JIMMA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
DEPARTMENT OF STATISTICS

Modeling Time-to-Good Control of Hypertension using Cox Proportional Hazard and Frailty Models: the case of Bahir-Dar Felege Hiwet Referral Hospital

By:

Ergoye Melese

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

October, 2014

Jimma, Ethiopia

Modeling Time-to-Good Control of Hypertension using Cox Proportional Hazard and Frailty Models: the case of Bahir-Dar Felege Hiwet Referral Hospital

MSc. Thesis

Ergoye Melese

Major Advisor: Yehenew Getachew (PhD)

Co-advisor: Dinberu Seyoum (PhD Scholar)

October, 2014

Jimma, Ethiopia

APPROVAL SHEET

JIMMA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

As these thesis research advisors, we here by certify that we have read and evaluated the thesis prepared by **ERGOYE MELESE** under our guidance and supervision, which is entitled **"Modeling Time-to-Good Control of Hypertension using Cox Proportional Hazard and Frailty Models: the case of Bahir-Dar Felege Hiwet Referral Hospital"**. We recommend this thesis to be submitted as it fulfills the requirements for the degree of Masters of Science (MSc) in Biostatistics.

Yehenew Getachew (PhD)	_____	_____
Major Advisor	Signature	Date
Dinberu Seyoum (PhD Scholar)	_____	_____
Co-advisor	Signature	Date

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

_____	_____	_____
Name of Chairman	Signature	Date
_____	_____	_____
Name of Major Advisor	Signature	Date
_____	_____	_____
Name of Co-advisor	Signature	Date
_____	_____	_____
Name of Internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the Degree of Masters of Science (MSc.) in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I earnestly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Name: ERGOYE MELESE

Signature

_____

Date

_____

Jimma University, Jimma

ACKNOWLEDGEMENT

First, and foremost, I would like to thank my creature ALLAH for his grace, wisdom, favor, faithfulness and wonderful help up to the present day.

My special thanks goes to Dr. Yehenew Getachew, my major advisor and instructor, for his kind support, advice, guidance and constructive comments beginning from the proposal development up to the final work of this thesis and will always be grateful for that.

Likewise, I am sincerely grateful to my co-advisor Mr. Dinberu Seyoum, for his valuable suggestions and comments beginning from the proposal development up to the final work of this thesis and my senior graduates particularly Mr. Sisay Wendaya and Mr. Amsayaw Tefera and all my friends for their continuous comment, advice and guidance.

I would like to thank Bahir-Dar Felege Hiwet Referral Hospital staff members for providing me the data used in this study more specifically Dr. Ayenew Gedis.

I would like to thank Jimma University for the financial support for this study, instructors and for all the support I was given at any time I asked for.

Finally, my deepest and warmest gratitude goes to my family that has been a source of pride and courage throughout my work, specially to my mother, husband and my sister who have been with me for giving me constructive ideas, encouragement and motivation during my studies.

ABSTRACT

Background: Hypertension is a chronic disease that has a major health problem over the centuries due to its significant contribution to the global health burden. According to World Health Organization, hypertension is the seventh leading cause of death in Ethiopia and it ranks as one of the first causes of cardiovascular related mortality.

Objective: The main objective of this study is to explore possible modeling approaches of time-to-good control of hypertension using Cox proportional hazard and frailty models, using data from Bahir-Dar Felege Hiwet Referral Hospital.

Methods: An institutional-based retrospective cohort study was conducted in June 2014. The study population consists of all hypertensive patients measured repeatedly at least three times with a three month follow up between1st January, 2009 to last December, 2013. Five hundred patients were selected using simple random sampling. The data were collected by trained data collectors using check list. SPSS version 16 and R software were used for data entry and processing of the data, respectively. First, single covariate analysis was done using Cox proportional hazard and univariate frailty models. Then all variables that are significant were included in the multi-variable analysis.

Results: The median survival time of hypertensive patients to attain good control is 48 months and the mean survival time is 43.6 months. Age and systolic blood pressure of patients have a negative relationship with outcome variable. However, fasting blood sugar has positive relationship with the outcome the variable. Moreover, the result showed that, the progression of outcome depends on patient's baseline socio-demographical characteristic such as age.

Conclusion: Cox proportional hazard based analysis reveled that the major factors that affect good control of hypertensive patients are age, systolic blood pressure, fasting blood sugar and creatinine. The result of univariate frailty analysis showed that there is unobserved heterogeneity between individuals in the study set-up, which indicates, there are unmeasured covariates. the clinicians should have give an attention to the younger age and lower systolic blood pressure group to attain good control of hypertension earlier like that of older age and higher systolic blood pressure groups.

Key Words: Hypertension; Time-to-Event; Cox proportional hazard; Frailty Model,

ACRONYMS AND ABBRIVATIONS

AIC	Akaikie Information Criterion
AFT	Accelerated Failure Time
AHA	American Heart Association
BMI	Body Mass Index
BUN	Blood Urea Nitrogen
CI	Confidence Interval
CPH	Cox Proportional Hazard
CVD	Cardio Vascular Disease
DAILYs	Disability Adjusted Life Years
DRIs	Dietary Reference Intakes
FBS	Fasting Blood Sugar
GFR	Glomerullie Filtration Rate
HBP	High Blood Pressure
HR	Hazard Ratio
HTN	Hypertension
IHRERC	Institutional Health Research Ethical Review Committee
ISH	International Society of Hypertension
JNC	Joint National Committee
KM	Kaplan-Meier
LR	Likelihood Ratio
mmHg	Millimeter of Mercury
mg/dL	Milligram per Deciliter
MOH	Ministry of Health

NHS	National Health Society
NICE	National Institute for health and Clinical Excellence
NSF	National Standards Framework
PHA	Proportional Hazards Assumption
PH	Proportional Hazard
SBP	Systolic Blood Pressure
SE	Standard Error
WHO	World Health Organization

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

1. INTRODUCTION

1.1. Background

Hypertension , also known as arterial hypertension or high blood pressure (HBP), is persistent elevation of the blood pressure above 140/90 mmHg and it is one of the most common diseases of the cardiovascular system. A normal blood pressure is 120/80. Its' progression is strongly associated with functional and structural cardiac and vascular abnormalities that damage the heart, kidneys, brain, vasculature, and other organs and lead to premature morbidity and death if not treated properly (Giles *et al.*, 2005). Elevated blood pressure means your heart is working harder than normal, putting both your heart and arteries under great strain. On average, people with uncontrolled hypertension are seven times more likely to have a stroke and six times more likely to develop congestive heart failure (Giles *et al.*, 2005)

If an individual has hypertension this means that their heart is pumping harder. If the heart is pumping harder it will consequently become enlarged, and that can be accompanied by congestive heart failure (Hoeger, 2009).

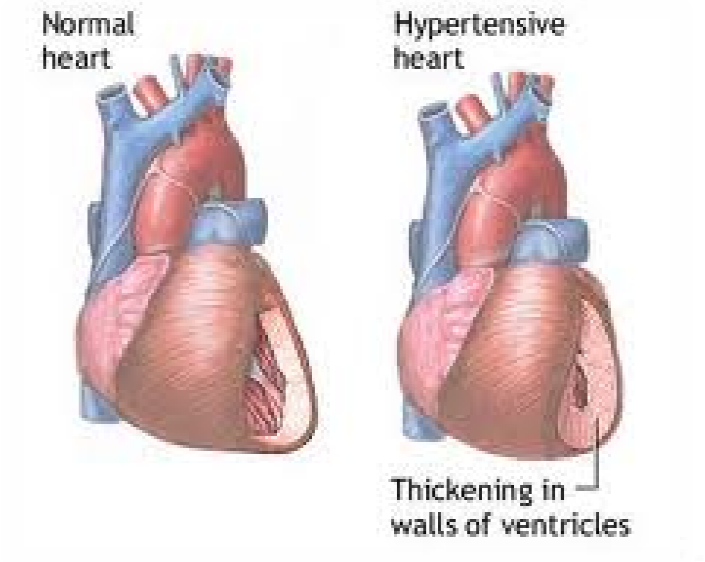

Figure 1: Difference thickening in walls of ventricles between normal and hypertensive heart

The long-term effects of this elevation can negatively influence the heart by causing cardiac hypertrophy. Cardiac hypertrophy occurs when the myocardium thickens. This thickening will decrease the size of the ventricular chambers, specifically the left ventricle, within the heart because the muscle will grow inwards, rather than outwards.

It is known that hypertension is associated with increased morbidity and mortality from stroke, myocardial infarction, cardiac failure, dementia, renal failure, and blindness (Lee & Cooper, 2009).

Hypertension is an important public health challenge which affects approximately one billion persons worldwide (Chobanian A. et al., 2003). According to the World Health Organization (WHO), hypertension is the leading risk factor for mortality (12.7% of deaths attributable) followed by tobacco use (8.7%) and high blood glucose (5.8%) (WHO, 2009). Each year at least 7.1 million people die as a consequence of hypertension (WHO, 2005). The overall average prevalence of hypertension in the world was estimated as 35% (37% in men and 31% in women) (Pereira M. et al, 2009). Blood pressure is measured by "the force in the arteries when the heart beats (systolic pressure) and when the heart is at rest (diastolic pressure)" (O'Brien et al., 2001).

According to the 1999 World Health Organization-International Society of Hypertension (WHO/ISH) guidelines for the management of hypertension (WHO, 1999), hypertension is defined as a systolic blood pressure (SBP) of 140 mmHg or greater and/or a diastolic blood pressure (DBP) of 90 mmHg or greater in subjects who are not taking antihypertensive medication. In general, the diagnosis of hypertension should be based on at least 2 blood pressure measurements per visit and at least 2 to 3 visits; although in particularly severe cases the diagnosis can be based on measurements taken at a single visit.

Recent studies have shown a wide linkage between high blood pressure and hypertension. Similarly, cardiovascular disease has been diagnosed to be related to high blood pressure components (Benetos *et al.,* 2001). A person is said to be experiencing high blood pressure if he/she has blood pressure above 140/90 mmHg measured on both arms. Hypertension with its complications is a real and important public health problem, especially for developed countries compared with the underdeveloped ones. It ranks as one of the first causes of cardiovascular mortality as reported by WHO (WHO, 2009).

It is found that hypertension affects 20-30% of the adult population. With age, the prevalence of the disease increases and reaches 50-65% in those over 65 years old. Thus, women taking oral contraceptives are more likely to develop hypertension. Main risk groups also are fertile women with obesity, smoking women, and aged women. With the development of hypertension, patients need to stop taking these drugs and dietary supplements. The decision to cancel other medications should be taken by their doctor.

There are few reports on the prevalence of hypertension in Ethiopia. According to the health and health-related indicators of (MOH 2000–2001), hypertension was the seventh leading cause of death in the country in 2001 (WHO ,2004). The prevalence of hypertension amongst bank employees in Addis Ababa was 18% with 13% in males and 5% in females (Teklu, 1983). A study on the hypertension prevalence and age-related changes in blood pressure in semi nomadic and urban Oromo's showed prevalence of 0.40% in the semi-nomadic and 3.15% in the urban population (Pauletto *et al.*, 1994).

Meta-analysis of 14 randomised trials for hypertension control by Collins et al. estimated that a long-term reduction of 5 –6 mmHg in blood pressure is associated with 35 –40% fewer strokes and 20 –25% less coronary heart diseases (Collins R., et al., 1990). The Seventh report of the Joint National Committee(JNC) on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure showed that a decrease of systolic blood pressure in the population by 5 mmHg would result overall in a 14% reduction in mortality due to stroke, a 9% reduction in mortality due to coronary heart diseases, and a 7% decrease in all-cause mortality (Chobanian A. et al., 2003).

Hypertension is a time to event case so survival analysis is the appropriate method of analyzing these types of cases.

Survival analysis is a statistical method for data analysis where the outcome variable of interest is the time to the occurrence of an event (Klembaum, D. G., 1996). The event can be death, occurrence of a disease, marriage, divorce, malaria, etc. The time to event or survival time can be measured in hour, days, weeks, months, years, etc. Hence, survival analysis is also referred to as "time-to-event analysis", which is applied in a number of applied fields, such as medicine, public health, social science, and engineering. In medical science, time to event can be time until recurrence in a cancer study, time to death, or time until infection. In

the social sciences, interest can lie in analyzing time to events such as job changes, marriage, birth of children and so forth (Lawless, J. F., 1982).

The developments from these diverse fields have for the most part been consolidated into the field of survival analysis. Because these methods have been adapted by researchers in different fields, they also have several different names: event history analysis (sociology), failure time analysis (engineering), duration analysis or transition analysis (economics). These different names do not imply any real difference in techniques, although different disciplines may emphasize slightly different approaches. Survival analysis is the name that is most widely used and recognized (Lee, E. T., and Wang, J. W., 2003).

The analysis of survival data is complicated by issues of censoring and truncation. Censored data arises when an individual's life length only is known to occur in a certain period of time. There are three different types of censoring but right censoring is the most common one in survival analysis.

The aim of this research is also to apply survival techniques to model time-to-event data in case of hypertension diseases based on the data obtained from the follow up record of the patients in Bahir-Dar Felege Hiwet Referral Hospital. Among the several survival models, Cox PH model is one of the most applied models in the field of survival analysis.

Cox Proportional Hazards Model: This is the most popular and widely used method for the analysis of survival data with the strong assumption, proportional hazard (PH) assumption, to evaluate the relationship between covariates and survival with the use of a mathematical model. This is called a semi parametric model because it does not assume any distribution for the baseline hazard (Perperoglou et *al.*, 2007). The most common approach to model covariate effects on survival is the Cox proportional hazards model by Cox (1972), which takes into account the effect of censored observations.

However the Cox PH model may not be appropriate in many situations and other modifications such as stratified Cox model (Klembaum, D. G. 1996) or Cox model with time-dependent variables (Collett, D., 2003) can be used for the analysis of survival data.

Frailty Model: In handling heterogeneity between individuals or within clustering groups, the choice of frailty distribution is very important. It is a random component designed to

account for variability due to unobserved individual-level factors that is otherwise unaccounted for by the other predictors in the model. A random effect is a continuous variable that describes excess risk or frailty for distinct categories such as individuals, families or herds. Frailty changes the individual hazard and is sometimes called *liability* or *susceptibility* in other settings. (Kleinbaum *et al,* 2007). The model is such that, events happen sooner for those who are more frail (Collett D., 2003). Genetic characteristics, growth and living environment are factors that caused difference between the subjects.

1.2. Statement of the Problem

Hypertension is one of the chronic diseases, which is a growing public health problem in both developed and developing countries. It has also been recorded as a common medical disorder and a silent killer to the general public. Each year at least 7.1 million people die as a consequence of hypertension (WHO, 2005). It is a potent risk factor for myocardial infarction, stroke, and heart failure, which are the leading causes of death and disability worldwide.

Hypertension is also controllable with interventions. Health care providers work with hypertensive individuals to control the condition and to prevent the deleterious side effects of uncontrolled hypertension**.**

Even though health professionals try to control blood pressure level, there are many questions which can be raised by everyone how the change is over time or does the change of blood pressure level has different pattern on different covariates and what are the factors that accelerate the blood pressure.

In Ethiopia, to the best of knowledge, there are virtually no studies that documented on time-to event data on hypertension case, except the studies about determinates of hypertension case in Ethiopia based on cross-sectional data.

The traditional Cox PH model has the potential to deal with aspects such as censoring as well as to investigate the effect of explanatory variables directly on the survival time. On the other hand, frailty modeling approach accounts for this problem by specifying independence among observed data items conditional on a set of unobserved or latent variables.

In this study, use both the Cox PH model and its extension individual frailty model to investigate the time-to-good control hypertensive patients using different covariates.

1.3. Objectives of the Study

1.3.1. General objective

The general objective of this study is to explore and identify survival techniques (such as Cox PH and frailty models) to model time-to-good control of hypertension, using data from Bahir-Dar Felege Hiwet Referral Hospital, Bahir-Dar, North Western Ethiopia.

1.3.2. Specific objectives

- To identify important risk factors or covariates which are significantly associated with time-to-good control of hypertension.
- To assess whether there is unobserved heterogeneity between individual hypertensive patients or not, and
- To construct a survival model based on two commonly used modeling approaches in survival analysis, namely Cox PH and frailty models using hypertensive patient's data set.

1.4. Significance of the Study

The findings will help to bring hypertension problem to the agenda of public health policy makers, researchers, and the public at large, so that appropriate treatment and control strategies are implemented along with a population wide surveillance intervention.

Also the outcomes of this study will help health care workers to anticipate and inform patients about the possible related risk factors of good control of hypertensive case they might encounter. In addition, clinicians can improve good control of hypertensive case among hypertensive patients by early diagnosis and appropriate intervention.

The results will also help donors and government to understand risk factors that influence good control of hypertension.

In addition this study will enrich literatures available on the issue and baseline for other studies and may trigger other researchers to conduct similar study in various parts of the country or further study on factors that determine good control of hypertension.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Definitions of Hypertension

Hypertension is a public health problem and a term used to describe HBP and it is usually defined by the presence of a chronic elevation of systemic arterial pressure above a certain threshold value. However, increasing evidence indicates that the cardiovascular risk associated with elevation of blood pressure above approximately 140/90 mm Hg increases in a log-linear fashion (Kannel, 1996). In the Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7) a category of ‘‘pre hypertension’’ was created using BP criteria of 120/80 mmHg to 139⁄89 mm Hg. This category did not emphasize that some individuals with pre hypertension already had the disease, hypertension, while others did not.

According to world health organization blood pressure can be categorized as optimal, normal, high normal and hypertensive. Optimal pressure is below 120/80 mmHg whereas normal is between 120/80 - 130/85 mmHg; readings between 120/80 and 139/89 is called pre-hypertension. High normal is considered to be between 130/85-139/89 mmHg whereas hypertensive is over 140/90 mmHg on repeated measurement and/or treatment with medication (WHO ,2004).

It has been called a silent killer as it is usually without symptoms. Hypertension takes a long time before diagnoses thereby causing major health problems as stroke and other cardiovascular diseases. Damage to organs as the brain, heart, kidneys and eye and so on are the long term effect of high blood pressure disease (Cunha, 2011).

Diagnosis of high blood pressure is usually measured with a device called sphygmomanometer. This consist of an inflatable rubber cuff, an air pump and a column of mercury or a digital readout reflecting pressure in an air column as well as electronic blood pressure machines. The readings are widely expressed in millimeters of mercury or mmHg.

2.2. Causes and Correlates of Hypertension

The main causes for hypertension are: Obesity, Acute pain or Stress, Aging, High intake of sodium (Salt), Hormonal changes in women, Smoking and too much consumption of alcohol, Genetics factors, Hereditary and family history for high blood pressure, Chronic kidney problem, Adrenal and thyroid disorder, Adrenal tumors, stress, Amphetamines, Birth Control pills, Cardiovascular disease, Cushing's disease, Hyperthyroidism and Kidney failure (Cunha et al., 2011).

It is known that high blood pressure usually develops in elderly women after menopause due to hormonal changes (Schofield et al., 1999). However, the occurrence is not a routine part of aging since there are other factors that influence the occurrence (Young L., 2011).

Oliveria *et al.* used a separate logistic regression models to examine the relationship between the baseline potential predictors and whether or not a participant was in target at the 12-month clinic visit. Each model included the predictor, an indicator variable for the SBP target group, and a term capturing the interaction between the predictor and the SBP target group. They found that, among socio-demographic variable age, income, sex, education level, place of residence and caste were significantly related with hypertension. But, family history and marital status were not significant. (Oliveria *et al.*, 2002)

Davarian *et al.* used a linear mixed model in longitudinal study to describe hypertension prevalence rates with increasing age and to examine the link between socio demographic and behavioral factors (including age, gender, education, residence, smoking, and BMI) and measures of blood pressure and overall hypertension in the Japanese population aged ≥28years.(Davarian *et al.*, 2013)

2.3. Signs and Symptoms of Hypertension

Clients may not have symptoms since the onset of hypertension, often called "**the silent killer**", is gradual. In some cases, hypertension is not diagnosed until the person experiences a major complication. Some minor symptoms may include: Consistent BP readings of 140/90 or higher, headache, flushed face, pulsing sensation in the head, dizziness, fatigue, insomnia and nervousness.

2.4. Diagnosis of Hypertension

An evaluation for hypertension by a medical doctor may include: A physical examination, including an accurate medical history, lab blood work (may include kidney profile, thyroid profile, and adrenal gland function), urine analysis, electro cardiogram and chest x-ray.

2.5. Risk factors associated with HBP

According to WHO, deaths as a result to non-communicable diseases as hypertension will increase by 17% over the next decade, with the greatest increase in the African region (27%) (Maher et al., 2010).

In 2003, a cross-sectional study conducted in Ghana recorded high prevalence in women (29.5%) compared to male (27.6%) and low level of awareness. However, focus has been on communicable diseases in developing countries until recently as that similar study conducted in 2006 still showed a high prevalence with 32.3% of participants not having knowledge of the disease (Amoah et al. 2006).

Previous studies have showed hypertension as one of the major causes of maternal death in Ghana (Ghana Maternal Health Survey, 2007). However, to a considerable extent, the growth and effectiveness of reducing maternal death by means of prevention and treatment of hypertension has not been effective even though it can be prevented.

In addition, a research conducted in Uganda concluded that approximately one in every three adult aged 20 years or older was hypertensive. Prevalence of 30.5% and female more hypertensive than males in this study suggested that advancing in ageing was a risk factor due to exposure to lifestyle risk factors of hypertension (Wamala et al., 2009). It is extremely important to investigate on the knowledge and attitude of this target group about the preventive measures of hypertension. Significant reduction in maternal mortality can thus be achieved (Danso et al., 2010).

Holmes and John demonstrated that the progression of hypertension is associated with current smoking, alcohol, physical activity, body mass index, marital status, level of education and age. Smoking is a risk factor in hypertension as it results in the constriction of the blood

vessels, increasing peripheral resistance, and hence elevating the blood pressure. Physical activity is known to lower blood pressure and to be protective against the development of hypertension. Exercise can reduce the obstacles to the flow of blood by increasing the elasticity of the arterial lumen, thus decreasing peripheral resistance (Holmes and John, 2013). Frederico *et al.* assessed the following variables: age, sex, race, urban life quality index, weight, height, and body mass index of hypertensive patients in order to evaluate risk factors associated with increased blood pressure in hypertensive patients using cross-sectional study. (Frederico *et al.*, 2004)

In the study done by Mancia *et al.,* in Tiruvallur district, South India, the following results were obtained. Using multivariate analysis, the variables considered were sex, age, category, education, occupation, body weight at initiation of treatment less than 35 kg, family history, smoking and drinking habits, type of drugs providers, whether patient took treatment under supervision in intensive phase and continuation phase.(Mancia *et al.,* 1999).

In a cross-sectional study by Akilew and Tadesse in Gondar fitted multiple logistic regressions and Odds ratios with 95% confidence intervals were calculated to identify associated factors. The following results were obtained. Age≥55 years, obesity, family history of hypertension, geographical difference, physical inactivity and self-reported diabetes were associated with hypertension. Hence, they recommend the design and implementation of community based screening programs (Akilew and Tadesse, 2012).

2.6. The Effect of HBP

High blood pressure is related to high occurrences of deaths. This is due to the fact it can cause life threatening illness as heart attacks, stroke as well as other disability adjusted life years (DAILYs). Globally, HBP accounted for about 7.6 million deaths (13.5%) in the year 2001. In addition, 92million of the population are globally affected with DAILY. This health burden was greatest for stroke and ischemic heart disease. (Lawes et al. 2001). According to WHO report in 2009 on mortality and burden of disease attributable to selected major risks, it was discovered that the risk of dying was more in low and middle income countries. However, considering the fact that it is a developing country, most participants for the study were unemployed. This could be justified that compliance to pharmacological treatment is unaffordable (Buabeng et al., 2004).

Additionally hypertension research papers report that recent statistics in Canada indicate that 20 percent of all Canadian adults currently suffer from high blood pressure (WHO, 1997).

2.7. Antihypertensive Treatments

Drugs administered for hypertensive patients that are intended to cardiovascular and renal mortality and morbidity. If blood pressure can be reduced, with treatment to <140/90 mmHg there will be a reduced risk of CVD complications. In persons with hypertension and diabetes or renal disease, the blood pressure goal is <130/80 mmHg. Trials have shown that successful antihypertensive treatment can decrease stroke incidence by 35% to 40%; myocardial infarction by 20% to 25%; and heart failure by >50% (Neal B. et al., 2000).

A large number of drugs are currently available for hypertension treatment. There are eight main groups of antihypertensive drugs: diuretics, beta-blockers, angiotensin converting enzyme inhibitors , angiotensin receptor blockers, calcium channel blockers, alpha 1 blockers, central alpha 2 agonists, and direct vasodilators .These eight groups have been tested largely in many clinical trials with the most important goal being to effectively reduce blood pressure and CVD events such as strokes, coronary heart diseases, and heart failure (Neal B. et al., 2000).

Most hypertensive patients cannot be controlled on one drug alone and will require two or more drugs selected from different classes. Once a satisfactory level of blood pressure control is achieved, patients can usually move to longer follow-up intervals for checking the stability of their blood pressure levels.

2.8. Prevention of Hypertension

The prevention and management of hypertension are major public health challenges worldwide. Prevention of hypertension may be pursued through healthy lifestyle changes. Widespread adoption of healthy lifestyles is critical for population prevention of high blood pressure as well as being an important part of treatment for those with hypertension. Healthy lifestyle determinants for preventing hypertension include weight reduction and maintaining normal body weight (BMI 18.5 –22.9kgm-2), moderate or vigorous physical activity, reduced salt intake, moderate alcohol consumption, and adopting a diet high in fruit and vegetables, and lower in dairy products, thus reducing intake of saturated and total

fat (Mancia G. et al., 2007). For overall cardiovascular risk reduction, smoking cessation is recommended for all smokers.

2.9. Cox PH and Frailty Models

A review of literature on survival analysis used in different journals reveals that the Cox PH model is the most widely used way of analyzing survival data in clinical research. Researchers in medical sciences often tend to prefer semi-parametric instead of parametric models because of fewer assumptions. The non-parametric method does not control for covariates and it requires categorical predictors. If the groups are similar, except for the treatment under study, then, the nonparametric methods can be used directly. But we cannot use multiple linear regression or logistic regression because they cannot deal with censored observations. We need another method to model survival data with the presence of censoring. One very popular model in survival data is the Cox proportional hazards model, which is proposed by Cox (1972).

The notation of frailty provides a covenant way to introduce random effect, association and unobserved heterogeneity into models for survival data. In its simplest form, a frailty is an unobserved random proportionality factor that modifies the hazard function of an individual, or of related individuals (Wienke, 2003).

CHAPTER THREE

3. MATERIALS AND METHOD

3.1. Study Area and Period

The study was conducted in Bahir-Dar Felege Hiwet Referral Hospital that is found in Bahir-Dar town, Amhara region, North Western Ethiopia. Bahir-Dar town is located 565 km away from Addis Ababa which is the capital city of Ethiopia. The hospital serves as a teaching and referral center for Bahir-Dar (including rural and urban) area community and, adjacent zones. This study was conducted from June 10, 2014 to June 19, 2014.

3.2. Study Design

An institutional-based retrospective cohort study was conducted.

3.3. Source and Study Populations

The source population were all hypertensive patients who receive treatment from Bahir-Dar Felege Hiwet Referral Hospital. The study population included all hypertensive patients whose age is 18 and above years and, who have followed at least three visits between 1st January, 2009 to last December, 2013 in Bahir-Dar Felege Hiwet Referral Hospital.

3.4. Inclusion and Exclusion Criterion

The study includes hypertensive patients who were recorded in the medical record room of Bahir-Dar Felege Hiwet Referral Hospital from 1st January, 2009 to last December, 2013 and those cards which have the vital data for this study. The patients with incomplete recording of baseline data for the study were excluded from the study.

3.5. Sample Size

From hypertensive patients starting anti-hypertensive treatment from 1st January 2009 to last December, 2013 I was taken 500 sample hypertensive patients.

3.6. Sampling Technique and Procedures

In this study secondary data from Bahir-Dar Felege Hiwet Referral Hospital registry were used to retrieve data of patients from initial date of entry to follow up. Simple random

sampling was used in this study. In this study, the sampling frames are those who had treated during 1st January, 2009 to last December, 2013 in Bahir-Dar Felege Hiwet Referral Hospital. A simple random sampling procedure was applied on the list by using lottery method. During the random sampling process, patients who were not eligible for the study were substituted by the next random number.

3.7. Data Collection Procedures

A data collection check list was used for the data collection. Information was collected from registration forms, follow-up forms and patient cards by three trained peoples and was supervised by the principal investigator. Patient socio-demographic data, blood pressure information were examined and collected carefully.

3.8. Study Variables

3.8.1. Response (Dependent) Variable

The response (dependent) variable is the survival time of hypertensive patients, that is, the length of time from anti-hypertensive drugs start date until the date of good control of hypertension (or censor) measured in months.

3.8.2. Predictor (independent/explanatory/covariate) Variables

Seven covariates (age, sex, residence, systolic blood pressure (SBP), creatinine, blood urea nitrogen (BUN) and fasting blood sugar (FBS)) were used for analyses. These covariates are described together with their values or codes in Table 1.

Table 1: Covariates of time-to-good control of hypertension

Name of covariates	Definitions	Values/Codes
Age	Age at start of treatment	In year
Sex	Sex of hypertensive patients	0=Female,1=Male
Residence	Place of patients	By kebele
SBP	Systolic Blood Pressure	In mmHg
Creatinine	Kidney Function	In mg/dL
BUN	Blood Urea Nitrogen	In mg/dL
FBS	Fasting Blood Sugar	In mg/dL
Time	Observed or Follow up time	In month
Overall assessment	Status of the patient	Event=1, Censor=0

Key: Event=Good control of hypertension through the follow up time.

Desirable range of: BUN is 7–18 mg/dL, Creatinine 0.6–1.2 mg/dL, FBS is 70-130mg/dL and good control of hypertension is SBP of 90-140mmHg.

3.9. Operational Definitions

- **SBP:** is a measurement of BP when the heart pumps the blood (or when the heart beats).
- **Good Control of Hypertension:** is keeping the systolic blood pressure of hypertensive patient between 90-140mmHg.
- **Censor:** When the individual does not experience the event within the specified period of time, or lost to follow-up, that is, an individual may drop out, transfer to other place or deaths due to causes of known or any other reasons.
- **Fasting Blood Sugar (FBS):** The measurement of blood glucose that is used for the diagnosis of diabetes based on plasma glucose levels obtained in the fasting state (greater than 126 mg/dL).
- **Blood Urea Nitrogen (BUN**): It is the measure of urea that is a nitrogenous waste product of protein and amino acid metabolism. The normal range is 7–18 mg/dL. With declining renal function, BUN levels increase. Measuring BUN therefore provides an indication of kidney health.

- **Creatinine:** is a product of muscle breakdown. The normal range is 0.6–1.2 mg/dL Serum creatinine is very indicative of renal function. The clearance of creatinine may be used to estimate GFR.

3.10. Data Quality Control

For the purpose of data quality control there were two days training for data collectors, data encoder and data clerk personnel. Intensive supervision was done by the principal investigator. All the data were cleaned, double entry and cross checked for their completeness before analysis. Random samples of registration forms were reviewed by principal investigator to conform reliability of data before data collection.

3.11. Methods of Data Analysis

3.11.1. Survival Analysis

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. Time ≡ Survival time (i.e. years, months, weeks, days, hours, minutes etc. from the beginning of follow-up of an individual until an event occurs) and Event ≡ Failure (i.e. death, disease incidence, relapse from remission, recovery or any designated experience of interest that may happen to an individual).

Survival analysis is the name for a collection of statistical techniques used to describe and quantify time to event data. In survival analysis we use the term ‘failure’ to define the occurrence of the event of interest. The term ‘survival time’ specifies the length of time taken for failure to occur.

In order to analyses the survival time (failure time random variable) we need the following:

i) Time origin, ii) Time scale and iii) Definition of an event

3.11.1.1. Censored Data

Most survival analyses must consider a key analytical problem called censoring. In essence, censoring occurs when we have some information about individual survival time, but we don't know the survival time exactly.

There are generally three reasons why censoring may occur:

- A person does not experience the event before the study ends;
- A person is lost to follow-up during the study period;
- A person withdraws from the study because of death (if death is not the event of interest) or some other reason (e.g., adverse drug reaction or other competing risk)

There are three types of censoring:

1) **Right censoring**: a subject is right censored if it is known that the event of interest occurs sometime after the recorded follow-up period. It is the most common types of censoring in survival analysis.

2) **Left censoring**: a subject is left censored if it is known that the event of interest occurs some time before the recorded follow-up period.

3) **Interval censoring**: a subject is interval censored if it is known that the event of interest occurs between two times, but the exact time of failure is not known.

In this study we use right censoring and non- informative (independent) censoring. Non-informative censoring means that the survival time is independent of the censoring time.

3.11.1.2. Survival Functions

Let T be a positive random variable representing the time until the relevant event occurs. In order to characterize the distribution of T one of the most often used functions is survivor function.

The survivor function, S (t): is defined for both discrete and continuous distribution as the probability that an individual survives beyond time t. i.e.

For a discrete random variable

$$S(t) = Pr(T \geq t), \quad 0 < t < \infty \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (3.11.1.2.1)$$

S(0)=1 and s(∞)=0 => 0≤ S(t) ≤ 1

For a continuous random variable T:

$$S(t) = \int_t^\infty f(x)dx \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.11.1.2.2)$$

Where; f(t) is the density function for continuous random variable T

$S(t) = 1 - F(t)$ where; $F(t) = Pr(T \leq t)$ and F(t) is the cumulative distribution function

- Survival time is a non-increasing function means (either decreasing or constant but not increasing).

The hazard function, $\lambda(t)$: sometimes called instantaneous failure rate, the force of mortality, or the age-specific failure rate, is defined as the conditional probability of failure at time t given that the individual has survived up to time t. It has any shape and is non-negative.

For continuous random variable

$$\lambda(t) = lim_{\Delta t \to \infty} \frac{1}{\Delta t} Pr(t < T < t + \Delta t | T \geq t) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.11.1.2.3)$$

$$\lambda(t) = \frac{f(t)}{S(t)}$$

The cumulative hazard function, $\Lambda(t)$: For continuous random variable is given by

$$\Lambda(t) = \int_0^t \lambda(x)dx \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.11.1.2.4)$$

$$\Lambda(t) = -\ Log\{S(t)\}$$

3.11.2. Non-Parametric Methods

Once we have collected time to event data, our first task is to describe it -usually this is done graphically using a survival curve. Visualization allows us to appreciate temporal pattern in the data. It also helps us to identify an appropriate distributional form for the data. If the data

are consistent with a parametric distribution, then parameters can be derived to efficiently describe the survival pattern and statistical inference can be based on the chosen distribution. Non-parametric methods are used when no theoretical distribution adequately fits the data. In epidemiology non-parametric (or semi-parametric) methods are used more frequently than parametric methods.

There are three non-parametric methods for describing time to event data:

- The Kaplan-Meier method,
- The life table method, and
- The Nelson-Aalen method

Kaplan-Meier method: The Kaplan-Meier (KM) estimator is the standard non parametric estimator of the survival function and is also called the Product-Limit estimator. KM method is based on individual survival times and assumes that censoring is independent of survival time (that is, the reason an observation is censored is unrelated to the cause of failure). Kaplan and Meier (1958) proposed an estimator called as Kaplan-Meier (K-M) Product Limit estimator which provides quick, simple estimates of the survival function or the cumulative distribution function (CDF) based on failure data that may even be multi censored. No underlying model (such as Weibull or lognormal) is assumed. Exact times of failure are required. Assume that we have n individuals on test and order the observed lifetimes for these n individuals from t_1 to t_n. Some of these are actual failure times and some are running times for individuals taken off test before they got an event. Suppose there r deaths have occurred, and the ordered death times are $t_{(1)},\ldots,t_{(r)}$, where $r \leq n$. The number of individuals who are not got an event just before time $t_{(j)}$, including those who are about to got an event at this time, will be denoted by n_j, j = 1,2,...,r, and d_j will denote the number who got an event at this time. The Kaplan-Meier estimate of the survival function, which is given by

$$\hat{S}(t) = \prod_{j=1}^{k} \frac{(n_j - d_j)}{n_j} \quad \text{..... (3.11.2.1)}$$

Life table method: The life table method (also known as the actuarial or Cutler Ederer method) is an approximation of the KM method. It is based on grouped survival times and is suitable for large data sets.

Nelson-Aalen method: An alternative estimator for the Kaplan-Meier estimator and

Log-rank Test: is the one commonly used non-parametric test for comparison of two or more survival distributions (curves). Instead of looking at fixed time points, we want to compare the whole survival function of different groups, H0: $S_1(t) = S_2(t) = \ldots = S_K(t)$

Since the true survival functions are unknown in each group, we go for a nonparametric test .Let $n_1$and n_2 be the number on individuals in the group 1 and 2, respectively and $n=n_1 + n_2$. Let n_{1i} and n_{2i} be the number of individuals at risk just prior to $t_{(i)}$ from the treatments 1 and 2 and d_{1i} and d_{2i} be the number of deaths at $t_{(i)}$ among the individuals in group 1 and group 2 and $d_{1i} + d_{2i} = d_i$; $n_{1i} + n_{2i} = n_i$. The log rank statistic is given by

$$X_{LR} = \frac{\left[\sum_{i=1}^{n}(d_{1i} - e_{1i})\right]^2}{\sum_{i=1}^{r} V_{1i}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.11.2.2)}$$

where $e_{1i} = {n_{1i}d_i}/{n_i}$, $V_{1i} = \frac{n_{1i}n_{2i}d_i(n_i - d_i)}{n_i^2(n_i - 1)}$, i=1,2,….,r

X_{LR} is the log rank statistic which is distributed approximately central chi-square distribution with one degree of freedom when the null hypothesis is true and the sample size is moderate.

3.11.3. Semi-Parametric Methods

3.11.3.1. Cox Proportional Hazard Model

The most common approach to model covariate effects on survival is the Cox proportional hazards model by Cox (1972), which takes into account the effect of censored observations. The Cox proportional hazards model is a semi-parametric model where the baseline hazard is allowed to vary with time. This model is semi-parametric because while the baseline hazard can take any form, the covariates enter the model linearly. In the proportional hazards models, we study the influence of covariates through the conditional hazard function. The Cox Proportional Hazards model by assuming the conditional hazard function of the lifetime T for covariate values $x_1, \ldots, x_p$ is given by;

$$\lambda(t|X) = \lambda_0(t) * \exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right)$$

$$= \lambda_0(t) * \exp(\beta^t x) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.11.3.1.1)}$$

where,

- $\lambda(t|X)$-is the hazard function at time t for a subject with covariate values $x_1, \ldots x_p$,

- $\lambda_0(t)$-is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero.
- $X = (x_1, x_2, \ldots, x_p)^t$ -is the values of the vector of explanatory variables for a particular individual, and
- $\beta^t = (\beta_1, \beta_2, \ldots, \beta_p)$- is a vector of regression coefficients.

So e^{β} describes the proportional change of the hazard due to the increase of X by one unit. We note that

$\beta > 0$ => hazard increases and $\beta < 0$ => hazard decreases.

This model, also known as the Cox regression model, makes no assumptions about the form of $\lambda_0(t)$ (non-parametric part of model) but assumes parametric form for the effect of the predictors on the hazard (parametric part of model). The model is therefore referred to as a semi-parametric model.

3.11.3.1.1. Hazard Ratio

The measure of effect is called hazard ratio. The Cox model is called a proportional hazards model since the ratio of the hazard rates of two individuals with covariate values X and X* is an expression that does not depend on t. Hazard ratio is given by;

$$HR(t) = \frac{\lambda(t|X)}{\lambda(t|X^*)} = \frac{\lambda_0(t)e^{\beta^t x}}{\lambda_0(t)e^{\beta^t x^*}} = e^{\sum_{i=1}^{p} \beta_i (X_i - X_i^*)} \quad \text{(3.11.3.1.1.1)}$$

This is a constant over time. HR compares the hazard of having an event with covariate value X to the hazard of having an event with covariate value X*. Often, the ratio is (inappropriately) called a relative risk. The hazard ratio states the effect of increasing the level of the covariate by one unit.

3.11.3.1.2. Assumptions of CPH Model

• The ratio of the hazard function for two individuals with different sets of covariates does not depend on time. It is very important to verify that the covariates satisfy the assumption of proportionality.

⇨ The Cox PH model assumes that the hazard ratio comparing any two specifications of predictors is constant over time. Equivalently, this means that the hazard for one

individual is proportional to the hazard for any other individual, where the proportionality constant is independent of time.

- Time is measured on a continuous scale.
- Censoring occurs randomly.

The assessment of the proportional hazards assumption can be done numerically or graphically, and a large number of procedures have been proposed over the years. Some authors recommend using numerical tests (e.g. Hosmer and Lemeshow ,1999) and others recommend graphical procedures since they believe that the proportional hazards assumption only approximates the correct model for a covariate and that any formal test, based on a large enough sample, will reject the null hypothesis of proportionality (Klein and Moeschberger ,1997).

The Cox model and its various generalizations are mainly used in the medical and bio-statistical field.

3.11.3.1.3. Estimation for CPH Model

Partial Likelihood estimate: is a technique developed to estimate about the regression parameters in the presence of nuisance parameters ($\lambda_0(t)$ in the Cox PH model). To estimate $\lambda_0(t)$ and β one approach is to attempt to maximize the likelihood function for the observed data simultaneously with respect to $\lambda_0(t)$ and β.

Let $t_1, t_2, \ldots, t_n$ be the observed survival time for n individuals and δ_i be the event indicator, which is zero if the i^{th} survival time is censored, and unity otherwise. Let the ordered event time of r individuals be $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$ and let $R(t_{(j)})$ be the risk set just before $t_{(j)}$ and r_j for its size. So that $R(t_{(j)})$ is the group of individuals who are not got an event and un censored at a time just prior to $t_{(j)}$. The conditional probability that the i^{th} individual got an event at $t_{(j)}$ given that one individual from the risk set on $R(t_{(j)})$ got an event at $t_{(j)}$ is:

P (individual i got an event at $t_{(j)}$| one event from the risk set $R(t_{(j)})$ at $t_{(j)}$)

$$= \frac{\mathbf{P(individual\ i\ got\ an\ event\ at\ t_{(j)})}}{\mathbf{P(one\ event\ at\ t_{(j)})}} \qquad \ldots (3.11.3.1.3.1)$$

Then the partial likelihood function for the Cox PH model for censored data is given by:

$$L(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(\beta^t X_i(t_i))}{\sum_{p \epsilon R(t_i)} \exp(\beta^t X_p(t_i))} \right]^{\delta_i} \quad \text{..... (3.11.3.1.3.2)}$$

Where R (t_i) is the risk set at time t_i.

The partial likelihood is valid when there are no ties in the dataset. That means there is no two subjects who have the same event time.

The full likelihood function for right censored data can be constructed as

$$L(\beta) = \prod_{i=1}^{n} h(t_i, x_i, \beta)^{\delta i} S(t_i, x_i, \beta) \text{.....(3.11.3.1.3.3)}$$

Where, $h(t_i, x_i, \beta) = h_o(t_i) \exp(\beta^T xi)$ is the hazard function for individual i and

$S(t_i, x_i, \beta) = (S_0(t_i))^{\exp(\beta^T xi)}$ is the survival function for individual i.

It follows that $L(\beta) = \prod_{i=1}^{n} (h_0(t_i) \exp(\beta^T xi))^{\delta i} (S_0(t_i))^{\exp(\beta^T x_i)}$

The full maximum likelihood estimator of β can be obtained by differentiating L(β) with respect to the components of β and the base line hazard. This implies that unless we explicitly specify the base line hazard, as in the case of parametric PH, we cannot obtain the maximum likelihood estimators from the full likelihood. To avoid the specification of the base line hazard, (Cox, 1972) proposed a partial likelihood approach that treats the baseline hazard as a nuisance parameter and removes it from the estimating equation. Instead of constructing a full likelihood, we consider the probability that an individual experiences an event at time t_i given that an event occurred at that time.

3.11.4. PH Assumption Checking

The main assumption of the Cox proportional hazards model is proportional hazards. Proportional hazards means that the hazard function of one individual is proportional to the hazard function of the other individual, i.e., the hazard ratio is constant over time. The assessment of the proportional hazards assumption can be done numerically or graphically, a great number of procedures have been proposed over the years. There are two general approaches for assessing the PH assumption:

i) Graphical approach and

ii) Goodness-of-fit test approach

i) Graphical Approach

We can obtain Cox PH survival function by the relationship between hazard function and survival function. The most popular graphical approach involves

- The use of “log–log” survival curves: The most popular of these involves comparing estimated log(–log) survivor curves over different (combinations of) categories of variables being investigated. By plotting estimated log(-log(survival)) versus survival time for two groups we would see parallel curves if the hazards are proportional. This method does not work well for continuous predictors or categorical predictors that have many levels because the graph becomes "clustered".
- The comparison of “observed” with “expected” survival curves: An alternative graphical approach is to compare observed with predicted survivor curves. The observed curves are derived for categories of the variable being assessed, without putting this variable in a PH model. The predicted curves are derived with this variable included in a PH model. If observed and predicted curves are close, then the PH assumption is reasonable.

⇨ If expected curves “look the same” as the observed, we deem the PH assumption to be okay. If they “look different”, we deem it to be false.

ii) Goodness-of-fit test Approach

It uses a test statistic or equivalent p-value to assess the significance of the PH assumption. This approach provides large sample Z or chi-square statistics which can be computed for each variable in the model, adjusted for the other variables in the model. A p-value derived from a standard normal statistic is also given for each variable. This p-value is used for evaluating the PH assumption for that variable. A non-significant (i.e., large) p-value, say greater than 0.10, suggest that the PH assumption is reasonable (satisfied), whereas a small p-value, say less than 0.05, suggests that the variable being tested does not satisfy the PH assumption.

A number of different tests for assessing the PH assumption have been proposed in the literature. We present the test of Harrel and Lee (1986), a variation of a test originally

proposed by Schoenfeld (1982) and based on the residuals defined by Schoenfeld, now called the Schoenfeld residuals.

Numerically: More conveniently, the cox.zph function calculates tests of the proportional-hazards assumption for each covariate, by correlating the corresponding set of scaled Schoenfeld residuals with a suitable transformation of time (the default is based on the Kaplan-Meier estimate of the survival function). The probability value of PH (p(PH)) and check the PH assumption is satisfied. If P(PH)<5% then the PH assumption violated else the PH assumption satisfied.

3.11.5. Model Diagnostics for CPH

After fitting a Cox's regression model to a practical data set, it is important to check whether the Cox's regression model is a good fit model for this data set. The use of diagnostic procedures for model checking is an essential part of the modeling process to check whether the fitted model is correct or not. In linear regression methods, residuals are defined as the difference between the observed and predicted values of the dependent variable. However, when censored observations are present and partial likelihood function is used in the Cox PH model, the usual concept of residual is not applicable. Among a number of residuals that have been proposed for use in connection with the Cox PH model, four major residuals are: Cox-Snell residual, martingale, deviance residual and Schoenfeld residual.

i) Cox-Snell residuals

The residual that is most widely used in the analysis of survival data is the Cox-Snell residual, so called because it is a particular example of the general definition of residuals given by Cox and Snell (1968). The Cox-Snell residual for the i^{th} individual, $i = 1,2,\dots,n$, is given by Properties and features of residuals, when survival outcome are modeled, have been extensively studied in the literature. The Cox -Snell residuals are commonly used for a direct assessment of excess events (i.e., to reveal subjects that are poorly fit by the model), and for evaluating whether the appropriate functional form for a covariate is used in the model. These Cox-Snell residuals are defined by:

$$r_i = \widehat{H}_0(t_i) * \exp(x_i\beta_i) \quad \text{(3.11.5.1)}$$

Where $\widehat{H}_0(t_i)$ is the Breslow estimator of the baseline cumulative hazard function at t_i.

If the Cox's regression model holds, we get that r_i is a censored sample of an exponential distribution with $\lambda = 1$. Therefore, we use a plot of $H(r_i)$ versus r_i to check the fit of the model. This gives a straight line with unit slope and zero intercept if the fitted model is correct.

Note: - the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be negative.

ii) Martingale residuals

Alternatively, by plotting the martingale residuals versus a covariate, we can verify whether the functional form is correct. For every individual, i = 1,…, n, the Martingale residual is defined by:

$$r_{mi} = \delta_i - \widehat{H}(t_i) = \delta_i - \widehat{H}_0(t_i) * \exp(x_i\beta_i) = \delta_i - \text{ri} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \quad (3.11.5.2)$$

Where $\widehat{H}$ is the fitted cumulative hazard function under the Cox's regression model and we note that

- The martingale residuals sum to zero
- In large sample, the martingale residuals are uncorrelated and have an expected value of zero.
- The martingale residuals take values between negative infinity and unity and
- δ_i = 1 for uncensored observation, δ_i= 0 for censored observation.

iii) Deviance residuals

The Deviance residual is defined by:

$$r_{Di} = sgn(r_{mi})[-2\{r_{mi} + \delta_i \log(\delta_i - r_{mi})\}]^{1/2} \ldots \ldots \ldots \ldots \quad (3.11.5.3)$$

Where r_{mi} is the martingale residual for the i^{th} individual, and the function sgn(.) is the sign function. This is the function that takes the value 1 if r_{mi} is positive and -1 if r_{mi} is negative. Thus, sign (r_{mi}) ensures that the deviance residuals have the same sign as the martingale residuals. The deviance residuals are a normalized transform of the martingale residuals. They also have a mean of zero but are approximately symmetrically distributed about zero when the fitted model is appropriate. It is known that the deviance residuals are symmetrically distributed about zero when the fitted model is adequate; Individuals with large positive or negative deviance residuals are poorly predicted by the model.

iv) Schoenfeld residuals

All the above three residuals are residuals for each individual. We will describe covariate wise residuals: Schoenfeld residuals. The Schoenfeld residuals were originally called partial residuals because the Schoenfeld residuals for i^{th} individual on the j^{th} explanatory variable X_j is an estimate of the i^{th} component of the first derivative of the logarithm of the partial likelihood function with respect to β_j. For each predictor in the model, Schoenfeld residuals are defined for every subject who has an event. Consider a Cox PH model with p predictors: $x_1, x_2, ..., x_p$. Then there are p Schoenfeld residuals defined for each subject who has an event, one for each of the p predictors.

The Schoenfeld residual for i^{th} individual on X_j is given by:

$$r_{pji} = \delta_i\{x_{ji} - a_{ji}\} \quad \text{..... (3.11.5.4)}$$

Where j=1,2, … , p and $a_{ji} = \frac{\sum_{l\in R(t_i)} x_{jl}\exp(\beta^t X_l)}{\sum_{l\in R(t_i)} \exp(\beta^t X_l)}$

- The Schoenfeld residuals sum to zero.

3.11.6. Strategies for Analysis of Non-Proportional Data

Suppose that statistical tests or other diagnostic techniques give strong evidence of non-proportionality for one or more covariates. To deal with this we will describe stratified Cox model.

3.11.6.1. Stratified Cox Model

The "stratified Cox model" is a modification of the Cox proportional hazards (PH) model that allows for control by "stratification" of a predictor that does not satisfy the PH assumption. Predictors that are assumed to satisfy the PH assumption are included in the model, whereas the predictor being stratified is not included.

We assume that we have k variables not satisfying the PH assumption and p variables satisfying the PH assumption. We denote the variables not satisfying the PH assumption by $Z_1, Z_2, ..., Z_k$, and the variables satisfying the PH assumption by $X_1, X_2, ..., X_p$. To perform the stratified Cox procedure, we define a single new variable, which we call Z^*, from the Z's to be used for stratification. We do this by forming categories of each Z_i, including those Z_i that are interval variables. We then form combinations of categories, and these combinations

are our strata. These strata are the categories of the new variable Z*. In general, the stratification variable Z∗ will have k∗ categories, where k∗ is the total number of combinations (or strata) formed after categorizing each of the Z's.

The general stratified Cox model is:

$$\lambda_g(t|X) = \lambda_{0g}(t) * \exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.11.6.1.1)$$

Where;

g-represents the number of stratum => g= 1, 2, …, k* strata defined from Z*.

Z*-not included in the model.

Different baseline hazard functions: $\lambda_{0g}(t)$, g=1,2,…, k* and Same coefficients: $\beta_1, \beta_2, \ldots, \beta_p$.

3.11.7. Frailty Models

Frailty: is a random component designed to account for variability due to unobserved individual-level factors that is otherwise unaccounted for by the other predictors in the model. A random effect is a continuous variable that describes frailty for distinct categories such as individuals, families or herds. Frailty changes the individual hazard and is sometimes called *liability* or *susceptibility* in other settings. The frailty approach is a statistical modeling concept which aims to account for heterogeneity, caused by unmeasured covariates.

The idea is that individuals have different frailties, and those who are most 'frail' will experience failure earlier than others. The inclusion of the frailty term in a Cox model allows for the possible correlation between the recurrence times of an individual. The concept of frailty provides a suitable way to introduce random effects in the model to account for association and unobserved heterogeneity. This heterogeneity is often referred to as variability, and it is one of the important sources of variability in medical, epidemiological and biological applications. In its simplest form, a frailty is an unobserved random factor that modifies multiplicatively the hazard function of an individual or a group or cluster of individuals. Vaupel et al., (1979) introduced the term frailty and used it in univariate survival models. Clayton, (1978) promoted the model by its application to multivariate situation on chronic disease incidence in families.

A random effect model takes into account the effects of unobserved or unobservable heterogeneity, caused by different sources. The random effect, called frailty and denoted here by a time-independent non-negative random variable *Z* is the term that describes the common risk or the individual heterogeneity, acting as a factor on the hazard function. Two categories of frailty models can be pointed out.

i) The class of univariate frailty models that consider univariate survival times.
ii) The class of multivariate frailty models that take into account multivariate survival times.

Sometimes because we do not know the values of the factor for each individual or sometimes we are not aware that there exists factors that we ought to include. Genetic variation, residence, heritability, and other properties of individual susceptibility to an event can now be analyzed using correlated frailty models (Yashin and Iachine, 1995a, b; Yashin and Iachine, 1997).

3.11.7.1. Univariate (Individual) Frailty Model

Univariate frailty models take into account that the population is not homogeneous. Heterogeneity may be explained by covariates, but when important covariates have not been observed, this leads to unobserved heterogeneity. Vaupel et al., (1979) introduced univariate frailty models (with a gamma distribution) into survival analysis to account for unobserved heterogeneity or missing covariates in the study population.

Suppose we have a sample of j observations where some observations are more failure prone due to reasons unknown (or unmeasured) but go ahead and estimate a garden variety model like this one:

$$h(t_j) = h_0(t) * \exp(\beta^t X_j)$$

In this model (a PH model) the hazard is increasing or decreasing as a function of x.

The Problem: If there are unmeasured or unobserved "frailties," the hazard rate will not only be a function of the covariates, but also a function of the frailties:

$$h(t_j) = h_0(t) * \exp(\beta^t X_j + \psi w_j) \quad \text{(3.11.7.1.1)}$$

Where w_j are the frailties and are assumed to be an independent sample from a distribution with mean 0 and variance 1 (Klein and Moeschberger 1997). (That is, they follow some distribution function).

Note-a couple of important things here:
1) If $\psi = 0$, then the standard proportional hazards model is obtained;
2) If the relevant factors comprising w_j could be measured, then ψ would go to 0.

An important point is that the frailty Z is an unobservable random variable varying over the sample which increases the individual risk if $Z > 1$ or decreases if $Z < 1$.

A tractable model to account for heterogeneity can be derived if one is willing to make some assumptions regarding the distribution of the frailty. To see this, let's rewrite our model to show how the frailties act multiplicatively on the hazard:

$$h_j(t|\beta^t X_j, Z_j) = h_0(t) * Z_j * \exp(\beta^t X_j) \quad \text{(3.11.7.1.2)}$$

(Note that $Z_j = \exp(\psi^t w_j)$.

For identification purposes, it is conventionally assumed that the mean of Z is 1 and the variance is unknown and equal to some parameter θ.

Note that we always make assumptions about Z in standard non frailty models. We assume Z to be 1 with probability 1! (Frailty may exist; we choose to ignore it.)
If the hazard is a function of the frailties, the survivor function must also be conditional on both the covariates and on the frailty term.

$$S(t) = \exp\left(-\int_0^t h(u|z)\,du\right) = \exp(-z\int_0^t h(u)\,du) \quad \text{(3.11.7.1.3)}$$

Many calculations can be done based on the Laplace transform. Hougaard (1984) demonstrated the importance of the Laplace transform for these calculations. The Laplace transform of a random variable Z is defined as

$$L(u) = \int \exp(-sz)\, g(Z)dz = E[\exp(-sZ)],$$

where $g(z)$ is the density of Z. The integral is over the range of the distribution, and the marginal survivor function is given by

$$S(t) = L[\exp(\int_0^t h(u)\,du)] \quad \text{.....(3.11.7.1.4)}$$

where L is the Laplace transformation Hougaard 2000 refers to this distribution as the "marginal survivor function" because it is the observed survivor function after Z has been integrated out. To derive the expected value of the survivor function, we need to specify a probability distribution for Z, call this g (z). These include: gamma, inverse Gaussian, log-normal, and power variance model. The gamma has most readily been adopted in applied research.

An important point is the identifiability of univariate frailty models. Univariate frailty models are not identifiable from the survival information alone. However, (Elbers and Ridder, 1982) proved that a frailty model with finite mean is identifiable with univariate data, when covariates are included in the model.

- The problem with ignoring frailties is seen in the hazard. In the PH models, the hazard is a multiplicative function of the measured covariates.
- With frailty, the hazard is also a function of z.

Now suppose that the gamma distribution is specified for g(z). We can define the gamma distribution as g (z, α, β) where α = 1/θ and β = θ. The density function for the gamma is then given by:

$$g(z, \alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha) Z^{\alpha-1} e^{-z/\beta}} \quad \text{.....(3.11.7.1.5)}$$

Where Γ(α) is the gamma integral ($\int_0^{\infty} Z^{\alpha-1} * e^{-z}$).

3.12. Limitations of the Study

The limitations of this study are:

- In Ethiopian context time to event data were not extracted well and it is very limited to specific area. As an option, it is preferred to extract data from medical cards of those already visited and registered at the respective hospital. There are many prognostic factors of hypertensive patients, such as; nutritional status, race, co-morbid

diseases, alcohol use, smoking status, weight, body mass index, level of education, marital status, exercise and others. In this thesis, it is limited only to seven covariates. Since all the necessary variables were not recorded on the patient's card.

- The other limitation is lack of published literature to compare and contrast the findings of this study obtained through modeling survival analysis of time to event data in the local context.
- Moreover, the investigator has both financial and time limitations in carrying out this study on a larger scale, even though a larger study would yield a better understanding of the problem.

CHAPTER FOUR

4. RESULT AND DISCUSSION

4.1. Descriptive Statistics and Non-Parametric Analysis

From a sample of 500 Hypertensive patients 205 (41%) patients have events (good control of hypertension) and 295 (59%) patients were censored observations. Among these sample patients 263 (52.6%) patients are females and 237 (47.4%) are males.

Table 2: Summary table for time to good control of hypertension

Covari-ates	Min.	Max.	Mean	Std. Error of Mean	Variance	Std. Deviation	Q1	Median	Q3
Time	9	60	31.99	0.635	201.491	14.195	21	30	42
Age	18	95	54.2	0.785	307.941	17.548	42	55	65
SBP	145	240	169.8	0.779	303.667	17.426	160	170	180
FBS	27	343	129.57	2.539	3223	56.775	94	114.5	155
BUN	8	310	42.2	1.337	893.859	29.897	23	35	56
Creatinine	0.1	10.6	1.356	0.039	0.776	0.8809	0.80	1.10	1.70

The mean and median follow up time of the patients are 31.99 and 30 months respectively with standard deviation of 14.195. The mean and the median age of the patients are 54.2 and 55 years. The average value of systolic blood pressure is 169.8 mmHg and the median is 170 mmHg.

The average fasting blood sugar level of the patients is 129.57 mg/dL and the median is 114.5 mg/dL. The mean blood urea nitrogen level of the patient is 42.2 mg/dL and the median is 35 mg/dL. The mean creatinine value of the patient is 1.356 mg/dL and the median is 1.1 mg/dL. The minimum survival time was 9 months and the maximum survival time was 60 months (Table 2).

Table 3: Cross tabulation of Sex, Age and SBP of hypertensive patients with Status and Time

Cova-riates	Group	Status		Time (Month)		
		Event (%)	**Censor(%)**	**≤ 30(%)**	**> 30(%)**	**Total (%)**
Sex	Female	108(52.7)	155(52.5)	133(50.6)	130(54.9)	263(52.6)
	Male	97(47.3)	140(47.5)	130(49.4)	107(45.1)	237(47.4)
	Total	205(100)	295(100)	263(100)	237(100)	500(100)
Age	≤ 45	75(36.6)	84(28.5)	93(35.4)	66(27.8)	159(31.8)
	>45	130(63.4)	211(71.5)	170(64.6)	171(72.2)	341(68.2)
	Total	205(100)	295(100)	263(100)	237(100)	500(100)
SBP	≤ 165	103(50.2)	121(41.0)	160(60.8)	64(27)	224(44.8)
	>165	102(49.8)	174(59.0)	103(39.2)	173(73.0)	276(55.2)
	Total	205(100)	295(100)	263(100)	237(100)	500(100)

Among 205 event observations, 108 (52.7%) patients are females and 97 (47.3%) are males and from 295 censored observations, 155 (52.5%) patients are females and 140 (47.5%) are males during the study period of time.

From 500 sampled cases, 263 patients have follow up time of ≤ 30 months and 237 patients have follow up time of >30 months, among these 133 (50.6%) patients are males and 130 (49.4%) are females who have ≤ 30 months follow up time and 130 (54.9%) male patients and 107 (45.1%) females have >30 months follow up time. (Table 3).

From the result obtained the median survival time of hypertension patients to have good control is 48 months (4 years).($\hat{S}_{(t=48)} = 0.5$) and the mean survival time is 43.6 months (3.63 years).

Table 4: Summary table of survival time for hypertensive patients

Covariates	Group	Event (number)	Mean (month)	Se(Mean) (month)	Median (month)
Sex	Male	97	43.2	1.19	45
	Female	108	43.9	1.15	51
Age (Years)	≤ 45	75	40.7	1.51	45
	>45	130	44.9	0.975	51
SBP(mmHg)	≤ 165	103	35.9	1.186	39
	>165	102	47	0.861	54
FBS (mg/dL)	≤185	177	42.8	0.892	45
	>185	28	47.1	2.036	57
BUN (mg/dL)	≤35	114	42.3	1.16	45
	>35	91	44.9	1.16	51
Creatinine (mg/dL)	≤ 1.5	125	44.8	1.03	51
	>1.5	80	41.3	1.36	42

For male group, the estimated mean and median survival times to attain good control of hypertension are 43.2 and 45 months respectively. For the female group, these figures are 43.9 and 51 months respectively. Comparison of the two estimated mean and medians reinforces the observation that the male is more effective overall than the female.

For Systolic Blood Pressure of ≤ 165 mmHg groups, the estimated mean and median survival times to attain good control of hypertension are 35.9 and 39 months respectively. For SBP > 165 mmHg group, these figures are 47 and 54 months respectively. Comparison of the two estimated mean and medians reinforces our previous observation that the patient with SBP ≤ 165 mmHg is more effective overall than the patient with SBP > 165 mmHg (Table 4).

4.2. Non-Parametric Analysis

4.2.1. Kaplan-Meier Survival Curves

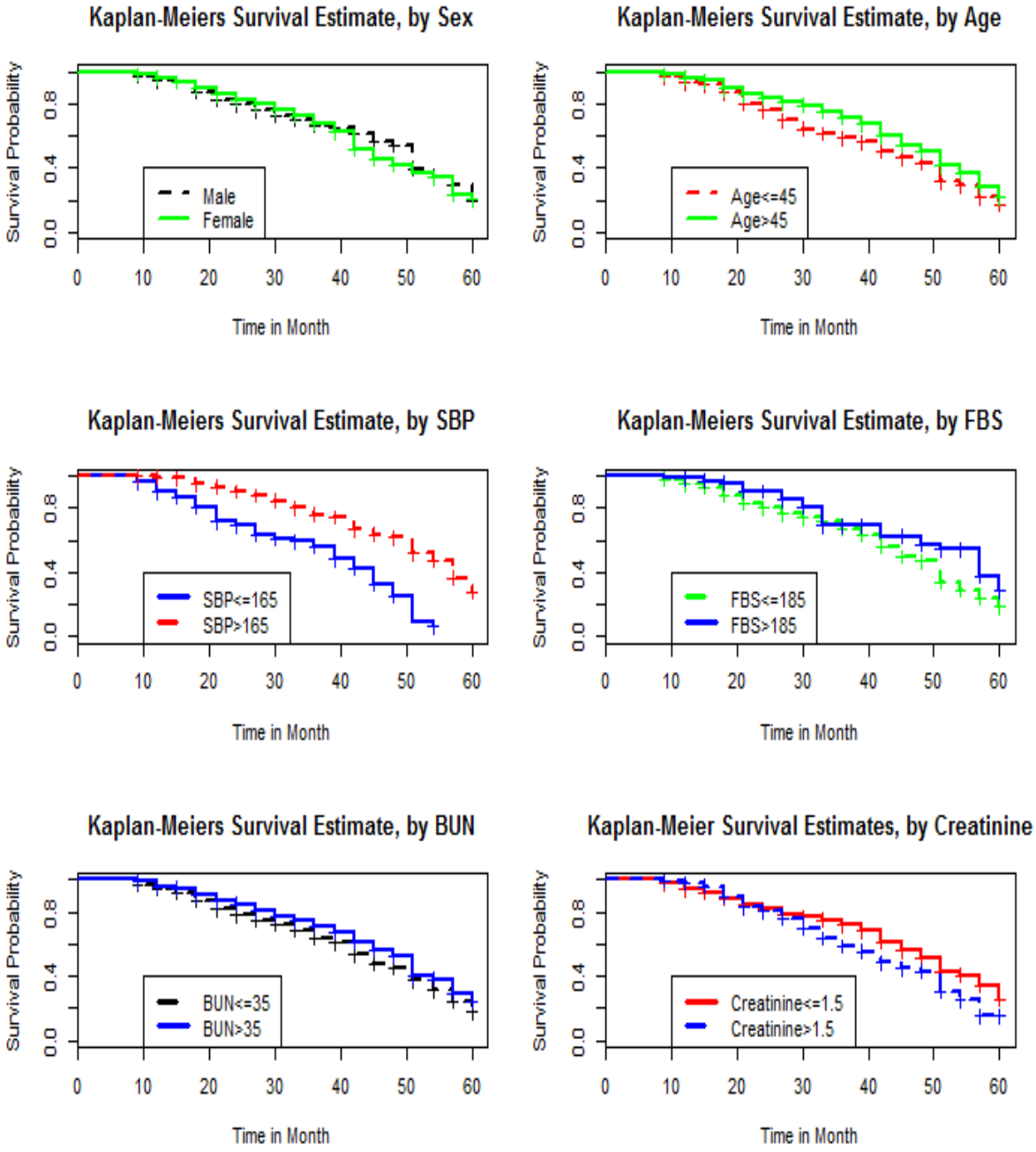

Figure 2: Kaplan- Meier estimates for Hypertension data set by Sex, Age, SBP, FBS, BUN and Creatinine

From the above KM curve, the pattern of female lying above male means that the female group attains good control of hypertension faster than male up to 40 month, and after 40 month, the male group attains good control of hypertension faster. In other words, until 40 months the proportion of subject attaining good control of hypertension is greater for females than males.

The pattern of patient age > 45 years lying above age ≤ 45 years means the age group > 45 years attain good control of hypertension faster, in other words, at any point in time the proportion of subjects attaining good control of hypertension is greater for age >45 years than age ≤ 45 years.

The pattern of patient SBP >165 mmHg lying above SBP ≤ 165 mmHg means the SBP group > 165 mmHg become good control of hypertension faster, in other words, at any point in time the proportion of subjects estimated to be good control of hypertension is greater for SBP >165 mmHg than SBP ≤ 165 mmHg.

4.3. Comparison of Survival Curves

4.3.1. Log-Rank-Test for Equality of Survival Functions

The null hypothesis of interest is no difference between survival curves.

The result indicates that there is a significant difference between the survival curves between two groups of age (Age ≤ 45 years and Age>45 years), SBP (SBP ≤ 165 mmHg and SBP>165mmHg), FBS (FBS≤185mg/dL and FBS>185mg/dL) and Creatinine (Creatinine ≤1.5mg/dL and Creatinine >1.5 mg/dL). On the other hand, there is no significant difference between the survival curves for the two groups of BUN (BUN ≤35 mg/dL and BUN>35 mg/dL) of patients and sex (male and female).

4.4. Standard Cox PH Model Building

Variable selection for model formulation are represented by the following table

Table 5: The effect of predictor variables on good control of hypertension

Covariates	Chi-Square	DF	P-value
Sex	0.3	1	0.583
Age (Years)	4.5	1	0.034
SBP (mmHg)	54	1	$2.05e^{-13}$
FBS (mg/dL)	3.9	1	0.0496
BUN (mg/dL)	2.5	1	0.112
Creatinine(mg/dL)	5.2	1	0.0232

DF: Degree of Freedom

The variable Sex and BUN have p-value of 0.583 and 0.112, which are > 5% level of significance indicates these two variables have no statistically significant effect on the outcome variable and the variables Age, SBP, FBS and Creatinine have a smaller p-value as compared to 5% level of significance indicates that these four variables have a statistically significant effect on good control of hypertension. In addition, these values can give an important hint for the following model selection and formulation (Table 5).

In the univariable analysis the predictor variables Age, SBP, BUN and Creatinine are statistically significant at 5% level of significance for good control of hypertension. Furthermore, using a modest level of significance 25%, the variable FBS included in the multivariable model for further investigation. But the variable sex was not statistically significant at any level of significance, and hence, not included in the multivariable analysis.

Table 6: Univariable and Multivariable Analysis of Cox Proportional Hazards

Covariates	Univariable Analysis				Multivariable Analysis			
	$\hat{\beta}$	P-Value	HR	95%CI for HR	$\hat{\beta}$	P-Value	HR	95%CI for HR
Age(Years)	-0.0121	0.0017	0.988	(0.9805, 0.9955)	-0.008050	0.0389	0.9920	(0.6339, 1.1390)
SBP (mmHg)	-0.0339	$4.3e^{-14}$	0.967	(0.9581, 0.9752)	-0.036521	$2.59e^{-13}$	0.9641	(0.2691, 0.4948)
FBS (mg/dL)	-0.00194	0.098	0.998	(0.9958, 1)	0.002912	0.0282	1.0029	(1.0088, 1.7786)
BUN (mg/dL)	-0.00501	0.049	0.995	(0.99, 1)	-0.001287	0.6213	0.9987	(0.6879, 1.2118)
Creatinine (mg/dL)	0.15	0.036	1.16	(1.01, 1.337)	0.124024	0.0683	1.1320	(0.5882, 1.3678)

$\hat{\beta}$: Coefficient for Covariates
P-Value: Probability Value
HR: Hazard Ratio
CI: Confidence Interval

The predictor variable BUN is statistically significant at 5% level of significant for good control of hypertension in the univariable analysis but it is not statistically significant in the multivariable analysis. Then by removing the variable BUN, a model that contains the covariates age, SBP, FBS and creatinine was fitted and all the covariates except creatinine were found to be statistically significant. This model is the best model as compared with the above and other multivariable models because it has smaller AIC value. Let us see how to choose the best multivariable model among different models as follows:

Table 7: Model selection for Hypertension data set

Models	Covariates in the model	nPar	AIC	LR	P-Value
1	Age, SBP, FBS, BUN, Creatinine	5	2125.019	76.05	$5.662\ e^{-15}$
2	Age, SBP, FBS, Creatinine	4	2123.271	75.8	$1.332e^{-15}$
3	Age, SBP, FBS	3	2124.25	72.82	$1.11\ e^{-15}$
4	Age, SBP, FBS, Creatinine,	7	2127.517	77.55	$4.341\ e^{-15}$

	Age*SBP,Age*FBS, Age*Creatinine					
5	Age, SBP, Creatinine, Age*Creatinine	4	2126.033	73.03	$5.218e^{-15}$
6	Age, SBP, Creatinine, Age*SBP	4	2127.607	71.46	$1.11e^{-14}$
7	Age, SBP, Creatinine, Age*FBS	5	2125.162	75.91	$5.995e^{-15}$
8	Age, SBP, FBS, SBP*FBS	4	2124.215	74.85	$2.109e^{-15}$

nPar: Number of Parameters
AIC: Akaike Information Criterion
LR: Likelihood Ratio
P-Value: Probability Value

Based on the above table result the following multiple covariate analysis for the best model was performed.

Table 8: Multiple Covariate Analysis for the best model of Cox Proportional Hazards

Covariates	$\widehat{\boldsymbol{\beta}}$	**SE($\widehat{\boldsymbol{\beta}}$)**	**P-Value**	**HR**	**95%CI for HR**
Age (Years)	-0.008161	0.003895	0.0361 *	0.991873	(0.9843,0.9995)
SBP (mmHg)	-0.036917	0.004932	$7.13e^{-14}$ ***	0.963756	(0.9545,0.9731)
FBS (mg/dL)	0.002905	0.001324	0.0282 *	1.002910	(1.0003,1.0055)
Creatinine(mg/dL)	0.124872	0.067934	0.0660.	1.133004	(0.9918,1.2944)

$\widehat{\boldsymbol{\beta}}$: Coefficient for Covariates
SE: Standard Error of Covariates
P-Value: Probability Value
HR: Hazard Ratio
CI: Confidence Interval

The best Cox PH multiple covariate model contains Age, SBP, FBS and Creatinine. From these, the variables Age and FBS are statistically significant at 5% level while SBP is statistically significant at any level of significance (Table 8).

Therefore, the **Fitted Cox PH model** for Hypertension data set can be represented as follows:

$$\boldsymbol{\lambda(t|X) = \lambda_0(t) * e^{-0.008161Age-0.036917SBP+0.002905FBS+0.124872Creatinine}}$$

4.5. Model Diagnostic

After fitting the above Cox's regression model to a hypertension data set, it is important to check whether the above Cox's regression model is a correct model for this data set by

using different residuals to assess goodness of fit. In survival analysis, several types of residuals can be determined as follows:

i) Cox-Snell Residuals

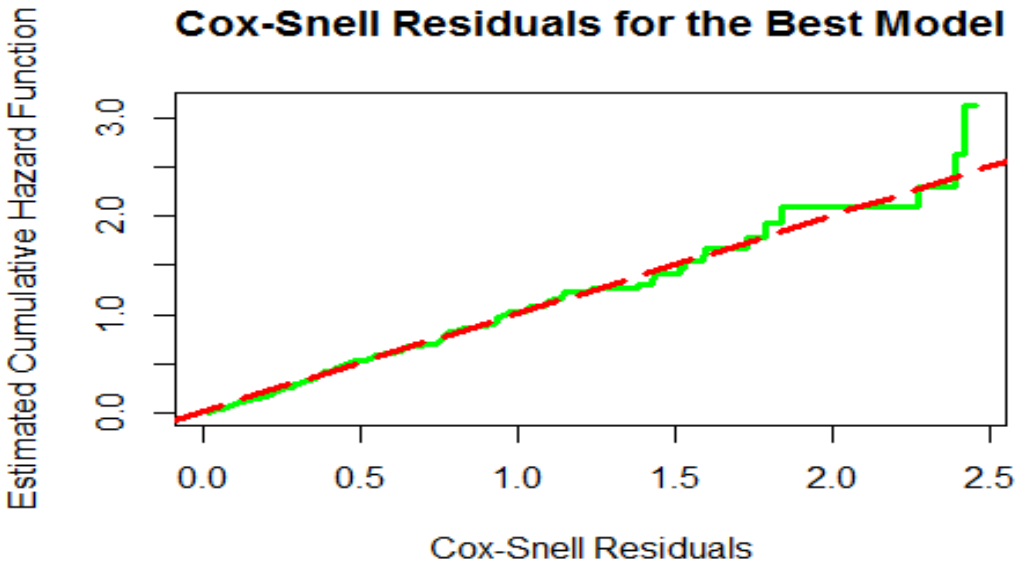

Figure 3: Cox-Snell Residual plot of the best fitted model for hypertension data set

If the model fit is adequate, then the points should follow a 45 degree line beginning at the origin. The above figure 3 presents an estimate of the cumulative hazard for these residuals, it lie along a straight line through the origin with a unit slop, since it fulfills the assumption. Therefore, the overall model fits the data reasonable well.

ii) Martingale Residuals

This could be used to determine the functional form of a covariate. To check which functional form is the best for a covariate, we can estimate several functional forms and compare them by the partial likelihood ratio test. Alternatively, by plotting the martingale residuals versus a covariate, we can verify whether the functional form is correct.

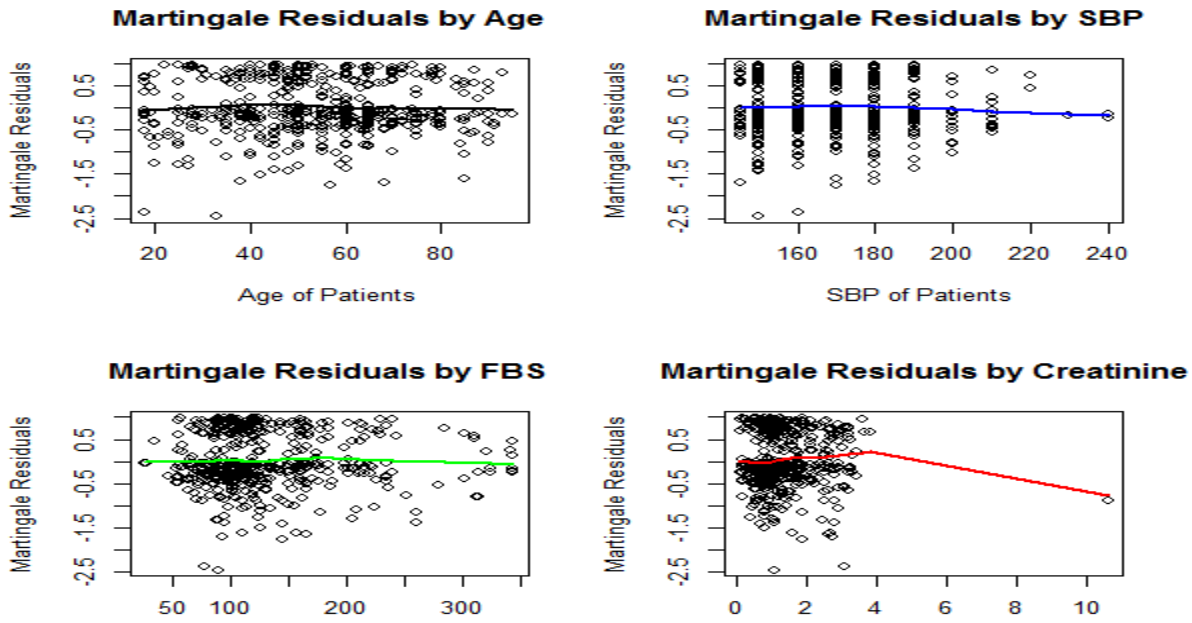

Figure 4: Martingale Residuals plot of the best fitted model for hypertension data set

By plotting the above martingale residuals versus a covariate, it can verify whether the functional form is correct. So as can be seen from the above figure 4: The entire above plot indicates a straight line means that the model is best model to fit the collected hypertension data set.

iii) Schoenfeld Residuals

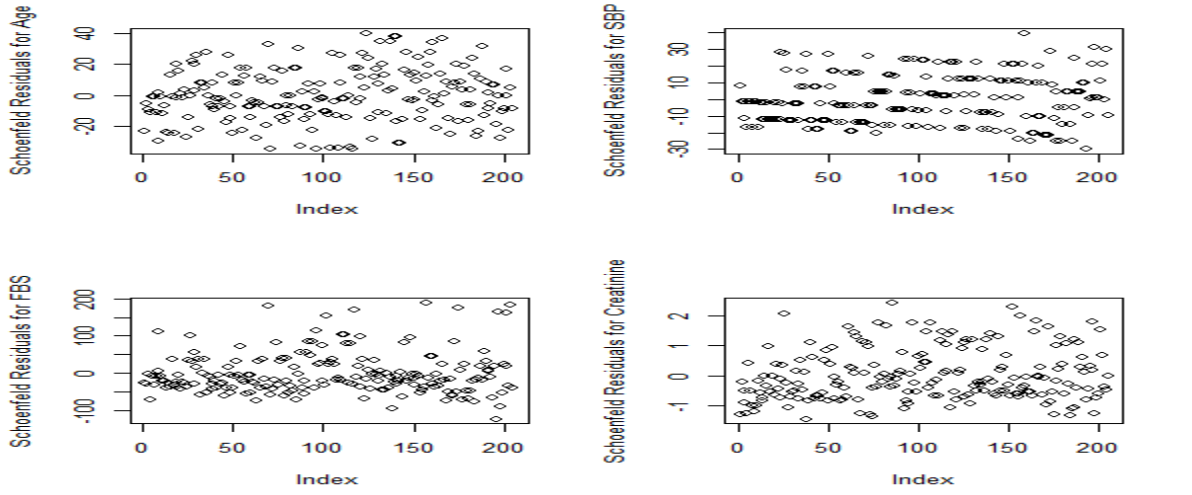

Figure 5: Schoenfeld Residuals plot of the best fitted model for hypertension data set

The above Schoenfeld residual plot shows that the functional form of each covariate and the plot is approximately a straight line indicates that each variable have a linear functional form.

iv) Deviance Residuals

It is a transform of martingale residuals, which is symmetric around zero. This is the only plot that is useful for checking outliers.

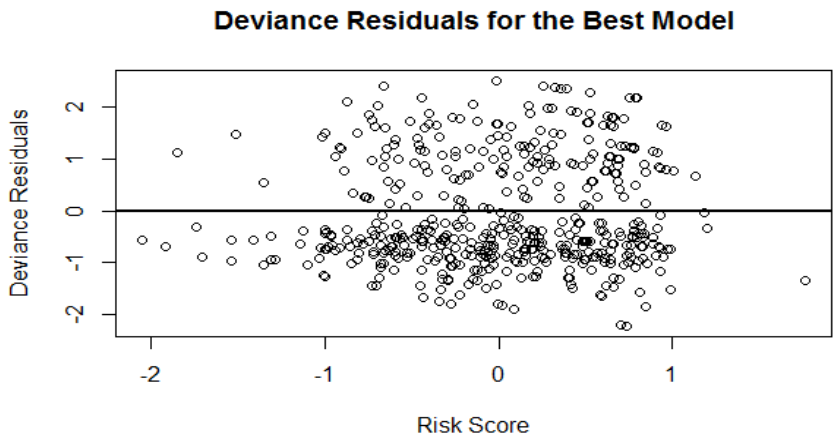

Figure 6: Deviance Residuals plot of the best fitted model for hypertension data set.

The above figure 6 shows that roughly symmetrically distributed around zero and it shows there is no outlier observation.

Therefore, the above four residuals plot shows that the selected final model for hypertension data set is good fit.

4.6. Proportional Hazard Assumption Checking

Tests and Graphs Based on the Schoenfeld Residuals

Testing the time dependent covariates is equivalent to testing for a non-zero slope in a generalized linear regression of the scaled Schoenfeld residuals on functions of time. A non-zero slope is an indication of a violation of the proportional hazard assumption. In a 'well-behaved' model the Schoenfeld residuals are scattered around zero and a regression line fitted to the residuals has a slope of approximately zero. The idea behind this test is that if the

proportional hazards assumption holds for a particular covariate, then the Schoenfeld residuals for that covariate will not be related to survival time.

Now let us see the test and Schoenfeld residual of each covariate included in the best model as follows:

Table 9: Statistical test for Proportional Hazards Assumption (PHA) of the covariates and their interaction with log of time and Schoenfeld residual

Covariates	rho	Chi-square	P-Value
Age	0.10267	1.950281	0.1626
SBP	0.08689	1.534951	0.2154
FBS	-0.00129	0.000369	0.9847
Creatinine	0.19292	5.202351	0.0226
GLOBAL	NA	9.341522	0.0531

Using the cox.zph function, rho is the Pearson product-moment correlation between the scaled Schoenfeld residuals and time. The hypothesis of no correlation is tested using Chi-square test. In the above table 9, the significant cox.zph test for Creatinine (P =0.0226<0.05) implies that the proportional hazards assumption has been violated for the Creatinine variable.

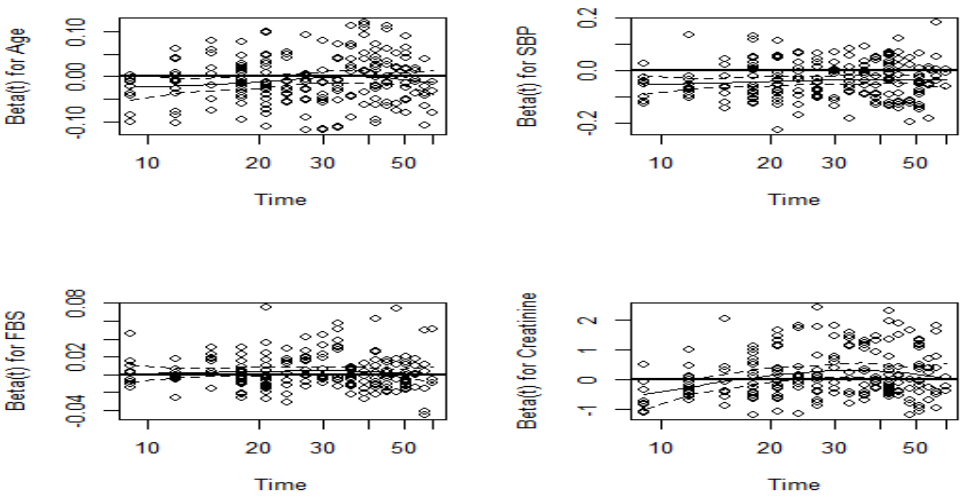

Figure 7: Plots of scaled Schoenfeld residuals for each covariate in a best model

The Schoenfeld residual plot (figure 7) shows also that the variable Creatinine violates the PH assumption because the plot is somewhat quadratic rather than straight line, but the plot of other three covariates included in the model shows a straight line which means that these variables fulfill the PH assumption.

As can also be seen from the previous KM curves in figure 1 the category of the variable Creatinine cross each other. This indicates that this variable violates the Proportional Hazard Assumption.

An alternative (and less sensitive) means of testing the proportional hazards assumption is to plot log [-log S(t)] versus time. If the proportionality assumption holds the two (or more) curve should be approximately parallel and should not cross.

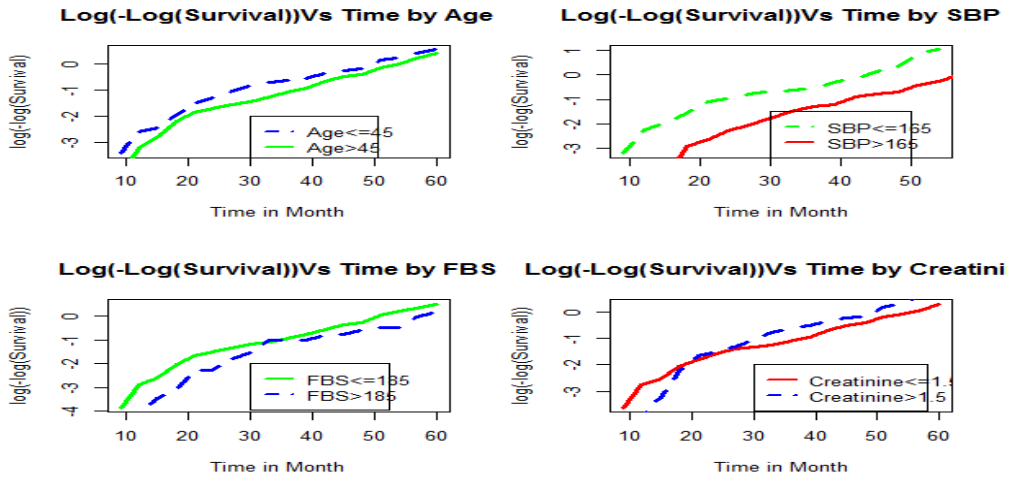

Figure 8: Plots of log [-log S(t)] versus time for each covariate in a best model

The above figure shows that the plot of a variable Creatinine cross to each other indicates that this variable violates the proportional hazard assumption but the remaining three covariates fulfill the PHA.

4.7. Strategies for Non-PH Assumption

From the above assumption checking by goodness of fit one predictor variable Creatinine violates the PH assumption. One method of dealing with this is to stratify the model by Creatinine. This means that to produce a separate baseline hazard function for each level of Creatinine. Note that by stratifying cannot obtain a hazard ratio for Creatinine since the

'Creatinine effect' is absorbed into the baseline hazard. After applying stratification of a model by creatinine, the results presented in table 10

Table 10: Stratified Cox PH model by creatinine

Covariates	$\widehat{\beta}$	SE($\widehat{\beta}$)	P-Value	HR	95%CI for HR
Age (Years)	-0.012746	0.005078	0.0121 *	0.987335	(0.9776,0.9972)
SBP (mmHg)	-0.036520	0.006166	$3.16e^{-09}$ ***	0.964138	(0.9526,0.9759)
FBS (mg/dl)	0.003643	0.001814	0.0445 *	1.003650	(1.000,1.0072)

$\widehat{\beta}$: Coefficient for Covariates
SE: Standard Error of Covariates
P-Value: Probability Value
HR: Hazard Ratio
CI: Confidence Interval

The results show that SBP of hypertension patients is statistically significant at 1% level and the two covariates age and FBS of patients are statistically significant at 5% level.

Hazard Ratio

The hazard ratio for the effect of SBP of patients adjusted for creatinine is given by the value 0.964138. This value can be obtained by exponentiation the coefficient -0.036520 of the SBP variable. The hazard ratio value can be interpreted to mean that the larger SBP group (for which SBP >165mmHg) has 0.964138 times the hazard as the smaller SBP group (for which SBP ≤ 165mmHg) to attain good control of HTN. With 95% confidence the hazard can go down as low as 0.9526 and can go up as high as 0.9759.

The value gives the estimated hazard ratio (HR) for the effect of Age; in particular, to see that the hazard for the larger age group (Age >45 years) is 0.987335 times the hazard for the smaller age group (Age ≤ 45 years) to attain good control of HTN. With 95% confidence the hazard can go down as low as 0.9776 and can go up as high as 0.9972 (Table 10).
A Modification of the Cox Proportional Hazard Model which is called **Stratified Cox Model** for Hypertension Data Set becomes:

$$\boldsymbol{\lambda(t|X) = \lambda_{Creatinine,0}(t) * e^{-0.012746Age-0.036520SBP+0.003643FBS}}$$

4.8. Univariate Frailty Model

The frailty may be individual-specific or group-specific. Models constructed in terms of group-level frailties are sometimes referred to as 'shared' frailty models because observations within a subgroup share unmeasured 'risk factors' that prompt them to exit earlier than other subgroups. Models based on individual-level frailties are simply called frailty models or individual-level frailty models.

Estimation of the frailty model can be parametric or semi-parametric (for gamma frailty models with a semi-parametric). If the number of subjects n_i is 1 for all groups, then the univariate frailty model is obtained. The null hypothesis is that there is no heterogeneity between individuals. If accepting the null hypothesis then no frailty model is needed.

4.8.1. Univariate Frailty Model Building

In the univariable frailty analysis the predictor variables Age, SBP, BUN and Creatinine are statistically significant at 5% level of significance for good control of hypertension, like univariable Cox PH analysis. Furthermore, using a modest level of significance 25%, the variable FBS is included in the multivariable model for further investigation. But the variable sex was not included in the multivariable analysis.

Table 11: Univariable and Multivariable Analysis of univariate frailty

Covariates	Univariable Analysis					Multivariable Analysis				
	$\widehat{\boldsymbol{\beta}}$	SE($\widehat{\boldsymbol{\beta}}$)	χ^2	DF	P-value	$\widehat{\boldsymbol{\beta}}$	SE($\widehat{\boldsymbol{\beta}}$)	χ^2	DF	P-value
Age	-0.0121	0.0039	9.82	1	0.0017	-0.0096	0.0044	4.77	1	$2.9e^{-02}$
SBP	-0.04	0.0052	58.7	1	$1.9e^{-14}$	-0.0405	0.0055	54.14	1	$1.9e^{-13}$
FBS	-0.0019	0.0012	2.74	1	0.098	0.0029	0.0015	4.13	1	$4.2e^{-02}$
BUN	-0.005	0.0026	3.86	1	0.049	-0.0015	0.0028	0.27	1	$6.0e^{-01}$
Creati-nine	0.15	0.072	4.41	1	0.036	0.1339	0.0791	2.86	1	$9.1e^{-02}$
Frailty (id)			0.01	0.01	0.71			59.39	52.9	$2.5e^{-01}$

$\widehat{\boldsymbol{\beta}}$: Coefficient for Covariates

P-Value: Probability Value
SE: Standard Error of estimate
χ^2: Chi-square
DF: Degree of Freedom

As can be seen from table 11, the predictor variable BUN is statistically significant in the univariable analysis based on frailty but it is not statistically significant in the multivariable analysis. Then by removing BUN, the model that contains covariates age, SBP, FBS and creatinine was fitted and all the covariates were found to be statistically significant. This model is the best model as compared with the above and other multivariable models because it has smaller AIC value. The final best univariate frailty model is shown in table 12.

Table 12: Parameter estimates (SE) in the univariate gamma frailty model applied to Hypertension data set.

Covariates	$\widehat{\boldsymbol{\beta}}$	**SE($\widehat{\boldsymbol{\beta}}$)**	**Chi-Square**	**DF**	**P-Value**
Age (Years)	-0.00962	0.00436	4.87	1	$2.7e^{-02}$
SBP (mmHg)	-0.04082	0.00543	56.61	1	$5.3e^{-14}$
FBS (mg/dL)	0.00296	0.00145	4.17	1	$4.1e^{-02}$
Creatinine (mg/dL)	0.13419	0.07862	2.91	1	$8.8e^{-02}$
frailty(id, dist = "gamma)			57.15	51.2	$2.6e^{-01}$
Likelihood ratio test=179	on	54.5 DF,	p-value=$3.44e^{-15}$		n= 500

$\widehat{\boldsymbol{\beta}}$: Coefficient for Covariates
SE: Standard Error of Covariates
DF: Degree of Freedom
P-Value: Probability Value

All three predictor variables included in the model have a statistical significant effect on good control of hypertension.

The smaller p-value of the likelihood ratio test on 54.5 degree of freedom indicates that rejecting the null hypothesis. The frailty effect was observed for those variables that were statistically significant at 0.10 level in univariate frailty analysis and indicates high frail. Which means that there is unobserved heterogeneity between individuals (some other important covariates have not been measured) (Table 12).

4.9. Discussion

In medical science, researchers are more interested in Cox PH model than other parametric models to estimate the survival model, mainly due to the less assumption required in the model. It requires some hypotheses to analyze the survival by means of Cox model. This model assumes that any change at independent variables level in the hazard function is independent of time. These hypotheses required for modeling Cox hazard model may not work in many conditions, especially in biomedical fields (Cox, 1984). If these hypotheses do not work, the results obtained from Cox model may be invalid. So, it is possible to use a model where stratification is done on the non-proportional predictors.

In this thesis, the hypertension data set was analyzed using Cox PH and univariate frailty models. After fitting the Cox PH model, the goodness of fit is assessed through residual plots.

The Cox PH analysis result suggested that SBP measures are decreasing over time. This supports the results of Yasin Negash (2013), who found that on average, SBP measures slightly decrease in a linear pattern over time. This implies that after the patients join the follow up program, their SBP decreases due to the therapy.

The median survival time of hypertension patients to attain good control is 48 months and the mean survival time is 43.6 months. Patients who have smaller SBP are more effective overall than patients having larger SBP to attain good control of hypertension.

The age group > 45 years attain good control of hypertension faster and SBP group > 165 mmHg attain good control of hypertension faster. Among the covariates, sex and BUN have no significant effect on the outcome variable but the remaining five covariates have significant effect on good control of hypertension. There is a significance difference between survival curves of the two groups of age, SBP, FBS, BUN and creatinine but no difference between survival curves of male and female patients.

In order to construct the best Cox PH model for the hypertension data set, first apply univariable analysis and choose the covariates which are statistical significant and then add these significant covariates to multivariable analysis. Finally, among different multivariable

models, the best fitted model for hypertension data set is the model having the covariates age, SBP, FBS and creatinine.

When to check the PH assumption except a variable creatinine all three covariates fulfill the PHA. The model was stratified by creatinine in order to manage non proportionality.

All the three covariates included in the model have a statistical significant effect on good control of hypertension. The p-value of likelihood ratio test with 54.5 degree of freedom indicates that rejecting the null hypothesis means that there is unobserved heterogeneity between individuals of hypertensive patients. Modeling a frailty effect is not only a function of unobserved heterogeneity but also of observed covariates.

Finally, the final fitted model showed that as age and SBP of patients increased, it takes a short period of time to attain good control of hypertension and vice versa. But there is a positive relationship between FBS and good control of hypertension.

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

This study is based on hypertensive patients' data set obtained from Bahir-Dar Felege Hiwet Referral Hospital. The objective of the study was to identify significant risk factors that affect good control time of hypertensive patients who have been under follow-up at Bahir-Dar Felege Hiwet Referral Hospital. For determining the risk factors for good control of hypertension and modelling the survival time, a total of 500 sample hypertensive patients were included in the study out of which 205 (41%) patients have an event (good control of hypertension) and the rest 295 (59%) patients were censored. Among 205 patients having an event, 108 (52.7%) are females and 97 (47.3%) are males.

This study revealed that after starting of anti-hypertensive treatment, hypertensive patients follow up time on average was nearly 32 months (2.67 years) with median survival follow up time estimated to be 30 months (2.5 years). Hypertensive patients attain good control of HTN on average 43.6 months (3.63 years) and the median survival time was 48 months (4 years).

The Cox PH analysis result showed that the major factors that affect good control of hypertensive patients are age, SBP and FBS. Among these, age and SBP have an inverse relationship with the outcome variable but FBS has a direct relationship with the outcome variable (i.e. good control of hypertension).

Furthermore, univariate frailty model was also applied for this data set. According to the result of univariate frailty model analysis, there is unobserved heterogeneity between individuals (i.e. there is another unmeasured covariates that affect good control of hypertension, but not included in this study).

5.2. Recommendation

Hypertension is the most serious disease in the world (silent killer). Modeling the survival time of this disease helps to identify the factors that affect the success of the therapy which helps to discover new treatment modality by considering the identified factors.

Thus further studies should be done in the area using these newly developed and most flexible methodologies by including additional covariates (social, economic, behavioral, nutritional, environmental, and the like) that may affect good control of hypertension.

Now a day's different health institutions are spread all over the country and provide different types of treatments for hypertensive patients. But it is not enough only giving a treatment to patients under a follow up clinic, but also it is important to know factors that contribute to the progression of the blood pressure.

This study revealed that the older age and higher SBP group of hypertensive patients attain good control of HTN earlier than the younger age and lower SBP groups. Therefore, the clinicians should have give an attention to the younger age and lower SBP group to attain good control of HTN earlier like that of older age and higher SBP groups.

In addition, governmental and non-governmental body should gives awareness for health workers to record all the necessary variables during follow up time to see the change of the disease.

6. REFERENCES

Akilew, A. and Tadesse, A. 2012. Prevalence and associated factors of hypertension among adults in Gondar, Northwest Ethiopia: a community based cross-sectional study, Ethiopia, Gondar University.

Amoah A. G., Addo J, Koram K. A. 2006. The changing patterns of hypertension in Ghana: a study of four rural communities in the Ga district. Available: (http://www.ncbi.nlm.nih.gov/pubmed).

Benetos, A., Thomas, M., Safar, K. and Bean, L. 2001. Systolic Blood Pressure, diastolic blood Pressure, and Pulse Pressure: An evaluation of their joint effect on mortality. *Journal of the American College of Cardiology*, 37(1): 163 – 168.

Buabeng K.O., Matowe L. & Plange-Rhule J. 2004. Unaffordable drug prices: the major cause of non-compliance with hypertension medication in Ghana. Available: http://www.ualberta.ca/~csps/JPPS7(3)/L.Matowe/hypertension.htm

Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Jr., Jones DW, Materson BJ, Oparil S, Wright JT, Jr. 2003. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*, 289(19):2560-2572.

Clayton, D.G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65, 141-51.

Collett, D. 2003. Modelling Survival Data in Medical Research. Chapman and Hall, London.

Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, Godwin J, Qizilbash N, Taylor JO, Hennekens CH. 1990. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet*, 335(8693):827-838.

Cox, D. R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society Series* B 34, 187.220.

Cunha J. P & Marks J, W. 2011. High blood pressure (hypertension). Available:http://www.medicinenet.com/high_blood_pressure/article.htm.

Danso, A. K. & Henry, S. Opare-Addo. 2010. Challenges associated with hypertensive disease during pregnancy in low-income countries. Available:(http://www.sciencedirect.com.)

Davarian, S., Crimmins, E., Takahashi, A. and Saito, Y. 2013. Socio demographic Correlates of Four Indices of Blood Pressure and Hypertension among Older Persons in Japan.

Elbers, C., Ridder, G. 1982. True and spurious duration dependence: the identifiability of the proportional hazard model. Review of Economic Studies XLIX, 403–409.

Frederico, D., Garcia1,S. and Aleyson , F. 2004. Evaluation of risk factors associated with increased blood pressure in children

Ghana Maternal Health Survey. 2007. Report on Survey. May 2009. Submitted by Ghana Statistical Service Accra, Ghana, Ghana Health Service Accra, Ghana and Macro International Inc. Calverton, Maryland, U.S.A.

Giles, T., Berk, B. and Black, H. 2005. On behalf of the Hypertension Writing Group. Expanding the definition and classification of hypertension, *Clin Hypertens*; 89:505–512.

Harrel, F. and Lee, K. 1986. Proceedings of the Eleventh Annual SAS User's Group International. 823–828.

Holmes, J. and John. H. (2013). Racial/Ethnic Variability in Hypertension Prevalence and Risk Factors in National Health Interview Survey, In American Health Research Institute.

Horvath K, Jeitler K, Siering U, Stich AK, Skipka G, Gratzer TW, Siebenhofer A. 2008. Long-term effects of weight-reducing interventions in hypertensive patients: systematic review and meta-analysis. *Arch Intern Med*, 168(6):571-580.

Hosmer, D.W., and Lemeshow, S. 1999. Regression Modelling of Time-To-Event Data. New York, John Wiley & Sons, Inc, 207.

Hougaard, P. 1984. Life table methods for heterogeneous populations. Biometrika 71, 75–83.

Kalbfleisch, J., Prentice, R. 2002. The Statistical Analysis of Failure Time Data. Wiley, New York.

Kannel, W. 1996. Blood pressure as a cardiovascular risk factor: prevention and treatment. *JAMA*. 275:1571–1576.

Klein, J.P. and Moeschberger, M.L. 1997. Survival Analysis: Techniques for Censored and Truncated Data. New York, Springer.

Klembaum, D. G. 1996. Survival Analysis: A Self learning text. Springer, New York.

Lawless, J. F. 1982. Statistical Models and Methods for Lifetime Data Analysis. Wiley, New York.

Lee & Cooper. 2009. Associated with increased morbidity and mortality from stroke, myocardial infarction, cardiac failure, dementia, renal failure, and blindness.

Lee, E. T., and Wang, J. W. 2003. Statistical Methods for Survival Data Analysis, 3rd ed. Wiley, New York.

Lock K, Pomerleau J, Causer L, Altmann DR, McKee M. 2005. The global burden of disease attributable to low consumption of fruit and vegetables: implications for the global strategy on diet. Bull World Health Organ, 83(2):100-108.

Maher D., Smeeth L. & Sekajugo J. 2010. Health transition in Africa: practical policy.

Mancia G, De Backer G, Dominiczak A, Cifkova R, Fagard R, Germano G, Grassi G, Heagerty AM, Kjeldsen SE, Laurent S *et al*: 2007. Guidelines for the management of arterial hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Eur Heart J* , 28(12):1462-1536.

Mancia, G., Grassi, G., Giannattasio, C. 1999.Sympathetic Activation in the pathogenesis of hypertension and progression of organ damage. *Hypertension*.34:724-728

Neal B, MacMahon S, Chapman N. 2000. Effects of ACE inhibitors, calcium antagonists, and other blood-pressure-lowering drugs: results of prospectively designed overviews of randomized trials. Blood Pressure Lowering Treatment Trialists' Collaboration. *Lancet*, 356(9246):1955-1964.

O'Brien, E., Beevers, G. and Gregory,Y. (2001).ABC of hypertension: Blood pressure measurement. Part III Automated sphygmomanometer: ambulatory blood pressure measurement. *BMJ*, **322**: 1110 – 1114.

Oliveria, S., Lapuerta, P., McCarthy, B. 2002. AschSM. Physician-related barriers to the effective management of uncontrolled hypertension. 162:413–420.

Pauletto, P., Caroli, M., Pessina, A. and Palu, C. 1994. Hypertension prevalence and age-related changes of blood pressure in semi-nomadic and urban Oromos of Ethiopia. *Eur J Epidemiol*.10 (2): 159–164.

Pereira M, Lunet N, Azevedo A, Barros H. 2009. Differences in prevalence, awareness, treatment and control of hypertension between developing and developed countries. *J Hypertens*, 27(5):963-975.

Proposals for primary care. Available: http://www.who.int/bulletin/volumes/.

Schoenfeld, D. 1982. Partial residuals for the proportional hazards model. *Biometrika*, 69: 51–55

Schofield I & Hazel H. 1999. Healthy Ageing: Nursing Older People. Mosby International Ltd, Lynton House, 7-12 Tavistock Square, London, UK. 81-85.

Teklu, B. 1983. Hypertension amongst bank employees in urban Ethiopia, *Ethp. Med. J.* (21).

Vaupel, J.W., Manton, K.G. and Stallard, E. 1979. The impact of heterogeneity on individual frailty on the dynamic of mortality. *Demography*, 16(3), 439-54.

Vietnam-National-Heart-Institute. 1992. Survey on hypertension in Vietnamese adult population. Hanoi: Vietnam National Heart Institute.

Wamala J. F., Karyabakabo Z., Ndungutse D., Guwatudde D. 2009. Prevalence factors associated with Hypertension in Rukungiri District, Uganda -Community Based Study Available: http://www.ncbi.nlm.nih.gov/pubmed. Accessed 10January 2011.

Whitworth JA: 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. *J Hypertens*, 21(11):1983-1992

Wienke, A., Lichtenstein, P. and Yashin, A.I. 2003. A bivariate frailty model with a cure fraction for modelling familial correlations in diseases. *Biometrics*, 59, 1178-83.

World Health Organization. 1999. International Society of Hypertension Guidelines for the Management of Hypertension. Guidelines Subcommittee. *J Hypertens*, 17(2):151-183.

World-Health-Organization. 2004. Ethiopia Strategy Paper. In: World Health Organization health action in crises. Available: www.who.int/entity/hac/crises/eth/Ethiopia_strategy_document.

World-Health-Organization. 2004. Global strategy on diet, physical activity and health In Geneva: World Health Organization. Available: http://www.who.int/dietphysicalactivity/publications/recsmarketing/en/index.html.

World-Health-Organization. 2009. Global Health Risks Summary Tables. In. Geneva, Switzerland: Health Statistics and Informatics Department. World Health Organization. Available: (http://www.who.int/evidence/bod).

Yashin, A. I. and I. A. Iachine. 1997. How Frailty Models can be used in the Analysis of Mortality and Longevity Limits. *Demography* 34:31-48.

Yashin, A.I. and I.A.Iachine. 1995a. How long can humans live? Lower bound for biological limit of human longevity calculated from Danish twin data using correlated frailty model. Mechanisms of Ageing and Development 80:147-159.

Yashin, A.I. and I.A.Iachine. 1995b. Genetic analysis of durations: Correlated frailty model applied to survival of Danish Twins. *Genetic Epidemiology* 12:529-538.

YoungL.H.2011.HeartDiseaseintheElderly.Available:(http://www.med.yale.edu/library/heartbk/21.pdf).

Yu Z, Song G, Guo Z, Zheng G, Tian H, Vartiainen E, Puska P, Nissinen A. 1999. Changes in blood pressure, body mass index, and salt consumption in a Chinese population. *Prev Med*, 29(3):165-172.