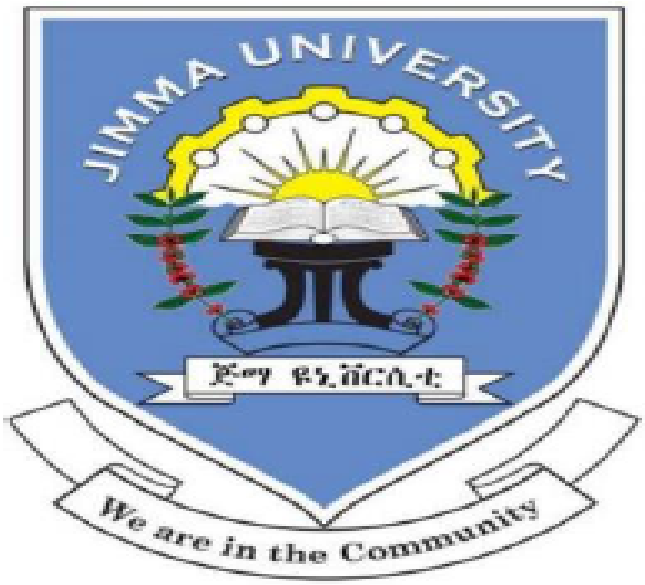

Joint Modeling Of Longitudinal International Normalized Ratio And Activated Partial Thromboplastin Time Test Of Deep Venous Thrombosis Patients Receiving Treatment In Jimma University Medical Center:Ethiopia

By:
Gezahagn Diriba

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

October 20, 2021
Jimma, Ethiopia

Joint modeling of Longitudinal International Normalized Ratio and Activated Partial Thromboplastin Time Test Of Deep venous Thrombosis Patients Receiving Treatment In Jimma University Medical Center:Ethiopia

MSc. Thesis

By:

Gezahagn Diriba

Main Advisor:Abdisa Gurmessa (PhD Scholar)

Co-Advisor: Shiferaw Befekadu (PhD Scholar)

October 20, 2021

Jimma, Ethiopia

Approval Sheet

As thesis research advisors, we herby certify that we have read the thesis prepared by **Gezahagn Diriba Tesemsa** our guidance, which is entitled "Joint Modeling of Longitudinal International Normalized Ratio and Activated Partial Thromboplastin Time Test Of Deep venous Thrombosis Patients Receiving Treatment In Jimma University Medical Center ", 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.

Abdisa Gurmessa (PhD Scholar)	.....	.....
Main Advisor	Signature	Date
Shiferaw Befkadu (PhD Scholar)	.....	.....
Co-advisor	Signature	Date

As the members of the board of examiners of MSc. thesis open defense examination of **Gezahagn Diriba Tesema**, we certify that we have read and evaluated the thesis and examined the candidate. We recommend that the thesis has been accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

.....	.....	.....
Name of External Examiner	Signature	Date
.....	.....	.....
Name of Internal Examiner	Signature	Date
.....	.....	.....
Name of Chairman	Signature	Date

DECLARATION

I declare that this thesis is my original work and that all source materials used for this thesis have been properly cited and acknowledged. This thesis has been submitted in partial fulfillment of the requirements for MSc. degree of Master of Science in Biostatistics at Jimma University. I earnestly declare that this thesis is not submitted to any other institution anywhere for the award of any academic degree, diploma, or certificate.

Gezahagn Diriba Tesema	.....	.....
Name	Signature	Date

This thesis has been submitted for examination with my approval as a University advisor and Co- advisor.

Abdisa Gurmessa (PhD Scholar)	.....	.....
Main Advisor	Signature	Date
Shiferaw Befekadu (PhD Scholar)	.....	.....
Co-advisor	Signature	Date

Acknowledgement

Above all, I would like to thank the almighty God and his mother St. Mary for the gift of health,wisdom and strength throughout any steps for the achievement of my thesis.

My deep gratitude goes to my advisor Mr.Abdisa Gurmessa (PhD Scholar) for his professional and deep comments, advice and guidance at any stage of this thesis. I appreciated his friendly contact and his punctuality for commented this paper. Simply, without his frequent supervision, this thesis may not reach at the stage it is now.

I sincerely appreciate to my co-advisor Mr.Shiferawu Befekadu (PhD Scholar) for his valuable suggestions and comments for the successful realization of this thesis. It is a great pleasure to express his commitment to share his experience, knowledge and his smooth contacts is also appreciable.

I wish to show my appreciation to Jimma University Medical Center for their permission to get the data for the study. My special appreciation and acknowledgment go to Ambo University for offering me to the opportunity and financial support during my study at Jimma University.

I would like to thank Mr.Bekane Bogale (MSc) and Mr.Yasin Negash (MSc) for their invaluable contributions and suggestions. And also, I thanks to my classmates, especially Mr.Fedasa Tesfaye who helped me in every aspect of this thesis preparation.

Finally, I am extremely grateful to my family for their love, supports and encouragement. And also, I thanks you everyone that I could not mention here, who have encouraged and helped me in different ways.

Abstract

Background: Deep Venous Thrombosis is a condition in which one or multiple blood clots form in the deep veins of the legs. People at high risk of deep venous thrombosis could take treatment of Anticoagulation to reduce their chance of developing serious conditions of stroke and heart attacks. The aim of this study is to address joint evolution and association of international normalized ratio and activated partial thromboplastin time over time and their associated risk factors.

Methods: The study was conducted at Jimma University Medical Center. The latest data were collected from Jimma University Medical Center from first, September 2018 to first, January 2021. A linear mixed-effects model was fitted for the international normalized ratio and activated partial thromboplastin time outcomes. Then, a joint mixed-effects model was fitted for the two endpoints.

Results: At baseline, the mean and standard deviation of both international normalized ratio and activated partial thromboplastin time were 1.378 and 0.6535 and 33.701 and 10.5738 respectively. In a separate analysis, the covariate smoking, alcohol user, prolonged immobilization, family history of DVT, and age are significant factors that affect measure of deep venous thrombosis. In joint analysis case, smoking, alcohol user, family history of DVT, and age are important significant factors of deep venous thrombosis patients.

Conclusion: The joint model fits the data well as compared to the separate model due to its smaller standard error. The results of the joint model suggested a very strong association between the evolutions of international normalized ratio and activated partial thromboplastin time. The evolution of association slowly increasing over time.

Recommendation: Based on the results of this thesis joint model is recommended for researchers to any types of multivariate response variables for evaluating correlation and rate of changes over time.

Key Words: Deep Venous Thrombosis, International Normalized Ratio, Activated Partial Thromboplastin Time, Linear Mixed Model, Joint Linear Mixed Model.

Contents

List of Tables

List of Figures

Acronyms

AIC	Akakike Information Criterion
AOE	Association Of Evolution
APTT	Activated Partial Thromboplastin Time
BIC	Bayesian Information Criterion
COCs	Combine Oral Contraception
COPD	Chronic Obstructive Pulmonary Disease
DVT	Deep Venous Thrombosis
EB	Empirical Prediction
EBLUPS	Empirical Best Linesr Unbiased Predictor
GLM	Generalized Linear Model
HIV	Human Immunodeficiency Virus
INR	International Normalized Ratio
IPC	Intermittent Pneumatic Compression
JUMC	Jimma University Medical Center
LMM	Linear Mixed model
ML	Maximum Likelihood
NCHS	National Center For Health Statistics
PE	Pulamonary Embolism
PH	Proportional Hazard Ratio
PT	Prothrombin Time
PTT	Partial Thromboplastin Time
REML	Restricted Maximum Likelihood
SSA	Sub Saharan African
UFH	Unfractionated Heparin
VTE	Venous Thromboembolism
WHO	World Health Organization

Definition of Symbols

Std.Dev	Standard Deviation
Estims	Estimates
Autorsive	Autoregressive
CompSymmetry	Compound Symmetric
s.e	standard error
CI	Confidence Interval
ρ	Correlation of the random effects
var(a_{10})	variance of random intercept for INR
var(b_{10})	variance of slope for INR
var(a_{20})	variance of random intercept for APTT
var(b_{21})	variance of slope for response APTT
σ_1^2	variance of international normalized ratio
σ_2^2	variance of activated partial thromboplastin time
corr(a_{10}, b_{11})	corrilation between random intercept and slope for INR
corr(a_{20}, b_{21})	corrilation between random intercept and slopes of APTT

1 Introduction

1.1 Background of the Study

Deep Venous Thrombosis is a condition in which one or multiple blood clots (thrombi) form in the deep veins of the legs. The deep veins are located within the muscles of the legs and are generally not visible on the surface as are superficial varicose veins. Because veins do not have a thick muscular wall, they are not able to help pump blood to different parts of the body as the arteries do. Instead, blood moves through the veins by either gravity or by the contraction of surrounding muscles, which squeezes blood in the veins back to the heart using a system of one-way valves in the veins. Also, it is the formation of a blood clot in a deep vein, that is most common in the legs or pelvis [1]. The known symptoms are pain, swelling, redness, and enlarged veins in the affected area, but some DVTs have no symptoms [2]. This illness is often "silent" and can mimic other common conditions such as heart attack, pneumonia, and anxiety [3]. People at high risk of deep venous thrombosis (blood clot) could take treatment of Anticoagulation to reduce their chance of developing serious conditions of stroke and heart attacks. Anticoagulation measured by international normalized ratio and Activated partial thromboplastin time because over-anticoagulation may result in bleeding and coagulation result in thrombotic complications, including stroke [4].

Globally, Deep Venous Thrombosis is a major health problem with high morbidity and mortality worldwide [5]. DVT represents a significant healthcare burden worldwide. The prevalence of DVT is reported to be approximately 100 per 100,000 people per year [6]. It is the third most common cardiovascular disease after acute coronary syndromes and strokes. DVT is predominantly a disease of the elderly with an incidence that rises with age [7] and It is a growing public health problem due to increasing in the ageing population [8] .

As the average life expectancy continues to increase, DVT will become a major health problem worldwide. Furthermore, elderly patients develop common internal illnesses, such as congestive heart failure, chronic obstructive pulmonary disease (COPD), acute infection, and atherosclerotic vascular disease [9]. It affects approximately about 0.1% of persons per year worldwide. Deep venous thrombosis(DVT), with an average annual incidence of 48 per 100,000 persons, which accounts for more than half in the United States and 52 per 100,000 persons in Australia [10]. In India, an overall incidence of confirmed DVT has been shown to be 17.46 per 100,000 patients with 64% occurring in non-surgical non-trauma patient [11]. An estimated 45 000 patients each year in Canada have DVT. Heparin is given until warfarin attains a therapeutic effect, as defined by a target international normalized ratio of 2.0 to 3.0 for 2 consecutive days [12].

The natural course of DVT therapy is associated with significant morbidity and mortality rates [13]. Standard therapy for acute DVT consists of heparin followed by oral warfarin. To monitor therapy with coumadin anticoagulants, prothrombin time (PT) was the standard until as recently as the past decade. Because the therapeutic range of PT depends on many factors, comparing patient results and assessing the significance of those results can be difficult. Therefore, the World Health Organization (WHO) developed an international standardization system known as the international normalized ratio (INR) [14] to detect blood clots. The international normalized ratio or INR calculation is based on the results of a PT that is used to monitor individuals who are being treated with the blood-thinning medication (anticoagulant) warfarin (Coumadin) [15].

Heparin is recommended as the initial treatment for acute deep vein thrombosis (DVT) [15] because such therapy improves survival after DVT, [16] and reduces asymptomatic DVT extension and possibly 3-month VTE recurrence. Activated partial thromboplastin time (APTT) monitoring and heparin dose adjustment to rapidly achieve and maintain an APTT therapeutic range corresponding to a plasma heparin level of 0.3-0.7 anti-Xa U/mL is also recommended and the heparin dose was monitored with a target range of activated partial thromboplastin time of 30-40 seconds [15, 17].

According to the national center for health statistics[NCHS,2016] reports that DVT is the most prevalent medical health problem with an annual incidence of 80 cases per 100,000. Each year in the united states, more than 200,000 people develop venous thrombosis; of those, 50,000 cases are complicated by PE. Lower-extremity DVT is the most common venous thrombosis, with a prevalence of 1 case per 1000 population. In addition, it is the underlying source of 90% of acute PEs, which cause 25,000 deaths per year in the United States.

In African, a study conducted in Nelson Mandela Hospital, Mthatha, shows that the prevalence of deep venous thrombosis was 12.5% among HIV-positive patients admitted to the medical wards [18]. Overall 200,000 patients of South African suffer from deep venous thrombosis each year due to most DVT is occult, the true incidence is unknown [19]. The variation of deep venous thrombosis surgery in Africa varied between 2.4% and 9.6%. In sub-Saharan Africa, as the study conducted in Nigeria [20] and Uganda [21] showed that the prevalence of DVT was, respectively, 2.4% and 5%. North African report showed that the prevalence of DVT in pregnant and postpartum women varied between 448 per 100 000 births per year [5] and 380 per 100 000 births per year in [22]. Patients in sub-Saharan Africa generally have poor anticoagulation control. As different conducted studies showed that there was a problem with access to medicines, international normalized ratio, and activated partial thromboplastin time monitoring, the lack of locally validated standardized dosing protocols, and low levels of anticoagulation knowledge among healthcare workers and patients. Anticoagulation patients in Sub-saharan African(SSA) are younger than those in high-income settings. For example, in Uganda and South Africa, the median age of patients attending five anticoagulation services was 56 years [23] and in Kenyan service, the mean age was 43 years [24].

In Ethiopia, different conducted studies showed that DVT was a common disease associated with sever complications and mortality. A recent report from Addis Ababa also showed that DVT was associated with malignancy, prolonged immobilization, pregnancy-related problems, and major trauma [25]. A prospective cohort study from

Ethiopia conducted in Jimma University [26] showed that the incidence density of deep venous thrombosis was approximately 2.99 per 1000 per day. Deep venous thrombosis management in Ethiopia depends on the national Standard Treatment Guidelines for General Hospital [27] and international clinical practice guidelines such as American college chest physicians DVT treatment guidelines [28].

The treatment of DVT aims to reduce morbidity and mortality. This is achieved by optimal therapy with anticoagulants to prevent thrombus extension and embolization. The major outcomes of venous thrombosis are death, recurrence, post-thrombotic syndrome, and major bleeding due to anticoagulation therapy [29].

longitudinal study is the study that individuals are measured repeatedly through time and 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 [30]. The measurements that measured 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 [30]. 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 [31].

This study would be used to build a mixed model methodology allows the longitudinal examination of International normalized ratio and activated partial thromboplastin time over time. The 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 [31].

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 [32]. These unobserved effects are then included in the model as random variables, or equivalently called, random effects. 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, International Normalized Ratio and Activated Partial Thromboplastin Time 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 INR and APTT can notice the physician to changes in the health of an individual [30].

Joint Longitudinal Study is the study at which the repeated observation of more than two outcomes is common in biomedical and public health research are measured simultaneous over time. Such experiments result in multivariate longitudinal data, which are unique in the sense that they allow the researcher to study the joint evolution of these outcomes over time. In many area of study, more than one response variable is followed longitudinally, and analyzing all jointly may be beneficial [30] . Measures of INR and APTT are highly related and changes in either often affect changes in the other. The separate analyses would not able to examine the correlation or association between the two outcomes. Therefore, it is more desirable to jointly modeling of the two outcome variables together [33]. There are two types of correlations in joint longitudinal models. Those are Serial correlation and cross correlation.

Serial correlation is a correlation which found between the observations at different time points within a subject and others. Also 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 [34]. One of the most important of modeling joint

longitudinal observations with differing outcome is used to answers the question of how the evolution of one response is related to the evolution of another response [35]. A flexible solution is to model the association between the different responses using random effects.

The organization of this thesis was built from five chapters. Chapter one is background of the study which describes the deep venous thrombosis, linear mixed model , joint linear mixed model, statements of the problem , objectives of the study and significance of the study, which is used as our tool to validate the thesis statement. Chapter two is a literature review about previous related studies on deep venous thrombosis, linear mixed model and joint linear mixed model. Chapter three provides a brief description on the study area, design, source of data, study population and statical methods of data analysis, which are used as our main material to analysis of the data. Chapter four contains result of the data analysis using descriptive statistics, linear mixed model and joint linear mixed model. Chapter five contains discussion, conclusion and recommendation based on the results.

1.2 Statement of the Problem

DVT is the third most common cardiovascular disease after acute coronary syndromes and stroke, affecting 2 million individuals in the United States each year. DVT is mentioned as a major health problem and one of the most common preventable causes of hospital deaths in the western world where the incidence is one case of DVT per 1000 population per year [36].

In Ethiopia, there were no sufficient published literature that documented on joint modeling of longitudinal INR and APTT test of deep venous thrombosis except the studies about determinates of deep venous thrombosis, treatment outcome of deep venous thrombosis based on cross-sectional retrospective and prospective cohort study by using multiple linear regression, Cox proportional hazards and logistic regression model to identify the factors that are associated with deep venous thrombosis over time without considering the responses INR and APTT jointly, correlations within the two outcomes, and subject-specific random effects. So, the purpose of this thesis is to

fill this gap.

In longitudinal data, with two outcomes, there is also a correlation between them and the correlation due to repeated measures over time. But their separate modeling of the international normalized ratio and activated partial thromboplastin time 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 [35]. An interest is lies in how the evolution of INR is related to the evolution of APTT, as well as how the association changes over time. Therefore, cross-sectional study and separate modeling would not be able to examine the association or evolution of the two outcomes evolves over time, but joint modeling can do.

Finally, this study was going to answer the following basic questions.

1). Which factors affect the separate as well as joint evolution of INR and APTT measurements on deep venous thrombosis patients?

2). How the evolution of the international normalized ratio is related to the evolution of activated partial thromboplastin time?

3). How the association between international normalized ratio and activated partial thromboplastin time change, or evolves over time?

4). Do separate modeling would be able to examine the association or evolution of the two outcomes evolves over time?

1.3 Objectives of the Study

1.3.1 General Objectives

The aim of this study is to investigate the joint evolution and association of international normalized ratio and activated partial thromboplastin time measurements of deep venous thrombosis patients and identify their associated risk factors.

1.3.2 Specific Objectives

The specific objectives of this study are:

• To estimate a separate and joint mixed effect model for the INR and APTT and identify the associated risk factors .

• To evolute the association between the evaluation of international normalized ratio and activated partial thromboplastin time .

• To explore the evolution of international normalized ratio and activated partial thromboplastin time test of DVT patients how they change over time.

• To identify the best model comparing the results obtained in the separate and joint models.

1.4 Significance of the Study

At the end of this study the result helps:

• To identify the factors that influencing the separate as well as joint evolution of INR and APTT test in DVT. This will in turn help the respective policymakers of the health sector in the effort to design an appropriate intervention strategy.

• It is used to compare the different groups of patients within responding to the drug simultaneously; so that it serves as a base for further study for the question of what brings this variation and others.

• It can be used as a reference for those who want to apply separate and joint modeling techniques in two longitudinal continuous sequences.

• To understand the importance of attending clinic/ hospital in the early stage of DVT consistently and sustain follow up of taking repeated treatment with preferable drugs.

2 LITERATURE REVIEW

Deep vein thrombosis (DVT), which is the clotting of blood components in the veins located deep within the leg, is the third most common cardiovascular disease after a heart attack and stroke [37]. Deep vein thrombosis commonly affects the leg veins (such as the femoral vein or the popliteal vein) or the pelvis' deep veins, which made their common sources of serious complications. It is predominantly a disease of middle age and the elderly [38]. It is very rare before the age of 18 years [39]. Incidence rates are higher in women, especially during childbearing years. Clinical assessment of patients with suspected DVT is difficult and often inaccurate. Therefore, clearly, imaging evaluation of patients with suspected DVT provides important information that can help clinicians set up the correct diagnosis. Heparin and Warfarin are the treatment that is given to a person with a high risk of deep venous thrombosis. Both are a type of anticoagulation which measured with activated partial thromboplastin time(APTT) and international normalized ratio (INR) to check the treatment effects.

The major risk factors affecting deep venous thrombosis

An understanding of the risk factors for venous thrombosis is necessary in order to maximize the prevention of this disease in high risk individuals and groups of patients.

Age: Age was significantly associated with deep venous thrombosis. According to a study done in Sudan University of Science and Technology by Nadir Ahmed et al. [40], who reported that age showed a significant difference between the mean age of DVT patients and healthy controls. In addition, a cross sectional and case control study carried out in Senegal involving 105 cases and 200 controls. The mean age for cases was 42 years, ranging from 17 to 78 year. The mean age of the control population was 38 years, ranging from 18 to 65 years. DTV is especially prevalent in those aged >60 years, patients with active cancer and surgical populations this finding conducted by Fall et al. [41, 42, 43]. Furthermore, the study conducted in Leiden University by Marissa et al. [44], who stated that the incidence of venous thrombosis is more than 10-fold higher in age 80 to 89 years old than in middle-aged individuals

40-50 years. The other study done in Jimma University by Mulatu et al. [26], who used a prospective cohort study and stated that, patients whose age greater than 35 were significantly associated with deep venous thrombosis. Similary study was done in Ethiopia, at Tikur Anbessa Specialized Hospital by Haile et al. [25], conducting a prospective cross-sectional study and show that age was significantly associated with recurrence of deep venous thrombosis

Gender: According to study done in Minnesota by Heit et al. [45], who found that being male were significant risk predictor for increasing deep venous thrombosis. Another, prospective study conducted by Nagler et al. [46] and by Moreno et al. [47] both of them stated that patients with male sex were significantly increased the recurrence of deep venous thrombosis. Similarly, a Cox regression analysis study conducted in Jimma University by Mulatu et al. [26] , the finding showed that being female was statistically associated with DVT recurrence with (CHR,1.799; 95% CI: 0.828-3.905; P .038).

Alcohol user: Drinking alcohol was one of the most important risk factor in development of deep venous thrombosis. A cross-sectional study conducted in United States showed that high maximum alcohol use is the behavior most linked to increasing the risk of DVT recurrence which reported by Golomb et al. [48] This may be related to oxidative stress and mitochondrial toxicity from alcohol, which will cause tissue injury and cell death. This lead to triggers coagulation activation, via exposure at the cell surface of phosphatidylserine by Hansen et al. [49] Similarly, a study conducted in Ethiopia by Mulatu et al. [51], who found that patients who use the unspecified amount of regular alcohol consumption was linked with increases in the development of DVT recurrence.

Smoking cigarette: A meta analysis study conducted by Cheng et al. [50], have shown that cigarette smoking is associated with an increased risk for DVT. Similarly, a study conducted in Jimma University by Mulatu et al. [51], who found that Current Smoker patients was linked with increases in the development of DVT recurrence.

Family History of DVT: According to study done in University of Kartuoum by Gader et al. [5], the results show that family history of deep venous thrombosis was the higher risk factor of deep venous thrombosis. Another study conducted by Ms Bezemer et al. [52], show that the positive family history of DVT increased the risk of deep venous thrombosis more than 2-fold (odds ratio [95% confidence interval],(2.2 [1.9-2.6]) and up to 4-fold (3.9 [2.7-5.7]).

Immobilization: Prolonged immobilization is a significant predictors of deep venous thrombosis as study conducted in Dakar(Senegal) by Awa et al. [53]. Similarly,the study conducted by Engbers et al. [54], show that the highest risk of thrombosis was found for immobilisation during hospitalisation (OR 48.7; CI95 6.6-361.0), and the risk of thrombosis out-of hospital was 15-fold increased within the two weeks after hospital discharge, and the risk remained increased for 3 months after hospital discharge. Another study was conducted in Addis Ababa by Haile et al. [25], who found that prolonged immobilization was considered as risk factor of deep venous thrombosis. This is particularly why DVT is such a concern for people on bed rest (say, in a hospital setting), those who have medical conditions that prevent them from walking, and those who drive long distances or travel on long flights and are stationary for longer than four hours.

Living condition: According to a study conducted in Port Said University by Shahin et al. [55], using linear regression model for prevalence of deep vein thrombosis, found that living condition was significantly associated with deep venous thrombosis. A prospective cohort study conducted in Jimma found that living condition was significantly affect on patients living with deep venous thrombosis and about 86% of pateints live with immediate family [51].

Place of Residence: According to study conducted in Jimma University by Abera et al. [51] showed that there is statistically significant difference between rural and urban area of patients live with deep venous thrombosis and about 57.4% of participants were

residing in the rural area.

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 [34].

Theory of linear mixed model

Many longitudinal studies are designed to investigate change over time in a characteristic which is measured repeatedly for each patient [30]. 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 [31].

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 [31]. For example the units may be patients in a longitudinal study where a measurement of biological laboratory markers such as INR and APTT measures is taken at every weeks 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 [56]. 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 individuals response trajectories change over time. Mixed-effects models are highly attractive due to their ability to handle missing and unbalanced data reasonably well

Review on joint modeling of bivariate longitudinal outcomes

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.

Bivariate longitudinal data arise when a set of different responses on the same unit are measured repeatedly over time. An example of a research question for such data is how the evolution of one response is related to the evolution of another response ('association of the evolutions'). A seemingly related, but different question is how the association between responses evolves over time ('evolution of the association'). To answer such research questions a joint modelling strategy is needed. Multivariate longitudinal data arise when a set of different responses on the same unit are measured repeatedly over time. A joint modeling of such kind of data is necessary to quantify, firstly, the relationship between evolutions of different responses and, secondly, the evolution of the relationship between different response variables over time. Thus, a pair wise fitting approach has been used in this study as proposed by Fieuws et al. [57] and Fieuws et al. [58].

The introduction of the linear mixed model, devised for continuous data, has been followed rapidly by extensions to deal with non-linear data (non-linear mixed effects model, and with non-continuous measurements (generalized linear mixed model were found by laird et al. [30] and Breslow et al. [59]. In a joint-modelling 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 to joint models for responses of the same response type as well as models for responses of different types. The approach has been used in a non-longitudinal setting to validate surrogate endpoints in meta-analyses done by Buyes et al. [60] and Burzykowski et al. [61] or to model multivariate clustered data .

Gueorguieva et al. [62] used the approach for the joint modelling of a continuous and a binary outcome measure in a developmental toxicity study on mice. Also in a longitudinal setting, Chakraborty et al. [63] 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 References [64, 65].

Thi'ebaut et al. [65] investigated the bivariate random effects model between the evolution of CD4 and HIV RNA and he reported the bivariate random effects model was significantly better than two separate univariate random effects models with ($p-value <$ 0.0001). He found the highest correlations between the slopes of the two markers at the same period: (before 4 months and after 4 months). Similarly, the joint mixed effect model on evolution of occurrence and prevalence of antimicrobial resistant zoonotic agents were executed by Ferrari and Cribari-Neto et al. [66]. They used beta-regression to illustrate the joint evolution on both outcome variables and they reported that, the correlation was estimated to be 0.95, with 95% confidence interval [0.414, 0.997] showing that the correlation was positively significant. Thus, there was a strong correlation between percentage resistant and prevalence and that both were increased with time. That correlation however ignores the effect of time.

The study conducted in Virginia Common wealth University by Join,M [67], who 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.

Also, Negash et al. [68] used joint random effect model 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. Under this study 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. For this study fitted statistics showed that the joint model resulted in better fit to the data than the separate models .

Yemane [69] used linear mixed effect model and joint mixed effect model to investigate the joint evolution of pulse rate and respiratory rate of cognitive heart failure patients and identify the potential risk factors affecting the two endpoints in Ayder referral Hospital of Mekelle University. A linear mixed-effects model was fitted for the pulse rate and respiratory rate outcomes. Furthermore, a joint mixed-effects model was fitted for the two endpoints using SAS software PROC MIXED, and the potential risk factors affecting their joint evolution are identified. Finally, he recommended that

to identify associated effect fitting joint model is better.

In addition, Zucker et al. proposed a multivariate growth curve model to make inferences on the association between evolutions .Williams [33] used this approach to model simultaneously growth curves for systolic and diastolic blood pressure, height and BMI. However, such a modelling strategy is restricted to the combination of outcomes of the same type. The joint modelling approach investigated in this study was the bivariate longitudinal mixed effect models that included both fixed and random effects.Thus, in this study INR and APTT be the bivariate outcomes for the i^{th} subject measured at j^{th} times for outcomes 1 and 2.

3 METHODOLOGY

3.1 Study Area

The study would be conducted at Jimma University Medical Center. Jimma University Medical Center was one of the oldest public hospitals in Ethiopia. It was established in 1930 E.C by Italian invaders for the service of their soldiers. After the withdrawal of the colonial occupants, it has been governed by the name of "Ras Desta Damtew Hospital" and later "Jimma Hospital" during the Dergue regime and currently JUMC. This time the hospital provides services for more than 20 million patients with 800 bedded.

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

3.2 Study design

The study was a retrospective cohort study longitudinal process research design. because, it investigates the admitted follow-up of all deep venous thrombosis patients, who have followed at least three visits from first, September 2018 to first, January 2021 in Jimma University Medical Center is included.

3.3 Source of Data

The source of data for this study would be secondary data. This data was extracted from the follow-up of patients' cards by assigning an identification number (ID) per individual from the medical center and the card contains epidemiological, laboratory, and clinical information of all deep venous thrombosis patients. To extract the data two days of training for two workers and seven days for supervisions has been provided

on data extraction and organization in order to have relevant data. Two weeks has been taken for data extraction and cleaning; Then, after data extraction, data entry, data editing, data coding, and organization was conducted.

3.4 Study population

The study population of this study was all deep venous thrombosis patients in the JUMC, who were admitted to follow up in time interval of the first September 2018 to first January 2021.

3.5 Inclusion and Exclusion criteria

All deep venous thrombosis patients who have a follow-up at least three visits and whose ages are greater than 18 were included in this study. But deep venous thrombosis patients who have less than three follow-ups and whose ages less than 18 were excluded, because DVT is very rare before the age of 18 years [39] .

3.6 Study variables

3.6.1 Dependent Variables

- International normalized ratio
- Activated partial thromboplastin time

3.6.2 Independent variables or covariates

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

Table 1: Covariates used in the Separate and Joint Analysis of INR and APTT outcomes.

Name	Definitions	Codes
Age	Age of Patients	In year
Place	Place of patients	0=Urban, 1=Rural
Sex	Sex of DVT Patients	0=Female,1=Male
Time	Time interval of patients taking treatment	Measured by weeks
Alcohol user	patients who use alcohols	0=No, 1=Yes
Smoking cigar	patients who smoke cigar	0= No, 1 = Yes
Family History	Family History of DVT patients	0= No, 1= Yes
Immobilize	Immobilizing of patients	0=No, 1= Yes
Living condition	living condition of patients	0= live with family, 1= live alone

3.7 Statistical methods of data analysis

Different statistical analysis including both descriptive and inferential statistics, such as: summary statistics, data exploring and model comparison have been used in this study. Joint random effects with LMM have been modeled to infer the joint effect of bivariate longitudinal outcomes of DVT patients.

3.7.1 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 [30]. 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.

3.7.2 Descriptive statistics and Data exploring

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, the average evolution,profiles plot, the mean profile plot and exploring the random effects and other data exploratory analysis for the data sets have been considered.

3.7.3 Exploratory Data Analysis (EDA)

The first step in any model building process is exploratory data analysis. longitudinal studies address the relationship of a response with explanatory variables, often including time and following aspects of the data would be explaining; like Individual profiles and Mean structure (the average evolution of the responses over time) of between and within-subjects through the white nuisance error term and random effects respectively. In general data exploration is extremely helpful tool in the selection of appropriate models. Under the exploratory data analysis, we will have seen the following aspects: individual profile plot, mean profile plot.

3.7.4 Exploring the individual profile

Individual profiles are important to plotting observed profiles over time, helps to identify general trends within-subjects, may detect nonlinear change over time, and provides information about the variability at given times. In our case, the individual trajectory of the relation of INR and APTT on DVT patients over time will be explored to see the trends, to decide which random effects to be included in the final model and what the covariance structure of these random effects should have.

3.7.5 Exploring mean structure

Exploring mean structure of data was used to see how the mean profile changes over time, which is essential for specifying the functional form of the mean response of the model. To understand the possible relationships among means over time, graphical inspection were used by connecting the average values computed at each time point separately. Under this study, unbalanced data was expected and unequally spaced interval, so loess smoothing was used. This gave us some insight on how the mean profile evolves over time and help us to choose a fixed-effects structure for the linear mixed model like linear, quadratic and so on.

3.8 Statistical Model

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

3.8.1 Linear mixed effect model

A mixed linear model is a generalization of the standard linear model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and nonconstant variability. 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.

Unlike general linear model family, linear mixed models (LMM) are models that handle data where observations are not independent. Many longitudinal studies are designed to investigate change over time in a characteristic which is measured repeatedly for each patient [30]. In addition, modeling the true correlation structure becomes statistically significant in the presence of missing values when the number of observations per subject is not large. The units are patients in a longitudinal study where measure-

ments of biological marker are taken at every three-monthly visit.

In mixed-effects models, response variables are assumed to be a function of fixed effect, non-observable random effect, and error term [30]. 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 [30] can be written as:
The development of a linear mixed-effects model can be performed in two stages based on [70].

Stage 1: The first stage assumes that Y_i satisfies a linear regression model,

$$Y_i = Z_i\beta_i + \epsilon_{ij} \text{.....(3.1)}$$

where Z_i is an appropriate design matrix. This model shows how the response evolves over time for the i^{th} subject where β_i is a p- dimensional vector of unknown subject specific regression coefficients and ϵ_{ij} is a vector of the residual component , j=1,2,3,4,.....n_i. Usually assumed to be normally distributed with mean zero and covariance matrix Ri. The model is completed by specifying the covariance Structure, which can be homogeneous or heterogeneous. Commonly used homogeneous covariance structures are; identity, autoregressive, compound symmetry.

$R_i = \sigma^2 In_i$

Where In_i denoting the identity matrix of dimension n_i. This is so under the strong assumption that all repeated measurements are independent though repeated measurements within the same subject are rarely independent.

Stage 2: The second stage is a multivariate regression model of the form $\beta_i = K_i\beta + b_i$ which models variability between the subjects with respect to their subject specific regression coefficients, β_i, K_i is a $(q \times p)$ matrix of covariates, b_i is are assumed to be independent following a q dimensional normal distribution with mean zero and general

covariance structure D.

$$\beta_i = K_i\beta + b_i \text{.....(3.2)}$$

Where:

K_i is ($q \times p$) matrix of known covariates. Additional covariates other than time can be adjusted in this step.

β is a p dimensional vector of regression coefficients

$b_i \sim N(0, D)$ and is a matrix of random effects indicating the deviation in individual subjects' measurements from the population average. D is the covariance matrix capturing the between subject variability.

Generally, Substituting for equation $\beta_i = K_i\beta + b_i$ in above equation 3.1, we get the linear mixed model given below. In the second stage, between subjects variability is modeled. This is achieved by relating the estimated coefficients with known covariates. Thus, the linear mixed model from the above two stages were formulated as:

$$Y_{i(t)} = Z_{i(t)}(K_i\beta + b_i) + \epsilon_{ij} \text{.....(3.3)}$$

Where $Z_{i(t)}K_i = X_{i(t)}$

$Y_{i(t)} = X_{i(t)}\beta + Z_{i(t)}b_i + \epsilon_{ij}$ $\epsilon_{ij} \sim N(0, R)$

$b_i \sim N(0, D)$

$Y_{i(t)}$: Measurement of univariate response of i^{th} subject at time t.

$X_{i(t)}$ and $Z_{i(t)}$ are the fixed and random of covariates with p and q dimension, respectively.

β: Vector of unknown parameters associated with fixed effect (of dimension p)

b_i:Vector of unknown parameters associated with random effect for i^{th} subject (of dimension q)

b_i stands for the random effects

The model in equation (3.3) assumes both fixed and random effects. Furthermore, b_1,.....,b_N and ϵ_1,.....,ϵ_N are assumed to be independent. The elements for the variance components are in the matrices D and R are called error component models, ϵ_1, can be decomposed into two components representing both subject specific variation

and variation over time that is serial correlation. Therefore $\epsilon_i = \epsilon_{1_{(i)}} + \epsilon_{2_{(i)}}$ where $\epsilon_{1_{(i)}}$ is the measurement error associated with ith subject and $\epsilon_{2_{(i)}}$ is associated with serial correlation for i^{th} subject.

D and R are variance components

$$E\begin{bmatrix} b_i \\ \epsilon_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and

$$var\begin{bmatrix} b_i \\ \epsilon_i \end{bmatrix} = \begin{bmatrix} D & 0 \\ 0 & R \end{bmatrix}$$

Assumption for the response: $Y \sim N(X\beta, V), V = Z_i D Z_i^{'} + R$

The name mixed model indicates that the model contains both the fixed or the mean model component and the random component and variable effects are either fixed or random depending on how the levels of the variables that appear in the study are selected. It can also use for data with unequal number of measurements per subjects. These unobserved effects are then included in the model as random variables, or equivalently called, random effects. A fixed effect is considered to be a constant which we wish to estimate, but the random effect is considered as just an effect coming from a population of effects [31].

3.8.2 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 general linear mixed effects model.

(i) $\epsilon \sim N(0, \sigma^2 I_{ni})$

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. This assumption can be relaxed by allowing to model non constant variances or special within group correlation structures.

(ii) b $\sim N(0, D)$

The random effects are normally distributed, with mean zero and covariance matrix D (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 plots of the residuals, the fitted values, and the estimated random effects.

3.8.3 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 [71].

3.8.4 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:

$$L = l(\beta, V, Yi) = \prod_{i=1}^{N} \{\ 2\pi^{\frac{-1}{2}} \det(V)^{\frac{-1}{2}} exp(\frac{-1}{2}(Y_i - X_i\beta)^T V^{-1}(Y_i - X_i\beta))\}$$

Where β is a vector of fixed-effects parameters and V is a vector containing the variance parameters. the likelihood function of the equation can be compared with the log of the likelihood function generally used in practice. Its maximum value coincides with that of the likelihood function. The log-likelihood of the model parameters, is defined as:

$$logL = l(\beta, V, Yi) = -\frac{N}{2} log(2\pi) - \frac{N}{2} log \det(V) - \frac{1}{2} \sum_{i=1}^{N} (Y_i - X_i\beta)^T V^{-1}(Y_i - X_i\beta))$$

$$K - \frac{N}{2} log \det(V) - \frac{1}{2} \sum_{i=1}^{N} (Y_i - X_i\beta)^T V^{-1} (Y_i - X_i\beta))$$

Where $K = \frac{N}{2} log(2\pi), V = Z_i D Z_i^{'} + R$

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 V. 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 [72]. The MLE of β on combining all the information from all the N subjects equals.

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

Where det refers to the determinant and the elements of V_i the matrix are functions of the covariance parameters in $\ominus$ [73].

3.8.5 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, $(Y_i - X_i\beta)$. It may be noted that

the full residuals are a linear combination of Y and furthermore $(Y_i - X_i\beta)$ and β are independent . From these facts, the joint-likelihood for and the variance parameters, , may be express as a product of the likelihoods based on $(Y_i - X_i\beta)$ and β.

$$L(V, \beta; Y_i) = L(V; Y_i - X_i\beta)L(\beta; \beta, V)$$

The REML is defined as follow equation

$$L(V; Y_i - X_i\beta) = \det(X_i' V^{-1} X_i)^{\frac{-1}{2}} \det(V)^{\frac{-1}{2}} exp\{\frac{-1}{2}(Y_i - X_i\beta)' V^{-1}(Y_i - X_i\beta)\}$$

3.8.6 Checking model assumptions for independent mixed models

After fitting an LMM, it is important to carry out model diagnostics to check whether distributional assumptions for the residuals are satisfied and whether the fit of the model is sensitive to unusual observations. The process of carrying out model diagnostics involves several informal and formal techniques. 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 [34] . 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. In general, model diagnostics should be part of the model-building process throughout the analysis of a clustered or longitudinal data set. In this case diagnostics only for the final model fitted has been considered.

Residual Diagnostics:- Informal techniques are commonly used to check residual diagnostics; these techniques rely on the human mind and eye, and are used to decide whether or not a specific pattern exists in the residuals. In the context of the standard linear model, the simplest example is to decide whether a given set of residuals plotted against predicted values represents a random pattern or not. These residual vs fitted

plots are used to verify model assumptions and to detect outliers and potentially influential observations. In general, residuals should be assessed for normality, constant variance, and outliers. In the context of LMMs, we consider conditional residuals and their studentized versions.

Diagnostics for Random Effects:- The natural choice to diagnose random effects is to consider the empirical Bayes (EB) predictors. EB predictors are also referred to as random-effects predictors or, due to their properties, empirical best linear unbiased predictors (EBLUPs). [73] . Recommend using standard diagnostic plots (e.g., histograms, Q_Q plots, and scatter plots) to investigate EBLUPs for potential outliers that may warrant further investigation. In general, checking EBLUPs for normality is of limited value, because their distribution does not necessarily reflect the true distribution of the random effects.

3.8.7 Joint model for bivariate continuous longitudinal data

The linear mixed model can be easily extended to include bivariate response variables by further stacking the data and defining a specific variance-covariance structure for the random effects. Bivariate linear mixed models are useful when analysing longitudinal data of two associated markers. In this paper, a bivariate linear mixed model including random effects and independent measurement error for both INR and APTT was presented. Longitudinal data are often collected in epidemiological studies, especially to study the evolution of biomedical markers. 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.

Extension to bivariate Case: Now under bivariate set-up two endpoints or measure of DVT (INR and APTT) as outcome variables are observed in each occasion. The two end points were longitudinally measured as a vector of responses, $Y_i(t)$, at each occasion and thus the following model was used:

$$Y_{i(t)} = X_i(t)\beta + Z_i(t)b_i + \epsilon_{i(t)}$$

where,

$$\epsilon_i = [\epsilon_i(t_1) \quad , \epsilon_i(t_2) \quad , ...\epsilon_i(t_n)] \sim MVN(0, R_i)$$

$b_i \sim MVN(0, D)$ $\qquad\qquad$ $\text{cov}(b_i, \epsilon_i) = 0$

$R_i = I_{n_i} \otimes \Sigma_{2\times 2}$, Where, $\Sigma_{2\times 2}$ is the variance covariance matrix of 2 endpoints (INR and APTT) conditional on b_i.

Let y_{ikj} represent the j^{th} observation from i^{th} the subject for the k^{th} response variables, where $i = 1, ...S$, $j = 1, ...n_{ik}$ and $k = 1, ...k$ .For this thesis k=1 and 2, where $N_K = \sum_{i=1}^{s} n_{ij}$ and $N = \sum_{k=1}^{k} N_k$. The vector $y_{ij} = [y_{i1k}, y_{i2k}...y_{ink}]$ then represents the n_{ik} observation of the k^{th} response variable from the i^{th} subject the vector $Y_k = [Y_{1k}, Y_{2k}...Y_{sk}]'$ representes the N_k observation from the k^{th} response variable across all subjects.

Assume: $Z_{i1}(t)b_{i1} = a_{i1}+b_{i1(t)}$ and $Z_{i2}(t)b_{i2} = a_{i2}+b_{i2(t)}$ are the subject-specific random effects from each linear mixed model.

In the context of modeling two response variables, the linear mixed-effects models for each response variable for subject i taken at time t can be specified as [35].

$Y_{i1(t)} = \mu_{1(t)} + a_{i1} + b_{i1(t)} + \epsilon_{i1(t)}$

$Y_{i2(t)} = \mu_{2(t)} + a_{i2} + b_{i2(t)} + \epsilon_{i2(t)}$

where, $\mu_{k(t)}$ refers to the average evolution (of the k^{th} response over time) and is a function of the fixed effect. The subject specific random intercepts a_{ik} and slopes $b_{ik(t)}$ describe how the subject specific profiles deviate from the average profile for the k^{th} response.

where $\mu_{1(t)}$ and $\mu_{2(t)}$ refers to the population means at time t. We assume that random effects are jointly distributed as follows.

$$\begin{bmatrix} b_{i1} \\ b_{i2} \end{bmatrix} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ b_{i1} \\ b_{i2} \end{bmatrix} \sim N(0, D)$$

where, D, the covariance matrix of the random effects, has the following structure

$$cov(D) = \begin{bmatrix} \sigma^2_{a_1} & \sigma_{a_1}\sigma_{a_2} & \sigma_{a_1 b_1} & \sigma_{a_1 b_2} \\ & \sigma^2_{a_2} & \sigma_{a_2 b_1} & \sigma_{a_2 b_2} \\ & & \sigma^2_{b_1} & \sigma_{b_1 b_2} \\ & & & \sigma^2_{b_2} \end{bmatrix}$$

3.8.8 Association of the evolution (AOE)

The main 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 as follow. Clearly, the correlation between the evolution of Y_1 and Y_2 is given by:

$$r_E = \frac{Cov(b_1, b_2)}{\sqrt{var(b_1) \times var(b_2)}} = \frac{\sigma_{b_1 b_2}}{\sqrt{\sigma^2 b_1 \times \sigma^2 b_2}}$$

3.8.9 Evolution of the association (EOA)

Similar, we use evolution of the association in joint mixed effects model to investigated the association between the responses variables evolved over time (evolution of the association). Assuming uncorrelated errors, the marginal correlation between $Y_1$and Y_2 at time t is given by [35]: And the marginal correlation between $Y_1$and Y_2 at time t is given by:

$$r_M(t) = \frac{Cov(Y_{i1}(t), Y_{i2}(t))}{\sqrt{var(Y_{i1}(t)) \times var(Y_{i2}(t))}} = \frac{\sigma_{a_1 a_2} + t\sigma_{a_1 b_2} + t\sigma_{a_2 b_1} + t^2\sigma_{b_1 b_2} + \sigma_{12}}{\sqrt{(\sigma^2_{a1} + 2t\sigma_{a_1 b_1} + t^2\sigma^2_{b1} + \sigma^2_1) \times (\sigma^2_{a2} + 2t\sigma_{a_2 b_2} + t^2\sigma^2_{b2} + \sigma^2_2)}}$$

It is not difficult to comprehend that as the number of response variables (or the dimension of multivariate response) increases, the number of covariance parameters increase exponentially and the problem of estimation of covariance parameters becomes more and more difficult. when t= 0 the marginal correlation converges to;

$$r_M(t) = \frac{\sigma_{a_1 a_2} + \sigma_{12}}{\sqrt{(\sigma^2_{a1} + \sigma^2_1) \times (\sigma^2_{a2} + \sigma^2_2)}}$$

which is essentially the correlation between the two random intercepts. If fact, when the error components are small, the closer the marginal correlation at t= 0 approximates the correlation between the random intercepts.

3.9 Joint Model Estimation

In the particular context of random-effects models, so-called adaptive quadrature rules can be used [74], 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 and the Gaussian Quadrature approximates the integral of a function, with respect to a given kernel, by a weighted sum over predefined abscissas for the random effects. To illustrate the main ideas, we consider Gaussian and adaptive Gaussian quadrature, designed for the approximation of integrals of the form

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

for a known function $f(z)$ and for $\phi(z)$ 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 equal.

$$f_i(y_i|\beta, D, \phi) = \int \Pi_{i=1}^{s} f_i(y_i|b_i, \beta, \phi) f(b_i|D) db_i$$

- where, b_i is $q \times 1$ dimensional vector with random effect $b_i \sim (0, D)$
- β is a vector of fixed-effects parameters and ϕ is a vector containing the variance parameters.

3.9.1 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 [74]. The correlation between two within-group errors ϵ_{ij} and $\epsilon_{ij}^{'}$ is assumed to depend on some distance between them, and is a vector of correlation parameters. Olkin and Tate 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: $corr(\varepsilon_{ij}, \varepsilon_{ij}^{'}) = h(\rho)$

Where h(ρ) 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. In CS structure the variances are homogeneous. There is a correlation between two separate measurements, but it is assumed that the correlation is constant regardless of how far apart the measurements are.Thus, the corresponding correlation model are:
$corr(\varepsilon_{ij}, \varepsilon_{ij}^{'}) = \rho$

General (Unstructured): An unstructured 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 an unstructured correlation model tends to over parameterized model. It is useful for few observations per subject that leads to precise correlation with observations.

Autoregressive (AR): The AR (1) structure has homogeneous variances and correlations that decline exponentially with distance. It also means that two measurements that are right to next to each other in time are going to be pretty correlated (depending on the value of p), but that as measurements get farther and farther apart they are less correlated. 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.
$\varepsilon_t = \phi_1\varepsilon_{t-1} + . \quad . \quad . \quad + \phi_p\varepsilon_{t-p} + \mu_t \qquad |\phi_p < 1|$
p is called the order of the autoregressive model, which is denoted by AR(p). There are p correlation parameters in an AR (p) model, given by $\phi = (\phi_1, \quad \phi_2 \quad . \quad . \quad .\phi_p)$. The AR (1) model is the simplest and one of most useful autoregressive model. The correlation function of AR is given as:
$h(k;\phi) = \phi^k$ k = 0,1,. . . , where, k-distance between time point
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

3.10 Model Comparisons or Selection Techniques

Model selection technique is one of the most frequently encountered problems in data analysis. In this thesis the most commonly known model selection criteria are Akaike Information Criterion (AIC) [75], the Bayesian Information Criterion (BIC) [30] and Log-likelihood ratio test are use.

Akaike's information criterion (AIC) and Bayesian information Criterion (BIC): are a measure of goodness of fit of an estimated statistical model. It is not a test on the model in the sense of hypothesis testing; rather it is a tool for model selection

$$AIC = -2\log L + 2k$$

$$BIC = -2\log L + k\log(N)$$

Where,k denotes the number of parameters in the model and N the total number of observations used to fit the model. Under these definitions, smaller is better. That is, if we use AIC to compare two or more models for the same data, we prefer the model with the lowest AIC and BIC.

Likelihood-ratio test: is constructed by comparing the maximized log likelihoods for the Saturated (full) model and reduced models, respectively, and the test statistic is expressed as:

$$LRT = -2\ln(L_0 - L_m)$$

Where,$L_0$and L_m are the maximum likelihood estimates that maximize the likelihood Functions of the reduced and full or saturated model, respectively

4 RESULTS AND DISCUSSION

4.1 Data descriptions and Descriptive statistics

In this study, socio-demographic and clinical data of 281 deep venous thrombosis patients who were above 18 years and who receive treatment for symptoms of DVT from first September 2018 to first January, 2021 in Jimma University Medical Central at baseline were considered. The two measure of DVT such as: International Normalized Ratio and Activated Partial Thromboplastin time have been used. These longitudinal response variables were measured for at least 3 visits.

A total of 281 adults of deep venous thrombosis patients with a minimum of three and maximum of six measures of INR, APTT and other covariates per individual of deep venous thrombosis patients were included for this study. The baseline characteristics and descriptive statistics of patients are displayed in table 4.1 below. Out of 281 deep venous thrombosis patients, about 126(44.8%) of patients were females and 155(55.2%) of patients were males. Regarding to place of 281 deep venous thrombosis patients, about 163(58%) of patients were living in urban area, and 118(42%) of patients were living in rural communities (district). Among these patients, about 174(61.9%) of the patients had no genetic effect or family history of DVT, 107(38.9%) of patients had family history (previously diagnosed of DVT). Among 281 deep venous thrombosis patients, more than half 164 (61.9%) of patients were no had prolonged immobilization and 117 (38.1%) of patients were had prolonged immobilization. Similarly, about 158(56.2%) of patients were not Smokers and 123(43.8%) of patients were Smokers. Out of 281 deep venus thrombosis patients, about 107(37.7%) of them were live with their family and 175(62.3%) of them were live alone. Likewise, about 139(49.5%)of participants were not drank alcohols and 142(50.5%) of participants were drank alcohols.

Table 4.1: Frequencies and Percentages for baseline categorical covariates of INR and APTT for each category of Deep venous thrombosis patients data.

Name of Variables	Categories	Frequency	Percent
Gender	Male	155	55.2%
	Female	126	44.8%
Residence	Rural	118	42%
	Urban	163	58%
Immobilize	Yes	117	41.6%
	No	164	58.4%
Alcohols Users	Yes	142	50.5%
	No	139	49.5
Family History	No	174	61.9%
	Yes	107	38.1%
Smoking status	No	158	56.2%
	Yes	123	43.8%
Living condition	With family	106	37.7%
	Live Alone	175	62.3%

The Mean and Standard deviation of continuous covariates and two outcomes of this study displayed in the Table 4.2 below. The age at first visit ranged from 20 to 76 years with average value equal to 37.48 years (with a standard deviation of 13.850 years). The longitudinally measured symptoms of deep venous thrombosis using INR and APTT per second were considered as bivariate responses. They were measured approximately every week at the study entry, and again a common measuring time is used for all patients. Hence, the average number of baseline INR and APTT was 1.378 and 33.701 per seconds with standard deviation of 0.6535 and 10.5738, respectively.

Table 4.2: Mean and Standard deviation of continuous covariates and two outcomes at baseline.

Name of Variables	Mean	Std.Dev
International Normalized Ratio	1.378	0.6535
Activated partial thromboplastin time	33.701	10.5738
Age of patients	37.48	13.850

4.2 Separate analysis of international normalized ratio and activated partial thromboplastin time test

The two jointly measured outcome analyzed separately by using linear mixed model. 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 data analysis

Exploratory analysis comprises techniques to visualize patterns in the data. Data analysis begins by making displays that expose patterns relevant to the scientific question. Below very important tools were used to visualize patterns of International Normalized Ratio and Activated Partial Thromboplastin Time test over time.

4.2.2 Exploring individual profile plot of INR and APTT over time

The longitudinal outcomes, INR and APTT, were measured at irregular time intervals with one, two, three or four a weeks gap. The following figure 4.1 shows that the individual profile plots of INR and APTT over time. The variability of INR between individuals was lower at baseline and it appeared to increase slowly through time. Similarly, The variability of APTT between individuals was lower at baseline and it appeared to increase slowly through time but the variability was higher in INR than in APTT. There is also a variabilty with in each subjects. Likewise, there was between and within subject's variability in both INR and APTT. In general between and within subject specific differences could not be ignored. Also the following figure 4.1 shows that the individual profile plots of INR and APTT along with average profile using loess smoothing can observed. The loess smooth in both plots suggest that the average profiles, the INR and APTT, have linear relationships over time which are increasing over time, but with different evolution over time.

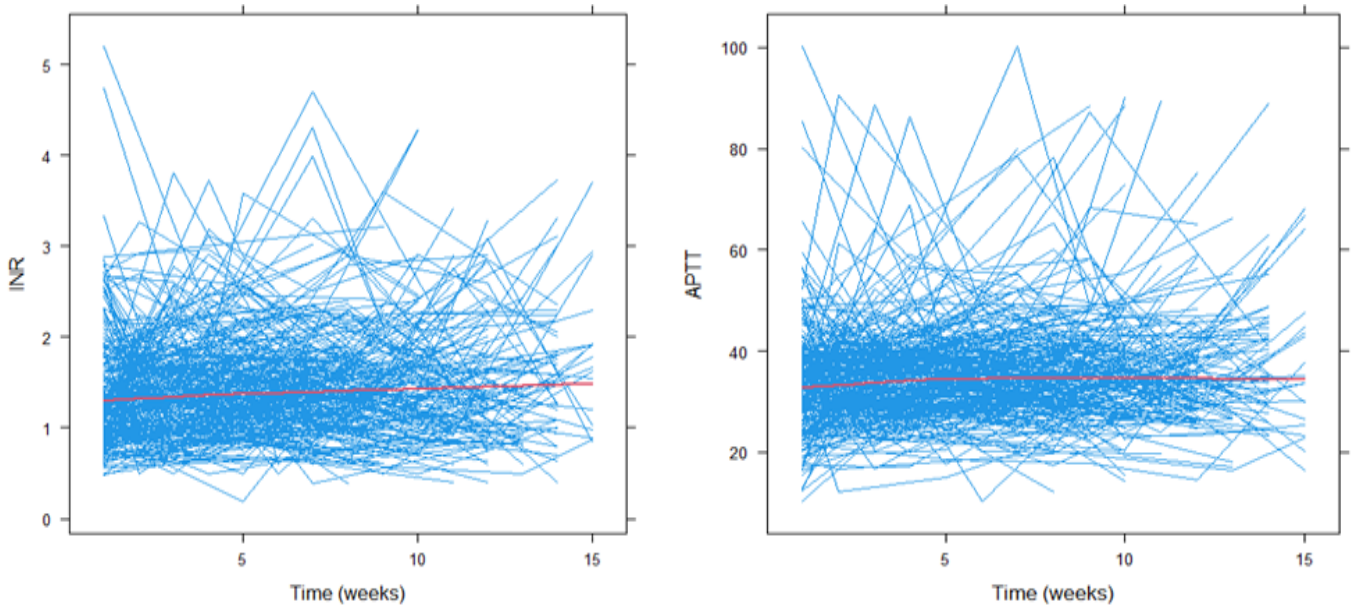

Figure 4.1: Individual profile plot of international normalized ratio and activated partial thromboplastin time

4.2.3 Mean profile Plots of INR and APTT of deep venous thrombosis patients:

The mean effect profile plot of the longitudinal measure of deep venous thrombosis INR and APTT of deep venous thrombosis patients shows that the rate of change of change over time is somewhat linear. So, linear with time random effect should be included in the model. The rates of change of changes over time increase step by step or from week to week. Mean profile plots are given below by using loess smooth curve. The loess smooth curve suggests that the average profiles of both the INR and APTT have a linear relationship over time. It indicates both INR and APTT shows a slowly increase pattern over time, but the rate of INR highly increase as compared to APTT.

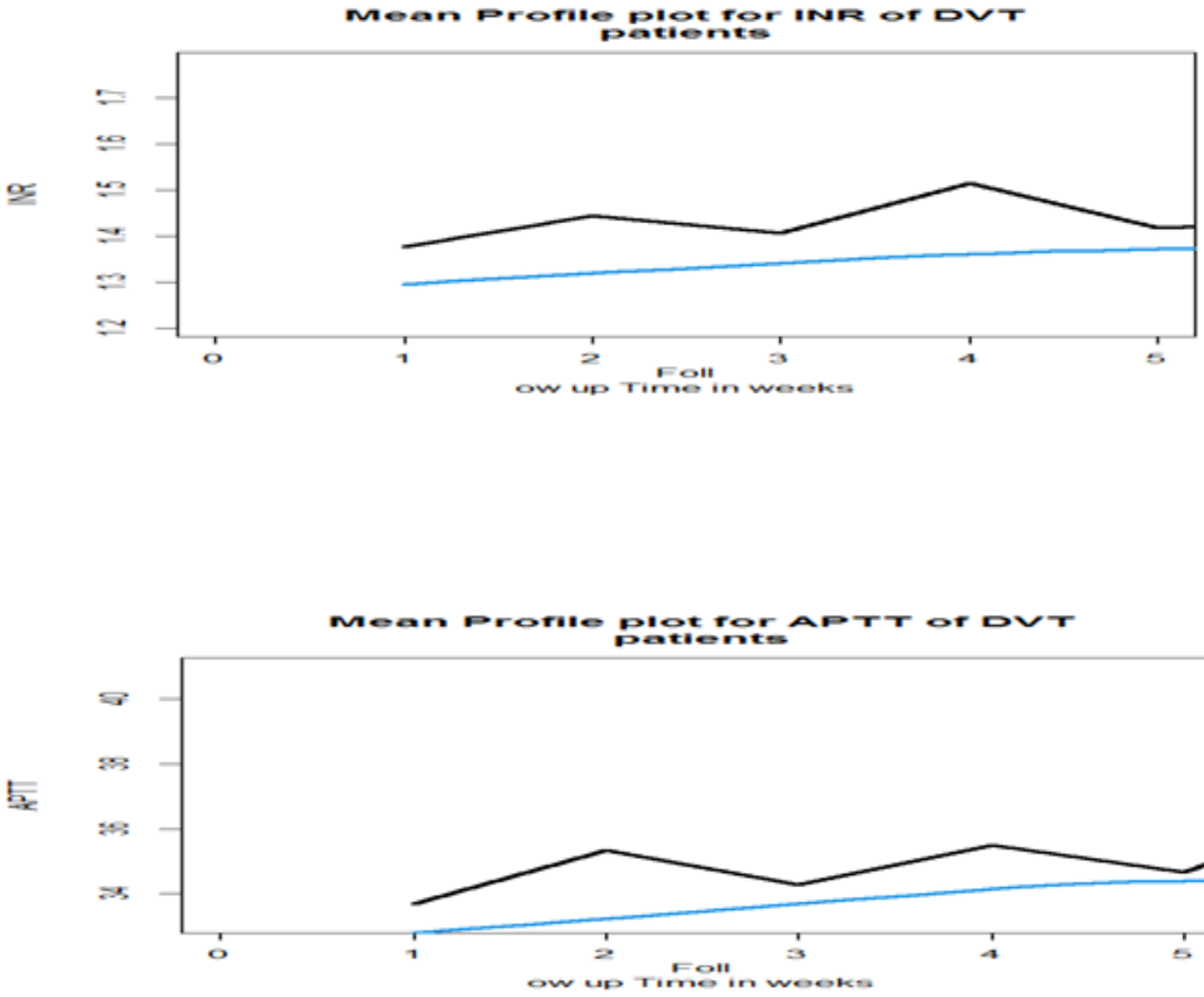

Figure 4.2: Mean profile Plots of INR and APTT of deep venous thrombosis patients

4.3 Model Selection

A primary goal of model selection is to choose the best model that provides the best fit to the observed data well. There may be several choices concerning which fixed and random effects should be included in an LMM. There are also many possible choices of covariance structures for the D and R_i matrices.

4.4 Linear Mixed Effects Models and Its Results

4.4.1 Linear mixed model for international normalized ratio

4.4.2 Selection of fixed effects for international normalized ratio

To select the fixed effect components of the response variable for international normalized ratio. A simple linear regression model using base line information without considering any random effect and neglecting any correlation between and within subject were fitted with all covariate and including same interaction terms as follow.
$INR_{ij} = \beta_{10} + \beta_{11}T_{ij} + \beta_{12}Sm_i + \beta_{13}Al_i + \beta_{14}Sex_i + \beta_{15}Im_i + \beta_{16}pl_i + \beta_{17}Fh_i + \beta_{18}A_i + \beta_{19}Lic_i + (\beta_{20}Sex_i + \beta_{21}pl_i + \beta_{22}Sm_i) * T_{ij} + \epsilon_{ij}$
Let INR_{ij} denote the j^{th} international normalized ratio of the i^{th} patient at time t_{ij},. Where i indexes the subjects i= 1, 2 ,.., 281 and j indexes the time visit for subject i, j= 1, 2, ..., n_i. ni represents the overall visits of subject i. Where : INR_{ij}- is International Normalized Ratio for i^{th} subjects. Al_i- Alcohol users $\beta_{10}, \beta_{11}, . \ . \ . \ \beta_{22}$ - Are the fixed effect coefficient parameters
A_i- Age of patients for i^{th} subjects. pl-place of patients for i^{th} subjects
Sex_i- Sex of patients for i^{th} subject Lic_i-Living condition of patients for i^{th} subjects.
Im_i-prolonged Immobilize of patients for i^{th} subjects, Sm_i- Smoking cigarette .
Fh_i- Family history of DVT patients for i^{th} subjects. ϵ_{ij}- Random term
T_{ij}-Time at which International Normalized Ratio measured.

From the outputs in table 6.1 (Appendix A), we can observe that some covariate are statistically significant, but living condition ,place of residence are more insignificant and the interaction term time by place of residence are statistically insignificant. Thus, the insignificant terms should be removed from the model starting with the most insignificant one of which is the place of residence and interaction term place of residence by time. The model was then refitted after removing the place of residence and interaction term place of residence by time and the AIC dropped from 2263.245 to 2257.245 indicating a better fit. The model was fitted again and the categorical covariate living condition was still insignificant. The next step is to remove the covariate living condi-

tion. Then the model was fitted again as follow. Hence, the fixed effects model with linear time effect for INR measurement is given by:

$INR_{ij} = \beta_{10}+\beta_{11}T_{ij}+\beta_{12}Sm_i+\beta_{13}Al_i+\beta_{14}Sex_i+\beta_{15}Im_i+\beta_{17}Fh_i+\beta_{18}A_i+(\beta_{20}Sex_i+\beta_{22}Sm_i)*T_{ij}+\epsilon_{ij}$

4.4.3 Selection of random effect for international normalized ratio

The random effect of the international normalized ratio is the rate that shows, the rate of change of INR measure over time including all potential covariates. The aim of this section is to select the random effect model of the rate of change of INR measure over time for the fitted models. 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. There are four different models with different random effects starting from a simple linear regression model (no random effects) have been fitted for international normalized ratio. Table 4.3 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 2257.003 to 2241.827, which indicates that model with intercept and slope, was a better fitting model for international normalized ratio measures of DVT.

Table 4.3: Selection of appropriate random effect models for international normalized ratio.

Random effect	AIC	BIC	logLik
No Random Effects	2257.003	2397.494	-1100.502
Random Intercepts	2246.871	2397.369	-1094.435
Intercept and Linear Slope	2241.827	2392.378	-1089.913
Intercept linear and Quadratic Slope	2242.488	2398.030	-1090.244

As it is shown in table 4.3 above, among different random effects mentioned, the model with random intercept and linear slopes was selected as the best model for INR among the four random effects with respective small values of AIC and BIC, 2241.8271 and 2392.378 ,the model was selected with log-likelihood ratio test with p_value of 0.0110. Let INR_{ij} denote the j^{th} international normalized ratio of the i^{th} patient at time t_{ij},.

Where i indexes the subjects i= 1, 2 ,.., 281 and j indexes the time visit for subject i, j= 1, 2, ..., n_i. ni represents the overall visits of subject i. Finally, the fixed effects model with linear time effect for INR measurement is given by:

$INR_{ij} = \beta_{10}+\beta_{11}T_{ij}+\beta_{12}Sm_i+\beta_{13}Al_i+\beta_{14}Sex_i+\beta_{15}Im_i+\beta_{17}Fh_i+\beta_{18}A_i+(\beta_{20}Sex_i+\beta_{22}Sm_i)*T_{ij}+R_i(T_{ij})+\epsilon_{ij}$

Where:- $R_i(T_{ij})$=$a_{10}+b_{11}*T_{ij}$, Here, $R_i(T_{ij})$ includes the random effects for intercept and linear time slopes, where the $b_i = (a_{10}, b_{11})^{'} \sim N(0, D)$. The vector $(\beta_{10}, \beta_{11}, , , \beta_{18})$ of fixed effects describes the average evolution of international normalized ratio and the vector $(a_{10}; b_{11})$ of random effects describes how the profile of the i^{th} subject deviates from the average evaluation of INR.

Therefore, the model which 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.

4.5 Linear mixed model for activated partial thromboplastin time test

4.5.1 Selection of fixed effects for activated partial thromboplastin time test

Similar to the response international normalized ratio, after passing through different procedures of model selection criteria's, the model with random interaction term and linear slop chosen as the best model fitting the data for activated partial thromboplastic time test. This model selected after removing living conditoin, place of residence and place of residence with interaction of time from the model with random interaction and linear slope, the model was fitted below.

Let $APTT_{ij}$ denote the j^{th} Activated Partial Thromboplastin Time Test of the i^{th} patient at time t_{ij},. Where i indexes the subjects i= 1, 2 ,.., 281 and j indexes the time visit for subject i, j= 1, 2, ..., n_i. ni represents the overall visits of subject i. Hence, the fixed effects model with linear time effect for APTT measurement is given by:

$APTT_{ij} = \beta_{20} + \beta_{21}T_{ij} + \beta_{22}Sm_i + \beta_{23}Al_i + \beta_{24}Sex_i + \beta_{25}Im_i + \beta_{27}Fh_i + \beta_{28}A_i + (\beta_{30}Sex_i + \beta_{32}Sm_i) * T_{ij} + \epsilon_{ij}$

Where :

$APTT_{ij}$- is Activated Partial Thromboplastin Time for i^{th} subjects.

A_i- Age of patients for i^{th} subjects.

Sex_i- Sex of patients for i^{th} subject

Im_i- prolonged Immobilize of patients for i^{th} subjects,

Fh_i- Family history of DVT patients for i^{th} subjects. Lic_i- Living condition of patients for i^{th} subjects.

Sm_i- is patients who Smoking cigar for i^{th} subject.

T_{ij}-Time at which Activated Partial Thromboplastin Time Measured.

ϵ_{ij}- Random term

4.5.2 Selection of random effect for activated partial thromboplastin time

The random effect of the activated partial thromboplastin time is the rate that shows, the rate of change of APTT measure over time including all potential covariates. The aim of this section is to select the random effect model of the rate of change of APTT measure over time for the fitted models. 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.

There are four different models with different random effects starting from a simple linear regression model (no random effects) have been fitted for Activated Partial Thromboplastin Time. Table 4.4 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 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 8559.255 to 8530.349, which indicates that model with intercept and slope, was a better fitting model for activated Partial thromboplastin Time measures of DVT.

Table 4.4: Selection of appropriate random effect models for Activated Partial Thromboplastin Time.

Random effect	AIC	BIC	logLik
No Random Effects	8559.255	8699.745	-4251.628
Random Intercept	8538.924	8685.892	-4240.462
Intercept and Linear Slope	8530.349	8684.432	-4234.175
Intercept Linear and Quadratic Slope	8532.699	8688.242	-4235.350

As it is shown in table 4.4 above, among different random effects mentioned, the model with random intercept and linear slopes was selected as the best model for APTT among the all of random effects with respective small values of AIC and BIC, 8530.349 and 8684.432, the model was selected with log-likelihood ratio test with p_value of 0.0019.

Let $APTT_{ij}$ denote the j^{th} Activated Partial Thromboplastin Time of the i^{th} patient at time t_{ij},. Where i indexes the subjects i= 1, 2 ,.., 281 and j indexes the time visit for subject i, j= 1, 2, ..., n_i. ni represents the overall visits of subject i. Finally,the fixed effects model with linear time effect for APTT measurement is given by:

$APTT_{ij} = \beta_{20} + \beta_{21}T_{ij} + \beta_{22}Sm_i + \beta_{23}Al_i + \beta_{24}Sex_i + \beta_{25}Im_i + \beta_{27}Fh_i + \beta_{28}A_i + (\beta_{30}Sex_i + \beta_{32}Sm_i) * T_{ij} + R_i(T_{ij}) + \epsilon_{ij}$

Where:- $R_i(T_{ij})$=$a_{20} + b_{21} * T_{ij}$, Here, $R_i(T_{ij})$ includes the random effects for intercept and linear time slopes, where the $b_i = (a_{20}, b_{21})^{'} \sim N(0, D)$. The vector $(\beta_{20}, \beta_{21}, , , \beta_{28})$ of fixed effects describes the average evolution of activated partial thromboplastin time and the vector $(a_{20}; b_{21})$ of random effects describes how the profile of the i^{th} subject deviates from the average evaluation of APTT.

Therefore, the model which including the random effects model are similar to those of the fixed effect model in terms of magnitude.

4.5.3 Model selection with correlation structure for the best model

Selection of best correlation structure for both response

In longitudinal study selecting best model is not selecting model with only the best mean structure, but also correlation structure is important. In linear mixed effect model covariance structures should be carefully selected in order to obtain valid inferences for

the parameters of fixed effects in the model. Ignoring important correlations increase probability of type I error, and underestimates standard errors of an estimate [76]. In fitting the linear mixed effect model, a series of covariance structures of the longitudinal INR and APTT of deep venous thrombosis patients were considered. From the possible covariance structures, the one with the smallest AIC and BIC with convergence of the model in REML and ML were considered. There are different correlation functions or correlation covariance structure , 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 autoregressive structure function is preferable for both international normalized ratio model and activated partial thrombplastic time test model.

Table 4.5: Selection of appropriate Correlation structure for both INR and APTT.

	INR			APTT		
correlation structure	AIC	BIC	logLik	AIC	BIC	logLik
CompSymmetry (CS)	2243.827	2404.387	-1089.913	8532.349	8692.909	-4234.175
Unstructured(UN)	2245.196	2450.914	-1081.598	8545.527	8741.210	-4233.764
Autorsive (AR(1))	2241.743	2402.3038	-1088.872	8526.072	8686.632	-4231.036

From the above table 4.5, among different correlation structure mentioned, the model with correlation structure of autoregressive (AR-1) structure was preferred for INR and APTT model with respective small values of AIC and BIC of 2241.743 and 2402.3038 for INR and 8526.072 and 8686.632 for APTT model . As shown from the table autoregressive (1) is the best correlation structure for international normalized ratio and activated partial thromboplastin time test model. An autoregressive (AR-1) structure, specified by type=AR, allows the correlations to diminish over time. $Corr(y_{ij}, y_{ij}^{'}) = \rho|t_{ij}, t_{ij}^{'}|$, Where, t_{ij} and $t_{ij}^{'}$ are the observation times for y_{ij} and $y_{ij}^{'}$. The autoregressive structures express the intra-subject correlations in terms of a single parameter ρ. So, this makes the autoregressive correlation structure is the best correlation structure. Therefore, for the data set of this study linear mixed model with autoregressive correlation can be considered as best final model for both response variables.

Table 4.6: Parameter estimates and standard errors for the separate linear mixed effects models of the INR and APTT outcomes for the final model.

			INR				APTT
Parameter	Estms(s.e)	p-value	95%CI	parameter	Estims(s.e)	p-value	95%CI
β_{10}	1.750(0.248)	0.0000	(1.263,2.239)	β_{20}	34.464(4.223)	0.0001	(26.179,42.799)
β_{11}	0.047(0.016)	0.0042	(0.0151,0.0796)	β_{21}	0.698(0.271)	0.011	(0.160,1.229)
β_{12}	0.181(0.087)	0.0389	(0.009,0.352)	β_{22}	2.960(1.416)	0.0375	(0.172,5.745)
β_{13}	0.044(0.049)	0.3622	(-0.051, 0.140)	β_{23}	2.003(0.841)	0.0179	(0.347, 3.659)
β_{14}	-0.176(0.086)	0.0403	(-0.345,-0.008)	β_{24}	-2.004(1.391)	0.1498	(-4.748,0.730)
β_{15}	-0.095(0.048)	0.0498	(-0.190,-0.0001)	β_{25}	-1.644(0.834)	0.0494	(-3.290,-0.005)
β_{17}	0.163(0.081)	0.0451	(0.004,0.322)	β_{27}	2.810(1.323)	0.0343	(0.209,5.414)
β_{18}	0.250(0.095)	0.0090	(0.063,0.438)	β_{28}	3.083(1.545)	0.0470	(0.041,6.125)
β_{20}	0.027(0.013)	0.0340	(0.002,0.052)	β_{30}	0.472(0.211)	0.0264	(0.056,0.889)
β_{22}	-0.032(0.013)	0.0144	(-0.058,-0.007)	β_{32}	-0.459(0.216)	0.0354	(-0.884,-0.032)
var(a_{10i})			0.1172	var(a_{20i})			28.0265
var (b_{11i})			0.0033	var (b_{21i})			0.9265
$\sigma^2(residual)$			0.3325	$\sigma^2(residual)$			91.0339
$corr(a_{10i}, b_{11i})$			-0.622	$corr(a_{20i}, b_{21i})$			-0.403

The result in the above table showed that both model contains the same main effect variable but there is a single effect term different and also not significant at 0.05 levels of significance. Removing it doesn't improve model. So, it is retained as it is.

The estimated parameter for the intercept of INR is 1.750 with a standard error of 0.248 represents an estimate of the average level of INR during the first follow-up time and excluding all covariates in the model. There is a significant difference between patients who did not smoking cigar and who smoking a cigar with the parameter estimate of 0.181, which indicates that as a unit of patients who smoke cigar increase the international normalized ratio test of deep venous thrombosis disease increase by 0.181 times smoking cigar than that of the person who did not smoke a cigar (reference group) at baseline. The parameter estimate of the interaction for a smoking cigar with time is -0.032, which implies that the average rate of international normalized ratio test of DVT disease is increasing inversely related to smoking cigars. The significant gender effect and the value estimate is -0.176 indicates that as a unit of males patients of deep

venous thrombosis disease increase the international normalized ratio test of DVT disease decreases by 0.176. In addition, the interaction of sex by time is significant and the parameter estimate is 0.027, indicating that the rate of change of the average INR in males is nearly higher by 0.03 as compared to females patients with deep venous thrombosis. The average intercepts for the patients who had prolonged immobilize and who did not have prolonged immobilize of deep venous thrombosis patients are statistically different with the parameter estimate -0.095, which indicates deep venous thrombosis patients who had prolonged immobilize have lower INR measure than those who did not had prolonged immobilize (reference group) at baseline. Patients who have a family history of deep venous thrombosis disease increase the international normalized ratio test of deep venous thrombosis disease by 0.163 units. The average intercepts for the age of deep venous thrombosis patients are statistically different with the parameter estimate 0.250, which indicates that as a unit of age increase the international normalized ratio test of deep venous thrombosis disease increase by 0.250 times the age of the patients.

In a similar manner, the estimated parameter for the intercept of APTT is 34.464 with a standard error of 4.223 represents an estimate of the average level of APTT at time = 0 (during the first follow-up time). There is a significant difference between patients who did not smoking cigar and who smoking cigar with the parameter estimate 2.960, which indicates that as a unit of patients who smoking cigar increase the activated partial thromboplastin time test of DVT disease increase by 2.960 times than the patients who did not smoking cigar (reference group) at baseline. The parameter estimate of the interaction for smoking cigars and time is -0.459, which implies that the average rate of APTT increase is inversely related to patients who smoking cigars. Deep venous thrombosis patients who drink alcohol and who did not drink alcohol are statistically different with the parameter estimates of 2.003, which indicates that as a unit of drink alcohol increase the activated partial thromboplastin time test of deep venous thrombosis disease increase by 2.003 than that who did not drink alcohol (reference group) at baseline. sex is not significant at baseline; this means that there was no statistically significant difference in APTT measures at baseline for male and

female of deep venous thrombosis patients and the interaction of sex by time is significant and the parameter estimate is 0.472, indicating that the rate of change of the average of APTT in males is nearly higher by 0.5 as compared to females patients of deep venous thrombosis disease. The average intercepts for the patients who are prolonged immobilization and not prolonged immobilization of deep venous thrombosis patients are statistically different with the parameter estimate -1.644, which indicates that on average of deep venous thrombosis patients who prolonged immobilize have lower APTT measure than not prolonged immobilization (reference group) at baseline. A patient who has a family history of deep venous thrombosis, increases activated partial thromboplastin time test of deep venous thrombosis disease by 2.810 as family history of DVT increases units. The average intercepts for the age of deep venous thrombosis patients are statistically different with the parameter estimate 3.083, which indicates that on average the activated partial thromboplastin time test of deep venous thrombosis disease is increased by 3.083 as age increase a unit.

The intercept of the random effects for both INR and APTT indicates that there is variability between subjects at baseline. And the slope of random effects for both INR and APTT indicates that there is variability within subjects over time. The correlation -0.622 and -0.403 indicates, there is a negative correlation between intercept and slope of linear time effect for the random part for INR and APTT, respectively. In addition, from the random effects, the residual terms $\sigma_1^2 = 0.3325$ and $\sigma_2^2 = 91.034$, indicates that variation within the deep venous thrombosis patients in different time of INR and APTT measurements, respectively.

4.6 Joint analysis of international normalized ratio and activated partial thromboplastin time test

In many situations, joint modelling of the multivariate longitudinal profiles is needed or has additional advantages over the separate analyses of the different outcomes. First, the association structure can be of importance. A possible question might be how the association between outcomes evolves over time or how outcome-specific evolutions are related to each other [35]. In a second situation, the aim can be to improve the results of a discriminate analysis by using more than one longitudinally measured outcome. In another situation, interest may be in comparison of average trends for different outcomes.

A joint mixed effect model for the two measure (endpoints) of DVT measures INR and APTT was fitted with an Autoregressive variance-covariance structure. This model is the same as the separate model 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 that was selected as a fixed effect in the separate linear mixed model. The subject specific random intercepts and random slopes were fitted to account for within-subject correlations.

Table 4.7: Parameter estimates and standard errors for the joint linear mixed effects models of the INR and APTT outcomes for the final model.

			Separate for INR				Joint INR
Parameter	Estms(s.e)	p-value	95%CI	parameter	Estims(s.e)	p-value	95%CI
β_{10}	1.751(0.247)	0.0000	(1.265,2.237)	β_{10}	1.805(0.249)	0.0001	(1.314,2.295)
β_{11}	0.048(0.016)	0.0041	(0.015,0.079)	β_{11}	0.046(0.015)	0.0083	(0.012,0.079)
β_{12}	0.182(0.087)	0.0387	(0.009,0.352)	β_{12}	0.198(0.086)	0.0256	(0.024,0.371)
β_{13}	0.045(0.051)	0.3600	(-0.051, 0.140)	β_{13}	0.045(0.050)	0.4050	(-0.056,0.137)
β_{14}	-0.176(0.085)	0.0401	(-0.345,-0.008)	β_{14}	-0.174(0.083)	0.0752	(-0.325,0.016)
β_{15}	-0.096(0.047)	0.0486	(-0.190,-0.001)	β_{15}	-0.096(0.047)	0.0541	(-0.189,0.002)
β_{17}	0.163(0.082)	0.0447	(0.004,0.322)	β_{17}	0.160(0.080)	0.0667	(-0.010,0.311)
β_{18}	0.252(0.094)	0.0089	(0.063,0.438)	β_{18}	0.261(0.092)	0.0161	(0.068,0.521)
β_{20}	0.027(0.014)	0.0336	(0.002,0.052)	β_{20}	0.027(0.013)	0.0663	(-0.002,0.050)
β_{22}	-0.032(0.013)	0.0142	(-0.058,-0.007)	β_{22}	-0.039(0.012)	0.0045	(-0.065,-0.012)
var(a_{10})	0.110(0.033)	0.0095	(0.039,0.226)	var(a_{10})	0.119(0.034)	0.0002	(0.073,0.226)
var(b_{11})	0.002(0.001)	0.0067	(0.001,0.005)	var(b_{11})	0.003(0.001)	0.0000	(0.002,0.006)
σ_1^2	0.353(0.021)	0.0000	(0.315,0.397)	σ_1^2	0.323(0.018)	0.0000	(0.291,0.361)
			Separate for APTT				Joint APTT
Parameter	Estms(s.e)	p-value	95%CI	parameter	Estims(s.e)	p-value	95%CI
β_{20}	34.490(4.209)	0.0000	(26.208,42.771)	β_{20}	33.500(4.211)	0.0000	(25.216,41.785)
β_{21}	0.695(0.290)	0.0109	(0.162,1.228)	β_{21}	0.757(0.285)	0.0084	(0.196,1.318)
β_{22}	2.960(1.415)	0.0373	(0.175,5.745)	β_{22}	3.358(1.410)	0.0213	(0.503,6.212)
β_{23}	2.004(0.837)	0.0175	(0.354, 3.652)	β_{23}	2.023(0.834)	0.0156	(0.388, 3.671)
β_{24}	-2.009(1.390)	0.1494	(-4.745,0.727)	β_{24}	-2.548(1.424)	0.0748	(-5.352,0.257)
β_{25}	-1.647(0.830)	0.0483	(-3.282,-0.012)	β_{25}	-1.483(0.826)	0.0737	(-3.109,0.143)
β_{27}	2.811(1.321)	0.0341	(0.213,5.410)	β_{27}	3.405(1.320)	0.0122	(0.745,6.060)
β_{28}	3.083(1.543)	0.0467	(0.045,6.121)	β_{28}	2.923(1.517)	0.0649	(-0.182,6.027)
β_{30}	0.472(0.221)	0.0260	(0.057,0.888)	β_{30}	0.565(0.219)	0.0105	(0.134,0.995)
β_{32}	-0.458(0.216)	0.0349	(-0.883,-0.033)	β_{32}	-0.523(0.223)	0.0200	(-0.963,-0.083)
var(a_{20})	28.134(8.948)	0.0299	(7.396,68.941)	var(a_{20})	30.703(9.085)	0.0004	(18.518,60.583)
var(b_{21})	0.466(0.207)	0.0122	(0.229,1.421)	var(b_{21})	0.938(0.222)	0.0000	(0.620,1.584)
σ_2^2	91.034(5.7781)	0.0000	(86.630,109.44)	σ_2^2	88.959(4.857)	0.0000	(80.157,99.302)
comman parameter							
Corr random effect		ρ			0.7440	0.0052	(0.1462,0.8690)
$-2log-likelihood$			10765.7		10452.5		
AIC			10781.7		10476.5		

The above result shows that same of the parameters are significant at 5% level of significance except the sex, interaction term sex by time, prolonged immobilization and alcohol user for INR and and sex and prolonged immobilization for APTT. Thus, the variable age, smoking cigar are identified as positively associated with change in INR. Smoking cigar, alcohol user and family history are the variable which are identified as a positive risk factor for the change in APTT.

Similar to separate model interpretation previously the parameter estimate of joint model interpretation is also the same. The fixed-effect intercept coefficient $\beta_{10} = 1.805(S.E = 0.249)$ represents an estimate of the average international normalized ratio at baseline time and excluding all covariates in the model. Similarly, the fixed effect intercept coefficients $\beta_{20} = 33.500(S.E = 4.211)$ represents an estimate of the average of activated partial thromboplastin time test at baseline time and excluding all covariates in the model. There is a significant difference between patients who smoking a cigar and those who did not smoke a cigar at 5% level of significance for both INR and APTT with parameter estimate $\beta_{12} = 0.198(S.E = 0.086)$ and $\beta_{22} = 3.358(S.E = 1.410)$ for INR and APTT respectively, which indicates that patients who smoking cigar are more affected with deep venous thrombosis disease increasing both INR and APTT measure than patients who did not smoke a cigar (reference group) at baseline. The interaction of smoking cigars by time is significant for both INR and APTT responses. The parameter estimates are $\beta_{22} = -0.039(S.E = 0.012)$ and $\beta_{32} = -0.523(S.E = 0.223)$ for INR and APTT respectively, which implies that the average rate of INR and APTT test of deep venous thrombosis disease are increase inversely related to patients who smoking a cigarette. The average intercepts for the age of deep venous thrombosis patients are statistically significant with the parameter estimate 0.261, which indicates that as a unit of age increase the INR increase by 0.261, which means age was directly related with international normalized ratio test of deep venous thrombosis disease. But age for APTT is not significant at baseline; this means that age is not significantly related to APTT. Deep venous thrombosis patients who drink alcohol and who did not drink alcohol are statistically different with the parameter estimate $\beta_{23} = 2.023(S.E = 0.834)$,

which indicates that as a unit of patients who drink alcohol increases, the APTT test of deep venous thrombosis disease increase by 2.023 than patients who did not drink alcohol (reference group) at baseline. But alcohol users for INR are not significant at baseline; this means that there was no statistically significant difference in INR measures at baseline for patients who drink alcohol and do not drink alcohol. It is statistically different between patients who have a family history of deep venous thrombosis and who did not have a family history of deep venous thrombosis with the parameter estimates of $\beta_{27} = 3.405(S.E = 1.320)$, which indicates that patients who have a family history of DVT, increase activated partial thromboplastin time test of deep venous thrombosis disease than patients who did not have a family history of DVT (reference group) at baseline. But the family history of deep venous thrombosis for INR is not significant at baseline; this means that there was no statistically significant difference in INR measures at baseline for patients who have a family history and who did not have a family history. The interaction of sex by time is significant and the parameter estimate is $\beta_{30} = 0.565(S.E = 0.219)$, indicating that the rate of change of the average of APTT in males is nearly higher by 0.6 as compared to females patients of deep venous thrombosis.

4.6.1 Evolution of association and Association of evolution

Evolution of association and association of evolution used to answer the question how the evolution of the INR is associated with the evolution of the APTT is typically derived from the covariance matrix of the random effects and correlation matrix of random effects. Using SAS PROC MIXED for joint model the estimated variance covariance matrix and the estimate of correlation matrix for random effects and slopes of both the INR and the APTT are provided in table 6.3 (Appendix A) and tabel 4.8 respectively.

Table 4.8: Estimation of Correlation matrix for the final joint model.

		INR		APTT	
		Intercept	Slope	Intercept	Slope
INR	Intercept	1.0000	-0.7063	0.7163	-0.6452
	Slope	-0.7063	1.0000	-0.6897	0.7440
APTT	Intercept	0.7163	-0.6897	1.0000	-0.6445
	Slope	-0.6452	0.7440	-0.6445	1.0000

The estimated Variance covariance matrix is given in table 6.3 (Appendix A) depicts that, there is greater variability in APTT than INR. The random intercept in APTT has a variance of 30.7031 while INR has a variance of 0.1189. There is negative covariance between any intercept with random slope. Meaning that those who have larger intercept encounter a lesser random slope. However, the covariance between random slopes is smaller than the random intercept which is about 0.04. Eventually, there is smaller variability in the international normalized ratio random slope. This is because a change in time increases and a greater change was observed in APTT and INR shows a little change. Also, the covariance's for both APTT and INR are positive, which is indicative of a positive correlation, which is seen in the estimated correlation matrix. With a joint mixed-effects model is possible to investigate how the evolution of INR is associated with the evolution of APTT. The AOE can be determined by reading the correlation between the two slopes directly from the estimated correlation matrix (Table 4.8) or it is determined from the variance-covariance matrix of a joint model by using the equation as:

$$AOE = \frac{Cov(b_1, b_2)}{\sqrt{var(b_1) \times var(b_2)}} = \frac{0.0420}{\sqrt{0.0034}\sqrt{0.9376}} = 0.743995 \cong 0.7440$$

Here the AOE between the random slope for INR and the random slope for APTT is 0.7440. Thus, the larger positive value suggests a positive strong association between the evolution of international normalized ratio and activated partial thromboplastin time tests.

It is also possible to determine how the association between INR and APTT evolves over time, the evolution of the association (EOA). The EOA can be used for investigating the association between the INR and APTT how they evolved over time. The EOA can be determined, and then visualized, using the marginal correlation between INR and APTT.Thus, the EOA at t=0 is determined by using the equation below:

$$r_M(t) = EOA = \frac{0.7163}{\sqrt{1+0.323} \times \sqrt{1+88.959}} = 0.06565877 \cong 0.0656$$

From results we see that the weakest correlation is 0.0656 at baseline, and this association slightly increases over time. Also, it can be determined by plotting marginal correlation over time. The below marginal correlation plot depicts that, there is slowly increasing of evolution of the association between INR and APTT over time. Generally, there is evidence that time has reasonable effect on association of evolution of both outcomes.

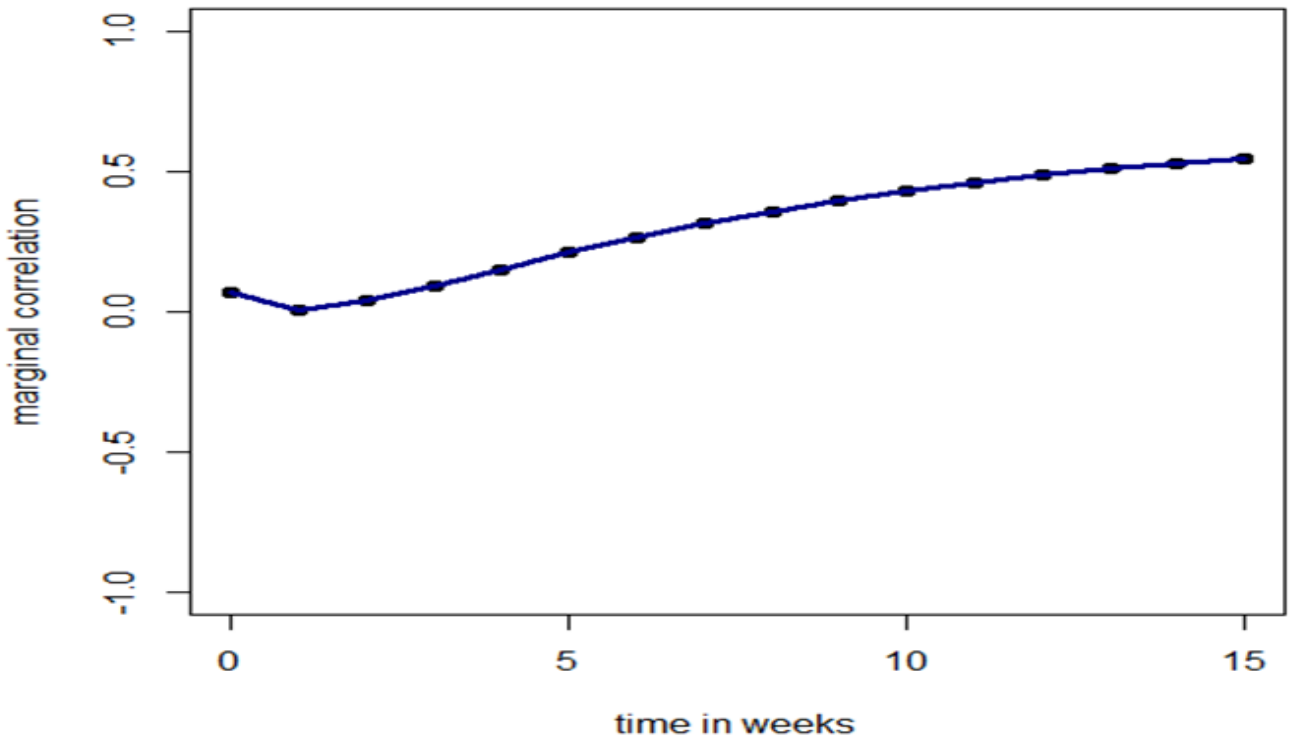

Figure 4.3: Evolution of the association for response variables over time

4.7 Comparison of separate and joint model

Most statistical models for repeated data are restricted to analysis of only single outcome variable. Those approaches are not flexible when the research question focuses on: the association structure of different outcomes, to test homogenous effect of a covariate across different outcomes and to model joint model for different outcomes. In order to answer such type of research question, [35] provided detailed explanation on joint modeling.

In this study, both separate and joint mixed effect models have been considered and parameter estimates for the separate and joint models are summarized in table 4.7. Technically, the separate models were fitted for two outcomes together, but assuming that $\rho = 0$, which entirely equivalent to fitting the two independent models separately as results were shown in above table 4.6. It also allows for a single likelihood for the model parameters enabling direct comparison with the correlated bivariate model fitted subsequently. Clearly, INR and APTT show a strong positive relationship as evidenced by the correlation of the random effects in joint mixed models. In addition, likelihood comparison shows a convincing improvement in model fit, when random effects are allowed to correlate.

Comparing the separate and joint models, although parameter estimates for both outcomes are nearly equivalent, small changes are observed in parameters of some covariate and insignificant cases. By comparing the results from the separate settings to the results from joint settings, there are several points of interest. 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 10765.7. The -2 log-likelihood value for the joint model is 10452.5. A likelihood ratio test indicated that the joint model provided a significantly better fit than the two separate models ($X^2 = 5030.21$, df =8, p-value < 0.0001). With regards to Akaike's information criterion (AIC), the joint model (AIC = 10476.5) is also indicated as a better fit than the separate model (AIC = 10781.7). Comparing the separate and joint models all parameter estimates for both continuous

response variables are almost the same, except for small changes. Notice how the joint model two measure of deep venous thrombosis i.e. INR and APTT seem to decrease the variability in the random effects, this can be seen in table 4.7. Taking into account the standard errors for the variance and covariance estimates, the joint model in general allowed for more accurate prediction (small errors) of the variability in the random effect.

Comparing the fixed effects for the separate and joint mixed models have some important things in the case for the two measure of DVT patients. Comparing the covariates between two types of models will yield further information of interest. Both the separate and joint models found a significant relationship between baseline age with INR. The age was positively associated with both INR models ($\beta_{18} = 0.252$ compared to 0.261), however, the SE (0.094 compared to 0.092) is smaller for the joint model, hence the 95% CI is also tighter for the joint model. Similarly, both the separate and joint models found a significant relationship between patients who smoking cigars and the measure of DVT. Patients who Smoke cigars were positively associated with INR ($\beta_{12} = 0.182$ compared to 0.198) and the 95%CI (0.009, 0.352 vs 0.024, 0.371) was also equally tighter for both models and Who smoke cigar was also positively associated with APTT ($\beta_{22} = 2.960$ compared to 3.358) and the 95%CI (0.175, 5.745 vs 0.503, 6.212) was also equally fitted for both models. Patients who drink alcohol were positively associated with APTT ($\beta_{23} = 2.004$ compared to 2.023), hence, the 95%CI (0.354, 3.652 vs 0.388, 3.671) was best fitted for joint model relatively to patients who did not use alcohols. Family history who have a history of DVT found to be positively associated with both separate and joint models with APTT ($\beta_{27} = 2.811$ as compare to 3.405), hence, the 95%CI (0.213, 5.410 vs 0.745, 6.060) was also tighter for joint model relatively to patients who did not have a family history of deep venous thrombosis disease.

4.8 Assumption of model diagnostic checking

The most useful methods that are used for diagnostic checking of the model are plots of the residuals vs fitted values, QQ-Plot, and the probability plot of estimated random effects. The primary quantities used in diagnostic checking were the within-group residuals which are defined as the difference between the observed response and the within-group fitted values. Different diagnostic checking plots for the final separate mixed linear models of international normalized ratio and activated partial thromboplastin time test measures of deep venous thrombosis are presented in **Appendix B**. According to figure 6.1 and figure 6.2, the plot of fitted versus standardized residuals for INR and APTT respectively, even if there are some outliers, it was indicated that the variability of the errors in both INR and APTT were almost nearly constant. That means the errors did not far deviate from each other. The horizontal line passes to the center of the residual vs fitted point. The point above and below horizontal lines are almost constant. The plot shows that for both INR and APTT normality assumptions of the error term are satisfied. Similarly, according to the probability plots those which shown in figure 6.3 and figure 6.4, even if the points were compacted at the two end tails for both outcomes INR and APTT, the normality assumption was supported through the upward nearly straight line of normal plots. Similarly, based on the normal probability plots of random effects with the subject (ID) specific random intercepts and random slopes those are shown in figure 6.5 and figure 6.6, even if it seems a slight deviation of normality at the bottom tail on the random slope (Time) for INR that is not that much worse deviation. Hence, there is no problem with normality assumptions of both random intercepts and random slopes for both INR and APTT models and the normality assumption of model diagnostic checking are almost fulfilled.

4.9 Discussion

This study conducted to investigate the relation between international normalized ratio and activated partial thromboplastin time test of deep venous thrombosis patients. There were two methods of model were considered to fitting the 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.

Before fitting the linear mixed model and joint model for the two response INR and APTT test of deep venous thrombosis patients, 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 INR and APTT within and between individuals. The exploratory analysis result for mean structure also suggested that on average, both INR and APTT measures slightly increesing in a linear pattern over time.

In this paper, the separate linear mixed analysis of the international normalized ratio and activated partial thromboplastin time test, fixed and random effect components were selected to include in the model. In both separate model, linear mixed model with intercept and slope were selected as best model with small values of AIC. The variables included in the model are determined using backward variable selection methods. Then, nine covariates and some interaction term by time were considered in the model, of these, living condition and place of residence and place of residence with its interaction term by time were not found to be statistically significant for INR. Similarly, living condition and place of residence and place of residence with its interaction term by time were not found to be statistically significant for APTT models. Those insignificant terms could be removed from the final model. Thus, the final model fitted

with both continuous and categorical covariates such as: time, age, sex, family history, immobilize, alcohol users, smoking cigar and interaction of time with sex and smoking cigar. The correlation structure used in this model were 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, autoregressive structure of order one which had small AIC value of 2241.743 for INR and for 8526.072 APTT was the most appropriate covariance structure for both INR and APTT of the model.

In this study, 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. The aim of the joint model able to answer additional important questions about the association in the evolutions of the responses as well as the evolution of the associations [35]. Two aspects of the relation were investigated: the association between the evolutions and the evolution of the association. Results of the joint model suggested a very strong association between the evolutions and a slowly increase evolution of the association between INR and APTT. Therefore, the joint mixed effect model was better fit than two separate random effect models. This finding is consistent with the previous literatures that was studied by Thibauta [65] on bivariate mixed effect model or first-order autoregressive process and independent measurement error for both markers of CD4 and HIVRNA in HIV patients with ($p - value < 0.0001$). This finding was also similar with the study of Ferrari and Cribari-Neto [66] who studied on application of joint models for resistance and prevalence a strong correlation between percentage resistant and prevalence and that both increase with time. The correlation is estimated to be 0.95, with 95% confidence interval [0.414, 0.997] showing that the correlation is significant. The finding indicated that the correlation between two response was positively significant association of evolution.

The covariates considered in this thesis were time, age, sex, family history, immobilize, alcohol users, smoking cigar, living condition and place of residence . Then, after removing insignificant covariate the final model fitted with time, age, sex, family history, immobilize, alcohol users, smoking cigar and interaction of sex and smoking cigar with time in both separate and joint analysis. Descusion of all the covariate with related studies were.

The result of this study revealed that age of patients were significantly associated with both INR and APTT, which indicates that age was positively affects patients with deep venous thrombosis. This finding was supported with study done in jimma university [26] using a prospective cohort study shows that, the patients whose age greater than 35 were significantly associated with recurrence of DVT. A similar result was found in Addis Ababa [25]. Similarly, our study was supported with the study done at Sudan University of Science and Technology [40] from the study age was a significant difference between the mean age of DVT patients and healthy controls. Also Yang et al. [78] agree with current study which reported that increasing age was a risk factor for postoperative DVT after spine surgery. This finding was supported with study done by Eman Saleh et. [79] who stated that age were associated significantly with prevalence of deep vein thrombosis.

The findings of this study implied that alcohol consumption were linked with increasing DVT by increasing APTT two times. The finding was supported by studies done in Jimma University by Abera et al. [51], the study found that being unspecified amount of regular alcohol consumption were increased in the development of DVT recurrence, approximately two times. Similar finding were found from San Diego, California, USA [48], where high maximum alcohol use is the behavior most linked to increased risk of DVT recurrence. It is evidenced that alcohol promoting oxidative stress and mitochondrial toxicity are associated with tissue injury and cell death. This triggers coagulation activation, via exposure at the cell surface of phosphatidyl serine [49].

This study also showed that among gender male patients of deep venous thrombosis had a negative effect on the INR, which indicates that male were major risk factor of deep venous thrombosis. This finding in line with study done at Minnesota by Heit et al. [45] and another one prospective cohort study done by Moreno et al. [47] did not agree with the finding, the result found that being male were significant risk predictor for increasing DVT occurrence and another finding Research and Practice in Thrombosis and Haemostasis using prospective cohort study at hematology department of Switzerland university by Nagler et al. [46], not agree with the this study, which showed that male sex increased risk of DVT occurrence. This difference is most probably reflected due to different sample size, method of data analysis and effects of covariates used in the study.

Similarly, the finding of this study show that prolonged immobilization had significant effect on both INR and APTT, which implies that patient who had a prolonged immobilization was negatively associated with the risk of deep venous thrombosis . This finding is not agree with the a study done at Dakar (Senegal) by Awa.O et al. [53]. Similarly, the study not agree with study done by Engbers et al. [54], who stated that the highest risk of thrombosis was found for immobilisation during hospitalisation (OR 48.7; CI95 6.6-361.0), and the risk of thrombosis out-of hospital was 15-fold increased within the two weeks after hospital discharge, and the risk remained increased for 3 months after hospital discharge. In addition, the finding of this study not supported with the study conducted in Addis Ababa by Haile et al. [25], who found that prolonged immobilization was considered as risk factor of deep venous thrombosis. A similar study done by partsch [77] who says that immobilization had significant association with recurrence of DVT. This difference is most probably reflected due to different sample size, method of data analysis and effects of covariates used in the study.

In this study, patients with family history of DVT had increased both INR and APTT than patients with deep venous thrombosis who were not had family history of DVT. The study also identifies that patients who had family history of DVT were associated with increasing INR and APTT than patients who did not had family history of

DVT.This implies that patients with family history of deep venous thrombosis easily affected with deep venous thrombosis. This is consistent with the findings of another study done in University of Khartoum by Gader et al. [5], who stated that family history of DVT is the higher risk factor of deep venous thrombosis. This study is also consistent with the study conducted by Ms Bezemeret al. [52], who stated that positive family history of DVT increased the risk of deep venous thrombosis more than 2-fold (odds ratio [95% confidence interval], 2.2 [1.9-2.6]) and up to 4-fold (3.9 [2.7-5.7]).

Furthermore, our study found that patients who smoking cigarette had increased INR and APTT than patients with deep venous thrombosis who were not smoking cigarette. Which implies that patients who smoke cigarette were more affected by deep venous thrombosis. This finding was similar with the study done by Cheng YJ et al. [50], the Studies had shown that cigarette smoking is associated with an increased risk for DVT. Similarly, a study conducted in Jimma University by Mulatu et al. [51], who found that Current Smoker patients was linked with increases in the development of DVT recurrence.

5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

In this study two methods of model were considered for fitting two response variables measured longitudinally. But, the main aim of this thesis was to develop joint mixed effects model for two measures of deep venous thrombosis (international normalized ratio and activated partial thromboplastin time test) as outcome variables. The joint model is the best model as 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 and BIC value which indicates that it fits the data better than the separate model. Hence, the joint model is not only the best model but also it gives a better fit to the data well. In addition, joint analysis output for association of evolution shows that, the two responses are strongly positively correlated for evolution and the correlation was statistically significant. Thus, the joint mixed effect model was preferred because the joint mixed effects model is more flexible in allowing separate fixed and random effects for each response INR and APTT through appropriate choice of potential risk factors (covariates) or fixed effect and random effects. Autoregressive of order one was preferred to fit both separate and joint mixed effect model because it has the smallest values of AIC and BIC in both model as compared to others covariance correlation structures.

In separate analysis of linear mixed model for international normalized ratio and activated partial thromboplastin time ,the covariate like time, smoking, sex, immobilization, family history of DVT, age and interaction of sex and smoking with time was significantly associated with INR and time, smoking, alcohol users, immobilization, family history of DVT, age and interaction of sex and smoking with time was statically associated with APTT. Out of those covariates time, smoking, family history of DVT and age were found to be positively associated with INR, but sex and immobilization was negatively associated with both INR and APTT. Similarly, time, smoking, alcohol user, family history of DVT and age were found to be positively associated with APTT.

Hence, those covariate were associated with the occurrence of deep venous thrombosis.

In joint analysis of both responses, the covariate time, smoking and age was positively associated with INR, but the interaction of smoking with time was found to be negatively associated with INR. time, smoking, alcohol users and family history of DVT were found to be positively associated with acivated partial thromboplastin time test of deep venous thrombosis. Out of those covariates smoking were found to be positively associated with both outcomes in both separate and joint mixed model. Finally, it is concluded that, joint modeling of longitudinal bivariate responses is necessary to explore the association between paired response variables of INR and APTT over time. A usual problem with the joint modeling is failing to convergence because of large number of association parameter to estimate.

5.2 Recommendation

- The joint model is address the same questions as separate model with more accuracy by addressing additional questions that may be of great interest to the researcher, such as the association of evolution and the evolution of association of the responses. So, fitting joint model is recommended for researchers to any types of multivariate response variables for evaluating correlation and rate of changes over time.

- Physicians should be encourage to carry out and stratify patients based on target long-term anticoagulant therapy to those patients at highest risk for DVT.

- Quit/Stopping drinking alcohol habits is recommended to individual who drink alcohol in order to reduce the cause of deep venous thrombosis.

- Based on the results from the present study, increasing physical activity is recommended to individuals with long periods of immobilization to decrease the risk of developing deep venous thrombosis.

- Patients who smoke cigarette habits more affected by deep venous thrombosis. Therefore, individuals should work against such bad habits and avoid places and

situations that associate with smoking. Hang out with non smokers or go places that don't allow smoking (like sport, museums, shops or libraries).

- Health workers should be record all the necessary variables during follow up time to see the change of the disease within and between subjects overtime.

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. Another some potential risk factors or covariates which may have high influence on deep venous thrombosis which were mentioned in some literatures may not measured in the data . In addition to this, there is no related published paper on this area in our country to the best knowledge of researcher, by using joint modeling to compare and contrast the results of this finding.

References

[1] Phillippe, H.M., 2017. Overview of venous thromboembolism. The American journal of managed care, 23(20 Suppl), pp.S376-S382.

[2] Kruger, P.C., Eikelboom, J.W., Douketis, J.D. and Hankey, G.J., 2019. Deep vein thrombosis: update on diagnosis and management. Medical Journal of Australia, 210(11), pp.516-524.

[3] Goldhaber, S.Z. and Fanikos, J., 2004. Prevention of deep vein thrombosis and pulmonary embolism. Circulation, 110(16), pp.e445-e447.

[4] Mouton, J.P., Blockman, M., SekaggyaWiltshire, C., Semakula, J., Waitt, C., Pirmohamed, M. and Cohen, K., 2021. Improving anticoagulation in subSaharan Africa: What are the challenges and how can we overcome them?. British Journal of Clinical Pharmacology.

[5] Gader, A.A., Haggaz, A.E.D. and Adam, I., 2009. Epidemiology of deep venous thrombosis during pregnancy and puerperium in Sudanese women. Vascular health and risk management, 5, p.85.

[6] Al-Hameed, F., Al-Dorzi, H.M., Shamy, A., Qadi, A., Bakhsh, E., Aboelnazar, E., Abdelaal, M., Al Khuwaitir, T., Al-Moamary, M.S., Al-Hajjaj, M.S. and Brozek, J., 2015. The Saudi clinical practice guideline for the diagnosis of the first deep venous thrombosis of the lower extremity. Annals of thoracic medicine, 10(1), p.3.

[7] Yamaki, T., Nozaki, M., Sakurai, H., Takeuchi, M., Soejima, K. and Kono, T., 2009. Presence of lower limb deep vein thrombosis and prognosis in patients with symptomatic pulmonary embolism: preliminary report. European journal of vascular and endovascular surgery, 37(2), pp.225-231.

[8] Goldhaber, S.Z. and Bounameaux, H., 2012. Pulmonary embolism and deep vein thrombosis. The Lancet, 379(9828), pp.1835-1846.

[9] Piazza, G., Seddighzadeh, A. and Goldhaber, S.Z., 2008. Deep-vein thrombosis in the elderly. Clinical and Applied Thrombosis/Hemostasis, 14(4), pp.393-398.

[10] Ho, W.K., Hankey, G.J. and Eikelboom, J.W., 2008. The incidence of venous thromboembolism: a prospective, communitybased study in Perth, Western Australia. Medical journal of Australia, 189(3), pp.144-147.

[11] Lee, A.D., Stephen, E., Agarwal, S. and Premkumar, P., 2009. Venous thromboembolism in India. European Journal of Vascular and Endovascular Surgery, 37(4), pp.482-485 .

[12] Hyers, T.M., Agnelli, G., Hull, R.D., Morris, T.A., Samama, M., Tapson, V. and Weg, J.G., 2001. Antithrombotic therapy for venous thromboembolic disease. Chest, 119(1), pp.176S-193S.

[13] Oger, E. and EPI-GETBO study group, 2000. Incidence of venous thromboembolism: a community-based study in Western France. Thrombosis and haemostasis, 83(05), pp.657-660.

[14] Yetkin, U., Karabay, . and nol, H., 2004. Effects of oral anticoagulation with various INR levels in deep vein thrombosis cases. Current controlled trials in cardiovascular medicine, 5(1), pp.1-4.

[15] Kearon, C., 2003. Duration of venous thromboembolism prophylaxis after surgery. Chest, 124(6), pp.386S-392S.

[16] Barritt, D.W. and Jordan, S.C., 1960. Anticoagulant drugs in the treatment of pulmonary embolism: a controlled trial. The Lancet, 275(7138), pp.1309-1312.

[17] Hirsh, J., Bauer, K.A., Donati, M.B., Gould, M., Samama, M.M. and Weitz, J.I., 2008. Parenteral anticoagulants: American College of Chest Physicians evidence-based clinical practice guidelines. Chest, 133(6), pp.141S-159S.

[18] Olubanwo, O.O., 2010. The profile of HIV/aids patients admitted with Deep Venous Thrombosis (DVT) at Nelson Mandela Hospital in Mthatha, South Africa (Doctoral dissertation, Stellenbosch: Stellenbosch University).

[19] Luciani, A., Clement, O., Halimi, P., Goudot, D., Portier, F., Bassot, V., Luciani, J.A., Avan, P., Frija, G. and Bonfils, P., 2001. Catheter-related upper extremity

deep venous thrombosis in cancer patients: a prospective study based on Doppler US. Radiology, 220(3), pp.655-660.

[20] Adeleye, A.O. and Ogun, G.O., 2016. Clinically diagnosed postoperative venous thromboembolism in a neurosurgery practice in Nigeria. World neurosurgery, 89, pp.259-265.

[21] Muleledhu, A.L., Galukande, M., Makobore, P., Mwambu, T., Ameda, F. and Kiguli-Malwadde, E., 2013. Deep venous thrombosis after major abdominal surgery in a Ugandan hospital: a prospective study. International journal of emergency medicine, 6(1), pp.1-5.

[22] Haggaz, A.R., Mirghani, O.A. and Adam, I., 2003. Venous thromboembolism in pregnancy and the puerperium in Sudanese women. International journal of gynaecology and obstetrics, 83(3), pp.309-310.

[23] Semakula, J.R., Mouton, J.P., Jorgensen, A., Hutchinson, C., Allie, S., Semakula, L., French, N., Lamorde, M., Toh, C.H., Blockman, M. and Sekaggya-Wiltshire, C., 2020. A cross-sectional evaluation of five warfarin anticoagulation services in Uganda and South Africa. PloS one, 15(1), p.e0227458.

[24] Karuri, S.W., Nyamu, D.G., Opanga, S.A. and Menge, T.B., 2019. Factors associated with time in therapeutic range among patients on oral anticoagulation therapy in a tertiary teaching and referral hospital in Kenya. East and Central African Journal of Pharmaceutical Sciences, 22(3), pp.85-95.

[25] Haile, L., Hawaz, Y. and Assefa, G., 2013. Risk factors of deep venous thrombosis in duplex and colour Doppler ultrasound at Tikur Anbessa specialized hospital, Addis Ababa, Ethiopia. East and Central African Journal of Surgery, 18(2), pp.61-69.

[26] Mulatu, A., Melaku, T. and Chelkeba, L., 2020. Deep Venous Thrombosis Recurrence and Its Predictors at Selected Tertiary Hospitals in Ethiopia: A Prospective Cohort Study. Clinical and Applied Thrombosis/Hemostasis, 26, p.1076029620941077.

[27] Food, M., 2014. Administration HC, Ethiopia CAo. Standard Treatment Guidelines for General Hospitals.

[28] Kearon, C., Kahn, S.R., Agnelli, G., Goldhaber, S., Raskob, G.E. and Comerota, A.J., 2008. Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians evidence-based clinical practice guidelines. Chest, 133(6), pp.454S-545S.

[29] Nicolaides, A.N., Fareed, J., Kakkar, A.K. and Breddin, H.K., 2006. Prevention and treatment of venous thromboembolism International Consensus Statement (guidelines according to scientific evidence). International Angiology, 25(2), p.101.

[30] Laird, N.M. and Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics, pp.963-974.

[31] McCulloch, C.E. and Neuhaus, J.M., 2005. Generalized linear mixed models. Encyclopedia of biostatistics, 4.

[32] Der, G. and Everitt, B.S., 2005. Statistical analysis of medical data using SAS. CRC Press.

[33] Williams, S.M., 2001. Modelling the association between increases in height and adiposity and changes in blood pressure between ages 7 and 18 years. In Proceedings of the 14th International Workshop on Statistical Modelling, Klein B, Korsholm L (eds). Odense, Denmark (pp. 521-524).

[34] Molenberghs, G. and Verbeke, G., 2001. A review on linear mixed models for longitudinal data, possibly subject to dropout. Statistical Modelling, 1(4), pp.235-269.

[35] Fieuws, S. and Verbeke, G., 2004. Joint modelling of multivariate longitudinal profiles: pitfalls of the randomeffects approach. Statistics in medicine, 23(20), pp.3093-3104.

[36] Geerts, W.H., Heit, J.A., Clagett, G.P., Pineo, G.F., Colwell, C.W., Anderson, F.A. and Wheeler, H.B., 2001. Prevention of venous thromboembolism. Chest, 119(1), pp.132S-175S.

[37] Abildstrom, S.Z., Rasmussen, S., Rosen, M. and Madsen, M., 2003. Trends in incidence and case fatality rates of acute myocardial infarction in Denmark and Sweden. Heart, 89(5), pp.507-511.

[38] Sotunmbi, P.T., Idowu, A.T., Akang, E.E. and Aken'Ova, Y.A., 2006. Prevalence of venous thromboembolism at post-mortem in an African population: a cause for concern. African journal of medicine and medical sciences, 35(3), pp.345-348.

[39] van Ommen, C.H., Heijboer, H., Bller, H.R., Hirasing, R.A., Heijmans, H.S. and Peters, M., 2001. Venous thromboembolism in childhood: a prospective two-year registry in The Netherlands. The Journal of pediatrics, 139(5), pp.676-681.

[40] Nadir Ahmed Ibrahim, Fathelrahman M Hassan, Mahmoud Mohamed Elgari, Sana Eltahir Abdalla.Risk factors for deep vein thrombosis of lower extremities in Sudanese women.Sudan University of Science and Technology, 2018 ;20(14): Pp 157164

[41] Liu, X., Phatak, H., Dillon, R. and Mitchell, S.A., 2013. Epidemiology and mortality of venous thromboembolism across patient populations: a systematic literature review. Value in Health, 16(3), p.A272.

[42] Fall, A.O., Proulle, V., Sail, A., Mbaye, A., Ba, P.S., Diao, M., Seck, M., Gadji, M., Gning, S.B., Diop, S. and Diye, T.N., 2014. Risk factors for thrombosis in an African population. Clinical Medicine Insights: Blood Disorders, 7, pp.CMBD-S13401.

[43] Viens, L.J., Henley, S.J., Watson, M., Markowitz, L.E., Thomas, C.C., D. Thompson, T., Razzaghi, H. and Saraiya, M., 2016. Human papillomavirusassociated cancersUnited States, 20082012. Morbidity and Mortality Weekly Report, 65(26), pp.661-666.

[44] Marissa E (2016): Conventional and age-specific risk factors for venous thrombosis in older people : the AT-AGE . s Leiden University. http://hdl.handle.net/1887/37409. accessed in June, 2016.Matsumoto M. H., Rodrigues L. C., Batalini L. G., Fonteles T. A. Bortoletto A. Oda T., Fuji T., Kato

Y., Fujita S. Kanemitsu N. (2000): Deep venous thrombosis after posterior spinal-surgery. Spine (Phila Pa 1976) 25, 296267

[45] Heit JA, Lahr BD, Petterson TM, Bailey KR, Ashrani AA, Melton LJ. Heparin and warfarin anticoagulation intensity as predictors of recurrence after deep vein thrombosis or pulmonary embolism: a population-based cohort study. Blood. 2011:blood-2011-05-357343.

[46] Nagler M, ten Cate H, Prins MH, ten CateHoek AJ. Risk factors for recurrence in deep vein thrombosis patients following a tailored anticoagulant treatment incorporating residual vein obstruction. Research and Practice in Thrombosis and Haemostasis. 2018;2(2):299-309.

[47] Moreno AF, Navarro MG, Snchez JO, Daz RM, Cerezo EM, de Ancos Aracil CL, et al. A risk score for prediction of recurrence in patients with unprovoked venous thromboembolism (DAMOVES). European journal of internal medicine. 2016;29:59-64.

[48] Golomb, B.A., Chan, V.T., Denenberg, J.O., Koperski, S. and Criqui, M.H., 2014. Risk marker associations with venous thrombotic events: a cross-sectional analysis. BMJ open, 4(3).

[49] Hansen-Krone IJ, Brkkan SK, Enga KF, Wilsgaard T, Hansen J-B. Alcohol consumption, types of alcoholic beverages and risk of venous thromboembolism: the Troms Study: Universitetet i Troms; 2011.

[50] Cheng YJ, Liu ZH, Yao FJ, et al. Current and former smoking and risk for venous thromboembolism: A systematic review and meta-analysis. PLoS Med. 2013;10(9):e1001515. doi:10.1371/journal.pmed.1001515

[51] Mulatu, A., 2018. Treatment outcome of deep veinous thrombosis and its predictors among hospitalized patients at selected tertiary hospitals in Ethiopia: a prospective cohort study (Doctoral dissertation).

[52] Irene D. Bezemer, MSc;Felix J. M. van der Meer, MD, PhD; Jeroen C. J. Eikenboom, MD, PhD.The Value of Family History as a Risk Indicator for Venous Thrombosis.Arch Intern Med. 2009;169(6):610-615

[53] Awa.O.T.fall,2014. Valerieproulle clinical medicine insights; blood disorders Risk factors for thrombosis in African population , Senegal Dakar.

[54] M.J. Engbers, J.W. Blom,M. Cushman,F.R. Rosendaal A. van,Hylckama Vlieg.The contribution of immobility risk factors to the incidence of venous thrombosis in an older population.J Thromb Haemost. 2014; 12: 290-6

[55] Eman S M Shahin,Samia Elgazzar,Mona Hassan(2017).Incidence and Prevalence of Deep Venous Thrombosis in Surgical Patients Mixed Research Design:(Sep. -Oct .2017), PP 21-26

[56] Yende Zuma, N., 2009. Modelling CD4+ count over time in HIV positive patients initiated on HAART in South Africa using linear mixed models (Doctoral dissertation)

[57] Fieuws, S. and Verbeke, G., 2004. Joint modelling of multivariate longitudinal profiles: pitfalls of the randomeffects approach. Statistics in medicine, 23(20), pp.3093-3104.

[58] Fieuws, S. and Verbeke, G., 2006. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. Biometrics, 62(2), pp.424-431.

[59] Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. Journal of the American statistical Association, 88(421), pp.9-25.

[60] Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D. and Geys, H., 2000. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics, 1(1), pp.49-67.

[61] Burzykowski, T., Molenberghs, G., Buyse, M., Geys, H. and Renard, D., 2001. Validation of surrogate end points in multiple randomized clinical trials with fail-

ure time end points. Journal of the Royal Statistical Society: Series C (Applied Statistics), 50(4), pp.405-422.

[62] Gueorguieva, R., 2001. A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. Statistical Modelling, 1(3), pp.177-193.

[63] Chakraborty, H., Helms, R.W., Sen, P.K. and Cohen, M.S., 2003. Estimating correlation by using a general linear mixed model: evaluation of the relationship between the concentration of HIV1 RNA in blood and semen. Statistics in medicine, 22(9), pp.1457-1464.

[64] Thibaut, R., Jacqmin-Gadda, H., Chne, G. and Commenges, D., 2002, July. 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 Modelling (pp. 521-524).

[65] Thibaut, R., Jacqmin-Gadda, H., Chne, G., Leport, C. and Commenges, D., 2002. Bivariate linear mixed models using SAS proc MIXED. Computer methods and programs in biomedicine, 69(3), pp.249-256.

[66] Ferrari, S.L.P. and Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions,J.App. Stat., 31: 799-815.

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

[68] Negash, Y., Kassahun, W., Gurmessa, A. and Birlie, B., 2016. Joint modeling of longitudinal systolic and diastolic blood pressure measurements of hypertensive patients receiving treatment. Electronic Journal of Applied Statistical Analysis, 9(2), pp.308-325.

[69] Fissuh, Y.H. and Muletav, G., 2015. A joint model for a longitudinal pulse rate and respiratory rate of congestive heart failure patients: at Ayder Referral Hospital of Mekelle University, Tigray, Ethiopia. J Biom Biostat, 6(5), p.1.

[70] Verbeke, G., 2000. Linear mixed models for longitudinal data. In Linear mixed models in practice (pp. 63-153). Springer, New York, NY.

[71] Verbeke, G., Lesaffre, E. and Brant, L.J., 1998. The detection of residual serial correlation in linear mixed models. Statistics in medicine, 17(12), pp.1391-1402.

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

[73] Brady, T., Kathleen, B. and Andrzej T.(2008) with contributions from Brenda W. Gillespie.Linear mixed models a practical guide using statistical software. Pp.348.

[74] Pinheiro, J.C. and Bates, D.M., 2000. Mixed-effects models in S and S-Plus.,(Springer-Verlag: New York). Mixed-effects models in S and S-Plus. Springer Verlag, New York.

[75] Sakamoto, Y., Ishiguro, M. and Kitagawa, G., 1986. Akaike information criterion statistics. Dordrecht, The Netherlands: D. Reidel, 81(10.5555), p.26853.

[76] Little, R. J. A. (2006). Modeling the drop out mechanism in repeated measures studies. Journal of the American Statistical Association, 90(431):1112 1121.

[77] Partsch H. Ambulation and compression after deep vein thrombosis: dispelling myths. Semin Vasc Surg2005; 18(3): 148-52.

[78] Yang S , LiuH and Sun Y(2015) :Prevalence and risk factors of deep vein thrombosis in patients after spine surgery: a retrospective case-cohort study : a retrospective case-cohort study. Sci. Rep.5, Dec; 94(48): e2205. 11834; doi: 10.1038/srep11834

[79] Eman Saleh M Shahin.(2017). Incidence and Prevalence of Deep Venous Thrombosis in Surgical Patients: Mixed Research Design. IOSR Journal of Nursing and Health Science (IOSR-JNHS) , vol. 6, no. 5, pp. 2126.

6 Appendix

6.1 Appendix A

Table 6.1: Parameter estimates of linear model for INR

Effect	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	1.5741447	0.1186139	13.271	$< 2e-16 ***$
LivingcoAlone	-0.0115356	0.0452443	-0.255	0.7988
Age	-0.0017944	0.0017314	-1.036	0.3003
Time	0.0148809	0.0129583	1.148	0.02511 *
Sexmale	0.1553022	0.0809772	1.918	0.0354 *
AlcoholYes	0.0473151	0.0416456	1.136	0.2561
FamilyHYes	-0.1722236	0.0742400	-2.320	0.0205 *
ImmobilizeYes	0.0826543	0.0409880	2.017	0.0440 *
SmokingYes	0.1405267	0.0827030	1.699	0.0496*
placeRural	0.0812740	0.0688878	1.180	0.2383
Time:Sexmale	0.0241490	0.0116952	2.065	0.0392 *
Time:FamilyHYes	0.0192337	0.0104569	1.839	0.0661
Time:SmokingYes	0.0275250	0.0120104	2.292	0.0221 *
Time:placeRural	-0.0003027	0.0099145	-0.031	0.9757

Table 6.2: Parameter estimates of linear model for APTT.

Effect	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	37.037402	1.995959	18.556	$< 2e-16$ ***
LivingcoAlone	-0.411119	0.761342	-0.540	0.58931
Age	-0.036148	0.029135	-1.241	0.21499
Time	0.200520	0.218055	0.920	0.35799
Sexmale	-2.292662	1.362633	-1.683	0.09275 .
AlcoholYes	1.873633	0.700785	2.674	0.00761 **
FamilyHYes	-2.583312	1.249264	-2.068	0.03889 *
ImmobilizeYes	1.415567	0.689719	2.052	0.04037 *
SmokingYes	2.779343	1.391674	1.997	0.04606 *
placeRural	0.573699	1.159200	0.495	0.62076
Time:Sexmale	0.482060	0.196799	2.450	0.01446 *
Time:FamilyHYes	0.158171	0.175962	0.899	0.36891
Time:SmokingYes	0.388140	0.202104	1.920	0.05505 .
Time:placeRural	0.007199	0.166835	0.043	0.96559

Table 6.3: Estimating variance-Covariance matrix for the final joint model.

		INR		APTT	
		Intercept	Slope	Intercept	Slope
INR	Intercept	0.1189	-0.0142	1.3687	-0.4788
	Slope	-0.0142	0.0034	-0.4892	0.0420
APTT	Intercept	1.3687	-0.4892	30.7031	-3.4581
	Slope	-0.4788	0.0420	-3.4581	0.9376

6.2 Appendix B

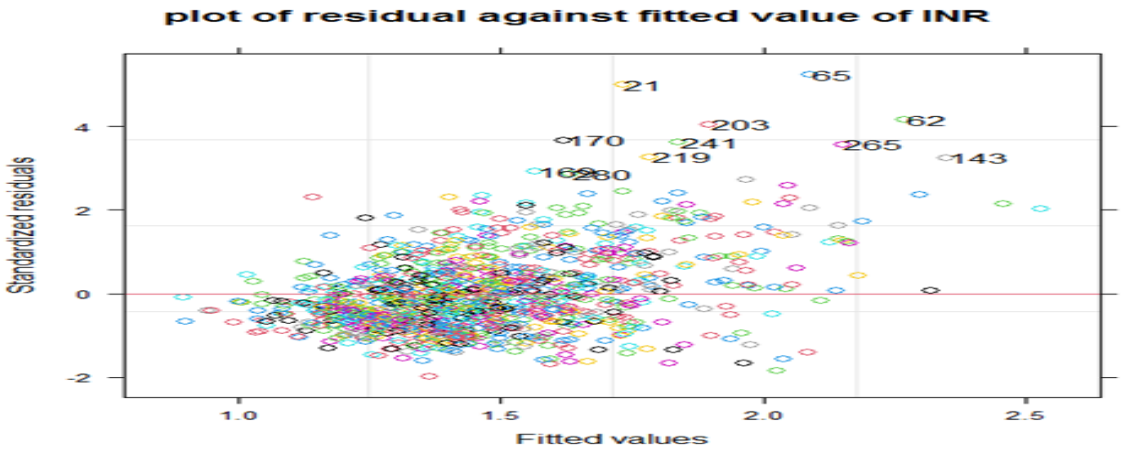

Figure 6.1: Plot of Residual versus Fitted Values for INR

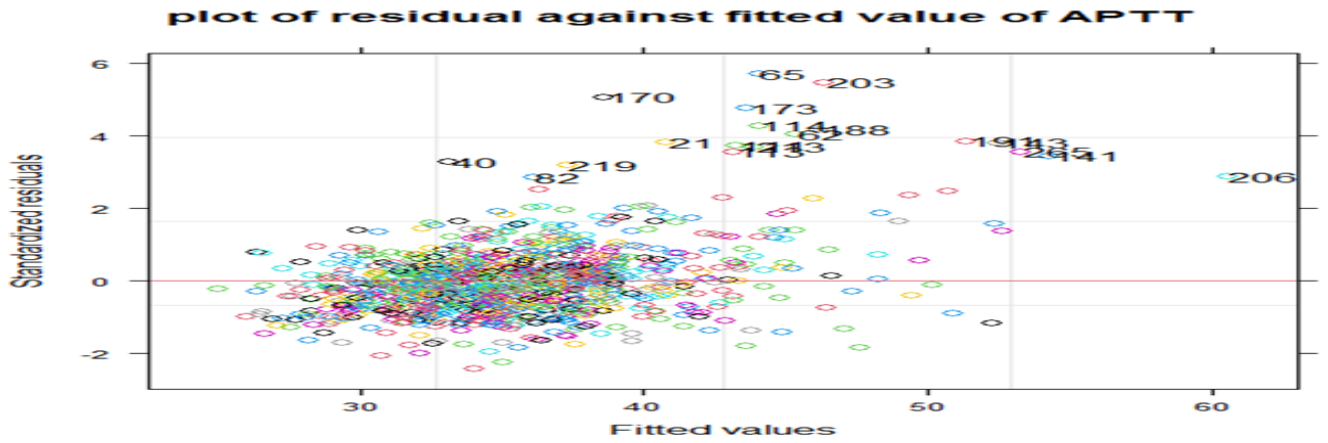

Figure 6.2: Plot of Residual versus Fitted Values for APTT

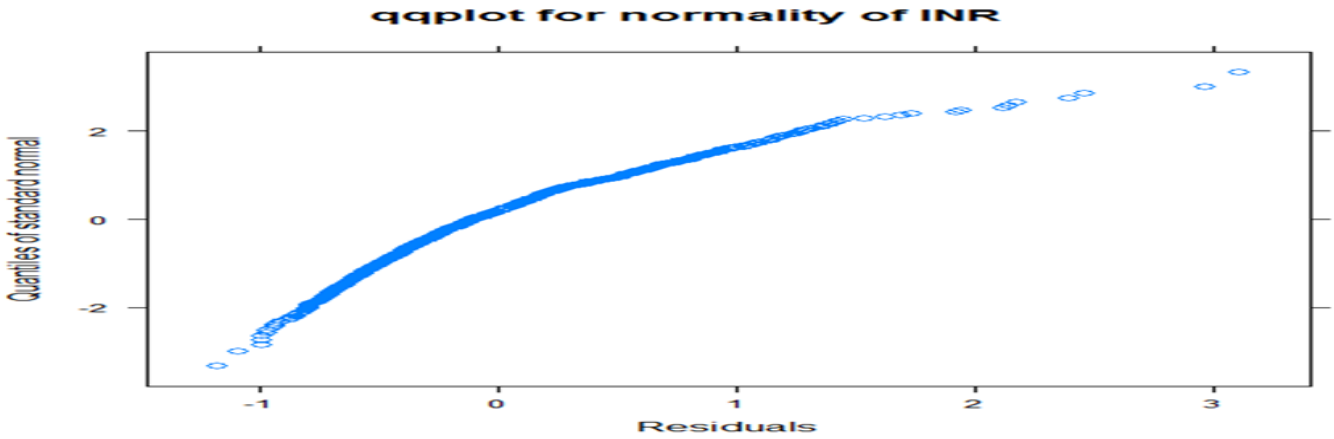

Figure 6.3: qq-Plot for Normality of INR

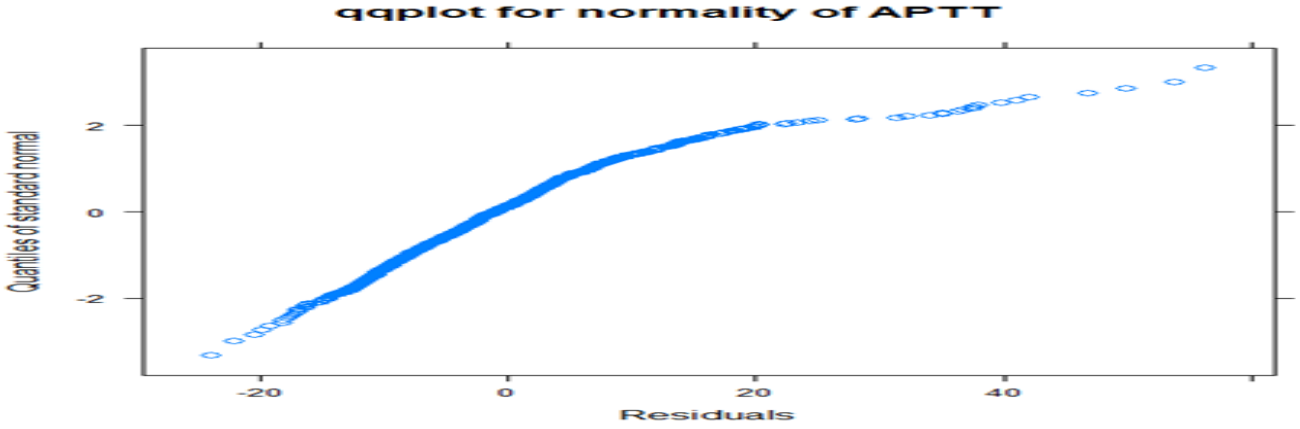

Figure 6.4: qq-Plot for Normality of APTT

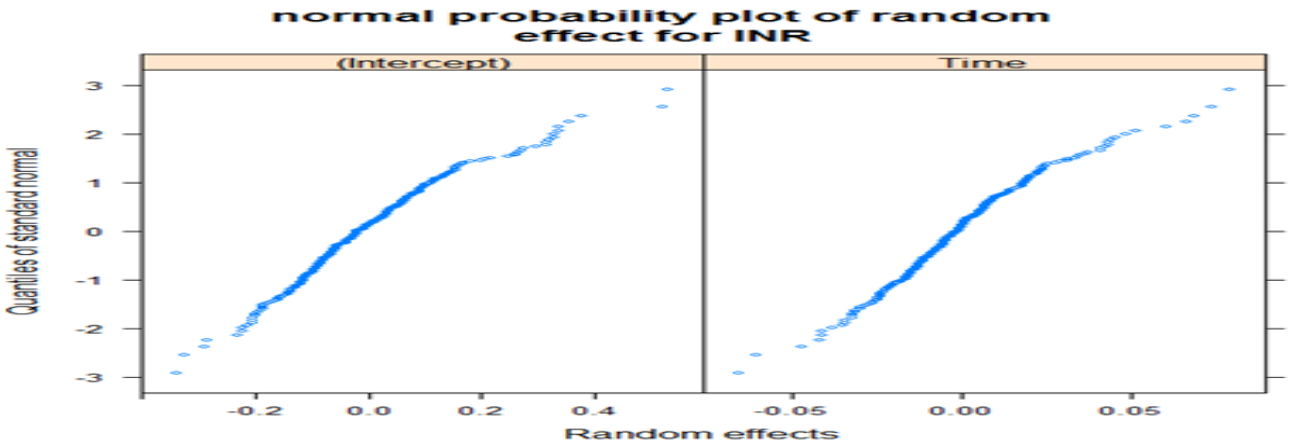

Figure 6.5: plot of normal Probability of Random Effect for INR

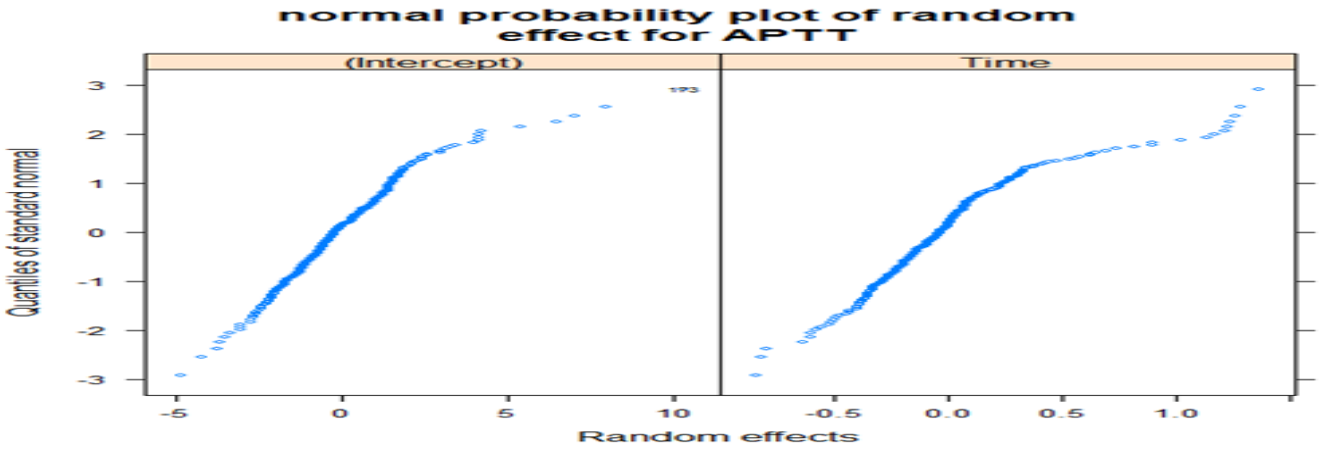

Figure 6.6: plot of normal Probability of Random Effect for APTT