state [11]. Those risks for CVD cause problems because they lead to atherosclerosis. Atherosclerosis is the narrowing and thickening of arteries and develops CVD for years without causing symptoms. The narrowing and thickening of the arteries are due to the deposition of fatty material, cholesterol and other substances in the walls of blood vessels. This can happen in any part of the body. For example around the heart, it is known as

coronary artery disease and in the legs, it is known as peripheral arterial disease [30].

2.1.2 Cardiovascular Disease Complications for Hypertensive Patients

Hypertension is a state of elevated systemic blood pressure that causes marked increment of cardiovascular risk. It is one of the major risk factors for CVD complications. According to Mcfarlane, S.I., Banerji, M. and Sowers, J.R., from various risk factors for CVD which includes diabetes, hypertension, dyslipidemia, central obesity, increased inflammation, and procoagulant state, hypertension is the major cause for increased CVD particularly in the high-risk populations including those with diabetes, minority population, elderly, and stroke victims [11].

CVD includes coronary heart diseases (CHD), stroke, heart failure, hypertensive heart disease, rheumatic heart disease, cardiomyopathy, abnormal heart rhythms, congenital heart disease, valvular heart disease, carditis, aortic aneurysms, peripheral artery disease, thromboembolic disease and venous thrombosis [31, 32]. In addition to other factors, about 54% of stroke and 47% of ischaemic heart disease (Coronary heart disease) are highly caused by hypertension [21]. Also, they account for 80% of deaths of CVD in males and 75% of deaths of CVD in females [31].

Coronary heart disease (CHD) involves the reduction of blood flow to the heart muscle due to build-up of plaque in the arteries of the heart. A common symptom is chest pain or discomfort which may travel into the shoulder, arm, back, neck, or jaw [33]. The other name for CHD is coronary artery disease (CAD) or ischemic heart disease (IHD) [34]. Strokes are brain attach, occurred when the blood supply to the brain becomes blocked . This may result from either blockage (ischaemic stroke) or rupture of a blood vessel (haemorrhagic stroke) [35]. Trouble speaking, paralysis, problems seeing in one or both eyes are the signs and symptoms happened after stroke begins [36].

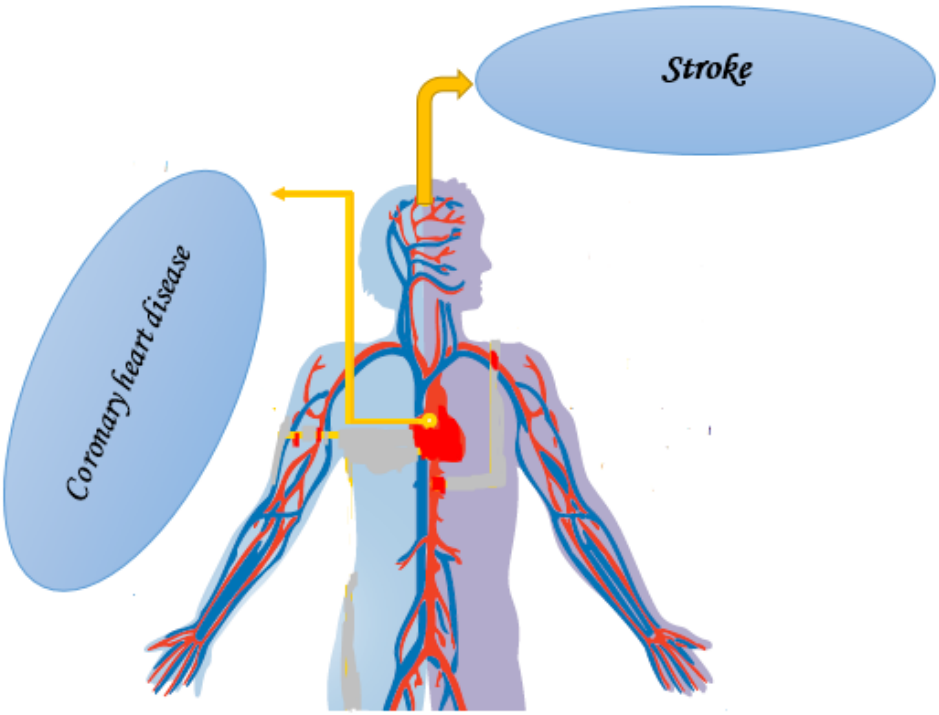


Figure 2.1: Types of CVD highly caused by hypertension

2.1.3 Impact of Cardiovascular Disease

Cardiovascular disease can affects low and middle-income countries in many ways [37]. The costs for CVD are, to the healthcare system and to the national economy. For example, in South Africa, 2% to 3% of the country's gross national income, or roughly 25% of South African healthcare expenditures, was devoted to the direct treatment of CVD [38]. There is also high proportion of CVD burden occurs earlier among adults of working age in developing countries. This can lead large impact on the country's economic availability [39].

Cardiovascular diseases (CVDs), including ischemic heart disease, stroke, and hypertensive heart disease are Ethiopia's second leading cause of premature death and disability (6,458 per 100,000). This creates high health care cost, and major disease and economic burden for the country [40]. CVD is not only an important public health problem, but it will also have a big economic impact

as a significant proportion of the productive population becomes chronically ill or die, leaving their families in poverty [41]. In order to prevent and control hypertension in the population, the country needs policies developed and implemented through a multi-sectoral approach involving the Ministries of Health and other sectors.

2.1.4 The Major Risk Factors

Patients with hypertension often have other major risk factors for cardiovascular disease (CVD). Little is known, and the followings are some of them [19, 42, 43].

Age

Older populations are at great risk to CVD complications. According to study from journal of cardiovascular development and disease on cardiovascular risks associated with gender and aging, age is an independent risk factor for CVD [44] . Another, retrospective cohort study conducted by Melaku Tadege, which used Cox-PH model showed that the risk of cardiovascular disease increased as the age of hypertensive patient increases [19]

Gender

While it may have long been seen as a man's disease, the risk of CVD in women has been underestimated, and symptoms may go unrecognized, complicating diagnosis and treatment [45]. Though CVD risk factors are shared by men and women, some may be more prevalent and/or more significant for one sex. For example, studies in Kenya and South Africa showed that females were more at risk than male [46].

Residence

Urbanization is increasing the burden of and CVD for hypertensive patients [43]. The relationship between urbanization and risk of CVD has been previously analyzed, revealing seemingly paradoxical results. While some studies suggest that with increased income, a Westernized diet and lifestyle would lead to increased risk of CVD [47], others conclude that increased resources available would lead to better access to preventative interventions and primary care [48].

Diabetes mellitus (DM)

Having diabetes, a condition that causes high levels of glucose in the blood, is a risk factors for cardiovascular disease. High glucose levels can damage the artery walls and make the buildup of fatty deposits (atheroma) more likely. Prospective study Wilson designed using relative risk method concluded, the great effects of DM in women relative to men for all cardiovascular events except congestive heart failure [49]. Using ANOVA and chi-square (χ^2) statistical analysis, Michael reported, CVD rate for hypertensive patients with history of diabetes was more than double to that of those without diabetes [50].

Proteinuria

Proteinuria is the presence of abnormal quantities of protein in the urine, possibly indicating damage to the kidneys. It is marker of renal and cardiovascular (CV) disease in hypertensive populations. Proteinuria is considered a risk factor for CVD and mortality in patients with hypertension. The study used hazard rate of logistic regression method and concluded effective BP control and proteinuria reduction are associated with more favorable Cardiovascular and renal outcomes [51].

According to study from Microalbuminuria in clinical practice, the presence of proteinuria in patients with treated essential hypertension varies between 4% and 16% in different series of treated hypertensive patients [52]. The INSIGHT Study also assessed the role of proteinuria as a risk factor in essential hypertension. The presence of proteinuria at baseline turned out to be a very potent predictor for the development of cardiovascular events and death in patients with essential hypertension and one or more associated cardiovascular risk factors [53].

Using Multiple Drug Type (Multi-drug Use)

Most patients with hypertension requires two or more anti-hypertensive drugs to achieve effective blood pressure control, and patients with hypertension may have one or more comorbidities, such as type 2 diabetes mellitus, that necessitate the use of additional medications [54, 55]. Multi-drug has a detrimental effect on adherence because many patients do not understand their complex regimens and have difficulty organizing their schedules to accommodate these regimens [56].

Most effective method for improving compliance with anti-hypertensive regimens was to simplify dosing [57]. Simplified dosing regimens resulted in increasing adherence by between 8%, 19.6% [58].

Baseline Complication

This describes if patients had CVD complications during the started time of the study. Existing cardiovascular disease or a previous cardiovascular event, such as a heart attack or stroke, is the strongest predictor of a future cardiovascular event [59].

Smoking Status

Smoking significantly increases the chance of developing CVD. It damages and narrows the arteries, making angina pectoris and heart attack more likely. Angina pectoris is condition characterized by pain or discomfort in the center of the chest, caused by heart muscle not getting enough blood [60, 61]. Nicotine also makes the heart beat faster and increases blood pressure, meaning the heart has to work harder to pump blood around the body [62]. The study which used percentage and adjusted relative risk for analysis concluded, the risk of cardiovascular disease in smokers is proportional to the number of cigarettes smoked and how deeply the smoker inhales, and it is apparently greater for women than men [63].

Hyperlipidemia

Hyperlipidemia is defined as elevations of fasting total cholesterol concentration which may or may not be associated with elevated triglyceride concentration. Several cardiovascular risk factors aggregate in patients with hypertension. For example, hyperlipidemia, glucose intolerance, and hyperinsulinemia are common in hypertension [64]. Diagnosing and managing hyperlipidemia as a way to prevent cardiovascular disease (CVD) is a common activity for primary care physicians [65].

Systolic and Diastolic Blood Pressures

Systolic blood pressure (SBP) and diastolic blood pressure (DBP) are the most commonly reported BP measures in clinical practice and research studies because they are well established cardiovas-

cular disease (CVD) risk factors and can be directly estimated. When considered separately, as meta-analysis of Prospective Studies Collaboration using age and cause specific death rate concluded, higher SBP and higher DBP are associated with increased CVD risk [66]. Study which used cox-PH and Joint model and conducted for comparing systolic, diastolic blood pressure, pulse and mean arterial pressure concluded, both SBP and DBP are independently associated to CVD events [67]. In contrast, in some studies, DBP has not been associated with CVD events after adjustment for SBP, especially in older populations [68].

2.2 Survival Analysis

Research on the statistical analysis of survival data from related individuals began in the mid-1970s with papers by Clayton, Holt and Prentice [69, 70]. For estimating conditional survival functions, non-parametric estimators can be preferred to parametric and semi-parametric estimators due to relaxed assumptions that enable estimation [71]. The Kaplan–Meier estimator which seems to have been first proposed by Bohmer is a non-parametric estimator [72]. It is used to estimate the survival distribution function from censored data. It was, however, lost sight of by later researchers and not investigated further until the important paper by Kaplan & Meier appeared [73].

Researchers in medical sciences often prefer semi-parametric models to the parametric ones because of their minimal assumptions [74]. A key reason to use this model is that even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest and adjusted survival curves can be obtained for a wide variety of data or we can say that cox-proportional hazard (PH) model is robust and will closely approximate the results of the correct parametric model [75].

A number of studies have been conducted to make a choice among several semi- and parametric survival regression methods, of which some proposed semi-parametric models as the most appropriate model. For example, Mohamad Amin Pourhoseingholi conducted retrospective study on gastric patients to make comparison between semi-parametric Cox and parametric models. The study made comparison in both univariate and multivariate analysis. According to the study, in

multivariate analysis, Cox-regression was concluded as the best model [76].

The parametric models give more precise estimates of the quantities of interest than semi-parametric or non-parametric models [25, 77]. Zhu, the study on the application of Weibull model proved the better performance of parametric Weibull model than semi-parametric Cox-PH model [78]. Wang used akaike information criterion (AIC) to compare the efficiency of both log-normal and Cox proportional hazard models. This indicated that, the log-normal model is a useful statistical model rather than Cox-PH model [79].

Ghorbani Gholiabad used gastric patients data to compare parametric and semi-parametric model. They used AIC criteria and concluded that weibull, log normal and log-logistic models were better than Cox model, and among all parametric models, the Log-normal was the best one [80]. Zare represented the parametric model outperformed in comparison to Cox model and among all parametric models, the Exponential and Weibull models were the best ones [81]. Georgousopoulou concluded, the worst performance of semi-parametric Cox proportional hazard model compared to parametric survival models with best performance of Weibull distribution from the parameters [82].

3 Methodology

3.1 Study Area

The study was conducted at Jimma University Teaching Hospital, which is one of the oldest public hospitals in Ethiopia. It was established in 1930 E.C by Italian invaders for service of their soldiers. After the withdrawal of the colonial occupants, it has been governed by the name of "Ras Desta Damtew Hospital" and later "Jimma Hospital" during Dergue regime and currently JUTH. The hospital is expected to provide health services for more than 20 million persons living in south western Ethiopia with 800 bedded [83].

The hospital is located in Jimma city and, Jimma is the largest city in South-western of Oromia Region at a distance of 355.2 Km from Addis Ababa, the capital city of Ethiopia. It has latitude and longitude of $7^0$40'N $36^0$50'E. Jimma has relatively cool tropical monsoon climate. The temperatures are in comfortable range, with the daily mean staying between 20^0C and 25^0C year-round. Jimma is the birth place of coffee and it represents about 11.8% of Ethiopians total coffee.

3.2 Study Design

The study is a retrospective cohort type because; it has investigated the time to CVD complication of hypertensive patients, since January, 2017 to December, 2019.

3.3 Target Population

The target population of this study is hypertensive patients admitted in 2017 and who were under the follow-up for anti-hypertension treatment at JUTH from 2017 to 2019.

3.4 The Data Set

Data set used for this study is survival data. It was secondary data and collected from Patients individual chart for investigation of time to CVD complications. The chart contains distinctive identification number for each patient. Patients are those who were visited JUTH and admitted

as hypertensive patients in 2017 (from January to December). For the data collection, one health professional and two experienced data collectors under the supervision of the researcher were contributed. A common characteristic of survival data is censoring, truncation, or combination of censoring and truncation.

In essence, censoring occurs when we have some information about individual survival time, but we don't know the survival time exactly. There are various categories of censoring, such as right censoring, left censoring and interval censoring. Right censoring is the most common form of censoring, where a subject's followup time terminates before the outcome of interest is observed. An observation is said to be left censored if individuals developed the event of interest prior to the beginning of the study. And, interval censoring is when event of interest occurs within an interval of time without the knowledge of when exactly happened. The data used for this study is right censored data and the censored was:

- Patients who died because of hypertension or other disease before developing CVD complications, since 2017 to 2019.
- Patients who dropout or referred to other hospital.
- At end time of the study, patients who were in the study but not developed the disease.

3.5 Inclusive and Exclusive Criteria

All patients who have 140/90 mmHg or more measures of BP, admitted in 2017 and visited the hospital for at least two times were included in the study [84]. The pregnant women were excluded, because they may not have similar factors with other non-pregnant. Also, patients who visited the hospital for only one time were excluded from the study.

3.6 Variables in the Study

3.6.1 Dependent Variable

Response variable in this study is the time (months) to CVD complications for hypertensive patients from patients started to followup for treatment. The event is CVD status with Yes (**1**) or No (**0**) response, when CVD developed or censored respectively.

Starting and End Time of Study

The time of the study is from 2017 to 2019, for three years or 36 months followup times. However, this study used only 2017 admitted patients. Entry of the data was considered from the admission date that patient started for treatment. The event is occurred when patients experienced CVD complications. The end time for the patients is when they developed the disease; they censored or when the study time is up in 2019. Diagnosis of CVD complications are known, by asking patients all related symptoms and using tests used to diagnose it [85].

3.6.2 Independent Variables

Covariates used as a risk factor in this study with their possible responses, codes or values were:

Table 3.1: Description of independent variables in this study.

Covariates	**Categories**	**Codes/Values**
Sex	Female	(0)
	Male	(1)
Residence place	Rural	(0)
	Urban	(1)
Diabetes mellitus	No	(0)
	Yes	(1)
Proteinuria	Absent	(0)
	present	(1)
Multi-drug use	No	(0)
	Yes	(1)
Baseline complication	No	(0)
	Yes	(1)
Hyperlipidemia	Absent	(0)
	Present	(1)
Smoking Status	No	(0)
	Yes	(1)
Age	Continuous	Years
Baseline systolic BP	Continuous	mm Hg
Baseline diastolic BP	Continuous	mm Hg

3.7 Statistical Methods

3.7.1 The Survival Models

Survival model is statistical model used for analysis of data which have survival time, censored observation and explanatory variables whose effect on the waiting time we wish to assess or control. The models examine the hazard rate, which is the conditional probability that an event occurs at a particular time interval (t). In other words, we examine how long it takes until the event of interest occurs. Generally, survival model in terms of hazard function is given by

$$h_n(t) = h_0(t)\exp(\beta' X) \tag{3.7.1}$$

Where n is a number of patients in this study, $h_0(t)$ denotes baseline hazard function at a given time t, X_n denotes covariates of CVD complications used in the model and β is regression coefficient from the models [26].

As described in the following sections this thesis used non-parametric method as a preliminary descriptive technique. The semi- and several parametric models were also applied to describe the relation between events and set of covariates. The baseline hazard function $h_0(t)$ have different assumptions in semi-parametric and parametric survival models. It is not specified in semi-parametric, and it assumed to the specific distributions in parametric survival models. Once the distribution is specified, the density function f(t), the survival function given by $S(t) = P(T > t) = \int_t^\infty f(u)du$ and hazard function h(t) have the following relationships [17].

$$h(t) = \frac{f(t)}{S(t)} = \frac{-d\ ln\ S(t)}{dt}; \qquad S(t) = \exp\left\{-\int_0^t h(u)du\right\} \tag{3.7.2}$$

3.7.1.1 Descriptive Approaches of Survival Analysis

Before proceeding with statistical procedures for inferences, it is better to apply some descriptive approaches which is used to summarize the main features of raw survival data. In this study, Kaplan-Meier (KM) for estimation survival function and log-rank test for comparison between two or more groups of categorical covariates were used. They are referred to as non-parametric

methods, because of they rely completely on empirical data without making any distributional assumptions [86]. KM estimator is sometimes referred to as the product-limit estimation, which estimates the survival function at time t. Survival function $\hat{S}(t)$ is the probability of experiencing the event (CVD complications) after time t. The median survival time, which is defined as the value at 0.5 survival function is the key estimate from the KM estimates. It is the preferred measure of central tendency when examining survival data. In the presence of censoring the median survival time is estimated by the earliest time at which the KM curve (i.e. survival estimates) falls below 0.50 [26]. The KM estimation for survival function at any particular time is given as

$$\hat{S}(t) = \prod_{i=1}^{j-1} P(T > t_{(i)} | T \geq t_{(i)}) \tag{3.7.3}$$

The log-rank test which is used for comparison of the survival curves of two or more categorical covariates also applied. Log-rank test is first proposed by Breslow, and it gives information on the significance of difference for the survival of two or more groups of patients [87].

3.7.1.2 Semi-parametric Cox-regression Model

Cox regression model is a model which describes the relation between the events, as expressed by the hazard function and a set of covariates. It needs the assumption of proportional hazards (PH) and linearity between covariates and log-hazards. The model is similar to model from equation 3.7.1 above, where the distribution of baseline hazard $h_0(t)$ is not specified. In Cox-regression model, partial likelihood (PL) method, which is introduced by Cox is used to estimate the regression coefficients [88]. Suppose that there are n number of hypertensive patients with time-to-CVD complications t_i, and that δ_i is CVD indicator, which is zero for censored and one for uncensored (experienced the disease), the partial likelihood function can be expressed in the form of

$$L_p(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(\beta' X_i)}{\sum_{l \in R(t_i)} \exp(\beta' X_l)} \right]^{\delta_i} \tag{3.7.4}$$

where, $R(t_i)$ is the risk at time t_i [17].

3.7.1.3 Parametric Survival Models

A parametric survival model is one in which survival time is assumed to follow a known distribution. It not only include assumptions about the distribution of failure or event times, but also about the relationship between the covariates and survival. Model for parametric survival models can be represented in similar form of the model at equation 3.7.1 above, where $h_0(t)$ follows some specific distributions [26]. Distributions of parametric model considered in this thesis were exponential, weibull, log-normal and log-logistic. Exponential distribution is a one-parametric distribution with constant baseline hazards λ. It is generalized to weibull distribution. Weibull distribution have two parameters λ and ρ. It allows the survival distribution of a population with increasing, decreasing, or constant risks. The scale parameter λ is reparameterized in terms of predictor variables and regression parameters. However, the parameter ρ called shape parametric from Weibull is held fixed.

Log-normal distribution is continuous probability distribution of a random variables whose logarithm is normally distributed. The general shape of the hazard rate for this model is quite similar to that of the log-logistic distribution. Regression models based on the log-normal distribution are very close to regression models based on the log-logistic model. The log-logistic distribution is the probability distribution of a random variable whose logarithm has a logistic distribution. The model has a hazard rate which is hump-shaped, that is, it increases initially and, then, decreases. It is similar in shape to the log-normal distribution, but has a more tractable form than that of the log-normal which makes it more convenient than the log-normal distribution. Log-logistic has a survival function and hazard rate that has a closed form expression, as contrasted with the log normal distribution which also has a hump-shaped hazard rate. Both are parametric AFT only [25].

Based on how covariates affect the hazard rate, exponential and weibull distributions can accommodate both the PH or accelerated failure time (AFT) assumptions. PH model is when the effect of covariates assumed as multiplicative with respect to the hazard whereas for AFT models, the effect of covariates is assumed as multiplicative with respect to survival time [77]. The parametric proportional hazard (PPH) have similar model representation with Cox-proportional hazards model.

It is also interpreted in the form of hazard ratio, and assumed proportionality of hazards. However, the baseline hazard function $h_0(t)$ in PPH is assumed to follow a known specific distributions, and coefficients are estimated using maximum likelihood (ML) method [89].

Accelerated failure (AFT) time model is one of parametric survival models that can be used as an alternative to PH model, especially to overcome the statistical problems due to the violation of PH assumption [90]. The model states that the survival function of an individual with covariate X at time t is the same as the survival function of an individual with a baseline survival function at a time $t\exp(\alpha' X)$. In other words, the model is defined by the relationship of $S(t) = S_0[t\exp(\alpha' X)]$. Using this relationship, the model at equation 3.7.1 above can be modified for AFT model as

$$h_n(t) = \exp(\alpha' X) h_0[t\exp(\alpha' X)] \tag{3.7.5}$$

where α is coefficients of covariates in the model and the factor $\exp(\alpha' X)$ is acceleration factor [25].

The effect of covariates in AFT model is directly accounted on survival times instead of the hazards rate as in the PH model. The coefficient α of this covariates are interpreted in terms of acceleration factor (ϕ), which is given by $\exp(\alpha^T X)$ and describes how a change in covariate values changes time to CVD complications from the baseline time scale. Whereas, coefficients from PPH are interpreted in terms of hazard ratio (HR), which is given by $\exp(\beta^T X)$. For coefficients of the same distribution with both parametric PH and AFT model, β and α have the following relationship.

$$\beta = -\alpha\rho \tag{3.7.6}$$

where ρ is shape parameter, and it is unit in exponential distribution [26].

The parameters, including regression coefficients in parametric survival models are estimated via maximum likelihood estimator (MLE). Suppose that there are n number of hypertensive patients with pair of $(t_i,\ \delta_i)$, $i = 1, 2, ..., n$ where δ_i is CVD indicator that takes zero for censored and one

for uncensored time-to-the disease complications t_i, the likelihood function is given as

$$L = \prod_{i=1}^{n} f_i(t_i)^{\delta_i} * S_i(t_i)^{1-\delta_i} \tag{3.7.7}$$

The parameterization of these parameters in the selected parametric survival distributions with baseline hazard $h_0(t)$ and survival functions $S_0(t)$ are given below [26].

Table 3.2: Summary table of the selected parametric survival distributions

Distribution	$h_0(t)$	$S_0(t)$	Parameterization
Exponential PH	λ	$\exp(-\lambda t)$	$\lambda = \exp(x\beta)$
Exponential AFT	λ	$\exp(-\lambda t)$	$\lambda = \exp(-x\beta)$
Weibull PH	$\rho\lambda t^{\rho-1}$	$\exp(-\lambda t^{p})$	$\lambda = \exp(x\beta)$
Weibull AFT	$\rho\lambda t^{\rho-1}$	$\exp(-\lambda t^{p})$	$\lambda = \exp(-\rho x\beta)$
Log-logistic	$\frac{\lambda\rho t^{\rho-1}}{1+\lambda t^{\rho}}$	$\{1+\lambda t^{\rho}\}^{-1}$	$\lambda = \exp(-x\beta)$
Log-normal	$\frac{-\frac{1}{2}(\ln t-\mu)^2/t\sigma\sqrt{2\pi}}{1-\Phi\{\frac{\ln t-\mu}{\sigma}\}}$	$1-\Phi\left[\frac{\ln t-\mu}{\sigma}\right]$	$\mu = (x\beta)$

Where $h_0(t)$= baseline hazard, $S_0(t)$= baseline survival, λ= scale, ρ= shape, σ= standard deviation in log-normal, PH= Proportional hazards, AFT = accelerated failure time, Φ= cumulative normal distribution and β= coefficients of covariates from the fitted model.

3.7.2 Comparison of Models

This thesis were made a choice from several semiparametric and parametric survival models. This is not a simple and the direct comparison is not suitable because, the scales of parameters in cox-model and parametric models are not similar with neither parameter estimates nor their estimated variances. This difference in scale of parameters are due to their different parametric estimation method, PLE for cox-model and MLE for parametric models . However, the two possible decision-path created by Melinda may helps to select the outperformed model. The first decision path is based on the assumption that the researcher have already undertaken a literature review and are armed with evidence based knowledge about how the process works. And the second is, trying

an alternative specifications and testing which model fits the data well by assessing the model fit. This thesis were used the second decision path, in which all models were fitted, and assessed using the overall goodness of fit model. The one way to assess if the model is adequately specified is using the diagnostic approach of the Cox-Snell residuals, via the construction of a residual plot that follows a unit exponential distribution with a hazard ratio of 1 [26]. The detail explanation of cox-Snell residual is given in section 3.7.3.3.

As parametric estimation in parametric survival model is through MLE, the scales of parameters in several parametric models are similar. In this thesis, model comparison among several parametric survival models were also done. It is well known that there is no single statistic that will definitely select the best model. However, akaike information criterion (AIC) will helps for the choice of appropriate model from several parametric models, that have a comparable covariates. For a given collection of models, AIC estimates the quality of each model, relative to other models and it estimates relative amount of information lost by a given model. The statistic is given as:

$$AIC = -2\log L + 2(p+k) \tag{3.7.8}$$

where -2log L is the -2 log-likelihood, p is number of covariates used in the model and k is some constant, k = 1 for the exponential model and k = 2 for the Weibull, log-logistic, and log-normal models [25].

Variable Selection Procedure

Variable selection process was applied to identify the potential subsets of covariates for the selected model. In this, all covariates were considered to be on an equal footing. When, individual covariate is added to or removed from the model, the new formed model respectively nests or nested in the original model. These alternative nested models formed can be compared by examining the change in the value of -2logL on adding terms into a model or deleting terms from a model. For this, the general strategy of variable selection process recommended by Collett's were applied.

In the first step, models that contain each of the variables one at time were fitted as a univariable analysis. The values of -2logL for these models are then compared with that for the null model to determine variables which significantly reduce the value of this statistic. Next, the important variable in the first step were fitted together. Consequently, variables that may cease to be important in the presence of certain covariates or not significantly increase the value of -2logL when they are omitted from the model then discarded. In the third steps, variables that were not important in the first step, and so were not under consideration in the second step were added to the model from step2 one at a time. This is because of, they may become important in the presence of others. Then, variables that reduce significantly value of -2logL were retained in the model. Finally, we determine whether interactions are needed in the model. This terms were added to the model in step 3 above, using hierarchic principle [17].

3.7.3 Model Diagnostics

3.7.3.1 Checking for Proportional Hazard Assumption

The proportional hazards assumption is so important to cox-regression that we often include it in the name (the cox-proportional hazards model). What it essentially means is that the ratio of hazard function for two individuals with different regression covariates, does not vary with time. It can be checked using statistical tests and graphical techniques. It is also required for PH specifications of parametric exponential and weibull survival models [25].

In this thesis, we have used plot of schoenfeld residuals to test the proportional hazards assumption. The residuals are essentially the observed minus expected values of the covariates at each failure time. The tests of proportional hazards assumption for each covariate were done by correlating the corresponding set of scaled Schoenfeld residuals with a suitable transformation of time, with the default being based on the KM estimate of the survival function. It produce a separate residual for each individual covariate. The plot of Schoenfeld residuals against time for any covariate should therefore not show a pattern of changing residuals for that covariate. If there is a pattern, then that covariate is time-dependent [26].

3.7.3.2 Parametric Baseline (Graphical Checking)

In this, we have focused on graphical checks for appropriateness of parametric models. This means, rejecting clearly inappropriate models, not proving that a particular parametric model is correct. The basic plot is made by estimating the cumulative hazard rate, and making plot for the models as follows [25].

Table 3.3: Linear $\hat{H}$ of parametric models in some function of time.

Model	Cumulative Hazard Rate ($\hat{H}$)	Plot
Exponential	λt	$\hat{H}$ versus t
Weibull	λt^{ϕ}	ln($\hat{H}$) versus ln(t)
Log-normal	-ln{1 - Φ[ln(t) - μ]/σ}	$\Phi^{-1}[1-\exp(-\hat{H})]$ versus ln(t)
Log-logistic	ln(1 + λt^{ϕ})	ln{ exp[$\hat{H}(t)$] - 1} versus ln(t)

λ= scale parameter, t= time, μ= an intercept from log-normal, σ= inverse of shape parameter, Φ= cumulative normal distribution, $\hat{H}$= estimated cumulative hazard

The plot of appropriate model is approximately linear. For example log-logistic distribution has a property that the ln{ exp[$\hat{H}(t)$] - 1} is linear with log of time. Where, $\hat{H}(t) = ln(1+\lambda t^{p})$ and ln{ exp[$\hat{H}(t)$] - 1} = ln(λ) + $p$ln(t) [25].

3.7.3.3 Cox-Snell Residuals

Cox-Snell residuals were plotted to asses the adequacy of all the fitted models via construction of a residual plot that follows a unit exponential distribution. When the survival function has been estimated from the model, the Cox-Snell residuals are defined as :

$$r_{Ci} = \hat{H}(T_i/Zj) \tag{3.7.9}$$

where, $\hat{H}(t_i)$ is the cumulative hazard function of the fitted model and Zj is covariates .

Cox-Snell Residuals for Cox Model

Suppose that the proportional hazards model $h(t/Z_j) = h_0(t)e^{\Sigma \beta_k Z_{jk}}$ has been fitted to the model and, assume that the data is $(T_j, \delta_j, Z_j), j = 1,2,...,n$. Where, $Z_j = (Z_{j1},...,Z_{jp})^t$ are all fixed

covariates. If the model fits the data correctly, Cox–Snell residuals follows the standard exponential distribution. This means that r_{Ci} are expected to have a mean of one, and it is given as

$$r_j = \hat{H}_0(T_j)\exp\{\hat{\beta}' Z_i\} \tag{3.7.10}$$

where $\hat{H}_0(t)$= Breslow's estimator of the baseline cumulative hazard rate, $\hat{\beta}$= estimated coefficients of covariate from cox-model [91].

To check whether the $r'_j s$ behave as a sample from a unit exponential, we compute the Nelson–Aalen estimator of the cumulative hazard rate of the $r'_j s$. If the unit exponential distribution fits the data, then, this estimator should be approximately equal to the cumulative hazard rate of the unit exponential $H_E(t)$ = t . Thus, a plot of the estimated cumulative hazard rate of the $r'_j s$, $\hat{H}_r(r_j)$, versus r_j should be a straight line through the origin with a slope of 1.

The Cox-Snell Residuals for Parametric Models

For the parametric regression problem, analogs of the residual plots described in semi-parametric can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates. The first such residual is the Cox–Snell residual that provides a check of the overall fit of the model. The Cox–Snell residual, r_j, is defined by $r_j = \hat{H}(T_j|Z_j)$, $\hat{H}$ is cumulative hazard of fitted model. If the model fits the data well, then the $r'_j s$ should have a standard (λ = 1) exponential distribution, so that a hazard plot of r_j versus the cumulative hazard of the $r'_j s$ should be a straight line with slope 1. For the parametric survival models considered in this thesis, the Cox–Snell residuals are [77]:

Models	**Cox-Snell residuals** (r_j)
Exponential:	$r_j = \hat{\lambda} t_i \exp\{\hat{\beta}' Z_i\}$
Weibull:	$r_j = \hat{\lambda} t_i^{\rho} \exp(\hat{\beta}' Z_i)$
Log-normal:	$r_j = \ln\left[1 - \Phi\left(\frac{ln(T_j) - \hat{\mu} - \hat{\gamma}' Z_j}{\hat{\sigma}}\right)\right]$
Log-logistic:	$r_j = \ln\left[\frac{1}{1+\hat{\lambda}\exp(\hat{\beta}^t Z_i) t_j^{\rho}}\right]$

3.7.3.4 The Quantile-Quantile Plot

The quantile-quantile plot or *Q-Q plot*, provides an exploratory method for assessing the validity of an accelerated failure time model for two groups of survival data. For any values of p in (0, 100) and the estimated survival function at a time $t(p)$ is $1 - \frac{p}{100}$, then the p^{th} percentile of the distribution $t(p)$ is given by

$$t(p) = S^{-1}(\frac{100-p}{100}) \tag{3.7.11}$$

Let $t_0(p)$ and $t_1(p)$ be the p^th percentiles estimated from the survivor functions of the two groups of survival data. The percentiles of the two groups may therefore be expressed as

$$t_0(p) = S_0^{-1}(\frac{100-p}{100}); \quad t_1(100) = S_1^{-1}(\frac{100-p}{100}) \tag{3.7.12}$$

where S_0 and S_1 are the survival functions in the two groups and it then follows that $S_1\{t_1(p)\}$=$S_0\{t_0(p)\}$, for all values of p. And, in acceleration failure time model the relationships of survival function between two groups are $S_1(t) = S_0(\phi t)$, which means $S_1\{t_1(p)\} = S_0\{\phi t_1(p)\}$. Now, using the above relationships we have

$$S_0\{t_0(p)\} = S_0\{\phi t_1(p)\}, \quad t_0(p) = \phi t_1(p) \tag{3.7.13}$$

If the accelerated failure time model is appropriate, a plot of the quantity of estimated $t_0(p)$ against $t_1(p)$, for suitably chosen values of p is a straight line through the origin. The slope of this line will be an estimate of the acceleration factor θ^{-1} [17].

4 Results

4.1 Descriptive Statistics

Out of all 343 hypertensive patients, 138 (40.23%) were developed cardiovascular disease complications with minimum and maximum time of 2 months and 35 months. The median time of patients developed the disease was 28 months with 95% confidence interval of (22, 23) months. For some group of patients, with no enough events of CVD complications, the upper limit of median survival time is not available. Due to the data given is skewed, this is common in survival analysis since there is no much enough events. 35 (23.49%) females and 103 (53.09%) males were experienced CVD complications with major participation about 194 (55.56%) and shorter median time of male patients. About 197 (57.43%) patients were from urban communities, and 96 (48.73%) of were experienced the disease. They have about half shorter median time, compared to patients from rural communities. Majority of patients about 232 (67.64%) had diabetes mellitus, of these 43% were experienced the disease with 23 months median time (Table 4.1).

About 184 (53.64%) proteinuria patients were participated, and approximately half of were experienced CVD complications. In comparison to patients with proteinuria (Absent), they had shorter median time. Relatively equal number of multi-drug users 168 (48.98%) and non-users 175 (51.02%) were involved, and of which 67 (39.88%) and 71 (40.57%) were developed CVD complications. Also, they have relatively similar values of median time. At baseline, 55 (16.03%) of patients were with CVD, and 39 (70.91%) of were re-experienced the disease with median time of 18 months. We observed that smoking status group is distributed with less number of smokers about 98 (28.57%) , of whom 40 (40.82%) were developed CVD complications with 22 months median time. Majority, about 93 (51.67%) of patients with hyperlipidemia participated in this study 180 (52.48%) were experienced CVD complications with median time of 20 months.

The median age of patients developed CVD complications were 65 years with 40 years and 85 years minimum and maximum respectively. Patients who experienced the disease have 170 mmHg median systolic and 110 mmHg median diastolic blood pressures (Table 4.1).

Table 4.1: Descriptive statistics of variables in this study.

For Categorical Covariates		Summary			
		No of Patients	Developed CVD(%)	Median	95%CI
Sex	Female	149 (43.44%)	35 (23.49%)	35	(33, _)
	Male	194 (56.56%)	103 (53.09%)	19	(16, 26)
Residence place	Rural	146 (42.57%)	42 (28.77%)	35	(32, _)
	Urban	197 (57.43%)	96 (48.73%)	19	(16, 26)
Diabetes Mellitus	No	111 (32.36%)	36 (32.43%)	34	(29, _)
	Yes	232 (67.64%)	102 (43.97%)	23	(19, 31)
Protienuria	Absent	159 (46.36%)	45 (28.30%)	34	(30, _)
	Present	184 (53.64%)	93 (50.54%)	19	(15, 25)
Multi-drug Use	No	175 (51.02%)	71 (40.57%)	27	(21, _)
	Yes	168 (48.98%)	67 (39.88%)	29	(21, _)
Baseline Complication	No	288 (83.97%)	99 (34.38%)	34	(26, _)
	Yes	55 (16.03%)	39 (70.91%)	18	(11, 27)
Smoking Status	No	245 (71.43%)	98 (40.00%)	29	(24, _)
	Yes	98 (28.57%)	40 (40.82%)	22	(18, _)
Hyperlipidemia	Absent	163 (47.52%)	45 (27.60%)	35	(31, _)
	Present	180 (52.48%)	93 (51.67%)	20	(16, 26)
		343	**138 (40.23%)**	**28**	**(22, 33)**
For Continuous Covariates		Descriptive			
	Mean	Median	SD	Min	Max
Age	64.22	65	9.039	40	85
Baseline SBP	171	170	10.46	140	190
Baseline DBP	109.6	110	10.12	88	137

Source: Jimma University Teaching Hospital 2017. CVD = Cardiovascular disease complications, CI= Median confidence interval, SD= Standard deviation, max= maximum and min= minimum value, SBP= systolic blood pressure, DBP= diastolic blood pressure.

4.1.1 Survival Function for Different Groups of Categorical Covariates

The aim is to compare the survival function of values in each categorical covariates, to know their significant difference on CVD complications of hypertensive patients. From the result of log-rank test (not exist here), there is significant difference between values of most covariates, except in multi-drug use and smoking. The survival function of those covariates were plotted to show visually how each values of categories are contributed to CVD complications. The plots are in

Appendix 3. The plots shows that, at all there is relatively similar patterns of survival, with rapidly descending estimated survival functions. The line for female, rural, proteinuria(Absent), hyperlipidemia(Absent) and baseline complications(No) lies above of male, urban, proteinuria(Present), hyperlipidemia(Present) and baseline complications(Yes) respectively, which indicates that they are survived longer from the disease. In terms of smoking status and multi-drug use factors, no clear difference was observed.

4.2 Statistical Analysis

Univariable and multivariable analysis was applied. In univariable analysis, the model which contains each covariates at a time were fitted to determine covariates that have the potential for being included in the multi-variable analysis. The values of -2logL for these models were compared with that for the null model. This was done for all proposed semi-parametric and parametric models, and there is similar conclusions in all. As an example, outputs from log-logistic model were drawn at (Appendix 1). Covariates like; age, sex, residence place, diabetes mellitus, proteinuria, baseline complications, hyperlipidemia, systolic blood pressure and diastolic blood pressure are significantly reduced the value of -2logL. This indicates that they have a power to be included in the multi-variable analysis. However, multi-drug use and smoking status were not significantly reduced the value of -2logL, and they were excluded from multi-variable analysis.

Similarly, multivariable analysis were also fitted using semi-parametric cox-regression and parametric model with exponential, weibull, log-logistic and log-normal distributions. R-software with the functions *coxph()* for cox-regression, *phreg()* for PPH and *survreg()* for AFT models were used. The outputs are drawn in (Appendix 2). To handle the tied of failures in cox-regression, the model were fitted using Breslow, Efron, Exact marginal and Exact partial likelihood and the model output with smallest AIC = 948.794 was selected.

At 5% level of significance, age, residence place, proteinuria and systolic blood pressure covariates are significant in all models. In proportional hazards models (both semi-and parametric), the confidence interval of hazard ratio of those significant covariates are not included 1. Also, the

confidence intervals of acceleration factor in AFT models are out of 1. This indicates that they are important predictors for CVD complications among hypertensive patients. Diastolic blood pressure is model based significant covariate for CVD complications. It is significant in only log-normal and log-logistic models. To give the full conclusions about the effect of covariates on CVD, the choice of the best fitted model and potential subsets of covariates are important.

4.2.1 Model Choice and Variable Selection

The fitted cox-regression model have no specified distributional assumptions of shape of survival function. Log-normal distribution is scaled by the standard deviation $\sigma = 0.681$ parameter, which is equivalent to inverse of shape parameter $\rho = 1.468$. Also, log-logistic distribution is scaled by $\sigma = 0.375$ and its shape parameter is $\rho = 2.669$. In all parametric models, the shape parameter is significant at 5% level of significance. This indicates that shape parameter is not zero at all. Weibull model have $\rho = 1.804$, which is more than 1, shows the hazard rate is monotonically increasing with the time. In exponential model it is fixed at 1.

Model estimation for cox-regression was done via partial likelihood, whereas estimation for parametric model was done via maximum likelihood method. With respect to the model choice, it was first done among semi-and parametric models using set of important covariates in univariable analysis. As we discussed in section 3.7.2, it is not simple and there is no straightforward criteria. The one way we have used is rejecting the model which did not fit the data well by assessing the overall goodness of fit, Cox-Snell residual diagnostic plots. In both models, the residuals are obtained from the fitted model, and follows unit exponential distribution if the correct model have been fitted. The plots and detail explanation of the plot is at section 4.3.1. In general, the plot suggested as there is a parametric survival model which can fit the data well, compared to cox-regression model.

In addition to Cox-Snell residual diagnostic plot, schoenfeld residuals plots were also applied to check the proportional hazards assumption for cox-regression model. It was plotted as schoenfeld residuals against time for all covariates and drawn in (Appendix 4). In some covariates, the line of the plot looks non-zero slope or deviates from (y= 0), which indicates a violation of the

proportional hazards assumption. This also shows again as cox-regression model is not fit the data well. Then after, the choice or comparison between several parametric models was done via AIC statistic.

Table 4.2: AIC and ρ of parametric survival models

Distributions	Degree of freedom	AIC	ρ
Exponential	10	1141.217	1.000
Weibull	11	1080.984	1.804
Log-logistic	11	1059.265	2.669
Log-normal	11	1062.319	1.468

AIC= Akaike's Information Criteria, ρ = shape parameter.
PH and AFT of both weibull and exponential have similar AIC and ρ values.

From Table 4.2, log-logistic model have the smallest value of AIC, which is AIC= 1059.265. This indicates that it is the appropriate and preferred model for CVD complications of hypertensive patients data compared to exponential, weibull and log-normal models. This is in accordance with the Cox-Snell diagnostic plots. Both AFT and PH of similar distribution have the same AIC value. For example, weibull PH and AFT have the same AIC = 1080.984. Since the selected is log-logistic model, no need to discuss more about parametric PH.

In the chosen log-logistic model above, variable selection process was done to identify the potential subsets of covariates. This process was done using Collete's recommended model selection approach, called purposeful variable selection. The univariable and multivariable analysis above can be the first and the second step for the variable selection process. In the next step; diabetes mellitus, baseline complications, sex and hyperlipidemia covariates, which do not significantly increase value of -2logL were consequently omitted from the set. Finally, multi-drug use and smoking status were added to the model, one at a time and, any of that do not significantly reduce value of -2logL which means, they are not retained in the model. Proceeding to this, the only significant interaction between systolic and diastolic blood pressure was determined to be included in

a model, using hierarchical principle. Therefore, ages, residence place, proteinuria, systolic blood pressure, diastolic blood pressure and the interaction of systolic and diastolic blood pressure are potential subsets of covariates for the final log-logistic model. Outputs of these covariates with their interpretations are as follows.

Table 4.3: Multi-variable analysis using the log-logistic parametric survival model

Variables	coef	se(coef)	z	ϕ (95% Conf. Interval)	p
Intercept	34.363	7.037	4.883	6.52e+04 (1.46e+04, 2.91e+05)	0.000*
Age	-0.037	0.006	-6.530	0.964 (0.953, 0.975)	0.000*
Residence					
Rural	Reference	–	–	1	–
Urban	-0.411	0.109	-3.774	0.663 (0.535, 0.821)	0.000*
Proteinuria					
Absent	Reference	–	–	1	–
Present	-0.261	0.110	-2.371	0.770 (0.621, 0.956)	0.018*
Baseline SBP	-0.162	0.042	-3.893	0.851 (0.784, 0.923)	0.000*
Baseline DBP	-0.229	0.065	-3.544	0.796 (0.701, 0.903)	0.000*
SBP:DBP	0.001	0.0004	3.354	1.001 (1.001, 1.002)	0.001*
log(scale)	-0.968	0.070	-1.104		0.000*
AIC =	*1044.522*	$\chi^2 =$	*244.810**	λ *=11.926e-16*	
Loglik(model) =	*-514.300*	$\sigma =$	*0.380*	ρ *=2.632*	

*Source: Jimma University Teaching Hospital 2017. z= wald statistic, se= standard error, p= p-value, *= p_value < 0.05, ϕ= acceleration factor, SBP = systolic blood pressure, DBP= diastolic blood pressure, SBP:DBP= combination of systolic and diastolic blood pressure, AIC= Akaike Information Criterion, χ^2 = Chi-square, ρ= shape parameter, σ= 1/ρ, λ= scale parameter*

The table above (Table 4.3) is shown for the coefficients, standard error of coefficients, acceleration factors ϕ and confidence intervals of ϕ for significant covariates from log-logistic model. The coefficient estimate of age is -0.037 with standard error of 0.006, which is significant at 0.05 level of significance. The acceleration factor is 0.964 with (0.953, 0.975) confidence interval, which

does not include 1. Holding the other covariates constant, increasing in one year age of patient is associated with 3.62% decrease in expected time to develop cardiovascular disease complications. Also, in estimates of residence place, we observed that the estimate is -0.411 with ($\phi = 0.663$) and confidence interval of (0.535, 0.821). This indicates that, patients from urban area is effective to fast the time of experiencing the disease, compared to patients from rural by factor of 0.663 (considering other covariates constant).

Turning to estimates of proteinuria, the estimated coefficient is -0.261, which is significant at 0.05 level. The acceleration factor for proteinuria is 0.770 with (0.621, 0.956) confidence interval. This indicates that the time for developing CVD is accelerated for patients with presence of abnormal quantities of protein in the urine (proteinuria) compared to those who have no proteinuria by estimated factor of 0.770. The acceleration factors for systolic and diastolic blood pressures are 0.851 and 0.796 with (0.784, 0.923) and (0.701, 0.903) confidence intervals respectively. However, they are not independently contributed to CVD complications, because their interaction with coefficient of 0.001 and acceleration factor 1.001 (1.001, 1.002) is significant at 0.05 level of significance. The coefficient of interaction effect is positive, smaller and not highly significant than the lower orders (partial of systolic and diastolic BP). This suggested that, the effect of systolic BP on CVD complications among hypertensive patient is increase with the effect of diastolic BP or the vice versa.

The final fitted log-logistic model with potential subsets of covariates have shape parameter (ρ = 2.632), which is greater than one. Therefor, the hazard function is uni-modal. It allows for some non-monotonic behavior, which means, the hazard increases to a maximum point at the beginning and then it decreases over time. The plots of hazard and survival function of final log-logistic model at mean of covariates are drawn at Figure 4.1 below.

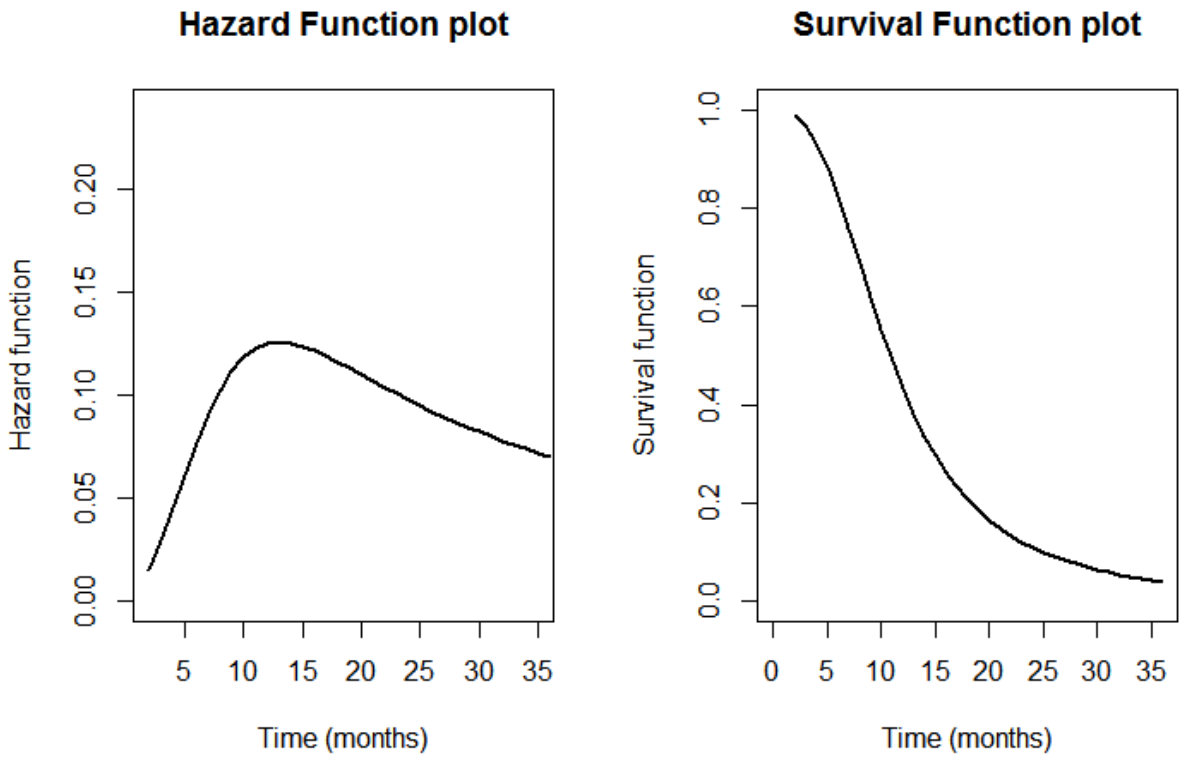


Figure 4.1: Hazard and survival function plot of log-logistic model at mean of covariates

4.3 Model Diagnostics

4.3.1 Cox-Snell Residual Plots

Cox-Snell residuals with their cumulative hazard functions had been obtained from the fitted cox-regression and several parametric models. It is plotted as cumulative hazard function versus Cox-Snell residuals (Figure 4.2). The plots rejects a choice of cox-regression, exponential, weibull and log-normal models, compared to log-logistic model. The line made by Cox-Snell residuals of log-logistic model is reasonably straight, and has approximately unit slope and zero intercept. Which means, approximately it follows a unit exponential distribution. This indicates that log-logistic model is adequate, efficient and appropriate model for analyzing CVD complications among hypertensive patients data at JUTH.

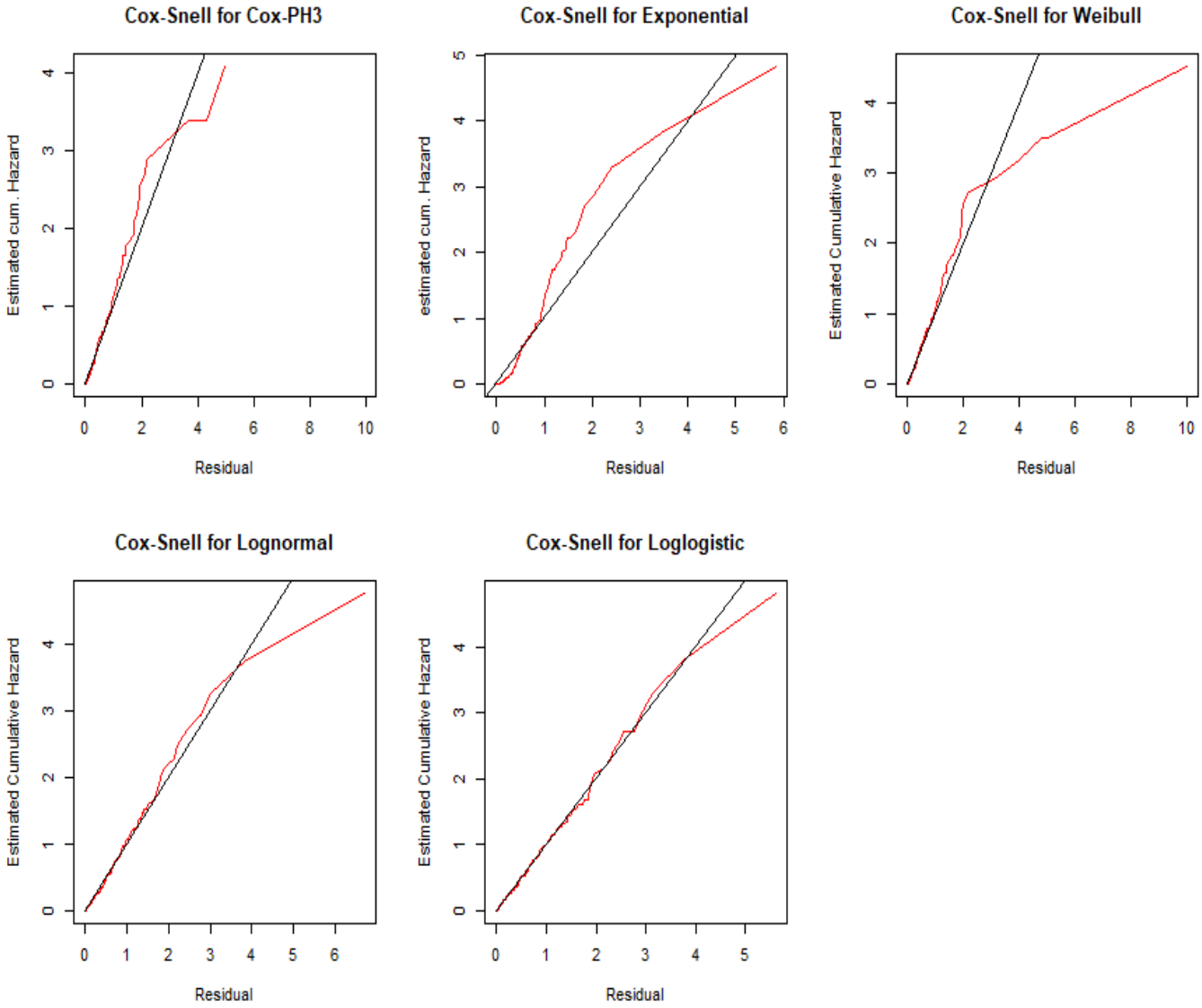


Figure 4.2: Cox-Snell residuals from the fitted cox-regression, exponential, weibull, log-normal and log-logistic models

4.3.2 Plots of Parametric Baselines for Several Parametric Models

The adequacy of parametric baselines for parametric survival models were checked using the appropriate plots. Exponential is plotted by cumulative hazard function with the time, weibull is plotted by the logarithm of cumulative hazard function with the logarithm of time, log-logistic is plotted by the logarithm of odds of failure with logarithm of time, and log-normal is plotted by probit of failures with logarithm of time (Figure 4.3).

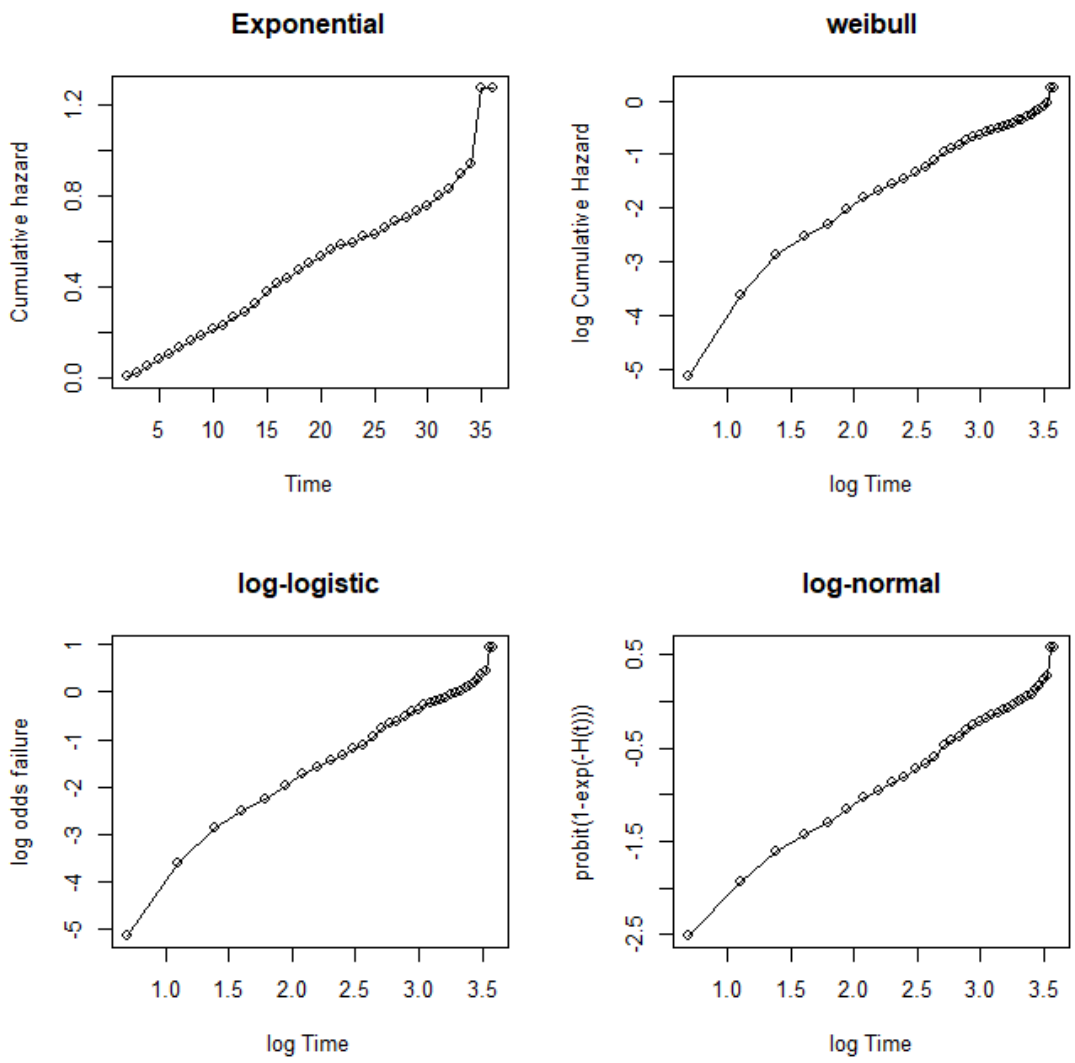


Figure 4.3: Graphical evaluation for parametric models

The plot of exponential model is relatively linear for most of observations, but few of are highly scattered at the end of the plot. Weibull, log-logistic and log-normal models have somewhat similar plots. Using parametric baseline plots, choice of the most appropriate model among weibull, log-logistic and log-normal is too difficult. However, they are more adequate than exponential model (Figure 4.3).

4.3.3 The Quantile-Quantile Plot

The quantile-quantile plot was made for assessing the validity of log-logistic accelerated failure time model using two different groups of patients. In this study we have used estimated percentile survival time of two different groups of patients, based on their residence place (urban and rural). The percentiles were estimated using the Kaplan-Meier estimator.

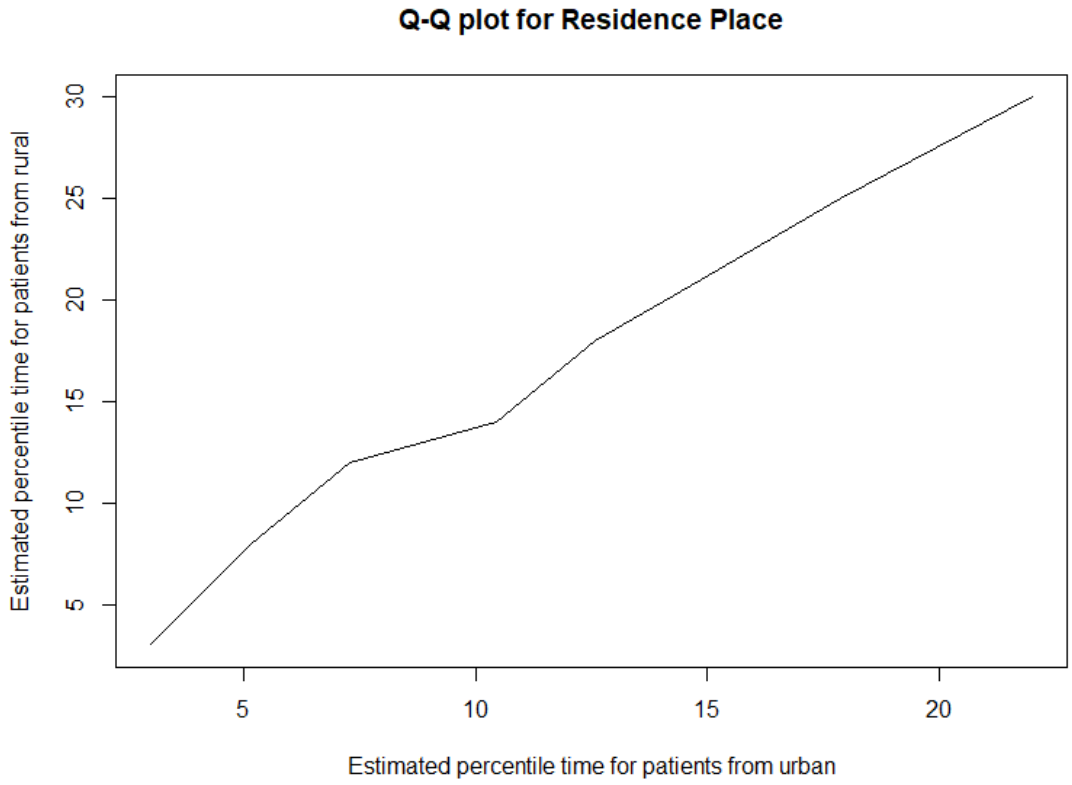


Figure 4.4: Q-Q plot of hypertensive patients data using survival times of Residence place

From the plot (Figure 4.4), we observed that the line of plot is approximately straight, and the slope of a straight line drawn through the line is approximately equals to the estimated acceleration factor 0.6385. Therefor, we can interpret it as the time of developing CVD complication is speed up by a factor of 0.6385 in hypertensive patients from urban, compared to those who were from rural. This suggested that the acceleration failure time model would be appropriate for the data.

4.4 Discussion

We have targeted on CVD complications event of hypertensive patients. The disease was number one cause of global deaths, taking an estimated 17.9 million lives each year and more than 75% are in low and middle income countries. Hypertension is contributed highly to CVD independently, or with the other associated risk factors. The main aim of the study was determining the major risk factors of CVD complications among Hypertensive patients at Jimma University Teaching Hospital, using appropriate survival model.

The choice between semi- and parametric survival models were done for the selection of appropriate fit model. This was applied using Cox-Snell residual goodness of fit test, after all models were alternatively fitted. Parametric model with reasonably straight line Cox-Snell residual plot, compared to semi-parametric model were chosen. Then after, the comparison between several parametric survival models were applied via AIC statistic. Finally, log-logistic model with approximately straight line Cox-Snell residual plot and smallest AIC = 1059.265 were selected as appropriate fit model for CVD complications of hypertensive patients in JUTH.

The most widely used survival model in medical research is cox-regression model, due to it is flexible and does not oblique to choose particular probability, unspecified baseline hazard and requires less model checking efforts. From the parametric part weibull is more popular because of it can estimates event rates (PH) and relative extension in survival time (AFT) [26]. In our context, log-logistic model was appropriate fit model. Adequacy of the model were checked using parametric baseline diagnostic plot. The plot suggested as it is appropriate model, with relatively similar to weibull and log-normal models. Cox-Snell residual diagnostic plots which were plotted for semi-parametric and all parametric models also agreed to the efficiency of log-logistic models for CVD of hypertensive patients data. Using urban and rural categorized groups of patients, quantile-quantile plot were made and assessed the validity of acceleration failure time model. The shape of hazard function plot for the fitted log-logistic model is hump-shaped, it increases first and then decreases with $\rho = 2.632$.

Hypertension itself can independently contributed to CVD complications. However, its influence is high with additional risk factors. Covariates considered as a risk factors in this thesis were age, sex, residence place, diabetes mellitus, proteinuria, multi-drug use, baseline complications, hyperlipidemia, smoking status, systolic BP and diastolic BP. In univariable analysis, which applied as the first step for variable selection process, multi-drug use and smoking status were insignificant factor for CVD complications, and they were not included to multivariable analysis. Purposeful variable selection process were used to select the potential subset of covariates in the model. Using the final log-logistic model for multivariable analysis age, residence place, proteinuria, systolic BP, diastolic BP and interaction of systolic and diastolic BP were significantly contributed to CVD complications of hypertensive patients.

Age which is one of significant variable in this study is the well known variable of interest in CVD complications. Aged patients had shorter time to develop CVD complications. It is confirmed to findings of Gesese MT. 2017 and Jennifer L, et al. 2019 [19, 44]. This is maybe when patients increase in age, their blood vessels become flexible, and this may makes harder for blood to move through blood vessels easily. Also, the fatty deposits (plaque) will be collected along their artery walls and it can slow the blood flow from the heart to different bodies. All this things including, poor nutrition and exercise habits happened due to increase in age can speedup the time of developing CVD complications [44].

Residence place is also a significant factor for CVD complications among hypertensive patients in JUTH. Rural lived patients had survived longer from CVD complications, compared to urban lived patients. This is in accordance with studies of Gesese MT. 2017 and Seedat Y, et al. 2018 [19, 43]. This fasting in time for urban lived patients maybe due to, poor quality of city living standard, which is decreasing in availability of safe, space for exercise or recreation, increasing pressures from mass marketing, and the availability of unhealthy and cheap food options in urban [47]. However, due to higher income and infrastructures in urban, which may helps to reduce the risk of CVD complications through healthier life style or better access for prevention and can prolong the time to develop the disease for urban lived patients, finding of Anand SS, et al. 2011 have different

conclusion to this study [48, 92].

Proteinuria is another important significant factor for CVD complications among hypertensive patients. The time to develop CVD complications in patients with proteinuria is accelerated, compared to those who have no proteinuria. Studies of Maione A, et al. 2009; Ruilope LM, et al. 1995 and Brown MJ, et al. 2000 [51, 52, 53] have similar conclusion to this finding. It is maybe when filters in kidneys (glomeruli), which allows only small amounts of protein found in the blood into urine are damaged in some way, and released too much protein into urine. This excess amount of protein released can be a causes for insufficiency of absorption or impaired filtration, when serum proteins are reabsorbed from urine, which may faster the time. Systolic and diastolic blood pressure of patients are also significantly contributed to CVD complications. However, their contribution in this study is based on the combination of both. In Sesso HD, et al. 2000 finding [67], similarly, both systolic and diastolic BP are strongly associated to CVDs, but independently. Dependent contribution in this study is maybe due to, the only baseline measure of blood pressures (both SBP and DBP) used in this study.

5 Conclusions and Recommendations

5.1 Conclusions

Log-logistic acceleration failure time model were chosen for determinants of cardiovascular disease complications among hypertensive patients at Jimma University Teaching Hospital, 2017.

More than 50% of hypertensive patients were from urban community, had diabetes mellitus, proteinuria and hyperlipidemia. Among these additional risk factors of hypertensive patients, proteinuria and urban residence place had great impact to determine CVD complications, through shortening the expected time to experience it. The overall median time of hypertensive patients develop CVD complications (40.23%) is about 28 months with minimum and maximum time of 2 months and 35 months respectively. SBP is another important prognostic covariate for CVD complications, and its effect in this study is high with DBP. Generally Age, residence place, proteinuria, SBP, DBP and combination of SBP and DBP are significant prognostic factors for CVD complications of hypertensive patients at Jimma University Teaching Hospital, 2017.

5.2 Recommendations

Based on this study, it is better to give recommendation for physicians, stakeholders, institutions or hospitals, governments or academicians. There are identified groups of patients who were at risk and experienced CVD complications within a very short time. Those are, oldest patients, who were from urban community, had proteinuria, had largest in measure of systolic and diastolic blood pressures. Physicians should give special attention for these patients to save them from the risk, especially for patients from urban and who had proteinuria. It is known that, process of aging cannot be changed, leading generally healthy lifestyle and at least doing cardiac activity will be recommended for the stakeholders. The ways of data organized or measured in Jimma University Teaching Hospital is not satisfactorily. The researcher suggested as it is better, if the data managing system is changed to digital form, and this will helps to extract much enough and qualified informations. Further study should be conducted with better study design, especially

the study which can address within variability of changes in systolic and diastolic blood pressures through time. Government body should have a responsible to facilitate this. For further analysis closed to this study, it is well if everyone uses log-logistic accelerated failure time model.

References

[1] British heart foundation - Cardiovascular heart disease. Accessed November 30, 2019 https://www.bhf.org.uk/informationsupport/conditions/cardiovascular-heart-disease

[2] Cooper RS, Amoah AG, Mensah GA. High blood pressure: the foundation for epidemic cardiovascular disease in African populations. Ethnicity and disease. 2003 Jan 1;13(2; SUPP/2):S2-48.

[3] Nkomo VT. Epidemiology and prevention of valvular heart diseases and infective endocarditis in Africa. Heart. 2007 Dec 1;93(12):1510-9.

[4] World Health Organization. Cardiovascular diseases (CVDs): fact sheet 2017.

[5] Keates AK, Mocumbi AO, Ntsekhe M, Sliwa K, Stewart S. Cardiovascular disease in Africa: epidemiological profile and challenges. Nature Reviews Cardiology. 2017 May;14(5):273-93.

[6] Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ, editors. Global burden of disease and risk factors. The World Bank; 2006 Apr 2.

[7] Bygbjerg IC. Double burden of noncommunicable and infectious diseases in developing countries. Science. 2012 Sep 21;337(6101):1499-501.

[8] The Federal Democratic Republic of Ethiopia Ministry of Health- Health Sector Transformation Plan, 2015 - 2019. Accessed January 14, 2020. https://www.globalfinancingfacility.org/sites/gff_new/files/Ethiopia-health-system-transformation-plan

[9] Ezega News. Cardiovascular Diseases Becoming Major Health Threat in Ethiopia. Accessed January 14, 2020 https://www.ezega.com/News/NewsDetails/3421/Cardiovascular-Diseases-Becoming-Major-Health-Threat-in-Ethiopia

[10] Centers for Disease Control and prevention(CDC), Global Health - Ethiopia. Accessed January 13, 2020.

[11] Mcfarlane SI, Banerji M, Sowers JR. Insulin resistance and cardiovascular disease. The Journal of Clinical Endocrinology & Metabolism. 2001 Feb 1;86(2):713-8.

[12] Kannel WB. Risk stratification in hypertension: new insights from the Framingham Study. American journal of hypertension. 2000 Jan 1;13(S1):3S-10S.

[13] Stuart Ali and Francesc Xavier Gomez-Olive Casas. Hypertension is the silent killer disease spreading across an Africa that isn't ready. Accessed January 13, 2020.

https://qz.com/africa/1036880/hypertension-and-other-cardiovascular-diseases-are-killing-millions-in-africa-every-year/

[14] Fuentes R, Ilmaniemi N, Laurikainen E, Tuomilehto J, Nissinen A. Hypertension in developing economies: a review of population-based studies carried out from 1980 to 1998. Journal of hypertension. 2000 Jan 1;18(5):521-9.

[15] Girmachew G. Ethiopia: Due Attention to Hypertension. Accessed January 13, 2020 https://allafrica.com/stories/201805310462.html

[16] Kibret KT, Mesfin YM. Prevalence of hypertension in Ethiopia: a systematic meta-analysis. Public Health Reviews. 2015 Dec;36(1):1-2.

[17] Collett D. Modelling survival data in medical research. CRC press; 2015 May 4.

[18] Misganaw A, Mariam DH, Araya T. The double mortality burden among adults in Addis Ababa, Ethiopia, 2006-2009. Preventing chronic disease. 2012;9.

[19] Gesese MT. Survival Analysis of Time to Cardiovascular Disease Complication of Hypertensive Patients at Felege Hiwot Referal Hospital in Bahir-Dar, Ethiopia: A Retrospective Cohort Study. Archives of Current Research International. 2017 Nov 4:1-9.

[20] Tesfaye B, Haile D, Lake B, Belachew T, Tesfaye T, Abera H. Uncontrolled hypertension and associated factors among adult hypertensive patients on follow-up at Jimma University Teaching and Specialized Hospital: cross-sectional study. Research Reports in Clinical Cardiology. 2017;8:21.

[21] Lawes CM, Vander Hoorn S, Rodgers A. Global burden of blood-pressure-related disease, 2001. The Lancet. 2008 May 3;371(9623):1513-8.

[22] Antonakoudis G, Poulimenos I, Kifnidis K, Zouras C, Antonakoudis H. Blood pressure control and cardiovascular risk reduction. Hippokratia. 2007 Jul;11(3):114.

[23] Richman IB, Fairley M, Jørgensen ME, Schuler A, Owens DK, Goldhaber-Fiebert JD. Cost-effectiveness of intensive blood pressure management. JAMA cardiology. 2016 Nov 1;1(8):872-9.

[24] World Health Organization (WHO). WHO Country Cooperation Strategy: Ethiopia. Accessed January 13, 2020.
https://www.afro.who.int/sites/default/files/2017-05/ethiopia_ccs-document-2014.pdf

[25] Klein JP, Moeschberger ML. Survival Analysis: Statistical Methods for Censored and Truncated Data. 2nd. 2003

[26] Pötter U. Melinda Mills, Introducing Survival and Event History Analysis. 2012

[27] Noble A, Johnson R, Thomas A, Bass P. Organsysteme verstehen-Herz-Kreislauf-System: Integrative Grundlagen und Fälle. Elsevier Health Sciences; 2017 Jun 14.

[28] Aaronson PI, Ward JP. The cardiovascular vascular system at a glance. Backwell Publishing. 3rd. 2007.

[29] Mohrman DE, Heller LJ. Cardiovascular Physiology, Lange Physiology; 2010 7thed .

[30] World Heart day (World Heart Federation). Accessed January 13, 2020. https://www.world-heart-federation.org/world-heart-day/world-heart-day-2019/cvds/

[31] Mendis S, Puska P, Norrving B, World Health Organization. Global atlas on cardiovascular disease prevention and control. World Health Organization; 2011.

[32] Naghavi M, Wang H, Lozano R, Davis A, Liang X, Zhou M. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015 Jan 10;385(9963):117-71.

[33] National Heart, Lung, and Blood Institute - Coronary Heart Disease. February 2019. https://www.nhlbi.nih.gov/health-topics/coronary-heart-disease

[34] Bhatia SK. Biomaterials for clinical applications. Springer Science & Business Media; 2010 Oct 14.

[35] Medical News Today - Stroke. Accessed February 12, 2020. https://www.medicalnewstoday.com/articles/7624.php#1

[36] Medical News Today - What to know about cardiovascular disease. Accessed February 12, 2020.
https://www.medicalnewstoday.com/articles/257484.php

[37] Di Cesare M, Khang YH, Asaria P, Blakely T, Cowan MJ, Farzadfar F, Guerrero R, Ikeda N, Kyobutungi C, Msyamboza KP, Oum S. Inequalities in non-communicable diseases and effective responses. The Lancet. 2013 Feb 16;381(9866):585-97.

[38] Pestana JA, Steyn K, Leiman A, Hartzenberg GM. The direct and indirect costs of cardiovascular disease in South Africa in 1991. South African Medical Journal. 1996;86(6).

[39] Gaziano TA. Cardiovascular disease in the developing world and its cost-effective management. Circulation. 2005 Dec 6;112(23):3547-53.

[40] WHO Regional Officer to Africa. Global HEARTS' Manual Adoption workshop held in Ethiopia. 2018
https://www.afro.who.int/news/global-hearts-manual-adoption-workshop-held-ethiopia

[41] Van de Vijver S, Akinyi H, Oti S, Olajide A, Agyemang C, Aboderin I, Kyobutungi C. 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. 2014 May 6;16(1).

[42] Weycker D, Nichols GA, O'Keeffe-Rosetti M, Edelsberg J, Khan ZM, Kaura S, Oster G. Risk-factor clustering and cardiovascular disease risk in hypertensive patients. American journal of hypertension. 2007 Jun 1;20(6):599-607.

[43] Seedat Y, Ali A, Ferdinand KC. Hypertension and cardiovascular disease in the sub-Saharan African context. Annals of translational medicine. 2018 Aug;6(15).

[44] Rodgers JL, Jones J, Bolleddu SI, Vanthenapalli S, Rodgers LE, Shah K, Karia K, Panguluri SK. Cardiovascular risks associated with gender and aging. Journal of cardiovascular development and disease. 2019 Jun;6(2):19.

[45] Cardiovascular disease risk factors. Accessed January 13, 2020. ada.com/cardiovascular-disease-risk-factor/

[46] Mugure G, Karama M, Kyobutungi C, Karanja S. Correlates for cardiovascular diseases among diabetic/hypertensive patients attending outreach clinics in two Nairobi slums, Kenya. The Pan African Medical Journal. 2014;19.

[47] World Heart Federation - Urbanization and cardiovascular disease Smith, S. et al., 2012. <http://www.worldheart.org/urbanization>

[48] Danaei G, Singh GM, Paciorek CJ, Lin JK, Cowan MJ, Finucane MM, Farzadfar F, Stevens GA, Riley LM, Lu Y, Rao M. The global cardiovascular risk transition: associations of four metabolic risk factors with national income, urbanization, and Western diet in 1980 and 2008. Circulation. 2013 Apr 9;127(14):1493-502.

[49] Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998 May 12;97(18):1837-47.

[50] Alderman MH, Cohen H, Madhavan S. Diabetes and cardiovascular events in hypertensive patients. Hypertension. 1999 May;33(5):1130-4.

[51] Maione A, Annemans L, Strippoli G. Proteinuria and clinical outcomes in hypertensive patients. American journal of hypertension. 2009 Nov 1;22(11):1137-47.

[52] Ruilope LM, Rodicio JL. Microalbuminuria in clinical practice. Kidney. 1995;4:211-6.

[53] Brown MJ, Palmer CR, Castaigne A, De Leeuw PW, Mancia G, Rosenthal T, Ruilope LM. Morbidity and mortality in patients randomised to double-blind treatment with a long-acting calcium-channel blocker or diuretic in the International Nifedipine GITS study: Intervention as a Goal in Hypertension Treatment (INSIGHT). The Lancet. 2000 Jul 29;356(9227):366-72.

[54] Chobanian AV, Bakris GL, Black HR, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. Jama. 2003 May 21;289(19):2560-71.

[55] Balkrishnan R, Rajagopalan R, Camacho FT, Huston SA, Murray FT, Anderson RT. Predictors of medication adherence and associated health care costs in an older population with type 2 diabetes mellitus: a longitudinal cohort study. Clinical therapeutics. 2003 Nov 1;25(11):2958-71.

[56] Murphy DA, Sarr M, Durako SJ, Moscicki AB, Wilson CM, Muenz LR. Barriers to HAART adherence among human immunodeficiency virus–infected adolescents. Archives of pediatrics & adolescent medicine. 2003 Mar 1;157(3):249-55.

[57] Schroeder K, Fahey T, Ebrahim S. Interventions for improving adherence to treatment in patients with high blood pressure in ambulatory settings. Cochrane Database of systematic reviews. 2004(3).

[58] Meredith P. Achieving quality 24-h blood pressure control with candesartan cilexetil. Blood Pressure. 2000 Jan 1;9(sup1):23-6.

[59] Tunstall-Pedoe H. Cardiovascular risk and risk scores: ASSIGN, Framingham, QRISK and others: how to choose. 2011

[60] American heart association Angina Pectoris-Stable Angina. Accessed January 13, 2020. https://www.heart.org/en/health-topics/heart-attack/angina-chest-pain/angina-pectoris-stable-angina.

[61] Southern cross medical library- Angina- cause, symptoms, treatment, prevention. 2017 - 2018 https://www.southerncross.co.nz/group/medical-library/angina-causes-symptoms-treatment

[62] British heart foundation - Smoking. Accessed February 15, 2020.

[63] Neaton JD, Wentworth D. Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease overall findings and differences by age for 316099 white men. Archives of internal medicine. 1992 Jan 1;152(1):56-64

[64] Castelli WP, Anderson K. A population at risk: prevalence of high cholesterol levels in hypertensive patients in the Framingham Study. The American journal of medicine. 1986 Feb 14;80(2):23-32.

[65] Nelson RH. Hyperlipidemia as a risk factor for cardiovascular disease. Primary Care: Clinics in Office Practice. 2013 Mar 1;40(1):195-211.

[66] Collaboration PS. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002;360(9349):1903-3.

[67] Sesso HD, Stampfer MJ, Rosner B, Hennekens CH, Gaziano JM, Manson JE, Glynn RJ. Systolic and diastolic blood pressure, pulse pressure, and mean arterial pressure as predictors of cardiovascular disease risk in men. Hypertension. 2000 Nov;36(5):801-7.

[68] Benetos A, Thomas F, Bean K, Gautier S, Smulyan H, Guize L. Prognostic value of systolic and diastolic blood pressure in treated hypertensive men. Archives of Internal Medicine. 2002 Mar 11;162(5):577-81.

[69] Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. 1978 Apr 1;65(1):141-51.

[70] Holt JD, Prentice RL. Survival analyses in twin studies and matched pair experiments. Biometrika. 1974 Apr 1;61(1):17-30.

[71] Wey A, Connett J, Rudser K. Combining parametric, semi-parametric, and non-parametric survival models with stacked survival models. Biostatistics. 2015 Jul 1;16(3):537-49.

[72] Böhmer PE. Theorie der unabhängigen Wahrscheinlichkeiten. InRapports Memoires et Proces verbaux de Septieme Congres International dActuaires Amsterdam 1912 Sep (Vol. 2, pp. 327-343).

[73] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American statistical association. 1958 Jun 1;53(282):457-81.

[74] Habibi D, Rafiei M, Chehrei A, Shayan Z, Tafaqodi S. Comparison of survival models for analyzing prognostic factors in gastric cancer patients. Asian Pacific journal of cancer prevention: APJCP. 2018;19(3):749.

[75] Grover G, Sreenivas V, Khanna S, Seth D. Estimation of Survival of Liver Cirrhosis Patients, in the Presence of Prognostic Factors Using Accelerated Failure Time Model as an Alternative to Proportional Hazard Model. International journal of statistics and applications. 2013;3(4):113-22.

[76] Pourhoseingholi MA, Hajizadeh EB, Abadi A, Safaee A, Moghimi Dehkordi B, Zali MR. Comparing cox regression and parametric models for survival analysis of patients with gastric cancer. Iranian Journal of Epidemiology. 2007 Sep 10;3(1):25-9.

[77] Kleinbaum DG, Klein M. Statistics for Biology and Health (survival analysis,). A Self-Learning Text: Springer. 2005.

[78] Zhu HP, Xia X, Chuan HY, Adnan A, Liu SF, Du YK. Application of Weibull model for survival of patients with gastric cancer. BMC gastroenterology. 2011 Dec 1;11(1):1.

[79] Wang BB, Liu CG, Lu P, Latengbaolide A, Lu Y. Log-normal censored regression model detecting prognostic factors in gastric cancer: a study of 3018 cases. World Journal of Gastroenterology: WJG. 2011 Jun 21;17(23):2867.

[80] Ghorbani Gholiabad S, Yazdani Charati J, Jan Babaie G. Evaluation of parametric and semi-parametric models in survival analysis of patients with gastric cancer. Journal of Mazandaran University of Medical Sciences. 2014 Dec 15;24(119):11-8.

[81] Ali ZA, Hosseini M, Mahmoodi M, Mohammad K, Zeraati H, Naieni KH. A comparison between accelerated failure-time and Cox proportional hazard models in analyzing the survival of gastric cancer patients. Iranian journal of public health. 2015 Aug;44(8):1095.

[82] Georgousopoulou EN, Pitsavos C, Yannakoulia CM, Panagiotakos DB. Comparisons between survival models in predicting cardiovascular disease events: application in the Attica study (2002-2012). Journal of Statistics Applications & Probability. 2015 Jul 1;4(2):203.

[83] HEALTH LOCAL NEWS-New Medical Center Inaugurated at Jimma University. 2018. Assessed
https://ethio.news/2018/12/08/new-medical-center-inaugurated-at-jimma-university/

[84] Food, Medicine and Health Care Administration and Control Authority of Ethiopia. Standard Treatment Guidelines for General Hospitals. 3rd edition. 2014.

[85] My Heart, My Life - The Heart Foundation: Common medical tests to diagnose heart conditions, 2012.
https://youtu.be/dGH9xD2w7uc

[86] Liu X. Survival analysis: models and applications. John Wiley & Sons; 2012 Jun 13. https://healthywa.wa.gov.au/Articles/A_E/Common-medical-tests-to-diagnose-heart-conditions

[87] Breslow NE. Analysis of survival data under the proportional hazards model. International Statistical Review/Revue Internationale de Statistique. 1975 Apr 1:45-57.

[88] Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological). 1972 Jan;34(2):187-202.

[89] Qi J. Comparison of proportional hazards and accelerated failure time models (Doctoral dissertation). 2009

[90] Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. Statistics in medicine. 1992;11(14-15):1871-9.

[91] Cox DR, Snell EJ. A general definition of residuals. Journal of the Royal Statistical Society: Series B (Methodological). 1968 Jul;30(2):248-65.

[92] Anand SS, Yusuf S. Stemming the global tsunami of cardiovascular disease. The Lancet. 2011 Feb 12;377(9765):529-32.

[93] Cholesterol-High Cholesterol Diseases
https://my.clevelandclinic.org/health/articles/11918-cholesterol-high-cholesterol-diseases]

APPENDIXES

Appendix 1. Log-logistic Univariable Analysis

Variables	coef	se(coef)	ϕ	[95% Conf. Interval] LCL	UCL	-2 logL
Intercept	7.0381	0.3776	1139.2018	543.5273	2387.7010	1132.6*
Age	-0.0659	0.0060	0.9363	0.9253	0.9474	
log(scale)	-0.7839	0.0700				
Intercept	3.6856	0.1224	39.8684	31.3626	50.6811	1244.8*
Sex (Male)	-0.7256	0.1413	0.4841	0.3670	0.6385	
log(scale)	-0.5714	0.0696				
Intercept	3.6037	0.1119	36.7342	29.5012	45.7407	1249.4*
Residence (Urban)	-0.6443	0.1335	0.5250	0.4042	0.6820	
log(scale)	-0.5788	0.0695				
Intercept	3.4638	0.1275	31.9372	24.8768	41.0015	1268.0*
Diabetes M (Yes)	-0.3308	0.1458	0.7183	0.5398	0.9560	
log(scale)	-0.5373	0.0697				
Intercept	3.5950	0.1084	36.4160	29.4461	45.0356	1247.6*
Proteinuria (Present)	-0.6589	0.1316	0.5174	0.3998	0.6697	
log(scale)	-0.5815	0.0696				
Intercept	3.2468	0.0957	25.7090	21.3115	31.0140	1273.4
Multi-drug Use (Yes)	-0.0220	0.1325	0.9782	0.7546	1.2682	
log(scale)	-0.5258	0.0696				
Intercept	3.3580	0.0811	28.7307	24.5083	33.6805	1258.2*
B-Complication (Yes)	-0.6350	0.1637	0.5299	0.3845	0.7304	
log(scale)	-0.5410	0.0693				
Intercept	3.2609	0.0827	26.0725	22.1726	30.6583	1273.0
Smoking Status (Yes)	-0.0894	0.1469	0.9145	0.6857	1.2196	
log(scale)	-0.5266	0.0696				
Intercept	3.5627	0.1115	35.2593	28.3394	43.8690	1254.0*
Hyperlipidemia (Present)	-0.5825	0.1352	0.5585	0.4285	0.7279	
log(scale)	-0.5571	0.0696				
Intercept	12.1692	0.8249	1.93e+05	3.83e+04	9.71e+05	1117.0*
Systolic BP	-0.0551	0.0049	0.9464	0.9374	0.9556	
log(scale)	-0.7848	0.0691				
Intercept	8.3223	0.6763	4114.5587	1093.0108	1.55e+04	1204.8*
Diastolic BP	-0.0486	0.0063	0.9526	0.9410	0.9643	
log(scale)	-0.6435	0.0690				

*Source: Jimma University Teaching Hospital 2017. Reference category-Sex(Female), Residence(Rural), Proteinuria(Absent), B-Complications (No) and Hyperlipidemia(Absent). se= standard error, ϕ= acceleration factor, LCL&UCL = Upper&Lower Confidence Level, LR= likelihood ratio test, *= p_value < 0.05, the -2*log *L for null model = 1273.4.*

Appendix 2. Multivariable Analysis for all the Candidate Models

1. Cox-regression model

						[95% Conf. Interval]	
Variables	coef	se(coef)	z	p	HR	LCL	UCL
Age	0.0515	0.0111	4.619	0.000	1.0528	1.0301	1.0761
Sex (Male)	0.3289	0.2276	1.445	0.148	1.3894	0.8894	2.1704
Residence (Urban)	1.0319	0.2113	4.884	0.000	2.8063	1.8547	4.2461
Diabetes M (Yes)	0.0272	0.2092	0.130	0.896	1.0276	0.6819	1.5484
Proteinuria (Present)	0.5624	0.1975	2.848	0.004	1.7550	1.1917	2.5844
B-Complications	-0.1586	0.2332	-0.680	0.496	0.8533	0.5403	1.3478
Hyperlipidemia (Present)	0.3371	0.2009	1.678	0.093	1.4009	0.9449	2.0769
Systolic BP	0.0454	0.0091	4.987	0.000	1.0465	1.0280	1.0653
Diastolic BP	0.0140	0.0083	1.693	0.090	1.0141	0.9978	1.0307
AIC =	*948.7943*						
Likelihood ratio test =	*213.86**						

Source: Jimma University Teaching Hospital 2017. Reference category-Sex(Female), Residence(Rural), Proteinuria(Absent), B-Complications (No) and Hyperlipidemia(Absent). z= wald statistic, se= standard error, p= p_value, HR=Hazard Ratio, LCL&UCL = Upper&Lower confidence level

2. Exponential AFT

						[95% Conf. Interval]	
Variables	coef	se(coef)	z	p	ϕ	LCL	UCL
Intercept	14.7848	1.2344	11.977	0.000**	5.56e+06	3.54e+05	8.71e+07
Age	-0.0454	0.0106	-4.293	0.000*	1.0465	1.0250	1.0684
Sex (Male)	-0.2022	0.2177	-0.929	0.353	1.2241	0.7990	1.8754
Residence (Urban)	-0.7421	0.1937	-3.832	0.000	2.1003	1.4369	3.0699
Diabetes M (Yes)	-0.0759	0.1988	-0.382	0.703	1.0789	0.7307	1.5929
Proteinuria (Present)	-0.4046	0.1861	-2.174	0.030*	1.4987	1.0407	2.1584
B-Complications	0.0786	0.2158	0.364	0.716	0.9244	0.6055	1.4112
Hyperlipidemia (Present)	-0.2236	0.1923	-1.163	0.245*	1.2506	0.8578	1.8233
Systolic BP	-0.0384	0.0088	-4.359	0.000*	1.0391	1.0213	1.0572
Diastolic BP	-0.0102	0.0079	-1.287	0.198	1.0103	0.9947	1.0261

3. Exponential PH

					[95% Conf. Interval]	
Variables	coef	se(coef)	p	HR	LCL	UCL
Age	0.0454	0.0106	0.000	1.0465	1.0250	1.0684
Sex (Male)	0.2022	0.2177	0.353	1.2241	0.7990	1.8754
Residence (Urban)	0.7421	0.1937	0.000	2.1003	1.4369	3.0699
Diabetes M (Yes)	0.0759	0.1988	0.703	1.0789	0.7307	1.5929
Proteinuria (Present)	0.4046	0.1861	0.030	1.4987	1.0407	2.1584
B-Complications	-0.0786	0.2158	0.716	0.9244	0.6055	1.4112
Hyperlipidemia (Present)	0.2236	0.1923	0.245	1.2506	0.8578	1.8233
Systolic BP	0.0384	0.0088	0.000	1.0391	1.0213	1.0572
Diastolic BP	0.0102	0.0079	0.198	1.0103	0.9947	1.0261
log(scale)	14.785	1.23	0.000			
AIC = 1141.217	$\chi^2 =$	*177.97**				
loglik(model) = -560.61	$\lambda =$	*3.7e-07*				

Source: Jimma University Teaching Hospital 2017. Reference category-Sex(Female), Residence(Rural), Proteinuria(Absent), B-Complications (No) and Hyperlipidemia(Absent). z= wald statistic, se= standard error, p= p_value, ϕ= acceleration factor, HR= hazard ratio, LCL&UCL = Upper&Lower confidence level, PH= proportional hazard, AFT= acceleration factor, AIC=Akaike Information Criterion, χ^2 = Chi-square, λ= scale parameter

4. Weibull AFT

Variables	coef	se(coef)	z	p	ϕ	[95% Conf. Interval] LCL	UCL
Intercept	10.5047	0.7748	13.56	0.000	3.65e+04	7991.5542	1.67e+05
Age	-0.0295	0.0061	-4.82	0.000	0.9710	0.9594	0.9827
Sex (Male)	-0.1594	0.1219	-1.31	0.191	0.8526	0.6715	1.0827
Residence (Urban)	-0.5801	0.1128	-5.14	0.000	0.5599	0.4488	0.6984
Diabetes M (Yes)	-0.0094	0.1121	-0.08	0.933	0.9907	0.7952	1.2342
Proteinuria (present)	-0.3390	0.1047	-3.24	0.001	0.7125	0.5803	0.8748
B-Complications (Yes)	0.1371	0.1238	1.11	0.268	1.1469	0.8999	1.4618
Hyperlipidemia (Present)	-0.1643	0.1080	-1.52	0.128	0.8485	0.6866	1.0485
Systolic BP	-0.0237	0.0050	-4.78	0.000	0.9766	0.9672	0.9861
Diastolic BP	-0.0070	0.0044	-1.59	0.113	0.9931	0.9846	1.0016
log(scale)	-0.5898	0.0650	9.08	0.000			

5. Weibull PH

Variables	coef	se(coef)	p	HR	[95% Conf. Interval] LCL	UCL
Age	0.0531	0.0108	0.000	1.0546	1.0325	1.0771
Sex (Male)	0.2875	0.2200	0.191	1.3331	0.8662	2.0517
Residence (Urban)	1.0463	0.2058	0.000	2.8470	1.9019	4.2618
Diabete M (Yes)	0.0169	0.2022	0.933	1.0170	0.6843	1.5116
Proteinuria (Present)	0.6114	0.1908	0.001	1.8430	1.2681	2.6785
B-Complication (Yes)	-0.2473	0.2249	0.272	0.7809	0.5026	1.2135
Hyperlipidemia (Present)	0.2963	0.1946	0.128	1.3449	0.9185	1.9694
Systolic BP	0.0427	0.0087	0.000	1.0436	1.0260	1.0616
log(scale)	10.505	0.775	0.000			
log(shape)	0.590	0.065	0.000			
AIC = 1080.984	χ^2=	*218.88**	λ	*= 5.91e-09*		
Loglik(model) = -529.5	$\hat{\sigma}$	*= 0.554*	ρ	*= 1.803674*		

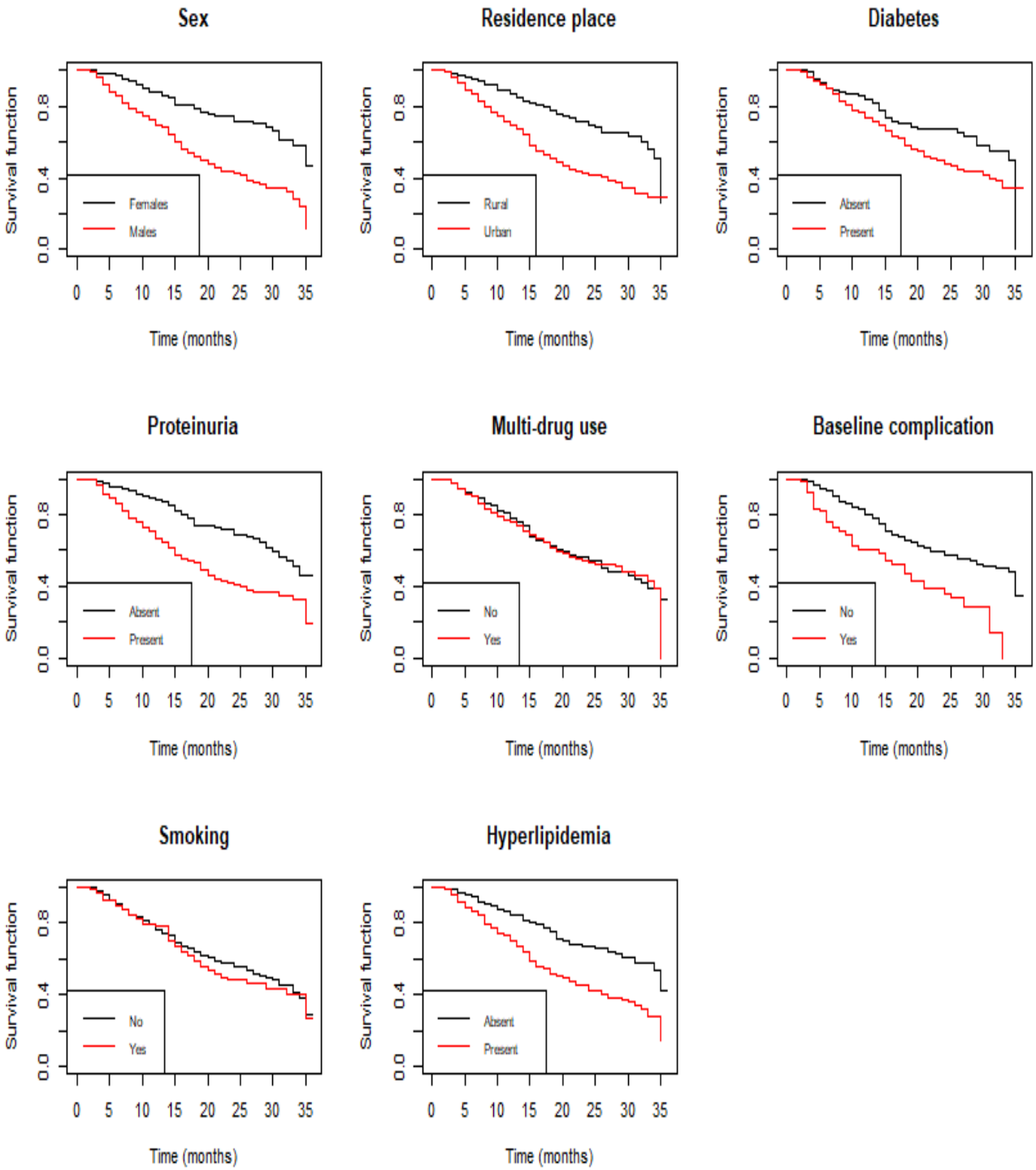
**Jimma University Teaching Hospital 2017. Reference category-Sex(Female), Residence(Rural), Proteinuria(Absent), B-Complications (No) and Hyperlipidemia(Absent). z= wald statistic, se= standard error, p= p_value, ϕ= acceleration factor, HR= hazard ratio, LCL&UCL = Upper&Lower confidence level, PH= proportional hazard, AFT= acceleration factor, AIC=Akaike Information Criterion, χ^2 = Chi-square, ρ= shape parameter, σ= 1/ρ, λ= scale parameter*

6. Log-normal AFT						[95% Conf. Interval]	
Variables	coef	se(coef)	z	p	ϕ	LCL	UCL
Intercept |	11.0320	0.7549	14.61	0.000	6.18e+04	1.41e+04	2.71e+05
Age	-0.0331	0.0060	-5.54	0.000	0.9674	0.9561	0.9788
Sex (Male)	-0.1331	0.1169	-1.14	0.255	0.8753	0.6961	1.1007
Residence (Urban)	-0.4534	0.1101	-4.12	0.000	0.6355	0.5121	0.7886
Diabetes M (Yes)	-0.0450	0.1143	-0.39	0.694	0.9560	0.7641	1.1961
Proteinuria (Prent)	-0.2983	0.1075	-2.77	0.006	0.7421	0.6011	0.9162
B-Complication (Yes)	-0.0845	0.1257	-0.67	0.501	0.9190	0.7183	1.1757
Hyperlipidemia (Present)	-0.1676	0.1078	-1.55	0.120	0.8457	0.6846	1.0446
Systolic BP	-0.0241	0.0050	-4.85	0.000	0.9762	0.9667	0.9857
Diastolic BP	-0.0123	0.0050	-2.49	0.013	0.9877	0.9782	0.9974
log(scale)	-0.3841	0.0609	-6.30	0.000			
AIC = 1062.319	$\chi^2 =$	*227.33**	$\mu =$	*11.03*			
Loglik(model) = -520.2	$\sigma =$	*0.6811*	$\rho =$	*1.468*			

7. Log-logistic AFT						[95% Conf. Interval]	
Variables	coef	se(coef)	z	p	ϕ	LCL	UCL
Intercept	11.0858	0.7632	14.53	0.000	6.52e+04	1.46e+04	2.91e+05
Age	-0.0339	0.0060	-5.67	0.000	0.9667	0.9555	0.9781
Sex	-0.1144	0.1180	-0.97	0.332	0.8919	0.7077	1.1240
Residence (Urban)	-0.4486	0.1071	-4.19	0.000	0.6385	0.5176	0.7877
Diabetes M (Yes)	-0.0297	0.1117	-0.27	0.791	0.9708	0.7799	1.2084
Proteinuria (Present)	-0.2794	0.1061	-2.63	0.008	0.7563	0.6142	0.9311
B-Complication (Yes)	-0.0843	0.1257	-0.67	0.502	0.9191	0.7185	1.1758
Hyperlipidemia (Present)	-0.1799	0.1053	-1.71	0.088	0.8353	0.6795	1.0268
Systolic BP	-0.0237	0.0050	-4.73	0.000	0.9766	0.9670	0.9862
Diastolic BP	-0.0134	0.0050	-2.70	0.007	0.9867	0.9771	0.9964
log(scale)	-0.9816	0.0693	-14.17	0.000			
AIC = 1059.265	χ^2=	*236.07**	$\lambda =$	*1.53e-05*			
Loglik(model) = -518.6	$\sigma =$	*0.3747*	$\rho =$	*2.6688*			

**Jimma University Teaching Hospital 2017. Reference category-Sex(Female), Residence(Rural), Proteinuria(Absent), B-Complications (No) and Hyperlipidemia(Absent). z= wald statistic, se= standard error, p= p_value, ϕ= acceleration factor, LCL&UCL = Upper&Lower confidence level, AFT= acceleration factor, AIC=Akaike Information Criterion, χ^2 = Chi-square, ρ= shape parameter, σ= 1/ρ, λ= scale parameter*

Appendix 3: Kaplan-Meier Estimated Survival Function Plot of Categorical Covariates



Appendix 4: Schoenfeld Residual Plots to Test Proportionality Hazards