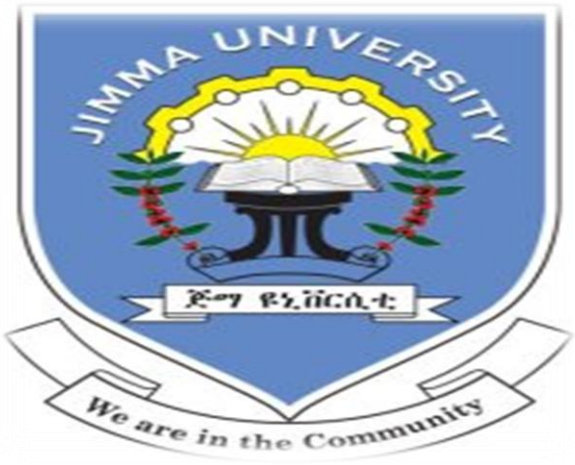

Joint Modeling of Longitudinal Systolic and Diastolic Blood Pressure Measurements of Hypertensive Patients Receiving Treatment in Jimma University Specialized Hospital

By:

Yasin Negash

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

January, 2014
Jimma, Ethiopia

Joint Modeling of Longitudinal Systolic and Diastolic Blood Pressure Measurements of Hypertensive Patients Receiving Treatment in Jimma University Specialized Hospital

MSc thesis

Yasin Negash

E-mail: yasinnegash51@yahoo.com

Mobile: +251 (0) 9 13 50 94 52 / 17 41 31 39

Main advisor: Wondwosen Kassahun (Assi. Prof, Ph.D Scholar)

Co-advisor: Abdisa Gurmassa (MSc)

Jimma, University

January, 2014

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

As thesis research advisors, we herby certify that we have read the thesis prepared by YASIN NEGAH under our guidance, which is entitled "**Joint Modeling of Longitudinal Systolic and Diastolic Blood Pressure Measurements of Hypertensive Patients Receiving Treatment in Jimma University Specialized Hospital**", in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Wondwosen Kassahun (Assi. Prof, Ph.D Scholar) Advisor	_____ Signature	_____ Date
Abdisa Gurmassa (MSc) Co-advisor	_____ Signature	_____ Date

As the members of the board of examiners of MSc thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science 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 Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Yasin Negash

Date: _____

Signature: _____

Jimma University, Jimma

DEDICATION

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

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. In developing countries, there is a rapid increase in hypertension prevalence, and in developed countries, the previous trend of an increase in hypertension prevalence is actually reversing. According to World health organization hypertension is the seventh leading cause of death in Ethiopia. Hypertension is also called high blood pressure, described by two numbers SBP and DBP. Hence, joint longitudinal model was used to address how the evolution of SBP is associated with the evolution of DBP.

Objective: The main objective of this study is to investigate the joint evolution and association of systolic and diastolic blood pressure measurements of hypertensive patients and identify the potential risk factors affecting the two end points in Jimma University Specialized hospital.

Methods: In this study secondary data was used from Jimma university specialized hospital in Hypertensive Outpatient Clinic. The study population consists of 354 hypertensive patients, measured repeatedly at least three times on each patient who are 18 years old or older those treated with antihypertensive drugs from September 2011 to July 2013 were used in this study. First, each of the outcomes is analyzed separately using linear mixed model. Then, a joint model is considered to study the joint evolution and identify the potential risk factors affecting the two end points.

Results: On average both SBP and DBP measures slightly decrease a linear pattern over time. In addition, the progression of both outcomes depends on patient's baseline socio-demographical characteristics. Fit statistics showed that the joint model resulted in better fit to the data than the separate models, implying a significant association among the two end points. Based on the joint model, sex, baseline age, and place of residence are the significant factor for the progression of blood pressure, but family history and all the interaction term except age by time, did not appear significant at 5% level of significance.

Conclusion: The results of the separate and joint model analysis are consistent. When the joint model is compared with the separate model, the joint model fitted the data better than the separate model. The result from the joint model suggested a strong association between the evolutions and a slowly increasing evolution of the association between SBP and DBP.

Key Words: Joint Modeling; Longitudinal Data Analysis; Linear Mixed Model;

ACKNOWLEDGEMENT

It is my pleasure to express my appreciation to those who have influenced this work .My special gratitude goes to Wondeson Kassahun (Assistant Professor, Ph.D Scholar), my main 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. Abdisa Gurmessa, head department of statistics, 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. Akalu Banbeta from Jimma University and Mr. Awol Seid from Haramaya University for their continuous comment, advice and guidance. And also, I thanks to my classmates, especially Mr. Zelalem Mehari and Mr. Yemane Hailu, who helped me in every aspect of this thesis preparation.

I would like to thank Jimma University Specialized Hospital Hypertensive outpatient clinic staffs especially Mr. Mohammed and Mr. Solomon who allowed and facilitate all the inputs for data collection. I am especially grateful to my sister Ayesha Negash, my brother Mustefa Negash, Jemal Awol, Ahmed Negash, and my Mother Shitu Siraje for their guidance and encouragement during thesis preparation.

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

Above all, I would like to thank almighty ALLAH for his grace, wisdom, favor, faithfulness and wonderful help up to the present day.

List of Acronyms

AIC:	Akaike Information Criteria
AIDS	Acquired Immunodeficiency Syndrome
AOE	Association of the evolution
AR:	Autoregressive
BIC:	Bayesian Information Criteria
BMI:	Body Mass Index
BP:	Blood Pressure
CV:	Cardiovascular disease
DBP:	Diastolic Blood Pressure
EOA	Evolution of the association
HTP	Hypertensive Patients
LMM:	Linear Mixed Effects Models
MA:	Moving Average
ML:	Maximum Likelihood
MOH	Minister of Health
REML:	Restricted Maximum Likelihood
SBP:	Systolic Blood Pressure
WHO:	World Health Organization

Table of Contents

List of Tables

List of Figures

CHAPTER ONE

1. INTRODUCTION

1.1. 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 and its role as a prominent risk factor for the development of a number of disease processes. 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).

Hypertension is also called high blood pressure described by two numbers: the systolic and the diastolic blood pressure. Systolic pressure is the maximum pressure in an artery at the moment when the heart is beating and pumping blood through the body. Diastolic pressure is the lowest pressure in an artery in the moments between beats when the heart is resting (O'Brien *et al.,* 2001). A person is said to be experiencing high blood pressure if he/she has systolic blood pressure above 140 mmHg or diastolic blood pressure above 90 mmHg measured on both arms on three occasions over a few weeks. Reduced systolic and diastolic blood pressure measures on the patients were indicative of positive response to treatment (Benetos *et al.,* 2001).

Previous study shows that, diastolic blood pressure has been considered the best predictor of cardiovascular disease risk in people with hypertension. But this paradigm began to change about 40 years ago when systolic blood pressure became the accepted predictor of risk (Zucker and Zerbe, 1995*).* Today, some researchers believe that a combination of measurements systolic and diastolic blood pressure paint a more complete picture of risk for a larger variety of patients.

In developing countries, there is a rapid increase in hypertension prevalence, and in developed countries, the previous trend of an increase in hypertension prevalence is actually reversing (Das *et al.,* 2005*).* The overall worldwide burden of hypertension in the year 2000 was estimated to be 972 million, 26.4% of the adult world population, with 333 million or 34.26% in developed and 639 million or 65.73% in developing countries (Hajjar *et al.,* 2006*).* It has been estimated that by the year 2025, 1.56 billion will have hypertension, an increase of 60% from the year 2000.

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).

Antihypertensive is a class of drugs therapy that is used to treat hypertension (high blood pressure) (Safar *et al.,* 2000*).*It seeks to prevent the complications of high blood pressure, such as stroke and myocardial infarction. Evidence suggests that reduction of the blood pressure by 5 mmHg can decrease the risk of stroke by 34%, of ischemic heart disease by 21%, and reduce the likelihood of dementia, heart failure, and mortality from cardiovascular disease.

The defining characteristic of a longitudinal study is that individuals are measured repeatedly through time. Longitudinal studies are in contrast to cross-sectional studies, in which a single outcome is measured for each individual. While it is often possible to address the same, scientific questions with a longitudinal or cross-sectional study (McCulloch *et al.*, 2008). Longitudinal studies have many advantages compared to cross-sectional designs, which study many observations at a given time. Investigators gather longitudinal data in order to study change in a response variable over time as well as to relate these changes in explanatory variables over time. Longitudinal studies consider both the between-subject and within-subject time-related variations, and provide more efficient estimators than cross-sectional designs with the same number and patterns of observations (Laird and Ware, 1982).

Analyses of multiple observations measured on the same individual over time are different from observations measured on different people. Responses measured repeatedly on the same unit or individual are correlated because they contain a common contribution from that unit (Fieuws and Verbeke, 2005). Moreover, measurements on the same individual close in time tend to be more correlated than measures far apart in time. Therefore it is important to try and model the correct correlation structure and this will yield more precise estimators of interest (Laird and Ware, 1982).

In addition, modeling the true correlation structure becomes significant in the presence of missing values and when the number of observations per subject is not large. There are two types of covariates in longitudinal studies in general. There are time invariant or baseline covariates (e.g. gender) and time varying covariates (e.g. weight). The linear mixed model has become the most commonly used tool for analyzing continuous repeated measures data from a sample of individuals in agriculture, biomedical, economical, and social applications(McCulloch *et al.*, 2008).

The approach that this study used to build a mixed model methodology allows the longitudinal examination of systolic and diastolic blood pressure over time. Mixed models provide a flexible and powerful tool for the analysis of data with complex covariance structure. A mixed model has two types of components, the systematic or fixed, or the mean model component and the random component (McCulloch *et al.,* 2008). The fixed component is a sub-model representing the contribution by fixed effects and the random component represents the contribution by random effects. A fixed effect is an effect where all levels of the variable are contained in the data and the effect is universal to the entire target population (Der and Everitt, 2006).These unobserved effects are then included in the model as random variables, or equivalently called, random effects. A random effects model means that the levels of the factor variable in the data being modeled comprise a random sample of levels in the target population.

Longitudinal studies are also comprise of a repeatedly response which consists of two or more elements are measured simultaneously on the same individual over time. For example, systolic blood pressure (SBP) and diastolic blood pressure (DBP) measures are collected simultaneously from a patient every time they visit the doctor's office. Together these measurements give the physician an indication of the health and functioning of an individual's circulatory system at a given time point, and longitudinal measures of SBP and DBP can alert the physician to changes in the health of an individual(Laird and Ware, 1982). Measures of SBP and DBP are highly related and changes in either often affect changes in the other. Separate analyses would not able to examine the correlation or association between the two outcomes. Therefore, it is more desirable to jointly modeling of two outcome variables together (Williams, 2001).

In a joint longitudinal model, there are two types of correlations. Serial correlation and cross correlation. Serial correlation, is between observations at different time points within a subject

and the other, cross correlation is between observations on different response variables at each time point. If different types of outcomes are measured at each time point, the correlation structure is more complicated and hence, more difficult for drawing inference(Olkin and Tate,1961*;* Molenberghs and Verbeke,2005)

There are different general approaches for modeling joint longitudinal observations with differing outcome. However; none of these approaches answers the question of how the evolution of one response is related to the evolution of another response (Fieuws and Verbeke, 2004*)*. A flexible solution is to model the association between the different responses using random effects. In applied sciences, random effects models have become the preferred tool to analyze various types of longitudinal data. With these models, the average evolution of a specific response is described using some function of time, and subject-specific deviations from this average evolution are introduced by using so-called random effects In a joint-modeling approach using mixed models, random-effects are assumed for each response process and by imposing a joint multivariate distribution on the random effects; the different processes are associated. This approach has many advantages and is applicable in a wide variety of situations. Indeed, the approach allows joining models for responses of the same response type as well as models for responses of different types (Laird and Ware, 1982).

Following part of this thesis is organized as follows: The statement of the problem and objectives of the study are presented next in this Chapter. Chapter 2 describes some literatures related to the associated factors for the progression of blood pressure and different joint modeling approaches. In Chapter 3, the data and the detail methods of data analyses are explained. Then, basic results of the study are presented in Chapter 4 and discussed in Sub -Section 4.5. Finally, some concluding remarks and recommendation are provided in Chapter 5.

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 is a potent risk factor for myocardial infarction, stroke, and heart failure, which are the leading causes of death and disability worldwide.

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 systolic and diastolic 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 published literatures that documented on this area except the studies about determinates of systolic and diastolic blood pressure control in Ethiopia based on cross-sectional data. They used multiple linear regression and logistic regression to identify determinants factors that progress systolic and diastolic blood pressure overtime separately without considering the correlations within the two outcomes and subject specific random effects.

In longitudinal data, with two outcomes there is also a correlation between them, in addition to the correlation due to repeated measures over time. But their separate modeling of the systolic and diastolic outcomes may not be appropriate, as the two are biologically correlated and mutually influential Joint modeling of the two responses, on the other hand, incorporates all information simultaneously and provides valid and efficient inferences (Fieuws and Verbeke, 2004*)*. And also the joint model is able to appropriately account for correlations within and between each outcome. For example, SBP and DBP are collected simultaneously from a patient every time .A great deal of interest then lies in how the evolution of SBP is related to the evolution of DBP, as well as how the association changes, or evolves, over time. Therefore, separate modeling would not able to examine the association or evolution of the two outcomes evolves over time, but joint modeling does.

In general, the motivation behind this study is to address the following major research questions:

1. How the average progressions of SBP and DBP in hypertensive patients treated with antihypertensive drugs changes over time?

2. What is the relationship or associations of the evolution between SBP and DBP look like over time?
3. What factors predict the evolution of Systolic and Diastolic blood pressures separately?
4. What factors predict the joint evolution of Systolic and Diastolic blood pressure?

1.3. Objectives of the Study

1.3.1. General Objectives

The main objective of this study is to investigate the joint evolution and association of systolic and diastolic blood pressure measurements of hypertensive patients and identify the potential risk factors affecting the two end points in Jimma University Specialized Hospital.

1.3.2. Specific Objectives

The specific objectives are to:

- Explore the mean evolution of SBP and DBP of hypertensive patients over time.
- Fit a separate mixed effect models for longitudinal systolic blood pressure and identifying the associated risk factor.
- Fit a separate mixed effect models for longitudinal diastolic blood pressure and identifying the associated factor.
- Fit a joint model for longitudinal systolic and diastolic blood pressure and identifying the associated factor for the progress of SBP and DBP jointly.
- Compare and contrast findings of the separate and joint model.

1.4. Significance of the Study

The results of this study will be useful in the development of an effective care and patient monitoring system on chronic and/or non-communicable disease. Specifically:

- It shades some light how systolic and diastolic blood pressure measurements are related with each other among hypertensive patients.
- It helps to identify the potential risk factors influencing the separate as well as joint evolution of systolic and diastolic blood pressure measurements in hypertensive patients. This will in turn help the respective policy makers of the health sector in the effort to design an appropriate intervention strategy.
- It can be used as a reference for those who want to apply separate and joint modeling techniques in two longitudinal continuous sequences.
- It is used to compare the different groups of patients how they respond to the drug simultaneously; so that it serve as a base for further study for the question what brings this variation and others.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Description of Hypertension

Hypertension 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 mm Hg 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. 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).

2.2. Factor Associated with the Progression of Blood Pressure and Related Study

Oliveria *et al.*, (2002) 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.

Davarian *et al.*, (2013) 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.

They obtained the following results: There was no significant difference in the prevalence of overall hypertension by age for men and women from ages 28-49 to 60+. Higher BMI and older age were linked to higher blood pressure and higher chance of having hypertension. More years of education and being female were associated with a lower likelihood of measured hypertension. Smoking, rural residence, and living alone were not significantly associated with the outcome measures.

Holmes and John (2013) 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.

Chenglin *et al.*, (2012) assessed the following variables: Age , BMI, Sex, Pulse rate , Alcohol, Previously diagnosed HTP, Smoking, Diabetes, Feeling stressed , Heart attack ,High cholesterol and Living alone in order to evaluate risk factors associated with blood pressure change over a period of time using linear mixed model to analyzing SBP and DBP individually and jointly. They obtained the following results: according to joint analysis; age, BMI, drinking alcohol, previously diagnosed HTP, smoking and filling stressed were positively associated with change in SBP, but pulse rate, heart attack and higher cholesterol level were negatively associated with change in SBP and sex and living alone were insignificant for SBP. Age, diabetes, heart attack, high cholesterol and living alone were negatively associated with change in DBP and all the rest variables except smoking, which is insignificant on the change of DBP, positively associated with the change in DBP. And also they conclude modeling SBP & DBP jointly has a better overall model fitting and produces a better estimate of correlation between SBP and DBP.

Kondo *et al.*, (2006) applied a general linear mixed model to demonstrate the usefulness of multilevel analysis to assess the association between aging and longitudinal blood pressure variations. They adopted both the general linear regression model and the general linear mixed model in statistical analyses where systolic and diastolic blood pressures (BP) were regressed by the observation period of one year, with the baseline body mass index, age, preference for salty

taste, daily alcohol consumption, smoking status, leisure time physical activity, and family history of hypertension treated as covariates. In the regression model, aging showed a significant relationship with the diastolic BP increase, but not with the systolic BP increase. In the mixed model, aging was found to be a significant predictor of the longitudinal rise in both systolic and diastolic BP. Random-effect analysis showed a significant inverse relationship between baseline BP and the slope of the regression line of longitudinal BP increase. These findings suggested that the effects of the regression to the mean could be separated at the upper level in the hierarchical model, thus resulting in improvement of the statistical power.

Frederico *et al.*, (2004) 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 cros-sectional study. Analysis of variance was used for comparison of means and the chi-square was used for comparison of proportions. Variables associated with increased blood pressure were included in a multiple regression model. They obtained the following results: According to univariate analysis, increased systolic and diastolic blood pressures were associated with high urban life quality index, white race and high body mass index. On multivariate analysis, body mass index, urban life quality index and height remained associated with increased systolic blood pressure; urban life quality index and age were associated with increased diastolic blood pressure.

Edwards and Fisher (2008) provided a linear mixed model example for a repeated blood pressure (BP) study. That showed strong association between repeated systolic and diastolic BP outcomes and a set of fixed effects. Using the same data, they constructed a repeated dichotomous outcome (controlled or uncontrolled BP), fit a GLMM with the same fixed effects used for the linear mixed model. Using the JNC VII classification of BP, BP is considered controlled if systolic BP is less than 140 mmHg and diastolic BP is less than 90 mmHg. they created a binary outcome that indicates whether a person's BP was controlled or uncontrolled at the time of measurement. Thus, for each subject, they have longitudinal binary data indicating controlled or uncontrolled. Then they fitted a GLMM with logit link and with random intercept and slope to this data to determine BP control over time.

Tomeckova and Stanovska (2002), in Czech Republic, were assessed the control of hypertension and survival analysis of the hypertensive patients in STULONG - longitudinal study of risk factors of atherosclerosis. The result show; The average values of blood pressure in hypertensive patients at the entry to the study were 149.35 and 95.98 mm Hg, systolic and diastolic BP respectively (n=289). The percentage of hypertensive patients was rising within the study (34.3 %, n=289, at the entry and 57.1 %, n=160, in the 20th year). The percentage of hypertensive patients on the medication was rising too (22.5 % at the entry and 76.2 % in the 20th year). In spite of aging, the average values of BP in hypertensive patients at the end of the study was lower (142.7 and 85.7 mm Hg, systolic and diastolic BP respectively) compared to the entry. Control of hypertension was dependant on the number of the visit in the whole study – better control was in hypertensive patients with more than twelve visits during the whole study compared to the group of patients with less than 12 visits.

In the study done by Mancia *et al.,* (1999), 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. Age greater than 45 years, previous history of treatment, alcoholism and body weight at initiation of treatment less than 35 kg were found to be risk factors for the progress of blood pressure during the treatment period. The other factors, namely sex, category, education, occupation and smoking, were not found to be significant for risk factor for the progression of blood pressure.

Another study that was done by Salehmohamed and Suheil (2007), in Mombasa, Kenya showed that prevalence of Hypertension in the population found to be 6.7%, increase in age and smoking were found to be a predisposing risk factor for hypertension. Smokers had significant risk ratio of 4 in acquiring hypertension. It was also noted that Chewing of Miraa, previous family history, living area and drinking alcohol had no significant association as a risk predisposing to progress hypertension. Occupation was not significantly associated with hypertension. The study unveiled that physical exercise had protective effect by decreasing the risk of having hypertension. High Body Mass index (BMI) and Weight to Hip ration (WHR) was a predisposing risk factor for having hypertension.

According to Akilew and Tadesse (2012) study in Gondar, Ethiopia, in a cross-sectional study they 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.

A study done by Kalkidan and Misra (2009) in Jimma, Ethiopia on healthy adult volunteers provided evidence that khat chewing induced a significant rise of arterial systolic and diastolic blood pressure and pulse rate in comparison with the baseline values. The peak effect on the arterial blood pressure and pulse rate was reached 3 hours after starting to chew, followed by a decline 1 hour after spitting the leaves corresponding with changes in plasma Cathinone levels an active ingredient of chat.

2.3. Antihypertensive Drugs

Howard *et al.,* (2006) proposed that the inadequate control and treatment of hypertension is believed to be the cause of 33% of all cases of stroke. To date researchers do not have a definitive answer for the cause of hypertension but there are factors that are known to elevate pressure, including body mass**,** age, diet, and family history. The identification, control, or removal of a specific causative factor is imperative for the millions of individuals with hypertension.

Cohen (1981) expressed whether the lowering of blood pressure by antihypertensive drugs might also increase the risk of cardiovascular events and mortality. The degree of reduction in risk for cardiovascular events conveyed by reduction in blood pressure may be related both to the mechanisms behind the pressure lowering and to the adverse effects of the various class of drugs. Whether antihypertensive treatment is effective in reducing the risks of hypertension many studies have been carried out to explore the association between high blood pressure and antihypertensive drug use, and the results have varied widely.

2.4. Longitudinal Data Analysis

Longitudinal data are a series of measurements of the same event taken from the same individual repeatedly over time. The most unique characteristic of longitudinal data is the ability to directly study change. The primary goal of most longitudinal studies is to characterize the change in response over time and the factors that influence this change (Molenberghs and Verbeke, 2008).

2.4.1. Theory of the Linear Mixed Model

Many longitudinal studies are designed to investigate change over time in a characteristic which is measured repeatedly for each patient (Laird and Ware, 1982). Analyses of multiple observations measured on the same individual over time are different from observations measured on different people. Investigators gather repeated measures or longitudinal data in order to study change in a response variable over time as well as to relate these changes in explanatory variables over time (McCulloch *et al.,* 2008).

In addition, modeling the true correlation structure becomes significant in the presence of missing value and when the number of observations per subject is not large. There are two types of covariates in longitudinal studies in general. There are time invariant or baseline covariates (e.g. gender) and time varying covariates (e.g. weight). The Linear Mixed Model (LMM) has become the most commonly used tool for analyzing continuous repeated measures data from a sample of individuals in agriculture, biomedical, economical, and social applications. Thus the term `individual' will have different interpretation or meaning for different areas of application. A special case of a linear mixed model is when there are no fixed effects leading to what is called a random effects model (McCulloch *et al.,* 2008).For example the units may be patients in a longitudinal study where a measurement of biological laboratory markers such as SBP and DBP measures is taken at every month visits. Thus the patient is measured repeatedly giving rise to a cluster of observations from each patient.

The linear mixed-effects model fits the mean response as a combination of population characteristics (fixed-effects) assumed to be shared by all individuals and subject-specific effects (random-effects) that are unique to a particular individual (Nonhlanhla, 2009). By including random-effects in the model, linear mixed-effects models are able to explicitly distinguish between within-subject and between-subject sources of variation. With a linear mixed-effects

model it is not only possible to estimate parameters that describe how the mean responses change over time, but it is also possible to predict how an individual's response trajectories change over time. Mixed-effects models are highly attractive due to their ability to handle missing and unbalanced data reasonably well.

2.4.2. Joint Modeling Approaches

Joint modeling of multivariate outcomes in longitudinal data analysis has been given great deal of attention in the past decades in many studies on a longitudinal outcome during follow-up time. Several approaches for the joint modeling have been proposed by various researchers.

John (2007), in Virginia Common wealth University, used Fels Longitudinal Study data to illustrate both separate and joint mixed-effects modeling strategies. Specifically, jointly modeled longitudinal measures of systolic (SBP) and diastolic (DBP) blood pressure during childhood (ages two to eighteen) were compared between participants who were diagnosed with at least three of the metabolic syndrome risk factors in adulthood (ages thirty to fifty-five) and those who were never diagnosed with any risk factors. On his analysis, he demonstrated the joint model is able to not only answer the same questions addressed as the separate model; it is also able to answer additional important questions about the association in the evolutions of the responses as well as the evolution of the associations. Furthermore, the additional information gained by incorporating information about the correlations between the responses was able to reduce the variability (standard errors) in both the fixed-effects estimates (e.g. differences in groups, effects of covariates) as well as the random-effects estimates

Fieuws and Verbeke (2004) used a joint random-effects model to evaluate hearing performance at two different frequencies measured repeatedly over time on subjects. The authors specified a bivariate longitudinal model for continuous responses with correlated random intercepts and slopes. Error terms were assumed to be independent conditional on the correlated random effects. The results indicated a discrepancy between the observed data and relations implied by the joint model. However, relaxing the conditional independence assumption by allowing the error terms to be correlated, improved model fit and revealed that the discrepancy was due to inappropriate modeling of the error covariance structure.

Bowman and Manatunga (2005) made inferences about the joint process and discussed the prediction aspect, which is an important part of data application. In their thesis they presented a joint model for a continuous response and an associated event risk that is both conceptually and computationally. The joint model consists of two components: observed HDS scores and computed withdrawal profiles. They also presented a mean estimator for both components of this joint process and estimate covariance parameters, including covariance between the two components. Their model provides inferences about the effect of treatment on serial measures of the joint process and a framework to predict levels of depression from updated patient histories

Thiebaut *et al*., (2002) used a random-effect bivariate model with correlated stochastic process to investigate the relationship between CD4 and beta-2-microglobulin, two important immunologic measurements in HIV/AIDS research. Another example of joint random-effect models used in psychometric studies is the work by MacCallum *et al.*, (1997). These authors used a multivariate three-level model specified in a fully Bayesian way to study the relationship between accuracy (binary measurement) and speed of test takers (continuous measurement) on response items clustered within subjects who were nested within groups

Chakraborty *et al*., (2003) obtained estimates of the correlation between blood and semen HIV-1 RNA by using a joint random-effects model. Other examples with longitudinal studies can be found in reference (MacCallum *et al.*, 1997). All of these examples refer to situations where the number of different outcomes is relatively low. Although the model formulation can be done irrespective of the number of outcomes to be modeled jointly, standard fitting procedures, such as maximum likelihood estimation, is only feasible when the dimension is sufficiently low or if one is willing to make a priori strong assumptions about the association between the various outcomes. Williams (2001) used this approach to model simultaneously growth curves for systolic and diastolic blood pressure, height and BMI. However, such a modeling strategy is restricted to the combination of outcomes of the same type.

CHAPTER THREE

3. METHODOLOGY

3.1. Data Source and Its Description

In this study the latest data from retrospective cohort follow up of all hypertensive patients whose age is 18 and above years and, who have followed at least three visits from September 2011 to July 2013 in Jimma University Specialized Hospital were used. Jimma university specialized hospital is located in south west of Ethiopia in Jimma town. It serves as a teaching and referral center for the Jimma area community and adjacent zones. This data was extracted from the follow up patients chart. This chart was recorded by assigning an identification number per individual and contains epidemiological, laboratory and clinical information of all hypertensive patients.

The data consists of 354 individuals with a minimum of three and maximum of thirteen SBP, DBP and other covariates were measured per individual of adult hypertensive patients. Patents' follow up time was one, two or three months gab according to the order of the doctor and the data were recorded on patients' medical follow up card by assigning an identification number per individual by health workers in the chronic follow up clinic, which helps to find the patients profile easily during his/her next visit time.

3.2. Variables

3.2.1. Dependent Variables

Two outcome variables were considered in this study; systolic and diastolic blood pressure for each individual measured at least three times.

3.2.2. Covariates

Five covariates were used for either the separate or joint analyses. Two of these covariates are continuous while three of them are categorical covariates. These covariates are described together with their values or codes in Table 1.

Table 1: Covariates used in the Separate and Joint Analysis of SBP and DBP outcomes

No.	Name	Definitions	Values/Codes
1.	Sex	Sex of hypertensive patients	0=Female,1=Male
2.	Pr	Place of residence	0=Urban, 1=rural
3.	Fh	Family history of hypertensive patients	0=No, 1=Yes
4.	Age	Age of patients at the start of anti-hypertensive drugs	In year
5.	Time	Observed or Follow up time	In month

Baseline age and time of the patients are continuous covariates. Gender of the patient's measure as categorical covariates with two levels: Male and Female. Place of resident is also another categorical covariate with two categories: Urban and Rural to examine whether there is a geographical effect on the elevation of hypertension or not. Finally family history was measured as a categorical variable to provide clear evidence of the pattern or evolution of certain diseases in a family (previously diagnosed HTP).

3.3. Statistical Methods of Data Analysis

3.3.1. Longitudinal Data Analysis

Longitudinal responses may arise in two common situations; one is when the measurements taken on the same subject at different times and the other is when the measurements taken on related subjects (clusters). In both of these cases, the responses are likely to be correlated (Laird and Ware, 1982). The term repeated measures is used to describe both the longitudinal and clustered data. One of the major objectives of statistical analysis is to address variations in the data. For longitudinal data, there are two sources of variations: within-subject variation; the variation in the measurements within each subject, and between-subject variation; the variation in the data between different subjects. Modeling within-subject variation allows studying changes over time, while modeling between-subject variation allows understanding differences between subjects.

3.3.1.1. Exploratory Data Analysis

The first step in any model building process is exploratory data analysis. Data exploration is a very helpful tool in the selection of appropriate models to visualize the patterns of data relative to research interests. Analyses of longitudinal data compare profiles over time and, indeed, time might be viewed as the primary systematic effect to be investigated. The aim of this process is to understand the data structure and determine the relevant modeling approaches suitable for it. Thus, this study assessed the nature of the data by exploring individual profiles, and the average evolution.

3.3.1.1.1. Exploring the Individual Profile

To explore the individual profile, plot of the response with time is used to show whether there is a noticeable pattern common to most subjects. These individual profiles can also provide some information on within and between subject variability. Plotting observed profiles over time helps to identify general trends within subjects and may detect change over time that provides information about the variability at given times.

3.3.1.1.2. Exploring the Mean Structure

Examining the data for clues about the likely nature of the mean structure, to see how the mean profile changes over time, is essential for specifying the functional form of the mean response of the model. So as to understand the possible relationships among means over time, for balanced data, graphical inspection can be used by connecting the average values computed at each time point separately. If the data is not balanced and unequally spaced interval loess smoothing can be used instead. In this study the data is unbalanced and also unequally spaced interval loess smoothing technique is used instead of mean structure over time. This will give idea as to how the mean profile evolves over time. The results of this exploration will be useful in order to choose a fixed-effects structure for the linear mixed model.

3.3.1.2. Linear Mixed Effect Model

Three classes of models are commonly used for analysis of longitudinal data; mixed effects model (or random effects model), marginal models (generalized estimating equations (GEE)

models) and transition models. Mixed-effects models provide a flexible and widely used model for the analysis of continuous longitudinal data introduced to incorporate or model the between-subjects variation and within subject correlation in the data. And it has been a popular method to handle both balanced and unbalanced scenarios, and allows the inclusion of covariates. In marginal models, the mean structure and the correlation (covariance) structure are modeled separately without distribution assumptions for the data while in the transitional models, the within subject correlation is modeled via Markov structures.

In mixed-effects models, response variables are assumed to be a function of fixed effect, non-observable random effect, and error term (Laird and Ware, 1982). When both the fixed and the random effects contribute linearly to the response, the model is called linear mixed-effects model. This model is described by Laird and Ware (1982) can be written as:

$$Y_i = Z_i(K_i\beta + b_i) + \varepsilon_i = Z_iK_i\beta + Z_ib_i + \varepsilon_i$$

Where $Z_iK_i = X_i$ and the final model becomes

$$Y_i = X_i\beta + Z_ib_i + \varepsilon_i \quad (1)$$

Where

- Y_i is the $N\times 1$ response vector for i^{th} subject: $Y_i = (Y_{i1}, Y_{i2}, \ldots Y_{iN})$
- Z_i is a $N \times q$ matrix of known covariates
- X_i is a $N\times p$ design matrix for the fixed effects
- β is a $p_{\times 1}$ dimensional vector of subject specific regression coefficients
- b_i is $q_{\times 1}$ dimensional vector of unknown random effects
- ε_i is $N\times 1$ error vector $\varepsilon_i \sim N(0, \Sigma_i)$, often $\Sigma_i = \sigma^2 I_{ni}$
- $b_i \sim N(0, G)$ i.e:- b_i has a q-variate normal density with mean vector 0 and a variance-covariance matrix G

In this model, $X_i\beta$ is the mean response and Z_ib_i incorporates the random effects part. The Z_ib_i can viewed as the true individual level of SBP or DBP trajectories after they have been adjusted for the overall mean trajectory and other fixed effects. The assumption $var(\varepsilon_i)=\sigma_\varepsilon^2 I$ can be relaxed by allowing to model non-constant variance or special within group correlation

structures. The random effects, b_i, and the with group errors, ε_i , are assumed to be independent for different groups and to be independent of each other for the same group. The columns of Z_{1i} are usually a subset of the columns of X_{1i}.

$$\text{And} \quad \text{Var}(Y_i) = \text{Var}(Z_i b_i) + \text{Var}(\varepsilon_i) = Z_i G Z_i' + \Sigma_i$$

In general, in mixed effects models, random effects b_i is introduced for each subject to incorporate the correlation between the repeated measurements within subject. Since each subject shares the same random effects, the measurements within subject are correlated. Moreover the random effects facilitate subject specific inference. A mixed effects model specifically incorporates both sources of variations: it uses random effects or subject effects to represent deviations of subject longitudinal trajectories from the population average.

Thus, a mixed effects model allows subject specific inference, in addition to standard population average inference.

Assumptions of Linear Mixed Effects Model: Before making inferences about a fitted mixed-effects model, we should check whether the underlying distributional assumptions appear valid for the data or not. There are two basic distributional assumptions for the linear mixed effects model.

i. $\varepsilon_i \sim N(0, \sigma^2 I_{nj})$

The within-group errors are independent and identically normally distributed, with mean zero and variance $\sigma^2 I_{ni}$, and they are independent of the random effects.

ii. $b_i \sim N(0,G)$

The random effects are normally distributed, with mean zero and covariance matrix G (Not depending on the group) and are independent for different groups. The most useful of methods for assessing the validity of these assumptions were based on Q-Q plots of the random effects.

3.3.1.2.1. Estimation of Fixed Effects

Both the maximum likelihood (ML) and restricted maximum likelihood (REML) were used for estimation of the parameters in this study. The maximum likelihood estimation method finds the parameter estimates that are most likely to occur given the data. The parameter estimates are derived by maximizing the likelihood function, which is a mathematical expression that describes the joint probability of obtaining the data expressed as a function of the parameter estimates (Verbeke *et al.,* 1998).

3.3.1.2.1.1. Maximum Likelihood Estimation

Suppose a random sample of N observations is obtained from a linear mixed effect model as defined above, then the likelihood of the model parameters, given the vector of N observations, is defined as:

$$L=l(\beta, \theta, Y_i)=\prod_{i=1}^{N}\left\{2\pi^{-1/2}|V|^{-\frac{1}{2}}\exp(-\frac{1}{2}(Y_i-X_i\beta)'V^{-1}(Y_i-X_i\beta))\right\}$$

Where: - β is a vector of fixed-effects parameters and θ is a vector containing the variance parameters. Given its simplicity in comparison to the likelihood function, the log of the likelihood function is generally used in practice. Its maximum value coincides with that of the likelihood function. The log-likelihood of the model parameters, is defined as

$$\text{Log } L=l(\beta, \theta, Y_i)= -\frac{N}{2}\log(2\pi)-\frac{1}{2}\log|V|-\frac{1}{2}(Y_i-X_i\beta)'V^{-1}(Y_i-X_i\beta)$$

$$=K-\frac{1}{2}\log|V|-\frac{1}{2}(Y_i-X_i\beta)'V^{-1}(Y_i-X_i\beta)$$

$$\text{Where } K=\frac{N}{2}\log(2\pi), \qquad V=Z_iGZ_i'+\Sigma_i$$

Now the values in the model parameters which maximize the log-likelihood may be determined. Estimates of the parameters are found by maximizing the log-likelihood given in above equation with respect to β and θ. One such method that may be used to maximize the log-likelihood function is the maximum likelihood (ML) method. The ML method first maximizes the log-likelihood with respect to the variance parameters, while treating the fixed-effects parameters, β, as constant. Upon determining the variance parameter estimates, the fixed-effects parameters are then determined by finding the values of β which maximize the log likelihood, while treating the variance parameters as constant. It is important to note, the maximum likelihood approach may

produce variance parameters that are biased downwards since they are based on the assumption that the fixed-effects parameters are known (Brown and Prescot, 1999).

Thus, $\frac{\partial l}{\partial \beta}$=-$X_i V^{-1} X_i \beta$+ $X_i' V^{-1} Y_i$

Then, the MLE of $\hat{\beta}$ on combining all the information from all the N subjects equals

$$\hat{\beta}=(\sum_{i=1}^{N} X_i' V^{-1} X_i)^{-1} \sum_{i=1}^{N} X_i' V^{-1} Y_i \quad (2)$$

3.3.1.2.1.2. Restricted Maximum Likelihood Estimation

This is another method that may be used to maximize the log-likelihood function. Sometimes this method is referred to as the restricted maximum likelihood method. It was developed in order to avoid biased variance component estimates that are produced by ordinary maximum likelihood estimation. This is because maximum likelihood estimates of variance components takes no account of the degrees of freedom used in estimating fixed effects. This means that ML estimates of variance component have a downwards bias which increases with the number of fixed effects in the model. For this approach, the fixed-effects parameters, β, are eliminated from the log-likelihood equation, such that it will only be defined in terms of the variance parameters. Then, a likelihood function based on the full residuals, (Y_i-$X_i\hat{\beta}$). It may be noted that the full residuals are a linear combination of y and furthermore (Y_i-$X_i\hat{\beta}$) and $\hat{\beta}$ are independent [8]. From these facts, the joint-likelihood for β and the variance parameters, θ, may be express as a product of the likelihoods based on (Y_i-$X_i\hat{\beta}$) and $\hat{\beta}$

$$L(\theta, \beta; Y_i)=L(\theta; Y_i\text{-}X_i\beta)L(\beta; \hat{\beta}, \theta)$$

Thus, yields the REML, defined as

$$L(\theta; Y_i\text{-}X_i\hat{\beta})=|X_i' V^{-1} X_i|^{-1/2}|V|^{-1/2}\exp\left\{-\frac{1}{2}\left(Y_i - X_i\hat{\beta}\right)' V^{-1}(Y_i - X_i\hat{\beta})\right\}$$

Therefore, the REML log-likelihood is defined as

$$\text{Log } L(\theta; Y_i\text{-}X_i\hat{\beta})=K\text{-}\frac{1}{2}\left\{\log|V| - \log|X_i' V^{-1} X_i|^{-1} + (Y_i - X_i\hat{\beta})' V^{-1}(Y_i - X_i\hat{\beta})\right\}$$

Despite $\hat{\beta}$ appearing in the REML log-likelihood in the above equation, it is present only as a function of the variance parameters. As with the maximum likelihood method, the variance parameters are now estimated by maximizing the REML log-likelihood with regards to the variance parameters. Given the nature of the REML likelihood, and its treatment of the fixed-effects as parameters, rather than as constants, the resulting variance parameter estimates are unbiased. In the same fashion of the maximum likelihood method, values of $\hat{\beta}$ are found by maximizing the REML log likelihood with regards to the fixed-effects parameters, while treating the variance parameters as fixed.

3.3.1.2.2. Model Checking Technique for Separate Linear Mixed Model

For linear mixed effects models, the assumption of normality needs to be assessed by looking at residual errors. It is assumed that the random effects are normally distributed and uncorrelated with the error term. Residual plots can be used visually to check normality of these effects and to identify any outlying effect categories. Examining the plot of the standardized residuals versus fitted values by any covariates of interest can give a better feeling (Molenberghs and Verbeke, 2008). The assumption of normality for the within-group error was assessed with the normal probability plot of the residuals by covariates. Similarly, Normality of the random effects is assessed using Normal Plot of each random effect. Normal plot of estimated random effects helps for checking marginal normality and to identify outliers.

3.3.1.3. Joint Model for Two Continuous Outcomes

The linear mixed model in equation [1] can be easily extended to include bivarate response variables by further stacking the data and defining a specific variance-covariance structure for the random effects. Consider for modeling the two response variables (Y_1 and Y_2) over time and incorporating random intercepts and slopes in order to model the correlations over time between responses.

Let y_{ijk} represent the ith observation, from the jth subject, for the kth response variable, where i = 1, …, n_{ij}, j= 1, …, S, and k = 1, …, K. For this thesis k=1 and 2. Also, define $N_k=\sum_{j=1}^{s} n_{ij}$ and $N=\sum_{k=1}^{k} N_k$.The vector $y_{ij} = [y_{1jk} \quad y_{2jk} \quad . \quad .. \quad . \quad . \, y_{njk}]$ then represents the n_{jk} observation of the kth response variable from the jth subject the vector $Y_k = [Y_{1k} \quad Y_{2j} \quad \dots \quad Y_{sj}]'$ represents the N_k

observation from the k^{th} response variable across all subjects. Finally, the vector $Y= [Y_1\ Y_2 \ldots\ Y_k]$ represents N observation across all response variables and subjects.

Fieuws and Verbeke (2004) were interested in the questions of how the evolution of one outcome is related to the evolution of another outcome ('association of evolutions') and how the association between outcomes evolves over time ('evolution of the association') for longitudinal multivariate data. To get flexible solutions to such questions, they investigated a joint model using a random effects approach. In this approach, random effects were assumed for each outcome and by adopting a joint multivariate distribution for the random effects, the different outcomes were associated.

In the context of modeling two response variables, the linear mixed-effects models for each response variable for subject j taken at time t can be specified as (Fieuws and Verbeke, 2004*)*.

$$Y_{j1}(t)=\mu_1(t)+a_{j1}+b_{j1}(t)+\varepsilon_{j1}(t) \qquad (3)$$

$$Y_{j2}(t)=\mu_2(t)+a_{j2}+b_{j2}(t)+\varepsilon_{j2}(t)$$

Where $\mu_k(t)$ refers to the average evolution (of the k^{th} response over time) and is a function of the fixed effects. The subject specific random intercepts a_{jk} and slopes $b_{jk}(t)$ describe how the subject specific profiles deviate from the average profile for the k^{th} response. The two response trajectories are joined together by assuming a joint distribution for the vector of random-effects, b_i, such as

$$b_i=\begin{bmatrix} a_{j1} \\ b_{j1} \\ a_{j2} \\ b_{j2} \end{bmatrix} \sim N(0,G)$$

Where the variance-covariance matrix for the random effects, G, has the following structure:

$$G=\begin{bmatrix} \sigma_{a1}^2 & \sigma_{a1b1} & \sigma_{a1a2} & \sigma_{a1b2} \\ \sigma_{b1a1} & \sigma_{b1}^2 & \sigma_{b1a2} & \sigma_{b2b1} \\ \sigma_{a2a1} & \sigma_{a2b1} & \sigma_{a2}^2 & \sigma_{a2b2} \\ \sigma_{b2a1} & \sigma_{b2b1} & \sigma_{b2a2} & \sigma_{b2}^2 \end{bmatrix} \qquad (4)$$

The error components for each response, which are independent of the random effects, can be taken to be uncorrelated (σ_{12} =0) and not associated with the random effects, such that the error components are defined as;

$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{21} & \sigma_2^2\end{bmatrix}\right) \tag{5}$$

Assuming σ_{12} =0 implies that, conditional on the random-effects, both response trajectories are independent. The assumption of conditional independence could alternatively be relaxed and the random errors could be taken to be dependent by allowing for a nonzero covariance between the error components ($\sigma_{12} \neq 0$).

3.3.1.3.1. Special Case of Variance Covariance Matrix

Special case can now be obtained by making specific assumptions for the variance covariance matrix G. Two such specific variance-covariance structures are described in the following subsections, a complete independence structure and a shared-parameters structure.

Complete Independence: The two response variable could be taken to be completely independent at any point in time, there by imposing the following structure for G (Howard, 2006):

$$\mathrm{G}=\begin{bmatrix}\sigma_{a1}^2 & \sigma_{a1b1} & 0 & 0\\ \sigma_{a1b1} & \sigma_{b1}^2 & 0 & 0\\ 0 & 0 & \sigma_2^2 & \sigma_{a2b2}\\ 0 & 0 & \sigma_{b2a2} & \sigma_{b2}^2\end{bmatrix} \tag{6}$$

Within a response variable, the random intercept and slope induce within-subject correlations in the repeated measures over time, while assuming independence between subjects. Moreover, this model assumes that the two responses are completely independent. The results for the model would be identical, in theory, to fitting two separate random-effect models.

Shared-Parameters: Now that a complete independence structure has been considered for the G matrix one may consider the other end of the spectrum where the two response variables could be taken to be completely dependent. In this case, the two responses essentially "share" the same

set of random effect parameters (intercept and slope) (Howard, 2006). When two parameters are completely dependent, the correlation between them is equal to one. This occurs when the covariance between the parameters is equal to the square root of the product of their respective variances. Most notations, however, define the model with a 2×1 vector of random effects, such as:

$$b_i=\begin{bmatrix} aj \\ bj \end{bmatrix} \sim N(0, G), \quad \text{with} \quad G = \begin{bmatrix} \sigma_a^2 & \sigma_{ba} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} \tag{7}$$

Clearly, the aforementioned structure imposes strong assumptions on the relationship between the two response variables. It is very unlikely that the two responses would exhibit complete dependence in the association between the random slopes and between the random intercepts. One advantage of this model, when the assumption is tenable, is that it drastically reduces the number of random effects that must be estimated when the number of response variables is large. For models with a large number of response variables, estimation would likely be impossible if the shared-parameters (or alternative approach) were not used.

3.3.1.3.2. Association of the Evolution (AOE)

One important question that may be addressed with a joint mixed-effects model is how the evolution of one response is associated with the evolution of another response ("association of the evolutions"). By definition, the correlation between the evolutions for the two random slopes is given by:

$$rE= \frac{Cov(b_1, b_2)}{\sqrt{Var(b_1)}\sqrt{Var(b_2)}} = \frac{\sigma_{b_1}\sigma_{b_2}}{\sqrt{\sigma_{b_1}^2}\sqrt{\sigma_{b_2}^2}} \tag{8}$$

3.3.1.3.3. Evolution of the Association (EOA)

A similar idea that may be investigated using a joint mixed effects model is how the association between the responses evolves over time ("evolution of the association"). Assuming uncorrelated errors, the marginal correlation between the two responses as a function of time is given by (Fieuws and Verbeke, 2004):

$$\text{rM(t)} = \frac{Cov(Y_{j1}(t), Y_{j2}(t))}{\sqrt{Var(Y_{j1}(t))}\sqrt{Var(,Y_{j2}(t))}}$$

$$= \frac{\sigma_{a_1,a_2} + t\sigma_{a_1,b_2} + t\sigma_{a_2,b_1} + t^2\sigma_{b_1,b_2} + \sigma 12}{\sqrt{\sigma_{a_1}^2 + 2t^2\sigma_{a_1,a_2} + t^2\sigma^2{}_{b_1} + \sigma_1^2}\sqrt{\sigma_{a_2}^2 + 2t^2\sigma_{b_2,b_2} + t^2\sigma^2{}_{b_2} + \sigma_2^2}} \quad (9)$$

3.3.1.3.4. Joint Model Estimation

In the particular context of random-effects models, so-called adaptive quadrature rules can be used (Pinheiro and Bates, 2000), were the numerical integration is centered on the estimates of the random effects, and the number of quadrature points is then selected in terms of the desired accuracy.

To illustrate the main ideas, we consider Gaussian and adaptive Gaussian quadrature, designed for the approximation of integrals of the form

$$\int f(z)\phi(z)dz$$

for a known function *f (z)* and for ϕ(z) the density of the multivariate standard normal distribution. Therefore first standardize the random effects such that they get the identity covariance matrix. Then, the likelihood contribution for subject i equals

$$f_i(y_i|\beta, G, \phi) = \int \prod_{j=1}^{ni} f_{ij}\left(y_{ij}\,|b_i, \beta, \phi\right) f(b_i|G)db_i \quad (10)$$

Where:

- b_i is $q_{\times 1}$ dimensional vector of unknown random effects, $b_i \sim N(0, G)$
- β is a vector of fixed-effects parameters and ϕ is a vector containing the variance parameters
- *f (z)* and for ϕ(z) denotes the density of the multivariate standard normal distribution

3.3.1.4. Correlation Structures

In longitudinal data analysis, when subjects are followed over time, there is a natural ordering of the data for each subject. Correlation structure are used to model dependence among observations, in mixed-effect model, it is used to model dependency among the within-group errors (Pinheiro, and Bates, 2000). The correlation between two within-group errors ε_{ij} , ε_{ij}' is assumed to depend on some distance between them, and ρ is a vector of correlation parameters. Olkin and Tate (1993) described the serial correlation structures in detail of the linear mixed-effects models; serial correlation structures are used to model dependency in the data observed sequentially over time and indexed by a one dimensional time vector. The general serial correlation model is defined as

$$\text{Cor}\,(\varepsilon_{ij}, \varepsilon_{ij'}) = h(\rho),$$

Where h(.)-indicates autocorrelation function. Some of the most common serial correlation structures used in practice are:

Compound symmetry:-It is the simplest serial correlation structure, which assumes equal correlation among all within-group errors of same subject. The corresponding correlation model is

$$\text{Cor}(\varepsilon_{ij}, \varepsilon_{ij}') = \rho$$

While the compound symmetry correlation model tends to be too simplistic for practical application

General (Unstructured):-The general correlation structure represents the other extreme in complexity to the compound symmetry structure. Each correlation is shown by a different parameter, the correlation function is $h(\rho) = \rho_k$; k = 1,2,. .. .While the general correlation model tends to over parameterized model. It is useful for few observations per subject that leads to precise correlation with observations.

Autoregressive (AR):- Box *et al*., (1994) described the family of correlation structure which includes different classes of linear stationary models: autoregressive models, moving average models, and mixture of autoregressive-moving average models. Autoregressive models express the current observation as a linear function of previous observation plus a homoscadastic noise

terms .Let ε_t indexes an observation taken at time t, μ_t indexes a noise term with $E[\mu_t] = 0$, and assumed independent of the previous observations.

$$\varepsilon_t = \phi_1\varepsilon_{t-1} + \ldots + \phi_p\varepsilon_{t-p} + \mu_t \quad |\phi| < 1$$

p is called the order of the autoregressive model, which denoted by AR(p). There are p correlation parameters in an AR (p) model, given by $\phi = (\phi_1, \phi_2, \ldots, \phi_p)$. The AR (1) model is the simplest and one of most useful autoregressive model. Its correlation function is

$$h(k; \phi) = \phi^k \quad k = 0; 1, \ldots$$

In the First-order autoregressive structure it is assumed that the correlation between time points decrease as the distances in time increase, this implies that the number of time intervals between pairs of observation increases, the correlation decreases and approaches to zero. Measurements that are closer in time have higher correlation than measurements with longer time between them. This structure will often be more realistic than the compound symmetry and has the same number of parameters which often makes it more preferable.

Variable selection technique:- To select significant variables, first the main effect and main effect by time interaction were incorporated to the initial candidate model. After that, avoid non significant variables one by one starting from the most non significant terms which is called backward variable selection technique(Pinheiro and Bates, 2002).

3.3.2. Model Comparisons or Selection Techniques

Model selection technique is one of the most frequently encountered problems in data analysis. In most observational epidemiological studies, investigators frequently attempt to construct the most desirable statistical model using the popular methods of forward, backward, and stepwise regression (Pinheiro and Bates, 2002). Of course knowledge of the subject matter plays an important role in model selection, but if based strictly on the data, model selection is often carried out using one of the automated procedures built into the software, of which the most popular method is perhaps stepwise model selection. These methods pose the problem of the arbitrary select ion of the significance levels in allowing a variable to enter into or to be dropped from the model during the selection process (Diggle *et al*., 1994). There is also the problem of multiple testing that comes with fitting and refitting the model. The issue

is made more complicated in the case of repeated o r longitudinal data where selecting the best model means not only to select the best mean structure but also the most optimal variance covariance structure for model selection criteria, like AIC, BIC and likelihood ratio test were used (Shah *et al.*,1997). . In this thesis the most commonly known model selection criteria are Akaike Information Criterion (AIC) (Sakamoto, 1986), the Bayesian Information Criterion (BIC) (Laird and Ware, 1982*)* and Log-likelihood ratio test were used.

$$AIC = -2\log L + 2p \qquad BIC = -2\log \text{Likelihood} + n \text{ Par} \log(N),$$

Where, -2 log L is twice the negative log-likelihood value for the model

P: - is the number of estimated parameters.

n par: -denotes the total number of parameters in the model

N: - is the total number of observations used to fit the model. Smaller values of AIC and BIC reflect an overall better fit.

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Baseline Information and Descriptive Statistics

A total of 354 adult hypertensive patients with a minimum of three and maximum of thirteen measures of SBP, DBP and other covariates per individual of hypertensive patients were included for this study. The baseline characteristics and descriptive statistics of patients are displayed in table 2 below. Out of these hypertensive patients, 173(48.87%) were females and 181(51.13%) were males. About 180(50.84%) of them were living in urban area (Jimma town), and 174(49.16%) of them were living in rural communities ("district") like Yebu, Seka,Lemu, Dedo, Serbo, Asendabo, Juma and adjacent to the towns according to a range of ecological and developmental contexts. About 201(56.78%) of the patients had no genetic effect or family history, 153(43.22%) of patients had family history (previously diagnosed HTP).

The age at first visit ranged from 18 to 86 years with average value equal to 50.198 years (with a standard deviation of 14.036 years). The average number of baseline SBP and DBP is 140.904 and 89.209 per mmHG with standard deviation of 18.583 and 12.727, respectively. The minimum and maximum SBP measurements were 90mmHG and 230mmHG, respectively, and the minimum and maximum DBP measurements were 50mmHG and 130mmHG, respectively.

Table 2: Frequencies and Percentages for baseline categorical covariates and with their baseline average value and standard deviation of SBP and DBP for each category of hypertensive patients' data

No.	Variable	Categories	n (%)	SBP	DBP
				Mean (s.d)	Mean(s.d)
1.	Gender	Male	181(51.13)	143.094(19.389)	88.343(13.271)
		Female	173(48.87)	138.613 (17.465)	86.994(12.397)
2.	Place of	Urban	180(50.84)	142.988(20.716)	88.621(12.963)
	residence	Rural	174(49.16)	138.888(16.062)	86.777(12.713)
3.	Family	No	201(56.78)	142.988(20.716)	88.621(12.963)
	history	Yes	153(43.22)	138.888(16.062)	86.777(12.713)

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

4.2. Separate Analysis of Longitudinally Systolic and Diastolic blood pressure

Both systolic and diastolic blood pressure measures are analyzed separately using the linear mixed models reviewed in Section 3.3.1.2. This is important in order to fully specify the mean response of the model and determine the fixed and random effects to be included in the model.

4.2.1. Exploratory Analysis

Plots are very important to visualize the pattern of SBP and DBP measures overtime before model building; different plots have been explored that expose the patterns relevant to the scientific question about the progress of SBP and DBP of hypertensive patients.

4.2.1.1. Exploring Individual Profile plots of SBP and DBP over time

As shown in figure 1a, the variability of systolic blood pressure between individuals seems higher at baseline and appears to decrease over time. Furthermore, considerable variability is observed within each subject. Similarly, figure 1b depicts a between and within subjects variability in diastolic blood pressure, both implying that the between and within subject specific differences cannot be ignored.

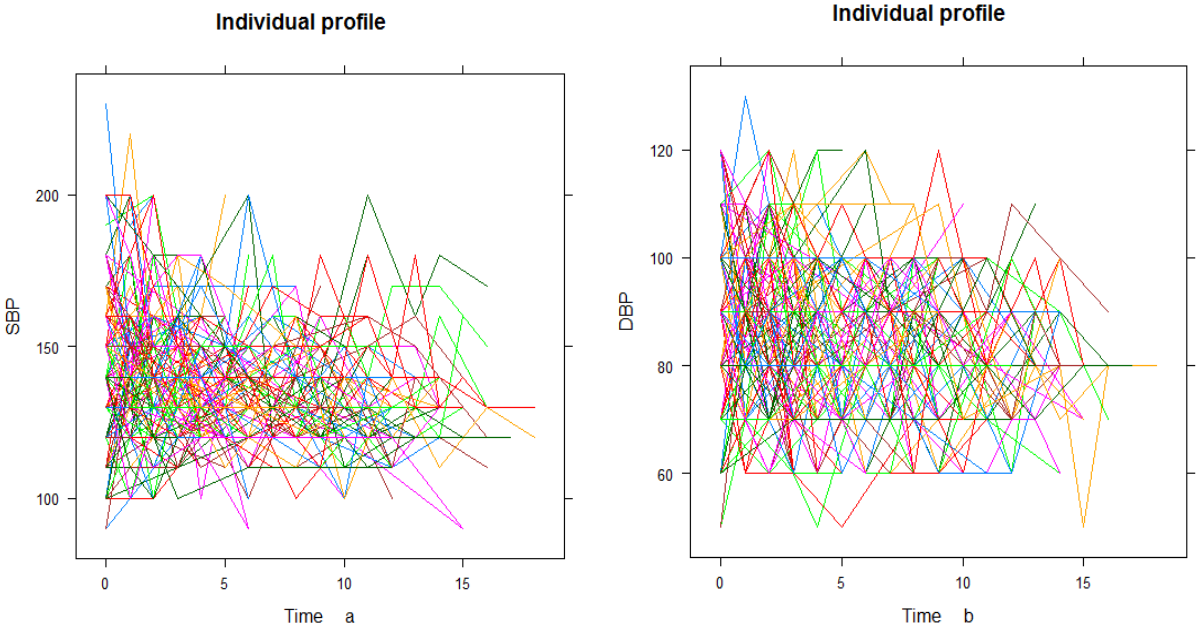

Figure 1: Individual profile plot of Systolic and Diastolic blood Pressure

4.2.1.2. Mean profile Plots of SBP and DBP of hypertensive patients

The loess smooth curve in the figures 2a and 2b suggests that the average profiles of both the SBP and DBP have a linear relationship over time. It indicates both systolic and diastolic blood pressure shows a decreasing pattern over time, but the rate of SBP highly decreasing as compared to DBP. And also it indicates the linear time effects may be included as fixed-effects in the model.

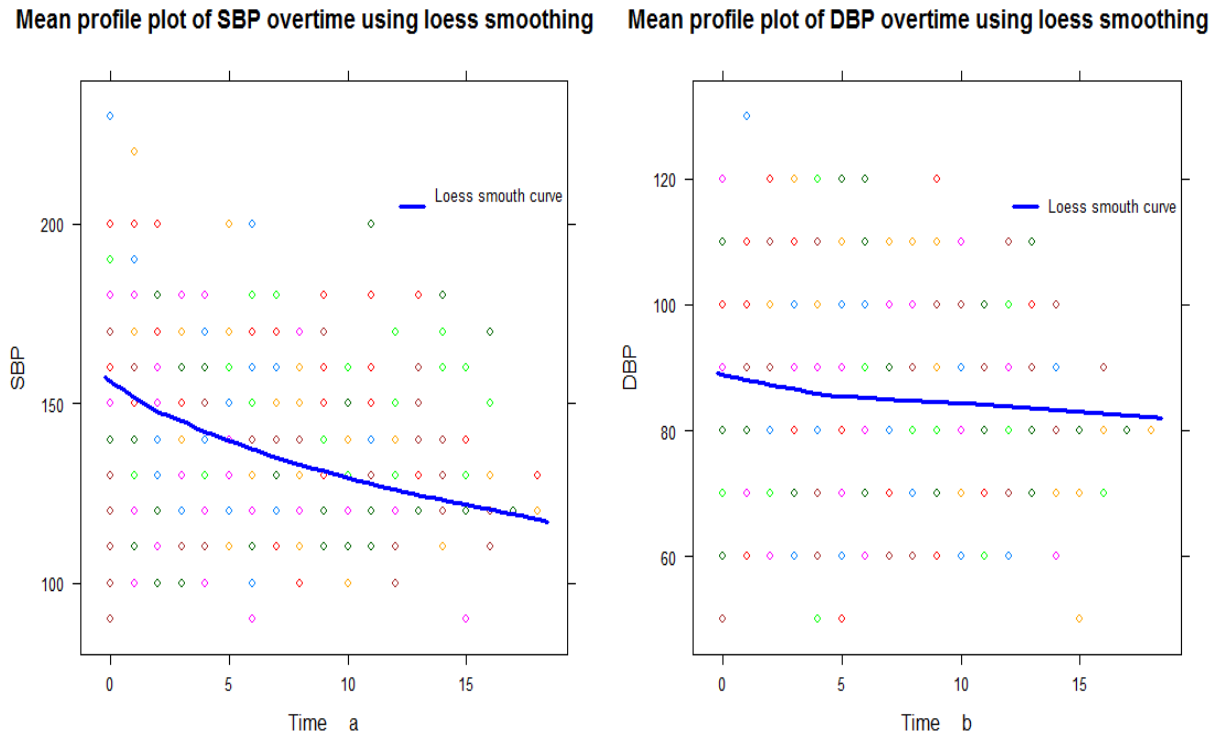

Figure 2: Mean profile plot of SBP and DBP of hypertensive patients

4.2.1.3. Mean Profile plots of SBP and DBP by different categorical covariates and Loess Smooth curve over time

Besides plotting systolic and diastolic blood pressure over time, it is also useful to include different subgroups graph to illustrate the relationship between both systolic and diastolic blood pressures and explanatory variables over time.

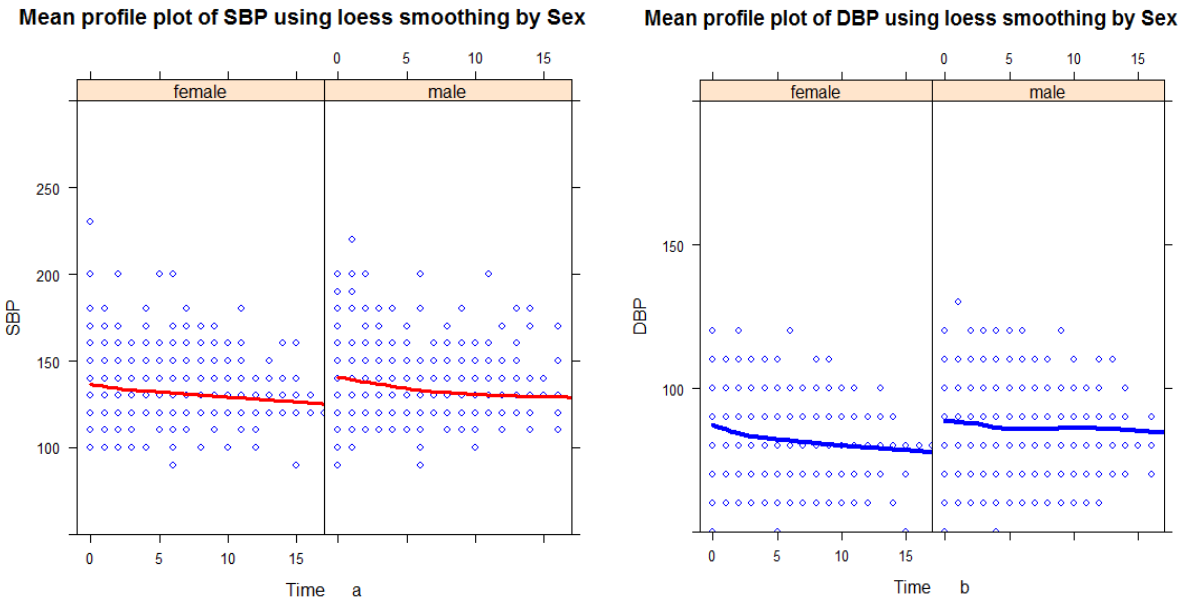

Figure 3: Mean profile plot of SBP and DBP by Sex

As shown in figure 3a and 3b, mean profile plot of SBP and DBP by sex using loess smoothing curve appears that both males and females have a decreasing systolic and diastolic blood pressure over time. However, both figures show that the slope for the males seems to be a bit higher than the slope for the females. It is also tends to indicate the mean profile for males was higher than that of females.

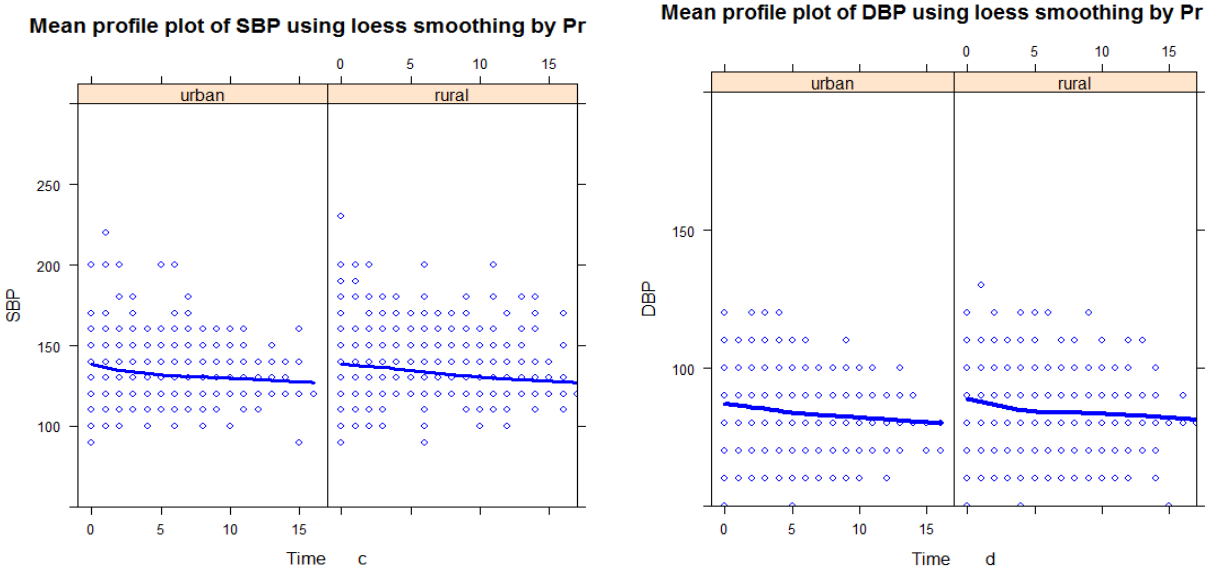

Figure 4: Mean profile plot of SBP and DBP by Place of residence

From figure 4c and 4d mean profile plot of SBP and DBP using loess smoothing curve of hypertensive patients in rural area looks almost the same as that of urban for both systolic and diastolic blood pressure at baseline and the two plots also showed almost the same overtime, to see whether there is significant difference or not the linear mixed effects model have been applied.

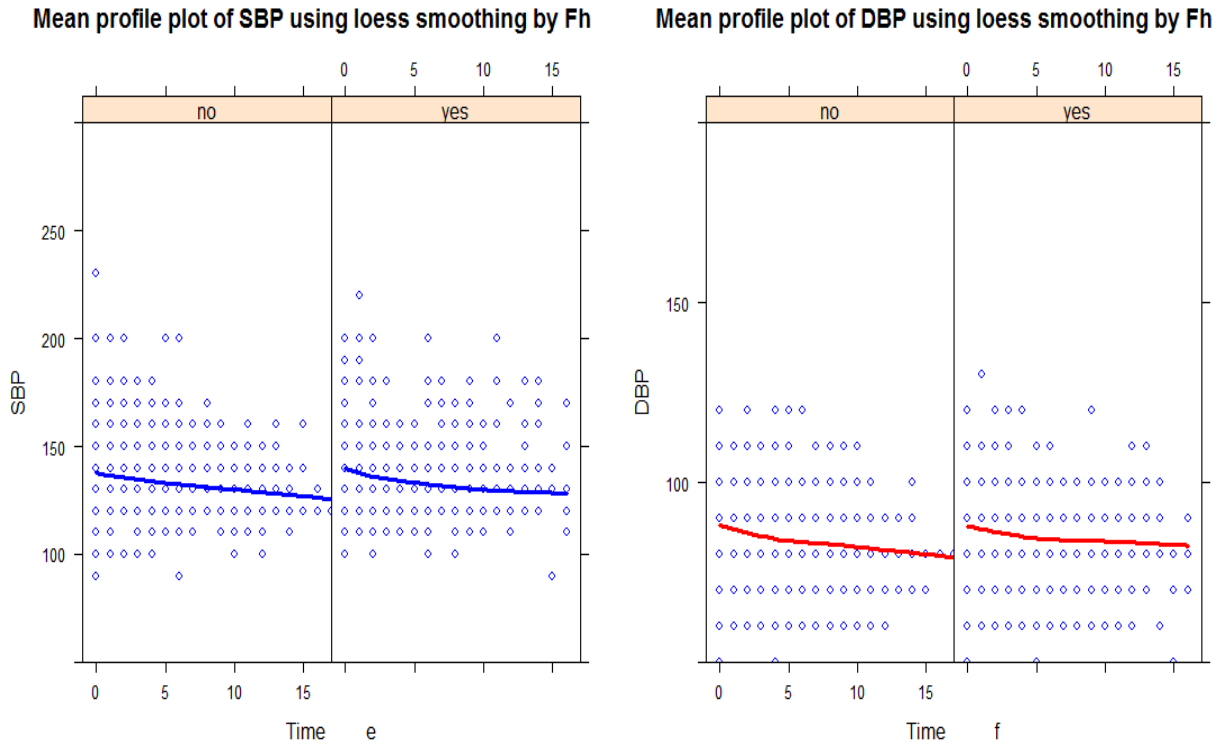

Figure 5: Mean profile plot of SBP and DBP by Family history

From figure 5e and 5f the mean profile plot of SBP and DBP using loess smoothing curve seems there were no difference between whether having family history or not over time as well as at the base line.

4.2.2. Linear Mixed Effects Models Results

Taking advantage of the fact that the linear mixed effects model (1) described by Laird and Ware (1982) in Section 3.3.1.2, has been used to analyze the repeated SBP and DBP measurements.

4.2.2.1. Separate Linear Mixed Model Analysis of Systolic blood pressure

The aim of this section is to select a set of fixed and random effects consecutively to fit a linear mixed model for systolic blood pressure.

4.2.2.1.1. Selection of Fixed Effects for Systolic blood pressure

To select the fixed effect components of the response variable, systolic blood pressure, including all covariates and interaction terms with time without considering the corresponding different random effects were fitted below:

Let SBP_{ij} denote the j^{th} systolic blood pressure of the i^{th} patient at time $t_{ij,}$. Where i indexes the subjects i= 1, 2 ,.., 354 and j indexes the time visit for subject i, j= 1, 2, ..., n_i. n_i represents the overall visits of subject i. Hence, the fixed effects model with linear time effect for SBP measurement is given by:

$$SBP_{ij} = \beta_{10}+\beta_{11}Sex_i+\beta_{12}Pr_i+\beta_{13}Fh_i+\beta_{14}A_i +\beta_{15}T_{ij} + (\beta_{16}Sex_i+\beta_{17}Pr_i+\beta_{18}Fh_i+\beta_{19}A_i)T_{ij} + \varepsilon_{ij}$$

Where:

SBP_{ij}: - Systolic blood pressure for i^{th} subjects A_i: - Age of i^{th} subjects

Sex_i : - Gender of i^{th} subjects Pr_i: - Place of residence for i^{th} subjects

$\beta_{10,}$ $\beta_{11,,,}$ $,\beta_{1p}$:- Are the fixed effect coefficient parameters

Fh_i: - Family history of i^{th} subjects $\varepsilon_{ij:}$ - Error term

T_{ij}: - Time at which systolic and diastolic blood pressure were measures

From the outputs in table 12 (Annex I), we can observe that all the covariate except family history are statistically significant, but all the interaction term except time by sex are statistically insignificant. Thus, the insignificant terms should be removed from the model starting with the most insignificant one of which is the interaction term place of residence by time with p-value of 0.912. The model was then refitted after removing the interaction term place of residence by time and the AIC dropped from 16253.08 to 16251.09 indicating a better fit. The model was fitted again and the categorical covariate family history was still insignificant. The next step is to remove the covariate family history with the p-value of 0.902. The model was fitted again and

the AIC dropped from 16251.09 to 16248.57. By following the same procedure the final fixed effect model for systolic blood pressure is given by:

$$SBP_{ij} = \beta_{10}+\beta_{11}Sex_i+\beta_{12}Pr_i+\beta_{14}A_i +\beta_{15}T_{ij} + (\beta_{16}Sex_i+\beta_{19}A_i)\, T_{ij} + \varepsilon_{ij}$$

Hence, in this study sex, place of residence, age, time and the interaction terms sex by time and age by time used as fixed effects in the model for systolic blood pressure.

4.2.2.1.2. Selection of Random Effects for Systolic blood pressure

In this section the aim is to select the random effect model of the rate of change of SBP measure over time including all potential covariates. In order to retain or remove the random effects from the model, it is better to fit the linear mixed effects model with different random effects.

Thus, four different models with different random effects starting from a simple linear regression model (no random effects) have been explored. Table 3 shows summary measures; Akaki information criteria (AIC), Bayesian information criteria and Log-likelihood ratio test for the models with different random effects. An appropriate random effect to the model was selected by using AIC value. The conclusion is consistent with the AIC and the BIC values for which smaller value is considered as better. That is, the AIC information criterion decreased from 16248.57 to 15940.16, which indicates that model with intercept and slope, was a better fitting model.

Table 3: Selection of Random Effects to be included in the Linear Mixed Effects Model for SBP

No.	Random Effects Included	AIC	BIC	Loglik
1.	No Random Effects	16248.57	16309.18	-8124.301
2.	Random Intercepts	16008.82	16075.66	-7992.407
3.	Random Intercepts and Linear Slopes	15940.16	16018.14	-7956.079
4.	Random Intercepts, Linear and Quadratic Slope	15940.90	16057.83	-7949.448

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

Therefore, both AIC and the BIC criterion suggests including the quadratic time slopes as random effects does not improve the model fit. As a result, the random quadratic time slopes are not included in the subsequent analyses.

Let SBP_{ij} denote the j^{th} systolic blood pressure of the i^{th} patient at time t_{ij}, j indexes the time visit for subject i, and n_i represents the overall visits of subject i. Hence, the full linear mixed effects model with linear time effect for SBP measurements is given by:

$$SBP_{ij} = \beta_{10} + \beta_{11}Sex_i + \beta_{12}Pr_i + \beta_{13}Fh_i + \beta_{14}A_i + \beta_{15}T_{ij} + (\beta_{16}Sex_i + \beta_{17}Pr_i + \beta_{18}Fh_i + \beta_{19}A_i)T_{ij} + w_{1i}(t_{ij}) + \varepsilon_{ij}$$

Where:- $w_{1i}(t_{ij}) = a_{10} + b_{11*}T_{ij.}$ Here, $w_{1i}(t_{ij})$ includes the random effects for intercept and linear time slopes, where the $b_i = (a_{10}, b_{11})' \sim N(0, G)$. The vector $(\beta_{10}, \beta_{11}, , , \beta_{1p})$ of fixed effects describes the average evolution of systolic blood pressure and the vector $(a_{10}; b_{11})$ of random effects describes how the profile of the i^{th} subject deviates from the average profile.

As shown in table 14 (Annex II), family history and interaction terms place of residence by time, family history by time and age by time are statistically insignificant at 5% level of significance. Initially we avoiding non significant variables one by one starting from the most non significant variable then compared the two nested models using AIC. First remove the categorical covariate, family history, having the most insignificant effect (p-value=0.721). The model was refitted after removing family history and the AIC dropped from 15940.16 to 15937.44 indicating a better fit. The model was fitted again and the term place of residence by time was still insignificant with a p-value of 0.703 thus removed from the model. The model was fitted again and the AIC dropped from 15937.44 to 15934.60. There were no other variables to be removed from the model. Hence the reduced model with small number of parameter is preferred. The final linear mixed model for systolic blood pressure is given by:

$$SBP_{ij} = \beta_{10} + \beta_{11}Sex_i + \beta_{12}Pr_i + \beta_{14}A_i + \beta_{15}T_{ij} + (\beta_{16}Sex_i + \beta_{19}A_i)T_{ij} + w_{1i}(t_{ij}) + \varepsilon_{ij}$$

Hence, the inferences for including the random effects model are similar to those of the fixed effect model in terms of magnitude. However, according to the AIC including the random effects component is better than the fixed effect model.

4.2.3.2. Separate Linear Mixed Model Analysis of Diastolic blood pressure

The linear mixed effects model (1) described by Laird and Ware (1982) again used to analyze the repeated DBP measurements. In this section to fit a model we need to select a set of fixed and random effects consecutively for diastolic blood pressure.

4.2.3.2.1. Selection of Fixed Effects for Diastolic blood pressure

To select the fixed effect for the response variables, diastolic blood pressure, for all covariates and interaction terms without considering the corresponding different random effects model were fitted below:

Let DBP_{ij} denote the j^{th} diastolic blood pressure of the i^{th} patient at time $t_{ij,}$. Where i indexes the subjects and j indexes the time visit for subject i, n_i represents the overall visits of subject i. Hence, the full fixed effects model with linear time for DBP measurement is given by:

$$DBP_{ij} = \beta_{20}+\beta_{21}Sex_i+\beta_{22}Pr_i+\beta_{23}Fh_i+\beta_{24}A_i +\beta_{25}T_{ij} + (\beta_{26}Sex_i+\beta_{27}Pr_i+\beta_{28}Fh_i+\beta_{29}A_i)T_{ij} + \varepsilon_i$$

Where: DBP_{ij} :- Diastolic blood pressure for i^{th} subjects

From the outputs in table 13 (Annex I), all the covariate except family history are statistically significant, but all the interaction term except time by age are statistically insignificant. Thus, the insignificant terms should be removed from the model starting with the most insignificant one of which is the interaction term family history by time (p-value=0.874). The model was then refitted after removing the interaction term family history by time the AIC dropped from 15150.77 to 15148.79 indicating a better fit. The model was fitted again and the interaction term place of residence by time was still insignificant .The next step is to remove the interaction term place of residence by time with the p-value of 0.7927. The model was fitted again and the AIC dropped from 15148.79 to 15145.46. By following the same procedure the final fixed effect model for diastolic blood pressure is given by:

$$DBP_{ij} = \beta_{20}+\beta_{21}Sex_i+\beta_{22}Pr_i+\beta_{24}A_i +\beta_{25}T_{ij} + (\beta_{26}Sex_i+\beta_{29}A_i)T_{ij} + \varepsilon_i$$

Hence, in this study sex, place of residence, age, time and the interaction terms sex by time and age by time used as fixed effects in the model for diastolic blood pressure

4.2.3.2.2. Selection of Random Effects for Diastolic blood pressure

After determining the fixed effects, we need to select a set of random effects which can help defining a model. In this section the aim is to select the random effect model of the rate of change of DBP measure over time including all potential covariates. We start by fitting the random effect model for all covariates and interaction terms consider the corresponding different

random effects model. In order to retain or remove the random effects from the model, it is better to fit the linear mixed effects model with different random effects.

Thus, four different models with different random effects starting from a simple linear regression model (no random effects) have been explored. Table 4 shows summary measures; Akaki information criteria (AIC), Bayesian information criteria and Log-likelihood ratio test for the models with different random effects. An appropriate random effect to the model was selected by using AIC. The conclusion is consistent with the AIC and the BIC values for which smaller value is considered as better. That is, the smallest AIC value14937.28 below indicates that model with intercept and slope was a better fitting model.

Table 4: Selection of Random Effects to be included in the Linear Mixed Effects Model for DBP

No.	Random Effects Included	AIC	BIC	Loglik
1.	No Random Effects	15174.35	15235.62	-7576.175
2.	Random Intercepts	14945.01	15011.85	-7460.504
3.	Random Intercepts and Linear Slopes	14937.28	15015.26	-7454.639
4.	Random Intercepts, Linear and Quadratic Slope	15017.30	15944.40	-73903.30

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

Therefore, both AIC and the BIC criterion suggests including the linear time effect as random effects does improve the model fit. The average evolution also augments a similar idea (figure 2b). Hence, the full linear mixed effects model with linear time effect for DBP measurements is given by:

$$DBP_{ij} = \beta_{20} + \beta_{21}Sex_i + \beta_{22}Pr_i + \beta_{23}Fh_i + \beta_{24}A_i + \beta_{25}T_{ij} + (\beta_{26}Sex_i + \beta_{27}Pr_i + \beta_{28}Fh_i + \beta_{29}A_i)T_{ij} + w_{2i}(t_{ij}) + \varepsilon_{ij}$$

Where:- $w_{2i}(t_{ij}) = a_{20} + b_{21*}T_{ij.}$ Here, $w_{2i}(t_{ij})$ includes the random effects for intercept and linear time slopes, where the $b_i = (a_{20}, b_{21})' \sim N(0, G)$.

As shown in table 15 (Annex II) , family history, place of residence and interaction terms sex by time place of residence by time, family history by time and age by time are statistically insignificant at 5% level of significance. Initially we avoiding non significant variables one by one starting from the most non significant variable then compared the two nested models using AIC. First remove the categorical covariate, family history, having the most insignificant effect

(p-value=0.811). The model was refitted after removing family history and the AIC dropped from 14937.28 to 14932.97 indicating a better fit. The model was fitted again and the term place of residence by time has still highly insignificant with a p-value of 0.8549 thus remove from the model. The model was fitted again and the AIC dropped from 14932.97 to 14929.18. By following the same procedure the final linear mixed model for diastolic blood pressure with small number of parameter, which is given by:

$$DBP_{ij} = \beta_{20} + \beta_{21}Sex_i + \beta_{22}Pr_i + \beta_{24}A_i + \beta_{25}T_{ij} + (\beta_{26}Sex_i + \beta_{29}A_i)T_{ij} + w_{2i}(t_{ij}) + \varepsilon_{ij}$$

Even though place of residence is insignificant, the AIC value of the nested model indicate that it is still existed in the final model. The inferences for the random effects model are similar to those of the fixed effect model in terms of magnitude and direction. However, according to the AIC the random effects model is better than the fixed effect model.

4.2.4. Selecting Correlation Structure for both SBP and DBP measures

In longitudinal study selecting best model is not selecting model with only the best mean structure, but also correlation structure. Among different correlation functions or correlation structure classes, in this study the most common correlation structures; unstructured covariance model, compound symmetric covariance models, and autoregressive structure of order one or AR (1) were used and compared. The small AIC value indicated that the model with unstructured covariance function is preferable for both response variable; systolic and diastolic blood pressure, as shown in table 5.

Table 5: Comparison of model with different correlation function for SBP and DBP

Model	SBP			DBP		
	AIC	BIC	logLik	AIC	BIC	logLik
Unstru.Model.	15934.60	15995.89	-7956.298	14929.18	14990.47	-7453.591
AR.Model.	15948.27	15997.14	-7962.137	14931.27	14993.13	-7454.134
Comsymm.Model	16165.25	16220.98	-8072.627	15085.51	15141.23	-7532.755

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

Unstru.Model=unstructured covariance model, AR.Model=autoregressive structure of order one or AR (1), Comsymm.Model=compound symmetric covariance models

Therefore, for the data set of this study linear mixed model with unstructured correlation can be considered as best final model for both response variables.

Table 6: Parameter estimates and standard errors for the separate linear mixed effects models of the SBP and DBP outcomes for the final model

SBP				DBP			
Para-Mater	Estimates (s.e)	p value	95% CI	Para-Mater	Estimates (s.e)	p value	95% CI
β_{10}	128.189(2.811)	0.0000	(122.7, 133.8)*	β_{20}	89.22(1.967)	0.000	(85.42, 93.163)*
β_{11}	5.121(1.539)	0.0010	(2.155, 8.216)*	β_{21}	2.926 (0.856)	0.007	(1.255, 4.639)*
β_{12}	3.011(1.099)	0.0062	(0.695, 5.034)*	β_{22}	0.805 (1.035)	0.436	(-1.411, 2.656)
β_{14}	0.144(0.054)	0.0086	(0.036, 0.253)*	β_{24}	-0.077 (0.037)	0.039	(-0.152, -0.003)*
β_{15}	-1.744 (0.308)	0.000	(-2.365, -1.135)*	β_{25}	-1.093(0.465)	0.0191	(-2.017, -0.170)*
β_{16}	-0.799(0.248)	0.0013	(-1.290, -0.312)*	β_{26}	-0.081 (0.176)	0.038	(-0.386, -0.038)*
β_{19}	-0.011(0.009)	0.0300	(-0.032,0-.007) *	β_{29}	-0.024 (0.005)	0.001	(-0.036, -0.013)*
Random effects							
$Var(a_{10i})$		130.093		$Var(a_{20i})$		53.666	
$Var(b_{11i})$		1.775		$Var(b_{21i})$		0.3615	
$Corr(a_{10i}, b_{11i})$		-0.715		$Corr(a_{20i}, b_{21i})$		-0.553	
σ_1^2		148.285		σ_2^2		94.112	

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

indicates insignificancy at 0.05 level of significance, AIC value for SBP=15934.6 and AIC value for DBP= 14929.18.

The estimated parameter for intercept of SBP is 128.189 with standard error of 2.811represents an estimate of the average level of SBP during the first follow up time. There is a significant difference between males and females at 5% level of significance. The significant gender effect and the value 5.12l indicates that on average males started with the higher SBP measure than female (reference group) at baseline. In addition the interaction of sex by time is significant and -0.799 indicates that the rate of change of the average SBP in males is nearly lower by 0.8mmHG as compared to females. The average intercepts for the rural and urban hypertensive patients are statistically different with the parameter estimate 3.011, which indicates that on average hypertensive patients living in rural area started with the higher SBP measure than living in urban area (reference group) at baseline. A parameter estimate age for SBP indicates a one year increase in age was associated with a normal increase of 0.144mmHg in SBP. The parameter estimate of the interaction for age and time is -0.011 which implies that the average rate of increase is inversely related to age. In other words, younger patients have the higher rate of change in SBP measure than older patients. The coefficient of time -1.744 (S.E = 0.308) indicates systolic blood pressure decreases overtime; this implies a unit increase in time was associated with 1.744mmHg decrease on SBP, after adjusting for sex and age of patients.

In the same manner the estimated parameter for intercept of DBP is 89.221 with standard error of 1.967 represents an estimate of the average level of DBP at time = 0 (during the first follow up time). The intercept for males is 2.926 greater than that for females and they are statistically significantly different, which indicates that on average male started with the higher DBP measure than female at baseline. In addition the interaction of sex by time is significant and -0.081 indicates that the rate of change of the average DBP in males is nearly lower by 0.08 units as compared to females. Place of residence is not significant at baseline; this means that there was no statistically significant difference in DBP measures at baseline for rural and urban hypertensive patients. A parameter estimate age for DBP indicates a one year increase in age was associated with a normal decrement of 0.077mmHg in DBP. The parameter estimate of the interaction for age and time is -0.024 which implies that the average rate of increase is inversely related to age. In other words, younger patients have the higher rate of change in DBP measure than older patients. The coefficient of time -1.093(S.E = 0.465) indicates diastolic blood pressure decreases overtime; this implies a unit increase in time was associated with 1.093mmHg decrease on DBP, after adjusting for sex and age of patients.

The intercept of the random effects for both SBP and DBP indicates there is variability between subjects at base line. And the slope of random effects for both SBP and DBP indicates there is variability within subjects over time. The correlation -0.715 and -0.553 indicates, there is a negative correlation between intercept and slope of linear time effect for the random part for SBP and DBP, respectively. In addition, from the random effects, the residual terms σ_1^2 =148.285 and σ_2^2 =94.112, indicates that variation within the hypertensive patients in different time of SBP and DBP measurements, respectively.

4.2.5. Model Diagnostics for Separate Linear Mixed Model of both SBP and DBP

Residuals versus observation ID number plots for final separate linear mixed model for both SBP and DBP are presented in figure 7 (Annex IV) suggested the residuals are symmetric around zero (i.e. positive and negative residuals are almost equal) and there is very few outlier for both response variables. Figure 9 in the same annex; also show the assumption of normality seems to satisfy for a random effect that is for intercept and slopes for both SBP and DBP. Q-Q plots for normality of random effects for both outcomes are also given in figure 8 of the same annex; which illustrates the random effects are normally distributed with mean zero and variance covariance matrix G. Residual versus fitted value plots in figure 10, we can see that the residuals seem to be randomly distributed with constant variance, since both plots does not show any systematic pattern. Thus, it meets the assumption of error term.

4.3. Joint Analysis of Systolic and Diastolic blood pressure

Previously, the two outcomes are analyzed separately for comparison purpose and identifying associated risk factors for the progress of SBP and DBP separately. The main focus of this study is to investigate the joint evolution and association of SBP & DBP, and associated risk factors for the progress of the two end points by considering a joint linear mixed effects model.

A joint linear mixed-effects model (3) was used to fit the two response variables, DBP and SBP, assuming an unstructured variance-covariance structure as discussed is section 3.3.1.3. This model is the same as the separate model discussed in the previous section, except the sets of random intercepts and slopes for each response are now correlated rather than independent. This model was fitted allowing for a linear time effect for each covariate and by considering all covariates as a fixed effect with all possible interaction terms, which is given in table 16 Annex.

As shown from the output in table 16 (Annex III), all the covariates except family history are statistically significant for SBP and all the covariates except family history and place of residence are statistically significant at 5% level of significance for DBP. The interaction term sex by time and age by time are significant for both SBP and DBP. Thus, the insignificant terms should be removed from the model and refitted after removing the insignificant terms, the AIC value dropped from 30865.4 to 30694.6 indicating a better fit, which is the final joint model.

Table 7: Parameter estimates and standard errors for the joint linear mixed effects models of the SBP and DBP outcomes for the final model

	SBP			DBP	
Effect	Estimates(s.e)	95%CI	Effect	Estimates (s.e)	95%CI
Intercept	128.46(2.801)	(122.98,134.12)*	Intercept	89.391(1.965)	(85.53, 93.27)*
SexMale	5.148(1.540)	(2.165, 8.262)*	SexMale	3.053(1.074)	(0.941, 5.164)*
PrRural	3.014(1.103)	(0.444, 4.785)*	PrRural	0.323(0.826)	(-1.302,1.948)
Age	0.140(0.055)	(0.032, 0.248)*	Age	-0.078(0.038)	(-0.153,-0.003)*
Time	-1.797(0.317)	(-2.424,-1.171)*	Time	-1.097(0.455)	(-1.990,-0.205)*
SexMale×Time	-0.809(0.25)	(-1.319, 0.294)	SexMale ×Time	-0.048(0.167)	(-0.378, 0.285)
Age×Time	- 0.011(0.008)	(-0.025,-0.0024)*	Age×Time	-0.025(0.006)	(-0.033, -0.012)*

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

** indicates significance at 0.05 level of significance AIC value=30694.6*

Here, the result shows all the parameters are significant at 5% level of significance except the interaction term sex by time for SBP and place of residence and the interaction term sex by time for DBP. Thus, the variable sex, place of residence and age are identified as positively associated with change in SBP, but time is negatively associated with SBP. Sex is the only variable which is identified as a positive risk factor for the change in DBP, but time and age are negatively associated with the change in DBP.

As interpreted in separate model previously the parameter estimate of joint model also gives the same interpretation. The estimated parameter for intercept of SBP is 128.46 with standard error of 2.801represents an estimate of the average level of SBP during the first follow up time and 89.391 is the estimated parameter for intercept of DBP with standard error of 1.965 represents an estimate of the average level of DBP at time zero. There is a significant difference between males and females at 5% level of significance for both SBP and DBP. The parameter estimate 5.148 and 3.053 for SBP and DBP respectively, indicates that on average males started with the higher SBP and DBP measure than females (reference group) at baseline. The interaction of sex by time is insignificant in joint model; this means that there was no statistically significant difference on the rate of change of the average SBP and DBP in males and females. The average intercepts for the rural and urban hypertensive patients are statistically different with the parameter estimate 3.014, which indicates that on average hypertensive patients living in rural area started with the higher SBP measure than living in urban area (reference group) at baseline. But place of residence for DBP is not significant at baseline; this means that there was no statistically significant difference in DBP measures at baseline for rural and urban hypertensive patients. A parameter estimate age for both SBP and DBP indicates a one year increase in age was associated with a normal increase of 0.140mmHg (SE = 0.055) in SBP and a normal decrease of 0.078mmHg (SE =0.038) in DBP. The parameter estimate of the interaction for age and time is 0.014 and 0.025 for SBP and DBP, respectively, which implies that the average rate of increase is inversely related to age. In other words, younger patients have the higher rate of change in both SBP and DBP measure than older patients. The negative sign for the coefficient of time indicates systolic and diastolic blood pressure decreases overtime. Hence, a unit increase in time was associated with 1.797mmHg of decreasing on SBP and 1.097mmHg of decreasing on DBP. In addition, standard error estimates for each parameter estimates in 'table 7' improves the performance of estimates and if the standard error value for a particular

parameter estimate is large it signify poor estimation of the parameter i.e less efficiency of the parameter.

Accordingly, the SAS PROC MIXED for joint model also provides the estimated variance covariance matrix, and the estimated correlation matrix for random effects of both the SBP and the DBP as determined in the form of using equation(4) from section 3.3.1.3 have been shown in Table 8 and Table 9, respectively.

Table 8: Variance-Covariance estimates for the final joint model

		SBP		DBP	
		Intercept	Slope	Intercept	Slope
SBP	Intercept	133.16	-11.5415	76.3829	6.2179
	Slope	-11.5415	2.0101	6.5604	0.8112
DBP	Intercept	76.3829	6.5604	55.3012	-2.8417
	Slope	6.2179	0.8112	-2.8417	0.4827

Table 9: Estimated Correlation for the final joint model

		SBP		DBP	
		Intercept	Slope	Intercept	Slope
SBP	Intercept	1.0000	-0.7054	0.8901	0.7756
	Slope	-0.7054	1.0000	0.6222	0.8236
DBP	Intercept	0.8901	0.6222	1.0000	-0.5500
	Slope	0.7756	0.8236	-0.5500	1.0000

From the random effects, it can be seen that variability is higher for SBP than DBP. The same may be said of the covariance for SBP and DBP; this means SBP appears to be more extreme. Also, the covariance's for both DBP and SBP are positive, which is indicative of a positive correlation, which is seen in the estimated correlation matrix. With joint mixed-effects model is possible to investigate how the evolution of DBP is associated with the evolution of SBP, the association of the evolutions (AOE). It is also possible to determine how the association between DBP and SBP evolves over time, the evolution of the association (EOA). The AOE can be determined by using equation (8) from section 3.3.1.3.2 or by reading the correlation between the two slopes directly from the estimated correlation matrix (Table 8). Here the AOE between the random slope for DBP and the random slope for SBP is 0.8236. Thus, the larger positive value

suggests a positive strong association between the evolution of systolic and diastolic blood pressures.

The EOA can be determined, and then visualized, using the marginal correlation between DBP and SBP, equation (9) from section 3.3.1.3.3. To visualize this, the implied correlation has been calculated and plotted over time using the marginal correlation between both response variables in Figure 6. Notice, that at its weakest correlation is 0.0075, at baseline, and this association slightly increases over time.

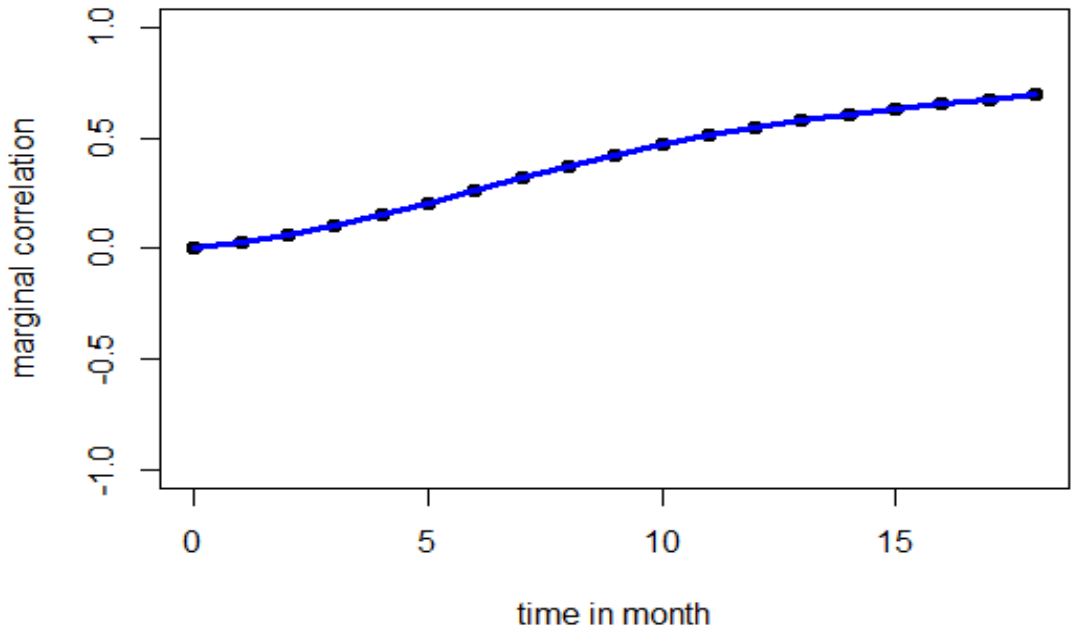

Figure 6: Associations of the Evolution

4.4. Comparison of Separate and Joint Model

Technically, the separate models were fitted for the two outcomes together anyway, but assuming that $\rho = 0$ (fit as a joint model with appropriate covariance terms equal to zero), which is entirely equivalent to fitting the models separately, using SAS PROC NLMIXED for both separate and joint model provides the following results.

Table 10: Parameter estimates and standard errors for separate and joint linear mixed effect model

Effect	Separate Model		Joint Model	
	Estimates(s.e)	95%CI	Estimates (s.e)	95%CI
Systolic blood pressure				
Fixed effects				
Intercept	128.19(2.814)	(122.66, 133.72)	128.46(2.801)	(122.89, 134.02)
SexMale	5.122(1.540)	(2.093, 8.151)	5.149(1.539)	(2.102, 8.197)
PrRural	3.012(1.105)	(0.837, 5.187)	3.0124(1.075)	(0.312, 5.713)
Age	0.145(0.055)	(0.037, 0.252)	0.146(0.020)	(0.032, 0.249)
Time	-1.777(0.305)	(-2.379, -1.174)	-1.797(0.317)	(-2.424, -1.171)
SexMale×Time	-0.799(0.259)	(-1.292,-0.308)	-0.809(0.257)	(-1.319, 0.293)
Age×Time	-0.011(0.009)	(-0.029,-0.007)	-0.014(0.008)	(-0.028,-0.007)
Random effects				
Var($\hat{a}_{10}$)	130.09(15.124)	(104.88, 165.68)	133.16(15.036)	(107.56, 169.18)
Var($\hat{b}_{11}$)	1.775(0.366)	(1.230, 2.786)	2.010(0.357)	(1.401, 3.126)
σ_1^2	149.28(5.684)	(138.74,161.08)	147.53(5.613)	(137.12, 159.19)
Diastolic blood pressure				
Fixed effects				
Intercept	89.349(1.956)	(85.502, 93.197)	89.351(1.967)	(85.482, 93.219)
SexMale	2.842(1.068)	(0.740, 4.944)	2.858(0.853)	(1.1840, 4.533)
PrRural	0.463(0.834)	(-1.177, 2.103)	0.423(0.828)	(-1.177, 2.083)
Age	-0.076(0.039)	(-0.151, -0.001)	-0.078(0.038)	(-0.153,-0.003)
Time	-1.093(0.468)	(-2.016,-0.170)	-1.097(0.475)	(-1.990,-0.204)
SexMale ×Time	-0.079(0.177)	(-0.296,-0.041)	-0.048(0.167)	(-0.378, 0.285)
Age×Time	-0.024(0.006)	(-0.036,-0.013)	-0.025(0.006)	(-0.037,-0.013)
Random effects				
Var($\hat{a}_{20}$)	53.698(7.498)	(41.705, 71.755)	55.301(7.455)	(59.594,93.171)
Var($\hat{b}_{21}$)	0.359(0.150)	(0.182, 1.013)	0.483(0.142)	(0.267, 1.127)
σ_2^2	95.179(3.625)	(88.456, 102.70)	93.936(3.575)	(9.026, 9.694)
Common parameters				
Corr. random effects	ρ	-	0.824(0.199)	(0.421, 1.202)
−2log-likelihood		30820.0		30670.6
AIC		30836.0		30694.6

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

ρ = correlation of the random effects, s.e= standard error, CI= confidence interval, AIC= Akaike information criteria

Here, the results from the separate and joint analyses are quite similar to each other but not identical. Clearly, systolic blood pressure and diastolic blood pressure show a strong direct relationship as evidenced by the correlation of the random effects in the joint model. The -2 log-likelihood value corresponding to the two separate models (fit as a joint model with appropriate covariance terms equal to zero) is equal to 30820.0.The -2 log-likelihood value for the joint model is 30670.6. A likelihood ratio test indicated that the joint model provided a significantly better fit than the two separate models (x^2=149.4 df =4, p-value <0.0001). With regards to Akaike's information criterion (AIC), the joint model (AIC = 30694.6) is also indicated as a better fit than the separate model (AIC = 30836.0). Comparing the separate and joint models all parameter estimates for both continuous response variables are almost the same, except small changes.

Comparing the continuous covariate age between the two types of models will yield further information of interest. The table 10 shows both the separate and joint model found a significant relationship between baseline age and SBP. Both were positively associated with SBP (slopes of 0.145 compared to 0.146), however, the SE (0.055 compared to 0.020) is smaller for the joint model, hence the 95% CI is also tighter for the joint model. Both models also concluded a normal decrease with regards to baseline age for DBP.

The estimated values for both the variance-covariance matrix (Table 8) and the correlation matrix (Table 9) also used for comparison purposes have been combined in Table 11, for both the separate models and the joint model. When comparing the results from the independent setting to the results from the dependent setting there are several points of interest. Notice how the joint model seems to decrease the variability in the random effects. Taking into account the SE's for the variance and covariance estimates, the joint model in general allowed for more accurate prediction (smaller errors) of the variability in the random effects, though just slightly.

Table 11: Estimates for the Covariance and Correlation Parameters in the Separate and Joint Models

	Covariance Estimates		Correlation Estimates	
	Separate Model	Joint Model	Separate Model	Joint Model
-2 Res	30820.0	30670.6	30820.0	30670.6
AIC	30836.0	30694.6	30836.0	30694.6
$\sigma_{a_1}^2$	130.09(15.124)	133.16(15.036)	1.000	1.000
$\sigma_{a_1a_2}$	-	76.383(8.566)		0.890
$\sigma_{a_2}^2$	53.698(7.499)	55.301(7.455)		1.000
$\sigma_{a_1b_1}$	-10.872(2.172)	-11.542(2.073)	-0.715	-0.705
$\sigma_{a_2b_1}$	-	6.560(1.368)		0.622
$\sigma_{b_1}^2$	1.775(0.366)	2.010(0.357)	1.000	1.000
$\sigma_{a_1b_2}$	-	6.218(1.321)		0.775
$\sigma_{a_2b_2}$	-2.446(0.964)	2.842(0.803)	-0.557	-0.550
$\sigma_{b_1b_2}$	-	0.811(0.199)		0.824
$\sigma_{b_2}^2$	0.359(0.151)	0.483(0.142)	1.000	1.000
σ_1^2	149.28(5.685)	147.53(5.613)	149.28(5.685)	147.53(5.613)
σ_2^2	95.179(3.625)	93.936(3.575)	95.179(3.624)	93.936(3.575)

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

(-2 Res = log-likelihood, AIC = Akaike's information criterion.)

In general, the joint model is preferred as it has a smaller log-likelihood and AIC than the separate model. Also, the SE's for the variance and covariance estimates are also evidence that the joint model is better than the separate models.

4.5. Discussion

In this thesis, two methods were considered for fitting two response variables measured longitudinally, a separate linear mixed effects model and a joint model. Since a joint model building usually starts from separate models for each component, initially each data are analyzed separately. Such separate analysis is preferred for several reasons. Firstly, it helps to specify the mean response of the model. Secondly, the random effects and fixed effects to be included in the linear mixed effect model can be easily determined, and thirdly initial values to be provided for the joint models can be obtained.

In separate linear mixed model separate analysis of systolic and diastolic blood pressure were carried out. Before fitting the linear mixed model separately for each out comes, exploring the data analysis have been explored to understand the data structure and determine the relevant modeling approaches. From individuals profile plot, we observed the existence of variability in both SBP and DBP within and between individuals. The exploratory analysis result for mean structure (loess smooth curve) also suggested that on average, both SBP and DBP measures slightly decreasing in a linear pattern over time, but the rate of decreasing is high in SBP than DBP. This supports the results of Tomeckova and Stanovska (2002) ,who found that the average values of BP in hypertensive patients at the end of the study was lower compared to the entry.

In the separate linear mixed analysis of the systolic blood pressure, fixed and random effect components were selected to include in the model. In linear mixed model with only intercept, intercept + time and intercept + time + time^2 were compared for the purpose of selecting the best random effect that enable to account the variability between individuals. The three models were compared using the AIC value and we got a model with intercept + time as random effect is the best and the variables to be included in the model were determined using backward variable selection method using R. Then, all the five covariates and their interaction term by time were considered, of these, family history with its interaction term by time and place of residence by time were extracted to be included in the final model. In this regard, more or less the findings from separate logistic regression model by Oliveria *et al*., (2002) supports this results, who found that among socio-demographic variable age, sex and place of residence were significantly related with hypertension. But, family history and marital status were not

significant. The correlation structure; unstructured covariance, compound symmetric covariance, and autoregressive structure of order one, AR (1) were used and compared using AIC in order to model dependence among observations. Then, the covariance structure having small AIC value was accepted to be the best. Thus, unstructured covariance which had small AIC value of 15934.6 was the most appropriate covariance structure of systolic blood pressure for the final model.

Turning to the separate linear mixed analysis of the diastolic blood pressure, the same procedures has been used as a linear mixed analysis of systolic blood pressure. In linear mixed model with only intercept, intercept + time and intercept + time + time^2 were compared using the AIC value, we got the same results as a linear mixed analysis of systolic blood pressure, and the variables to be included in the model were determined using backward variable selection method using R. Then, the five covariates and all their interaction term by time were considered, from these, family histories with its interaction term by time and place of residence by time were extracted to be included in the final model. There was no clear published literature relating to separate analysis of diastolic blood pressure in hypertensive patients to support this evidence. But, Mancia (1999) in Tiruvallur district, South India, were more or less support this evidence for blood pressure in general not for a single diastolic blood pressure. The correlation structure; unstructured covariance, compound symmetric covariance, and autoregressive structure of order one, AR (1) were also used and compared using. Then, the covariance structure having small AIC value was accepted to be the best. Thus, unstructured covariance which had AIC value of 14929.18 was the most appropriate covariance structure of diastolic blood pressure for the final model.

After separate analysis of each response variables, a joint model using random-effects was used in a bivariate setting with longitudinally measured continuous outcomes. The two outcomes were tied together by a common distribution for the random intercepts and slopes, implying independence conditional on the random effects. The aim of the joint model is able to not only answer the same questions addressed as the separate model; it is also able to answer additional important questions about the association in the evolutions of the responses as well as the evolution of the associations (Fieuws and Verbeke, 2004). Results of the joint model in this study suggested a strong association between the evolutions of SBP and DBP. This result is

supported by Edwards and Fisher (2008); they showed strong association between repeated systolic and diastolic BP outcomes. The joint model also suggested a slowly increasing evolution of the association over time. In fact, there was no evidence in the literature of estimates for the AOE of DBP and SBP for hypertensive patients treated with anti-hypertensive drugs. But findings of John (2007) used the same model, who found a strong association between the evolutions and a slowly increasing evolution of the association between DBP and SBP over time for children ages two through eighteen. Furthermore, the additional information gained by incorporating information about the correlations between the responses was able to reduce the variability (standard errors) in both the fixed-effects estimates as well as the random-effects estimates. Such result is consistent with the previously published data on hypertensive patients blood pressure measurements (Hai and Wanzhu, 2012) using semi-parametric mixed model.

Finally the two models were compared, the results from the separate and joint analyses were quite similar to each other for both systolic and diastolic blood pressure; all parameters were statistically significant, except sex by time insignificant in joint model for systolic and diastolic blood pressure. Based on AIC and log-likelihood value a joint model fitted a data better than comparing to separate linear mixed models. The covariates considered in this thesis were sex, baseline age, place of residence, time and family history. Using the joint analysis, sex, baseline age and place of residence were identified as positive risk factors for the change in SBP, but time is negatively associated with SBP. Sex is the only variable which is identified as a positive risk factor for the change in DBP, but time and baseline age are negatively associated with the change in DBP. The presence of family history (previously diagnosed HTP) did not have any association on the change of both SBP and DBP (Table 7). Some of the findings from joint linear mixed model by Chenglin et al., (2012) support this results, who identified age and previously diagnosed HTP were positively associated with change in SBP, but sex was insignificant for SBP. And they found age was negatively associated with change in DBP, but Sex and previously diagnosed HTP were positively associated with change in DBP.

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

In this study two methods were considered for fitting two response variables measured longitudinally. The result shows both the separate and joint analyses are consistent. But, the joint model is the simplest (less complex) model compared to the separate model because its standard error of the parameter estimates is smaller. And also, the joint model has a very smaller AIC value which indicates that it fits the data better than the separate model. Hence, the joint model is not only the simplest model but also it results a better fit to the data.

Based on separate analysis; the evolution of SBP and DBP measures were significantly differ with respect to time, sex, baseline age and time interaction with sex and age of hypertensive patients. Moreover, on average SBP and DBP measure decreases in a linear pattern over time after patients initiated anti- hypertensive drugs.

Based on joint analysis; sex, baseline age and place of residence were identified as positive risk factors for the change in SBP, but time was negatively associated with SBP. Sex was the only variable which is identified as a positive risk factor for the change in DBP, but time and baseline age were negatively associated with the change in DBP. The presence of family history did not have any association on the change of both SBP and DBP.

The joint model also suggested a strong association between the evolutions and a slowly increasing evolution of the association between systolic and diastolic blood pressure.

5.2. Recommendation

Now days' different health sectors 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, also it is important to know factors that contribute to the progression of the blood pressure. In this study, the progression of blood pressure was found to be different in all patients due to age, sex, place of residence and time. Further studies are required to improve the progression of blood pressure together with the necessary variables. In addition, governmental and non governmental body gives awareness for health workers to record all the

necessary variables during follow up time to see the change of the disease within and between subjects overtime using longitudinal data analysis.

Even though, separate model is most common practice for researchers to model several outcomes involved in a disease process, the joint model is also able to address the same questions as separate model with more accuracy (smaller standard errors) while addressing additional questions that may be of great interest to the researcher, such as the AOE and the EOA of the responses. Thus, fitting joint model is recommended for researches to any types of multivariate response variable. In this study, it is focus on only two response variables, for future work, one might want to look at modeling more than two response variables over time.

Limitation of the Study

This thesis is not done without limitation. In Ethiopian context repeated (longitudinally) measured data were not extracted well and it is very limited to specific area to obtain, in spite of this, 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; alcohol use, smoking status, body mass index, level of education, marital status, exercise and other. In this thesis, it is limited only to the four covariates. This is because all the necessary variables were not recorded on the patient's card, except those four covariates. Despite the above limitations, this study used two longitudinal response variables using joint model, but there were no related published literatures on the country to compare and contrast the findings of this study obtained through both the joint and separate modeling frameworks in the local context.

Reference

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.

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.

Bowman, F. and Manatunga, A.(2005). A joint model for longitudinal data Profiles and associated event risks with application to a depression study. *Applied statistics*, Vol. **54(2):** 301316.

Box, G., Jenkins, G. and Reinsel, G. (1994). Time Series Analysis: Forecasting and Control. Holden-Day

Brown, H. and Prescot, R. (1999). Applied Mixed Models in Medicine. John Wiley & Sons Ltd

Chakraborty, H., Helms, R. and Sen, P. (2003). Estimating correlation by using a general linear mixed model: evaluation of the relationship between the concentration of HIV-1 RNA in blood and semen. **22**:1457–1464.

Chenglin ,Y., Gary, F., Lisa, D. and Lehana, T.(2012). Analyzing diastolic and systolic blood pressure individually or jointly? In McMaster University

Cohen,J., Grimm,R. and Smith, W.(1981). Multiple risk factor intervention trial (MRFIT).VI.Intervention on blood pressure.*PreventiveMedicine*;**10**:501-518.

Das, S., Sanyal, K., and Basu, A. (2005). Study of urban community survey in India: growing trend of high prevalence of hypertension in a developing country. *Int J Med Sci*;**2**:70-78.

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.

Der, G. and Everitt, B.(2006). Statistical Analysis of Medical Data using SAS : Taylor & Francis Group, LLC.

Diggle, P., Zeger, S., and Liang, K.(1994). Analysis of Longitudinal Data : University Press Inc., New York.

Edwards, L., Fisher, M., Wolfinger, R., Qaqish, B. and Schabenberger, O. (2008). An R statistic for fixed effects in the linear mixed model, *Statistics in Medicine***; 27**, 6137-6157

Fieuws, S. and Verbeke, G. (2004). Joint Modeling of Multivariate Longitudinal Profiles: Pitfalls of the Random-Effects Approach. *Statistics in Medicine*, **23**; 3093-3104,

Fieuws, S. and Verbeke, G. (2005). Evaluation of the pairwise approach for fitting joint linear mixed models: A simulation study. Technical Report TR0527, *Biostatistical Centre*, Katholieke Universiteit Leuven, Belgium.

Fieuws, S. and Verbeke, G. (2006). Pair wise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics.* **62**, 424-431.

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

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.

Hai, L. and Wanzhu, T. (2012). A semi parametric regression model for paired longitudinal outcomes with application in childhood blood pressure development in Indiana University School of Medicine

Hajjar, I., Kotchen , J. and Kotchen, T. (2006). HYPERTENSION: Trends in Prevalence, Incidence, and Control. *Annu. Rev. Publ. Health.* **27**: 465-490

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

Howard, G. (2006). Racial and geographic differences in awareness, treatment, and control of hypertension. *Stroke*, **37**:1171-1178,.

John, M. (2007).Joint mixed-effects models for longitudinal data analysis: a application for the metabolic syndrome, in Virginia Common Wealth University

Kalkidan, H., and Misra, A. (2009). Khat as risk factor for hypertension: A systematic review, Ethiopia, Jimma university

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

Klimentidis,Y., Dulin-Keita,A., Casazza, K., Willig, A., Allison, D. and Fernandez,J.(2011). "Genetic admixture, social-behavioural factors and body composition are associated with blood pressure differently by racial-ethnic group among children," *Journal of Human Hypertension*, vol. **26**, no. 2, pp. 98–107.

Kondo,T., Watanabe,Y. and Ueyama,J. (2006). Application of a general linear mixed model to assess the association between aging and longitudinal blood pressure variations among healthy Japanese workers

Laird, N., and Ware, J. (1982). Random-effects models for longitudinal data. Biometrics

MacCallum, R., Kim, C., Malarkey, W. and Kiecolt-Glaser, J.(1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research* **99**;215–253.

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

McCulloch, C., Searle, S. and Neuhaus, J. (2008). Generalized, Linear, and Mixed Models: John Wiley & Sons, Inc.

Molenberghs, G. and Verbeke, G. (2005). Models for Discrete Longitudinal Data. Springer Science+Business Media, Inc.

Molenberghs, G. and Verbeke, G. (2008). A review on linear mixed models for longitudinal data, possibly subject to drop out. *Statistical Modelling.* **1**(4), 235-269.

Nonhlanhla, Y. (2009). Modelling CD4+ Count Over Time in HIV Positive Patients Initiated on HAART in South Africa Using Linear Mixed Models. MSC

O'Brien, E., Beevers, G. and Gregory,Y. (2001).ABC of hypertension: Blood pressure measurement. Part III Automated sphygmomanometry: 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.)

Olkin, I. and Tate, R. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann of Math Stat.* **32**, 448-465

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.

Pinheiro, J., and Bates, D. (2000). Mixed-Eects Models in S and S-PLUS. Springer Verlag New York.

Safar, M., Rudnichi, A., Asmar, R. (2000).Drug treatment of hypertension: the reduction of pulse pressure does not necessarily parallel that of systolic and diastolic blood pressure; **18**: 1159–1163

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). Akaike Information Criterion Statistics. D. Reidel Publishing Company

Salehmohamed, J.and Suheil, H. (2007). Risk factors for hypertension among urban males in Mombasa Kenya

Shah, A., Laird, N. and Schoenfeld, D.(1997). A random-eects model for multiple characteristics with possibly missing data.**92**:775–779.

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

Thiebaut , R., Commenges, D., Chˆene, G.(2002). Bivariate longitudinal study of CD4+ cell count and HIV RNA taking into account informative drop-out and left censoring of HIV RNA values. In Proceedings of the 17th International Workshop on Statistical Modeling, *Stasinopoulos M, Touloumi G* (eds).**67**: 521–524.

Tomečková, M. and Stanovská, Z. (2002), control of hypertension and survival analysis of the hypertensive patients in stulong - longitudinal study of risk factors of atherosclerosis

Verbeke, V. and Lesaffre, L.(1997). Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population. *Journal of the American Statistical Association*, Vol. **91**, 321–464

Verbeke, G., Lesare, E. and Brant, L. (1998). The detection of residual serial correlation in linear mixed models. *Statistics in Medicine* **17**, 1391-1402.

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

Williams, S.(2001). Modeling the association between increases in height and adiposity and changes inq blood pressure between ages 7 and 18 years. InProceedings of the 14th International Workshop on Statistical Modeling, Klein B, Korsholm L (eds). **21**:521–524

Zucker, D., Zerbe, G. and Wu, M.(1995). Inference for the association between coefficients in a multivariate growth curve model.**51**:413–424.

ANNEX

I: Fixed effects component with all covariates with the corresponding estimates and time interactions for SBP and DBP

Table 12: Fixed effects component with all covariates with the corresponding estimates and time interactions for SBP.

Effect	Estimate	S.e	DF	Pr > |t|	95% CL	
Intercept	128.93	2.146	1940	<.0001	(124.72. 133.14)	*
SexMale	4.137	1.075	1940	0.0001	(2.027, 6.247)	*
PrRural	3.093	1.051	1940	0.003	(1.032, 5.155)	*
FhYes	-0.132	1.054	1940	0.900	(-2.198, 1.936)	
Age	0.134	0.038	1940	0.003	(0.064, 0.216)	*
Time	-1.575	0.296	1940	0.0001	(-2.156, -0.993)	*
Time×SexMale	-0.443	0.190	1940	0.019	(-0.816, -0.071)	*
Time×PrRural	-0.020	0.187	1940	0.912	(-0.388, 0.347)	
Time×FhYes	-0.133	0.185	1940	0.471	(-0.496, 0.229)	
Age×Time	-0.008	0.007	1940	0.225	(0.023,-0.005)	

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

** indicates statistical significance at 0.05 level of significance . CI=Confidence Interval, S.e=standard error.AIC value= 16253.08*

Table 13: Fixed effects component with all covariates with the corresponding estimates and time interactions for DBP

Effect	Estimate	S.e	DF	Pr > |t|	95% CL	
Intercept	89.497	1.618	1940	0.0001	(86.323, 92.670)	*
SexMale	2.487	0.810	1940	0.002	(0.897, 4.077)	*
PrRural	0.810	0.792	1940	0.306	(-0.743, 2.364)	
FhYes	-0.319	0.795	1940	0.688	(-1.8776, 1.239)	
Age	-0.075	0.029	1940	0.009	(-0.133, -0.018)	*
Time	-0.974	0.393	1940	0.013	(-1.745,-0.203)	*
Time×SexMale	0.198	0.143	1940	0.166	(-0.082, 0.479)	
Time×PrRural	-0.036	0.141	1940	0.798	(-0.3132, 0.241)	
Time×FhYes	-0.022	0.139	1940	0.874	(-0.295, 0.251)	
Age×Time	-0.019	0.005	1940	0.0002	(-0.030,-0.009)	*

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

** indicates statistical significance at 0.05 level of significance . CI=Confidence Interval, S.e=standard error AIC value= 15150.77.*

II: Separate Linear Mixed effects model with all covariates with the corresponding estimates and possible time interaction for SBP and DBP

Table 14: Linear Mixed effects model with all covariates with the corresponding estimates and time interaction for SBP

Effect	Estimate	S.e	Pr > |t|	95% CL	
Intercept	128.44	3.062	0.0001	(122.42, 134.47)	*
SexMale	5.084	1.547	0.0011	(2.041, 8.127)	*
PrRural.	3.414	1.497	0.023	(0.469, 6.359)	*
FhYes.	-0.537	1.503	0.721	(-3.492, 2.418)	
Age	0.12	0.055	0.011	(0.033, 0.250)	*
Time	-2.062	0.369	0.0001	(-2.791, -1.333)	*
Time×SexMale	-0.795	0.252	0.002	(-1.293,-0.297)	*
Time×PrRural	-0.101	0.246	0.684	(-0.586, 0.386)	
Time×FhYes	0.119	0.246	0.626	(-0.364, 0.604)	
Age×Time	-0.012	0.009	0.210	(-0.029, 0.006)	

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

** indicates statistical significance at 0.05 level of significance . CI=Confidence Interval, S.e=standard error AIC value= 15940.16.*

Table 15: Linear Mixed effects model with all covariates with the corresponding estimates and time interaction for DBP

Effect	Estimate	S.e	Pr > |t|	95% CL	
Intercept	89.732	2.141	0.0001	(85.521, 93.942)	*
SexMale	2.778	1.075	0.0102	(0.663, 4.893)	*
PrRural	0.935	1.045	0.372	(-1.120, 2.989)	
FhYes	-0.251	1.048	0.811	(-2.311, 1.809)	
Age	-0.077	0.038	0.046	(-0.152, -0.001)	*
Time	-1.145	0.511	0.026	(-2.153, -0.137)	*
Time×SexMale	0.045	0.164	0.785	(-0.279, 0.369)	
Time×PrRural	-0.132	0.162	0.415	(-0.452, 0.188)	
Time×FhYes	0.065	0.159	0.685	(-0.251, 0.381)	
Age×Time	-0.025	0.006	0.0001	(-0.037, 0.013)	*

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

** indicates statistical significance at 0.05 level of significance . CI=Confidence Interval, S.e=standard error AIC value= 14937.28.*

III: Joint Linear Mixed effects model with all covariates with the corresponding estimates and time interaction for SBP and DBP

Table 16: Parameter estimates and standard errors for the Joint linear mixed effects model with all covariates and time interaction for both SBP and DBP

	SBP			DBP		
Effect	Estimate (S.e)	P. value	95% CL	Estimate (S.e)	P. value	95% CL
Intercept	128.44 (3.062)	0.0001	(122.42, 134.47)*	89.374(2.127)	0.0001	(85.190, 93.557)*
SexMale	5.083(1.547)	0.0011	(2.041, 8.127)*	2.808 (1.073)	0.0093	(0.698, 4.918) *
PrRural	3.414(1.497)	0.0232	(0.469, 6.359)*	0.822(1.041)	0.4299	(-1.224, 2.869)
FhYes	-0.537(1.503)	0.7209	(-3.493, 2.418)	-0.190(1.045)	0.8556	(-2.244, 1.864)
Age	0.142(0.055)	0.0107	(0.033, 0.250)*	-0.078 (0.038)	0.0435	(-0.153, -0.002)*
Time	-1.782(0.334)	0.0001	(-2.442, -1.122) *	-1.145(0.512)	0.0263	(-2.153, -0.136)*
Time×SexMale	-0.795(0.252)	0.0019	(-1.293, -0.297)*	0.029(0.163)	0.8553	(-0.293, 0.352)
Time×FhYes	0.119(0.246)	0.6263	(-0.365, 0.604)	0.045(0.159)	0.7760	(-0.269, 0.359)
Time×PrRural	-0.1004(0.246)	0.6842	(-0.586, 0.385)	-0.094(0.160)	0.5595	(-0.410, 0.223)
Time×Age	-0.011(0.009)	0.0102	(-0.029,-0.006) *	-0.025(0.006)	0.0001	(-0.037,-0.013)*

Source: Jimma University Specialized Hospital, Ethiopia; from September, 2011 to July, 2013

** indicates statistical significance at 0.05 level of significance . CI=Confidence Interval, S.e=standard error, AIC value= 30865.4*

IV: Model diagnostics for Separate Linear Mixed Model

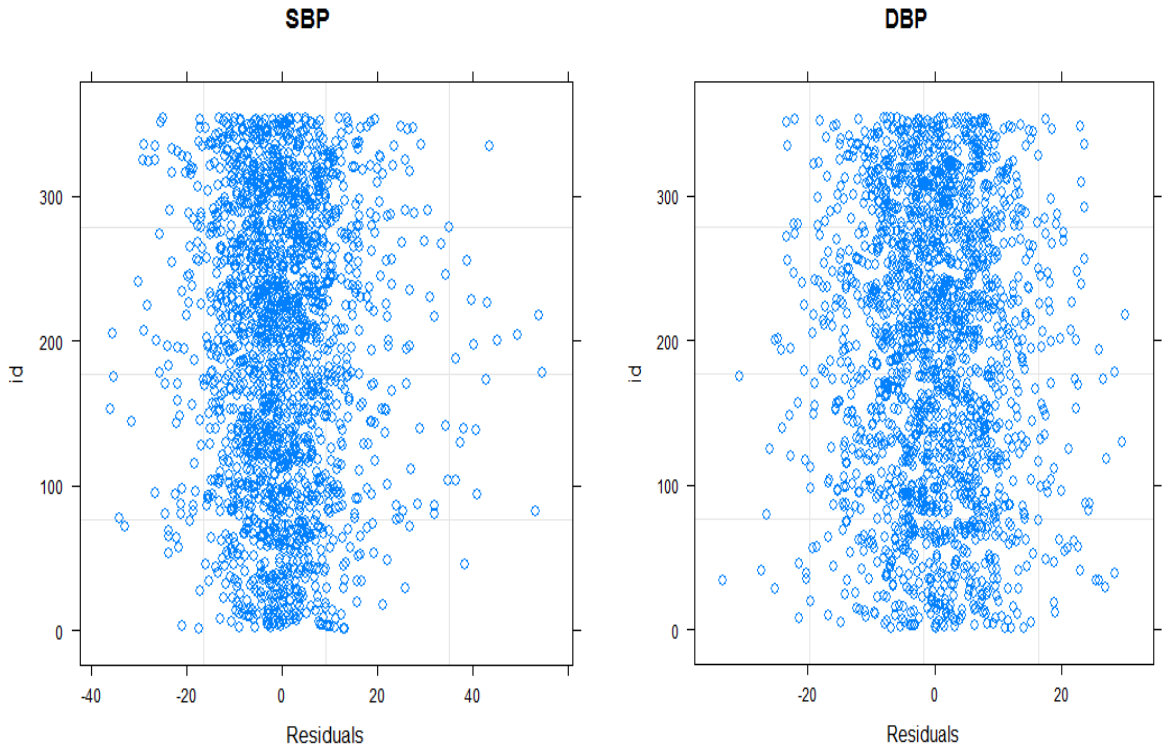

Figure 7: Residuals vs observed id numbers

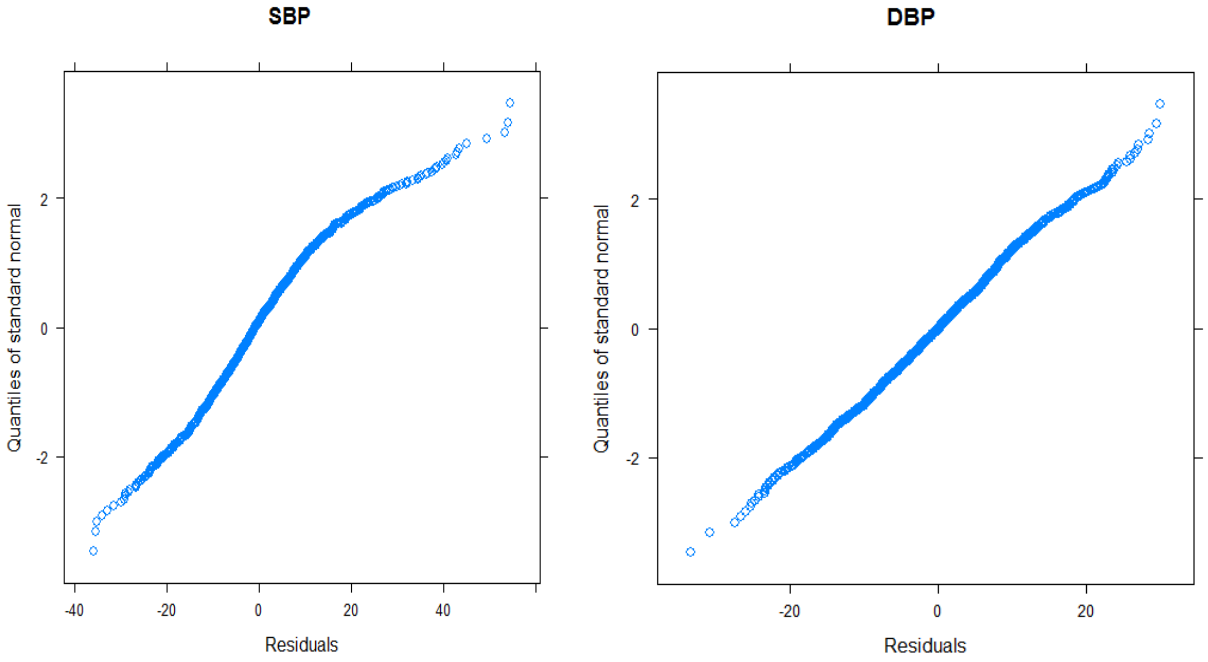

Figure 8: Normal Q-Q points for random effects

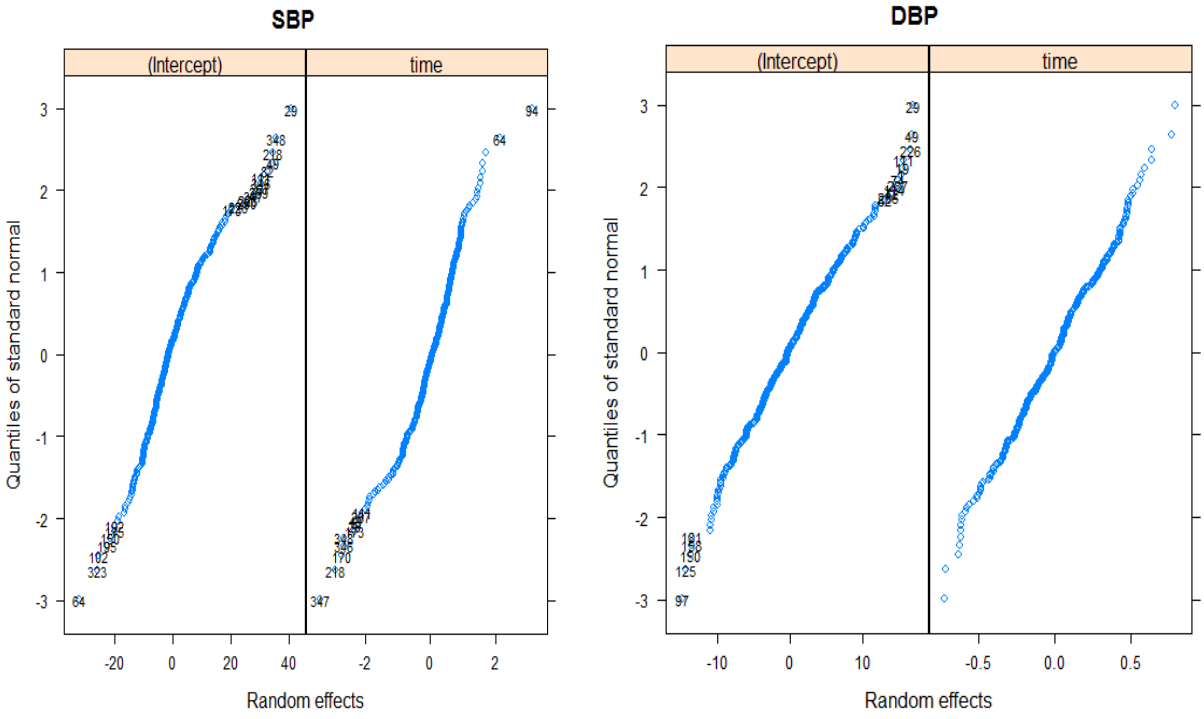

Figure 9: Q-Q plots for random intercept and slopes

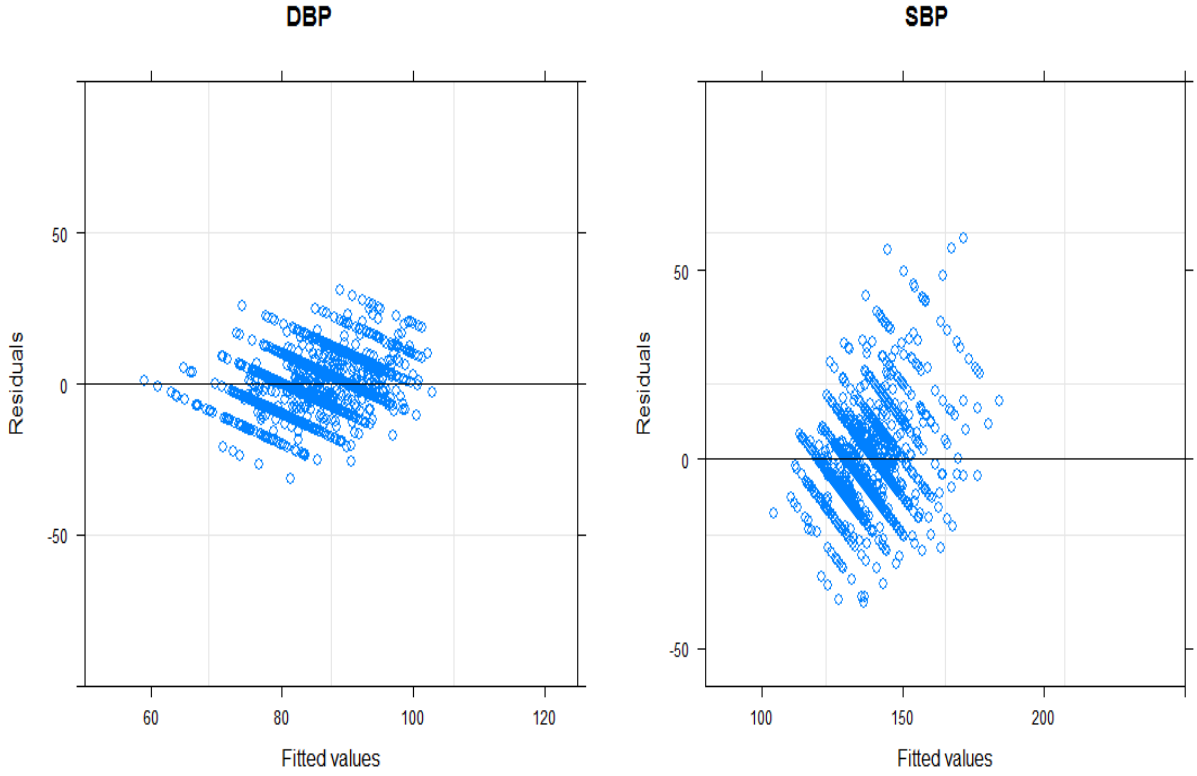

Figure 10: Residuals vs fitted value