2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Modelling the Evolution of CD4+ Cell Counts and Hemoglobin Concentration Level for HIV-1 Patients on Antiretroviral Therapy (ART) in Mildmay Uganda

Supervisor :
Prof. dr. Ziv SHKEDY

Supervisor :
Mr. LAWRENCE LUBYAYI

Alemu Takele Assefa

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Modelling the Evolution of CD4+ Cell Counts and Hemoglobin Concentration Level for HIV-1 Patients on Antiretroviral Therapy (ART) in Mildmay Uganda

Supervisor :
Prof. dr. Ziv SHKEDY

Supervisor :
Mr. LAWRENCE LUBYAYI

Alemu Takele Assefa

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

Abstract

Background: HIV attacks an immune cell called the CD4 cell which is responsible for the bodys immune response to infectious agents. As such, the number of CD4 cell per cubic millimeter of blood is widely used as an important biomarker for progression to AIDS when studying the efficacy of drugs to treat HIV-infected patients. Also Hemoglobin concentration of the blood is widely used as an aid in assessment of the state of health. People with HIV often experience low or declining hemoglobin levels. In order to monitor the immunity system of HIV infected patients, World Health Organization (WHO) guidelines suggest the use of simple laboratory tests such as hemoglobin (Hb) and total lymphocyte count (TLC) as an indicator of initiation of antiretroviral treatment (ART) and also as a surrogate marker to monitor immune response to therapy in symptomatic HIV patients.

Objectives: The study was aimed at describing the evolution of CD4+ cell counts and Hemoglobin concentration level and to make comparisons between selected antiretroviral (ART) regimens and patient characteristics.

Methods: A total of 1636 HIV-1 patients from Mildmay Uganda on first line antiretroviral treatment (ART) between January 1 2009 to December 31 2012 were included into the study. Thin-plate regression spline under General Additive Mixed Modelling (GAMM) framework was applied for each outcome to describe their evolution independently. Also joint parametric mixed effect modelling was implemented for the two outcomes together in order to explore the evolution of the association between them.

Results: The initiation of ART for HIV-1 patients scaled up the level of their CD4 cell counts as well as Hemoglobin level. CD4 counts level was increased in the first 5 to 9 months from ART initiation and then showed stable evolution around the threshold (350 cells/μL). Hemoglobin level was also increased in the first 7 to 13 months and then a stable evolution around the normal range (12.5 g/dL) on average was observed. The evolution of CD4 cell counts showed significant variation between some selected ART regimens, especially ART that had AZT + 3TC backbone showed relatively slow evolution (the time needed to reach to the threshold) compared to TDF + 3TC. Also gender, baseline age, and baseline WHO clinical stages were important predictors to describe the evolution. On the other hand, on average the evolution of Hemoglobin level also showed significant difference between ART backbones depending on the sex of patients. The joint analysis of the two outcomes also indicated a weak linear association between their evolution.

Key Words: ***Antiretroviral Treatment, CD4 cell Counts, Hemoglobin level, Mixed Effect Model, Thin-Plate Spline***

Contents

Dedication

This thesis is dedicated to my father Takele Assefa, my mother Aynalem Mekonen, my lovely wife Enatnesh Mengistu, and my son Abeselom Alemu.

Acknowledgment

It is a pleasure to thank many people who made this thesis possible. This work would not have been possible without the support and help of them.

First and foremost, I would like to thank the almighty GOD for having made everything possible. Secondly, I would like to express my sincere gratitude to my supervisor Prof. Dr. Ziv SHKEDY for the continuous support of my Masters study and thesis, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of analyzing the data and writing of this report. I would like also to express my sincere gratitude to my external supervisors Mr. Lawrence Lubyayi and Mr. Dan Kajungu for their continuous support to accomplish my thesis successfully.

I also owe my heartfelt gratitude to Dr. Barbara Mukasa (Executive director Mildmay Uganda), and all staff members of Mildmay Uganda Research department and Data Management center for their ultimate support. I would like also to owe my special thank to Esther Kawuma, Mary Odiit, Daniel Mwehire, Christine Kiru, and Vastine Birungi.

I would also like to thank all my professors at Hasselt University for their support to expand my knowledge, and also VLIR for providing me the scholarship. My thanks also goes to all my friends for all the fun we have had in the last two years. Finally but not the least, I would like to thank my beloved family for the support and encouragement throughout all the academic years.

Alemu Takele, Assefa

September 1, 2015

Diepenbeek, Belgium

1 Background of the Study

1.1 Introduction to HIV-1 and Blood Laboratory Tests for HIV-1 Patients

Human immunodeficiency virus (HIV) is the virus that causes acquired immunodeficiency syndrome, also known as AIDS. HIV kills or damages the cells of the body's immune system, destroying CD4 positive (CD4+) T cells, a type of white blood cell vital to fighting off infection. Because HIV compromises the immune systems, HIV-positive people are vulnerable to other infections, diseases, and complications. A blood test is used to confirm the presence of HIV in the body and once it is confirmed a number of laboratory tests are important for initial evaluation of HIV-infected patients upon entry into care; during follow-up if antiretroviral therapy (ART) is not initiated; and before and after initiation or modification of therapy to assess the virologic and immunologic efficacy of ART and to monitor for laboratory abnormalities that may be associated with antiretroviral (ARV) drugs [18].

On the other hand, Hemoglobin is a protein in RBCs that carries oxygen to the body. Normal hemoglobin levels are 12.0-16.0 grams per deciliter (g/dl) in women and 14.0-18.0 g/dl in men. People with HIV often experience low or declining hemoglobin levels, usually due to a decline in the number of RBCs produced by the bone marrow [17]. Because of the lack of laboratory technologies in resource-limited countries due to the high prices of tests such as immuno-phenotyping by flow cytometry or labeling with monoclonal, and plasma viral load testing antibodies in order to monitor the immunity system of HIV infected patients, World Health Organization (WHO) guidelines suggest the use of simple laboratory tests such as hemoglobin (Hb) and total lymphocyte count (TLC) as an indicator of initiation of antiretroviral treatment (ART) and also as a surrogate marker to monitor immune response to therapy in symptomatic HIV patients [4, 27].

Laboratory tests can give important clues about the health status of people with HIV. Some of these tests—specifically complete blood counts (CBC), chemistry screens, T-Cell counts and viral load tests—should be done shortly after someone finds out they are HIV positive to establish a "baseline" measure of immune status and viral activity. Establishing this baseline helps people and their health care providers monitor disease progression as well as the effects of treatments. Age, sex, stress, current therapies, active infections and other factors can affect the results of these tests, and test results should be interpreted with these other factors in mind [17].

1.2 HIV-1 in Africa, Uganda

HIV/AIDS is a major public health concern and cause of death in many parts of Africa. Although the continent is home to about 15.2% of the world's population,[34] Sub-Saharan Africa alone accounted for an estimated 69% of all people living with HIV [13] and 70% of all AIDS deaths in 2011. According to

the 2013 report on HIV/AIDS in Uganda, HIV epidemic in Uganda continues to be generalized, and has not changed pattern in the last three decades. The country achieved impressive success in the control of HIV during the 1990s, bringing down HIV prevalence among adults aged 15-49 years from a national average of 18.5% in 1992 to 6.4% as reported in the 2005 sero-survey [16]. The 2011 AIDS Indicator Survey in Uganda reported HIV prevalence at a national average of 7.3% and important variations by sex and in specific regions. Although Uganda continues to experience a high rate of new HIV infections; the trend over the last three years before the end period of this study shows a decline, from an estimated 162,294 in 2011 and 154,589 in 2012, to 140,908 in 2013 [16].

1.3 Antiretroviral Therapy (ART) for HIV-1

The introduction of highly active antiretroviral therapy (ART) as treatment for HIV infection has greatly improved mortality and morbidity for adults and adolescents living with HIV around the world [26, 27]. Standard antiretroviral therapy (ART) consists of the combination of at least three antiretroviral (ARV) drugs to maximally suppress the HIV virus and stop the progression of HIV disease. Huge reductions have been seen in rates of death and suffering when use is made of a potent ARV regimen, particularly in early stages of the disease [27]. According to the World Health Organisation (WHO) recommendation, adults and adolescents start with two nucleoside reverse-transcriptase inhibitors (NRTIs) plus a non-nucleoside reverse-transcriptase inhibitor (NNRTI) on a first line therapy [27] [32]. The most commonly used NRTIs in resource-limited countries, such as in Mildmay Uganda, are zidovudine (AZT) + lamivudine (3TC), tenofovir (TDF) + lamivudine (3TC), abacavir (ABC) + lamivudine (3TC), or stavudine (d4T). Also efavirenz (EFV), and nevirapine (NVP) are widely used types of NNRTIs [3, 17, 18].

In 2012, 68% of people living with HIV in sub-Saharan Africa had access to antiretroviral treatment (ART). 10 countries reported reaching universal access (at least 80% of adults eligible for ART) under the World Health Organizations (WHO) 2010 guidelines (those with a CD4 counts of 350 cells/μL or less) [28]. The WHO's 2013 guidelines have subsequently made many more people eligible for treatment by expanding the CD4 treatment initiation to 500 cells/μL or less for adults, adolescents and older children [32].

1.4 Objectives

The main objective of this retrospective observational study was to describe the progression of CD4 cell counts and Hemoglobin concentration level over time, for patients on first line antiretroviral therapy (ART) in Mildmay Uganda and to compare the progression between selected combinations of antiretroviral drugs (ARV) regimens. It was also aimed at determining whether the evolution depends on selected patient characteristics.

1.5 Motivations to the Study

The therapeutic benefits of ART are often limited by long-term toxicities and evolution of drug-resistant virus [5, 13, 34]. In resource-rich countries, HIV treatment is monitored routinely with laboratory measures such as blood chemistry, HIV viral load, and CD4 counts for early detection of side effects of medications and drug-resistant virus [5, 14]. Due to the lack of accessible and affordable laboratory services, routine laboratory monitoring is not feasible in most resource-limited countries [5, 13, 34]. Without laboratory monitoring, many patients may experience prolonged virologic failure and develop drug resistance mutations, which could ultimately limit second-line treatment options, increase morbidity, mortality and increase transmission of resistant viruses in the population [14, 28, 32]. The World Health Organization (WHO) recommends CD4 counts monitoring every six months and viral load testing only when the capacity exists [27]. There are limited data on laboratory monitoring of treatment from real world pediatric HIV clinics in resource-limited countries where there are frequent shortages of laboratory reagents, breakdown of equipment in addition to poor compliance with clinic appointments making testing at fixed intervals impossible [27]. In this retrospective study we tried to describe the long-term effect of selected ART regimens on the evolution of CD4 cell count that are directly related to the immune status of HIV-1 patients.

In addition, the advent of potent antiretroviral therapy has altered the expected natural history of human immunodeficiency virus (HIV) infection and of many previously associated opportunistic complications, including malignancies. At the same time, HIV suppression hasn't affected all of these complications equally and the longer expected survival of infected patients may allow the development of newer complications. Additionally, the use of potent antiretroviral combination therapy may itself lead to hematological toxicities [30]. Drug therapy for HIV infection or its subsequent complications is also a common cause of anemia [22]. Severe anemia, defined either as a hemoglobin level of less than 7.5 - 8.0 grams per deciliter or anemia that requires transfusion, can be seen in 24% of those who receive Zidovudine (AZT) 1500 mg daily [19]. This study also tries to describe the trajectory of Hemoglobin concentration level in HIV-1 patients treated with different ART regimens in Mildmay Uganda.

2 Review of Literatures

A research [20] was conducted aiming at describing the evolution of CD4 cell counts for patients on ART at Mildmay Uganda, which had almost the same study population with this study. They fitted a linear mixed effect model to describe the evolution of CD4 cell counts for patients on ART at Mildmay Uganda and from their result they discussed that the trajectory of CD4 cell count (in logarithmic scale) can be explained with cubic time effect. They also discussed that the effect of gender on the evolution depends on the patient's baseline CD4 category, also on the NNRTI drug that the patient uses. In their conclusion they added that patients who started ART at higher baseline CD4 counts evolved higher than those who started at lower CD4 counts.

Also, a retrospective study [1] was conducted to investigate the longitudinal analysis of change in CD4+ cell counts of HIV-1 patients on antiretroviral therapy (ART) based on secondary data from the HIV/AIDS Monitoring Program at the Builsa District hospital, in which patients were enrolled and their CD4+ cell counts were regularly monitored and thus generating repeated measures of their CD4+ cell counts. The purpose of the study was to investigate some plausible determinants of change in CD4+ cell count. They used Mixed effects modelling approach for modelling the CD4+ cell counts of the patients. Their results showed that, the initial CD4+ cell count of a patient, the duration of treatment and the drug type used in the treatment, were the factors that significantly determined a patients current CD4+ cell count. They also added that there is strong positive association between CD4+ count and duration of treatment (time). The effect of age on change in CD4+ cell count was also statistically significant at the 5% significance level. A patient has an average of 2.551 count disadvantage for every year older he/she is at the time of diagnosis. There was also significant gender differential among patients that were on treatment, that is, they found that the average CD4+ cell count for males is about 29 counts higher than that of their female counterparts.

On the other hand, [2] carried out a study to determine the prevalence of autoimmune haemolytic anaemia in HIV-infected patients and to compare the haematological/immunological characteristics of subjects with anaemia and those without. They used a total of 350 HIV-infected subjects attending the Lagos University Teaching Hospital who consented were recruited for the study. This included 250 subjects with anaemia (haemoglobin concentration <10 g/dl) as cases and 100 subjects without anaemia as controls. From their result they discussed that subjects with anaemia had lower mean CD4 cell count (284.3 cells/μl).

From a study [15] on the variation of Hemoglobin level with age and sex, it was discussed in general that the Hemoglobin concentration of the blood is widely used as an aid in assessment of the state of health. Therefore, they added that it is necessary to establish the trend of Hemoglobin values in relation to age and sex. From their result they discussed that men on average have higher level of Hemoglobin especially during younger age compared to women. Also, they mentioned that elderly people are usually

associated to lower level of Hemoglobin. Besides to this, they suggested that altitude of the location where subjects reside should be taken into account. However, in our study we did not consider altitude as a factor variable because all patients included in the study live almost in the same place which was in Kampala, Uganda and near by villages where the altitude variation is very small.

Another interesting research on the correlation between CD4 cell count and Hemoglobin level was [4] and we briefly discussed this research as follows. Lymphocyte CD4+count, a standard laboratory test for staging of HIV infection, is expensive and unavailable in resource-restricted countries. Total lymphocyte count (TLC) and hemoglobin (Hb) are recommended as simple & inexpensive surrogates. The aim of the study was to assess the correlation, sensitivity and predictive power of these parameters as substitutes for CD4 counts. One hundred HIV patients enrolled in this analytic descriptive study in Ahvaz, a city in the South of Iran, from 2005 to 2006. They were tested for CD4 counts, TLC, Hb, and hematocrit (Hct). The cutoffs were determined as: 200 cells/μL, 1200 cells/μL, 12 g/dl and 30%, respectively. The correlation coefficient established correlation between values. Sensitivity, specificity and positive predictive values were also calculated. From their result, a strong correlation was observed between CD4 counts and TLC (R = 0.645, P = 0.001), but no correlation was seen between CD4 counts and Hb or Hct (R= 0.451, P=0.056 and R= 0.375, P=0.816 respectively). This study shows that TLC is a suitable surrogate marker for CD4 counts. Hb and Hct are of limited value in predicting CD4 counts and should not be substituted for CD4 counts.

On the other hand, in [11] a possible association between serum neopterin concentrations and blood cell counts (CD4+ T cells, white blood cells, platelets, red blood cells) and hemoglobin and hematocrit in 94 HIV-1-seropositive individuals was investigated. They found significant negative correlations between neopterin concentrations and CD4+ T cells, hemoglobin, hematocrit and platelets. These correlations were also significant if either only the sub population or the entire set of data were considered for calculations. Finally they discussed that hematological abnormalities are associated with chronic immune activation in patients with HIV-1 infection.

3 Methodology

3.1 Study Population

Mildmay Uganda (MUg) opened in 1998 to provide palliative outpatient care for people living with HIV/AIDS, and to act as a teaching and training centre for HIV/AIDS health care personnel in Uganda. MUg is primarily a treatment centre for persons living with HIV/AIDS and their families. It currently offers family centered care and support to approximately 24,000 clients. MUg has one central clinic located in Lweza, 12km out-side Kampala, the capital of Uganda. MUg is also involved in a district health systems strengthening programme to build HIV/AIDS health care capacity in existing health centres in 18 local districts [12]. On this background, this study was conducted using routine data for patients who started ART between 2009 and 2012 at MUg, and were on first line regimen drugs. Blood parameters including CD4 cell counts and Hemoglobin are normally measured for patients on ART on routine basis every 6 months. The criteria for starting ART in Uganda followed the WHO guidelines. First-line ART comprised a Nucleoside Reverse Transcriptase Inhibitor (NRTI) backbone i.e. Lamivudine (3TC) plus one of Zidovudine (AZT), Stavudine (d4T) or Tenofovir (TDF), plus a Non-nucleoside Reverse Transcriptase Inhibitor (NNRTI) drug i.e. either Nevirapine (NVP) or Efavirenz (EFV). However, in this study the comparison of the evolution of CD4+ cell count and Hemoglobin level was solely focused on AZT + 3TC and TDF + 3TC from NRTI and NVP and EFV from NNRTI. Because of the high level and variability of CD4+ cell counts in children [17], patients with baseline age 12 and above were used to describe the evolution of CD4+ cell count where as patients from all age groups were used to describe the evolution of Hemoglobin. Informed consent was not obtained from individual patients but the analysis was done anonymously. It is worth mentioning however, that patients in care at Mildmay Uganda, consent and assent to use of their data for scientific research.

3.2 The Data and Patients Characteristics

Gender and baseline age are usually considered in many studies on HIV-1 and related areas. Some studies on the same topic such as [1, 20, 21] have found that these two patient characteristics are potential factors in influencing the level of CD4+ cell count also they have interaction effect with other factors. These factors are also recommended by many studies [15, 30, 31] to be considered when some one wants to study the concentration level of Hemoglobin. Also, the clinical staging and case definition of HIV for resource-constrained settings were developed by the WHO in 1990 and revised in 2007 [6, 27]. Staging is based on clinical findings that guide the diagnosis, evaluation, and management of HIV/AIDS, and it does not require a CD4 cell count. Clinical stages are categorized as 1 through 4 (labeled as I, II, III, and IV in this study), progressing from primary HIV infection to advanced HIV/AIDS. These stages are defined by specific clinical conditions or symptoms and the higher the stage the lower CD4 cell counts

is expected [6, 27], consequently we included WHO clinical stages at the start of ART into the study as patients baseline characteristics.

A total of 1636 patients (1422 for CD4 counts study and 1567 for Hemoglobin study) were eligible for the study and the characteristics of these patients with respect to selected study variables is shown in Table 2. Regarding the distribution of ART regimens among patients, the data was quite unbalanced, for example from Table 1 we can see that 1336 (81.7%) of patients were given ART which had AZT + 3TC backbone and only 300 (18.3%) patients were given ART which had TDF + 3TC backbone. From the same table we can also see that NVP was the dominant NNRTI (covering 80%) compared to EFV (covering 20%) for patients with ART backbone of AZT + 3TC, whereas the proportion of these two NNRTIs was nearly balanced among patients with ART that had TDF + 3TC backbone. Regarding patients baseline characteristics, in Table 2 it is shown that 67.5% of the study patients were female and the rest 32.5% were male. With respect to age composition, most of patients (75%) were between 12 to 45 years and the rest 25% of them were children and elderly patients. Similarly, 77.6% of them were on the first WHO clinical stage at the time of ART start, 21.3% of them were on the second stage, and the rest 1% of them were on the third and fourth WHO stages. 32 (2%) patients' baseline WHO clinical stage was not registered, or perhaps missed, when they start ART but not excluded from the analysis. Since the number of patients under WHO stage III and IV were considerably small, the analysis was based on a new group that combines these two groups as stage III and IV are associated to sever AIDS progression [27]. Additionally, patients entered to the study in different times between January 2009 to December 2012, as a result we classified patients according to the year they were entered to the study and shown in Table 2. 44% of the patients entered into the study in 2009, among these patients most of them (91%) were given ART that had AZT + 3TC backbone and also 78% of them were given ART that had NVP. Similarly, 20% of the patients were entered to the study in 2010, and the rest 24% and 13% were entered in 2011 and 2012 respectively.

Table 1: ART Regimens distribution

NRTI	NNRTI	Frequency	Percent	Conditional Percent
AZT+3TC	**EFV**	269	16.44	20.13
AZT+3TC	**NVP**	1067	65.22	79.87
Sub Total		1336	81.66	
TDF+3TC	**EFV**	157	9.6	52.33
TDF+3TC	**NVP**	143	8.74	47.67
Sub Total		300	18.34	
Total		1636		

Figure 1 shows the follow up summary of patients included in the study and it can be seen that most

of the patients were measured 2 to 4 times within the study period for Hemoglobin test and 5 to 6 times for CD4 cell count test. The time of laboratory test was not in a regular interval even if the institution (MUg) planned to visit each patient in 6 months interval. As a result, in this study time span between ARV start date and laboratory test was treated as a continuous measurement with minimum value 0 for measurements taken at the date patients started ART. Regarding the time of measurement, the data was considered as unbalanced longitudinal data as the time gap between laboratory measurements for the outcomes was not fixed [29].

Table 2: Patient characteristics stratified by NRTI and NNRTI

Variables	**NRTI**		**NNRTI**		**Total**
	AZT	**TDF**	**EFV**	**NVP**	
	n (%)	**n** (%)	**n** (%)	**n** (%)	**n** (%)
Gender					
Female	887 (54.22)	218 (13.33)	212 (12.96)	893 (54.58)	1105 (67.54)
Male	449 (27.44)	82 (5.01)	214 (13.08)	317 (19.38)	531 (32.46)
Baseline WHO Stage					
I	1014 (63.21)	231 (14.4)	335 (20.88)	910 (56.73)	1245 (77.61)
II	281 (17.52)	60 (3.74)	77 (4.8)	264 (16.46)	341 (21.26)
III	11 (0.69)	4 (0.25)	9 (0.56)	6 (0.37)	15 (0.94)
IV	2 (0.12)	1 (0.06)	6 (0.37)	3 (0.19)	3 (0.19)
Missing					32 (1.96)
Baseline Age Group					
≤ 1	29 (1.77)	0 (0.00)	1 (0.06)	28 (1.71)	29 (1.77)
1-4	75 (4.58)	0 (0.00)	7 (0.43)	68 (4.16)	75 (4.58)
5-11	106 (6.48)	4 (0.24)	33 (2.02)	77 (7.71)	110 (6.72)
12 -45	977 (59.72)	250 (15.28)	326 (19.93)	901 (55.07)	1227 (75.00)
≥ 45	149 (9.11)	46 (2.81)	59 (3.61)	136 (8.31)	195 (11.92)
ARV Start Year					
2009	654 (39.98)	67 (4.10)	155 (9.47)	566 (34.60)	721 (44.07)
2010	316 (19.32)	1 (0.06)	64 (3.91)	253 (15.46)	317 (19.38)
2011	291 (17.79)	90 (5.5)	88 (5.38)	293 (17.91)	381 (23.29)
2012	75 (4.58)	142 (8.68)	119 (7.27)	98 (5.99)	217 (13.26)
Total	1336 (81.66)	300 (18.34)	426 (26.04)	1210 (73.96)	1636 (100.00)

n = number

% = Percentage

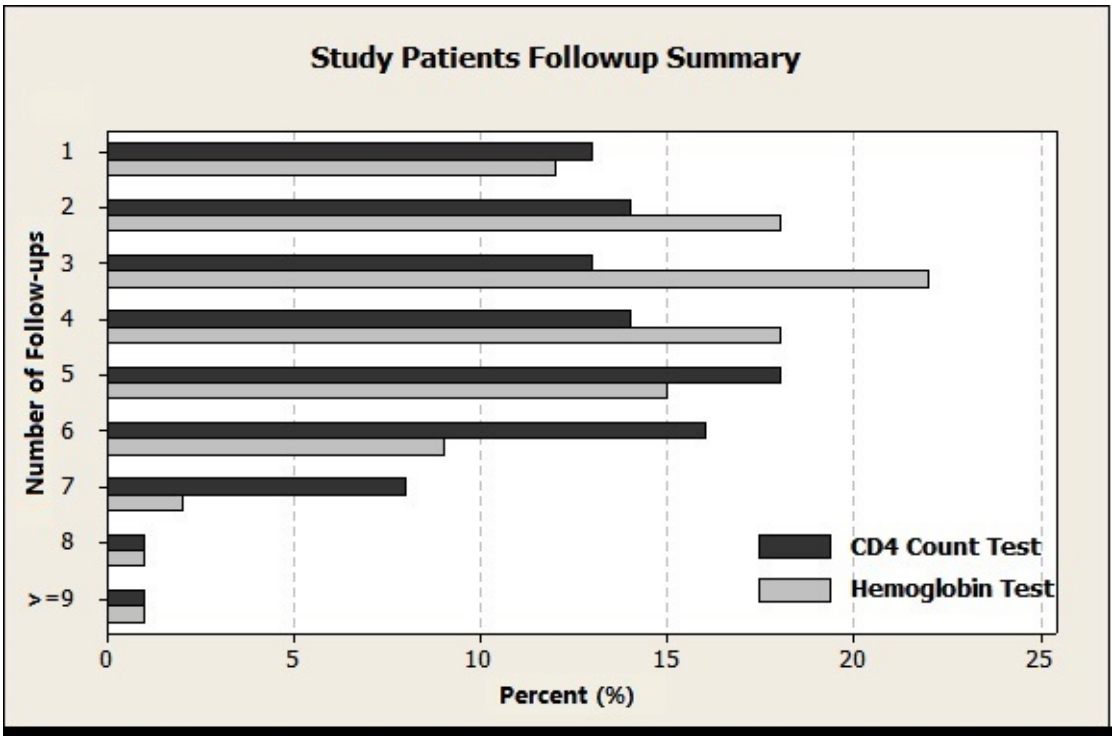


Figure 1: Follow-up statistics of patients within the study period (from Jan 1, 2009 to Dec 31, 2012)

3.3 Statistical Methods

The data used in this study includes longitudinally recorded response variables (CD4+ cell count (in cells/micro-liters) and Hemoglobin concentration level (in grams/deciliter)) from each participating subject. At least one measurement was taken from each participant and a maximum of up to 10 or 11 observations had been observed longitudinally on the same subject. From this fact we can deduce the presence of non-ignorable correlation within measurements of the same patient. As a result, for statistical analysis the usual assumption about independent observations of the response variable(s) may not be reasonable and a solution to this problem is to model the data with longitudinal regression models that take in to account the within subject correlation [8, 29]. Note that in this study two response variables CD4+ cell count and Hemoglobin concentration level are considered. These outcomes were measured from each patient longitudinally and hence one can expect association (correlation) between these outcomes. In this study we passed through two approaches: first modelling the outcomes separately assuming independence between them considering the suggestions from some researches [11], and secondly fitting a model under multivariate setting that takes into account the possible association between them. These two approaches are briefly discussed in the next sections and the reader can go to the cited references for their detailed theory and concept.

3.3.1 Data Exploratory Methods

Exploration of the structure of the data, important for guiding the appropriate modeling framework, was done using graphical techniques such as i) the individual profiles plot which gives us an idea on the within and between subject variability, ii) the mean profile plot that suggests the initial plausible assumption on the mean structure of model, and iii) the variance function that could tell a plausible initial assumption on the structure of random effects [29]. Since the data was not recorded in fixed time interval for each participant patient, it was mandatory to use smoothing techniques to explore the mean and variance structure, as a first step Lowess smoothing technique was used [7]. In order to get an insight on the evolution of the association between the two outcomes of interest over time, a scatter plot diagram along with Pearson's correlation coefficient was used.

3.3.2 Semi-parametric Mixed Effect Models

From several researches that tried to describe the evolution of CD4+ cell count such as [1, 5, 20], it has been noted that the evolution of CD4+ cell count and other blood parameters have a non-linear evolution which can be fitted with non-linear parametric longitudinal mixed effects model or linear mixed effect models with non-linear time effect as [1, 5, 20] had done. Although such parametric mean models enjoy simplicity, they have suffered from inflexibility in modeling complicated relationships between the response and covariates in various longitudinal studies [10]. Additionally, from the statistical point of view, the parameter estimation process for non-linear models is based on numerical algorithms which are dependent on initial values and then iteratively finds the estimates of the parameters. Also, non-linear parametric models may be too restrictive [24]. Semi-parametric models on the other hand offer flexibility and can capture non-linear curve shapes without assuming any parametric structure for the mean [24, 33]. Therefore, we proposed penalized Thin-Plate Regression Splines (TPRS) to analyze the data. The other advantage of penalized TPRS over the other classes of semi-parametric (splines) models is that it avoids knot selections and the placements of knots, as both of these emerge naturally from the mathematical statement of the smooth problem [33]. Interestingly, the smoothing parameters of the model can be estimated within the mixed model framework [10, 25, 33].

Let $\mathbf{Y}_i(t)$ be the response of interest (CD4+ cell count or Hemoglobin level) measured at time t on the i^{th} subject (i = 1, 2, ... N). The penalized spline model, with patient specific random effects $\boldsymbol{b_{0i}}$, $\boldsymbol{b_{1i}}$, ..., $\boldsymbol{b_{pi}}$ can be expressed as

$$\begin{aligned} \boldsymbol{Y}_i(t) &= \boldsymbol{X\beta} + S(t) + \boldsymbol{Zb} + \boldsymbol{\epsilon}_i(t) \\ S(t) &= \sum_{l=1}^{v} \gamma_l f_l(t) \\ \boldsymbol{Zb} &= \boldsymbol{b}_{0i} + \boldsymbol{b}_{1i}t + ..., + \boldsymbol{b}_{pi}t^p \end{aligned} \tag{1}$$

Where $\boldsymbol{X}$ is the design matrix for the covariates (fixed effects) such as NRTI, NNRTI, baseline age,

gender, and WHO clinical stages which are parametrically modelled with parameter vector $\boldsymbol{\beta}$. $f_i(t)$s are a set of thin plate spline basis functions, γ_l are the coefficients of the basis function [25, 33], $\mathbf{Z}$ is random effects (time) design matrix, $\boldsymbol{b}^T = [\boldsymbol{b_{0i}},\ \boldsymbol{b_{1i}},\ ...,\ \boldsymbol{b_{pi}}]^T \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$ and random error terms $\boldsymbol{\epsilon}_i(t)$ i.i.d. $N(0,\ \sigma^2_\epsilon)$. The patient specific random effects $\boldsymbol{b}_{pi}$'s accounts for the correlated nature of the repeated measurements of CD4+ cell count and Hemoglobin level. The need for these random effects was tested with a mixture χ^2 distribution which is an approximate likelihood ratio test[29]. Each smooth is treated as having a fixed effects (unpenalized) component, which can be absorbed into $\boldsymbol{X\beta}$, and a random effects (penalized) component, which can be absorbed into $\boldsymbol{Zb}$. The random effects component of the smooth also has an associated Gaussian distributional assumption, based on the wiggliness measure for the smooth, and having an unknown variance parameter, which is related to the smoothing parameter [10, 33].

The smoothing parameter λ can be estimated using optimization criterion such as generalized cross-validation (GCV) [24, 33]. In this study, we use the link between mixed model and splines. This is an advantage of fitting the thin plate regression splines within the mixed model framework [24]. Roughness is quantified by the integral of squared m-th order derivatives. We implemented the most commonly used roughness penalty, which is m=2. Note that, for m=2 and a given λ, $\hat{S}$ is a cubic smoothing spline [10, 24, 33]. For further properties and formulations of penalized TPRS model applied in this study, we refer the reader to [10, 24, 33].

Different mean structures $\boldsymbol{X\beta}$ were compared using Akaike Information Criterion (AIC) (for non-nested models), and/or Log-Likelihood Ratio test (for nested models) [29]. Model simplification was tried by assuming different structures for the covariance structure of subject specific random effects $\boldsymbol{\Sigma}$ such as Diagonal (0 covariance between random effects), Compound-Symmetric (equal covariance between random effects), and Unstructured. At a fixed mean structure these different assumptions were compared using AIC as well [23, 29]. AIC based comparisons and Likelihood ratio tests were done after setting Maximum Likelihood (ML) parameter estimation method. However the final model was based on Restricted Maximum Likelihood (REML) estimation method in order to avoid biased estimate of the variance components [29]. So far, it was assumed that the residual variance-covariance structure is simple, $\sigma^2_\epsilon\mathbf{I}$, i.e. assuming independence between $\epsilon_i(t)$, but this could be false as the observations are measured in a time sequence that result in the occurrence of serial correlation [8, 10, 29]. Hence, the presence of serial correlation was also tested using Likelihood based test with REML method because it reduces the well-known finite sample bias in the estimation of the covariance [29].

Further more, the thin plate spline smoothing was stratified by ART regimens (the stratum are the combinations of the levels of NRTI and NNRTI) allowing different smoothing of the response variables at the 4 combinations [33]. This was chosen because the intention of this study was to compare NVP and EFV given a particular backbone (AZT + 3TC or TDF + 3TC) and also vice versa. However, the evolution of the outcome was assumed to evolve similarly in women and men. This assumption also

holds to the levels of WHO clinical stages. Additionally, in order to investigate the rate of change in CD4+ count and Hemoglobin level, the first order derivative of the model (1) was examined graphically accompanied with 95% confidence band. In order to check the validity of the fitted model, the observed and predicted evolution of the outcomes for randomly selected 9 subjects were compared graphically, in this way a good model is recognized if the predicted evolution overlaps on the observed evolution with the least error. In addition, at a randomly selected 4 time points, again the predicted and observed values of the outcome were compared using scatter plot where a straight line pattern of the dots on a 45^o line indicates a good fit.

3.3.3 Joint Modelling of Multivariate Longitudinal Data

Anemia occurs frequently among patients seropositive for human immunodeficiency virus (HIV) but the etiology of anemia in HIV infection often remains unclear [22, 30]. Some studies such as [11] investigated a possible association between serum Hemoglobin level concentrations and blood cell counts (CD4+ T cells, white blood cells, platelets, red blood cells) in HIV-1-seropositive individuals and they discussed that there is a significant negative correlations between Hemoglobin concentrations and CD4+ T cells. In contrary to this, a non-significant correlation was observed in some studies such as [4]. This study also aims to investigate the association, we will call it correlation later, between the two outcomes CD4+ cell count and Hemoglobin concentration level overtime. The semi-parametric mixed effect models proposed previously model the evolution of these two outcomes independently even if both were measured from each patient longitudinally. Since the two outcomes were measured at time t from subject i (i=1,2, ... N) one can expect correlation between them. A joint modeling of such kind of data is necessary to quantify the evolutions of the two responses and at the same time the evolution of the relationship between them over time. It also ensures valid inferences as it appropriately account for the correlation among the outcomes [9, 10].

Random effects can be used to generate an association structure between the repeated measurements of a specific outcome [29]. The same idea can be used to construct multivariate longitudinal models [10]. Now under Multivariate set-up let $\mathbf{Y_{mi}(t)}$ be the m^{th} outcome (m=1, 2) of interest measured at time t from patient i, i=1,2, ... N, then

$$\begin{aligned} \boldsymbol{Y_{mi}(t)} &= \boldsymbol{X_m^T\beta} + \boldsymbol{Z_m^T b}_{mi} + \boldsymbol{\epsilon}_{mi}(t) \\ \boldsymbol{\epsilon}_{mi}(t) &\sim MVN(\mathbf{0}, \boldsymbol{R_i}) \quad \boldsymbol{b}_{mi} \sim MVN(\mathbf{0}, \boldsymbol{D}) \quad Cov(\boldsymbol{\epsilon}_{mi}(t), \boldsymbol{b}_{mi}) = 0 \end{aligned} \tag{2}$$

Where $\boldsymbol{X}_m\boldsymbol{\beta}$ is the design matrix for the fixed effect covariates that are modelled parametrically with $\boldsymbol{\beta}$ the matrix of coefficients that need to be estimated from the data. $\boldsymbol{Z}_m$ is also the design matrix for the random effect covariates, which was time of laboratory test (visit time) in this study, with random coefficients $\mathbf{b}_{mi}$. Also $\boldsymbol{\epsilon}_{mi}(t)$ represent the random error term at time t for response m. Note that here the assumption of conditional independence does not hold because, given $\boldsymbol{b}_{mi}$, the observations are

not totally independent. Given $\boldsymbol{b}_{mi}$, the observations measured at same occasion on same individual might be correlated. As a result, $\boldsymbol{R}_i = \boldsymbol{I}_{n_i} \bigotimes \boldsymbol{\Sigma}_{m\times m}$. $\mathbf{D}$ is the covariance matrix of the random effects $\boldsymbol{b}_{mi}$. Since the evolution of the association between the two outcomes was the main interest to employ joint modelling, two kinds of correlations can be mathematically obtained as follows. (i) The correlation between the evolution of log CD4+ cell count ($Y_{1i}(t)$) and Hemoglobin level ($Y_{2i}(t)$) is given by $r_E = \frac{Cov(\boldsymbol{b}_{1i},\boldsymbol{b}_{2i})}{\sqrt{Var(\boldsymbol{b}_{1i})\times Var(\boldsymbol{b}_{2i})}}$, where $\boldsymbol{b}_{mi}$ is the random slope for the m^{th} model (in order to express the equations easily we assumed here only random intercept and slope for linear time effect are included in the model), and (ii) the marginal correlation between log CD4+ cell count ($Y_{1i}(t)$) and Hemoglobin level ($Y_{2i}(t)$) at time t is also given by $r_M = \frac{Cov(Y_{1i}(t),Y_{2i}(t))}{\sqrt{Var(Y_{1i}(t))\times VarY_{2i}(t))}}$ [9, 10]

The model building strategies such as selecting the best mean structure, covariance structure, and serial correlation for the error terms were based on the strategies used in the univariate mixed effect models discussed above. However, the data used in this study was unbalanced and only very few subjects had measurements of both outcomes at the same time where as others had a gap between the time at which the measurements was done as well as some (not few) patients had measurement for only one outcome, as a result model convergence and related problems were obstacles to fit the joint model. The proposed solution to overcome this problem was discretization of the measurement time into 1 month of interval, the interval size was narrow enough to avoid multiple measurements for each patient at the same time point, and then the analysis was carried on with the discretized time point. Since discretization involves loss of information, the joint modelling was fitted with this limitation in mind.

3.3.4 Statistical Softwares

Data management and statistical analysis were done using `SAS 9.4`® and `R 3.1.3` software packages. Under `mgcv` package in `R 3.1.3` the function `gamm` was mainly used to fit semi-parametric mixed effect model. `MIXED` procedure from `SAS 9.4`® was also used mainly to fit multivariate mixed effect models. All statistical hypotheses were tested at 5% significance level, i.e. the probability of false effect finding was fixed to be 1 in 20 studies/experiments.

4 Results

4.1 Exploratory Data Analysis

Prior to fitting the model, we examined the nature of the data that could be used as a guide for the modelling framework. The CD4 cell counts data exhibited huge variability unlike the Hemoglobin level data and it was heavily skewed to the right. To get rid of the skewness in CD4 data log_e transformation was applied and the analysis was carried on with the transformed outcome. Figure 2 (a) and (b) show the individual profile over time and smoothed mean structure of log CD4 cell count and Hemoglobin level respectively. The mean structure shown in (a) reveals that the average evolution of CD4 cell counts, in log scale, seems to have a fast growth in the first about 3 months (about 100 days) after ART initiation and continued stable around 400 cells/μL (log CD4 = 6) after 3 months. Similarly, (b) shows that the mean concentration of Hemoglobin level rises faster in the first about 7 months (about 200 days), which was slightly slower compared to the average CD4 growth, and shows a constant trend around 12.5 g/dl after about 7 months since the beginning of ART treatment. The mean structure shown in this figure also depicts a non-linear evolution of both outcomes as the blue line (representing the smoothed mean structure) had a non-linear curve. On the other hand, the individual profiles indicated with gray color on both plots (a and b) reveal the presence of high within and between patients variability for both responses. Also, few extreme CD4 cell counts were observed as it can be seen in Figure 2 (a) and relatively less number of extreme Hemoglobin levels were also appeared as it can be seen in (b). These extreme measurements, however, were left untouched because patients may experience extremely small CD4 cell counts when they start ART at late WHO clinical stages (stage III and IV) and also patients with severe Anemia usually encounter very small level of Hemoglobin [19, 22, 30].

Patients had considerably different CD4 cell counts and Hemoglobin level at the start of ART (at time 0) and this may suggest the need for subject specific intercepts during mixed effect modelling. Also, some patients evolved differently, for example some showed faster growth and some others showed a declining CD4 counts or Hemoglobin level and hence it is statistically plausible to suggest the requirement of subject specific slopes that could capture the individual level evolutions under mixed effect modelling framework. Additionally, the smoothed mean structure for both responses showed a non-linear trajectory suggesting a non-linear (or possibly a high order polynomial) mean structure modelling is a good starting assumption. This kind of average evolution is actually in line with what was observed in some researches on the same area such as [1, 20]. On the other hand, the smoothed variance structure shown in Figure 16 (a) (in Appendix) depicts a slightly upward parabola curve for log CD4 counts outcome. This curvature statistically implies that it is plausible to start with a model that includes both random intercept and slope [29]. Secondly, the smoothed variance structure for Hemoglobin level also shown in Figure 16 (b) (in Appendix) shows a slightly polynomial curve which suggests to start model building with higher order random effects structure for Hemoglobin concentration level.

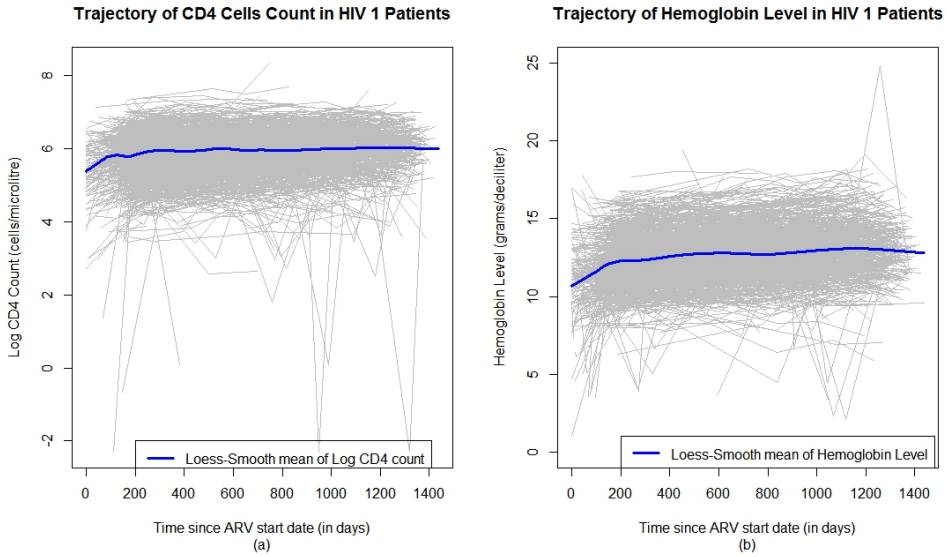


Figure 2: Smoothed mean structure (blue line) and patient specific trajectory (gray lines) of (a) Log CD4 counts and, (b) Hemoglobin concentration level

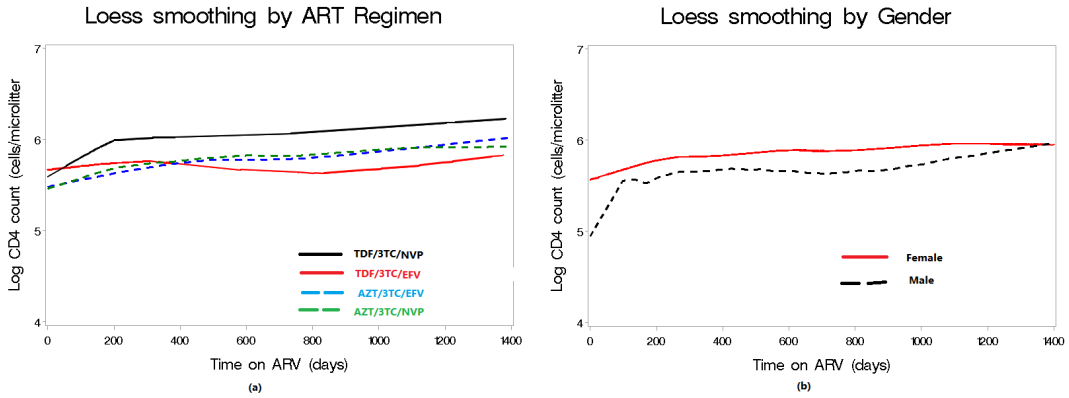


Figure 3: Smoothed mean structure for log CD4 counts stratified by (a) ART Regimen, (b) Gender

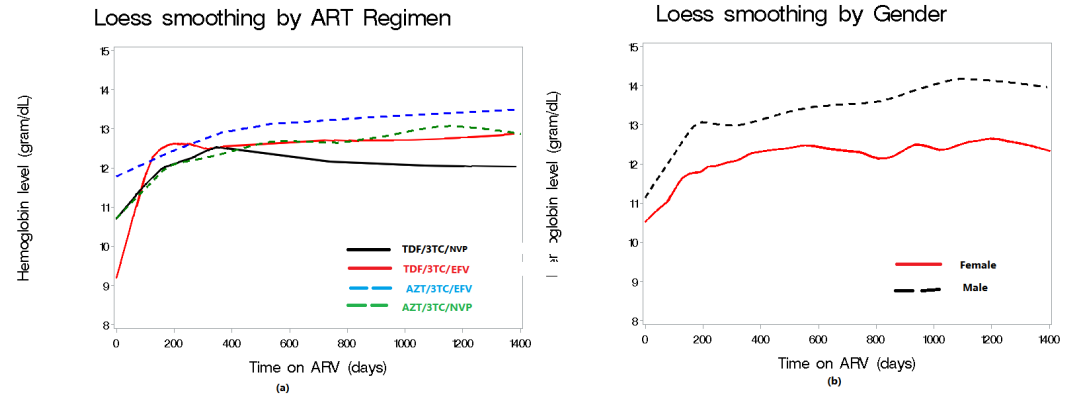


Figure 4: Smoothed mean structure for Hemoglobin level stratified by (a) ART Regimen, (b) Gender

The smoothed mean structure of log CD4 cell counts stratified by ART regimens and sex of patients is shown in Figure 3 (a) and (b) respectively. The plot depicts that NVP combined with TDF + 3TC backbone seems to result higher CD4 counts on average compared to all other regimens. Also, patients who received EFV combined with TDF + 3TC backbone seem to have quite lower CD4 counts compared to all other regimens. In addition, NVP and EFV seem to be equally efficient when they are administered with AZT + 3TC backbone. On the other hand, the level of CD4 cell count was larger for women compared to men patients but the difference became indistinguishable at the end period of the study. Figure 4 (a) and (b) also describes the smoothed mean trajectory of Hemoglobin level stratified by ART regimens and gender of patients respectively. The plot shows that patients taking NVP combined to TDF + 3TC seem to have lower hemoglobin level on average. Similarly, patients who received AZT/3TC/EFV regimen showed better Hemoglobin level compared to others. Unlike to the CD4 counts result shown in Figure 3 (b), Figure 4 (b) shows that women seem to encounter lower Hemoglobin level compared to men patients on average all over the time during the study.

The other interest of this study was to describe the association between the two outcomes overtime. The association was first investigated through a scatter plot along with marginal correlation as it can be seen in Figure 5 stratified by NRTI, NNRTI, and Gender (the stratification is indicated with different colors) by discretizing measurement time into 6 months interval. From a general view the dots did not show any systematic pattern which could indicate a particular type of relationship. The marginal correlations (indicated in the bottom right corner of each plot) also suggest weak linear relationship between Hemoglobin level and CD4+ cell counts at each discretized time point. Generally, the exploratory analysis on the association between the evolution of the two outcomes of interest did not show a positive sign but this needs to be verified using advanced modelling of multivariate longitudinal data as the descretized time point used in this primary analysis was not narrow enough to reduce information loss.

4.2 Penalized Thin Plate Regression Splines for CD4 Cells Count

The parameter estimates of the final model (eq. (3)) are shown in Table 4. The final model was obtained through comparing different mean structures that contain different possible combinations of the predictor variables using AIC and log-likelihood ratio test. The requirement of random effects ($\boldsymbol{b}_{ki}\ for\ k = 0, 1, ..., p$) into the model was tested using a mixture of χ^2 distribution and the results are shown in Table 7 (in Appendix). These tests were done at the same mean structure and also the estimation method was set to Restricted Maximum Likelihood. The need for random intercepts was significant ($\chi^2_{0:1}$ =1814.49, p-value <0.0001). In addition, the need for both random intercept, and random slope for linear time effect was also significant ($\chi^2_{1:2}$ =53.06, p-value <0.0001). Again, the need for both random intercept, random slope for linear time effect, and random slope for quadratic time effect was not significant ($\chi^2_{2:3} = 1.64$, p-value = 0.5454). This result led us to include random intercept ($\boldsymbol{b}_{0i}$) and random slope for linear time effect ($\boldsymbol{b}_{1i}$) in the random part of model (1). Also, unstructured

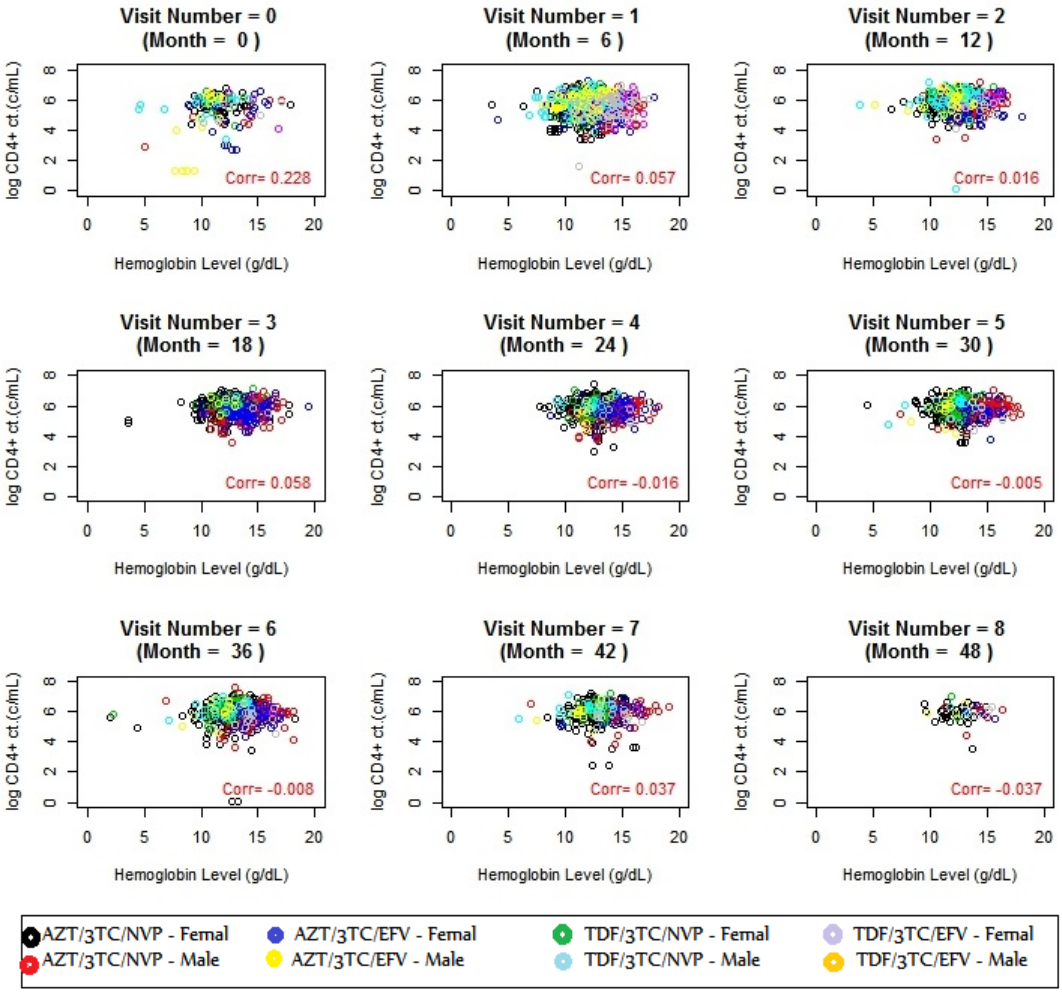


Figure 5: Scatter plot diagram of Hemoglobin level versus CD4+ cell count at each discretized time point. The gap between the visits is 6 months also the indicated correlation coefficients are marginal all over the factors.

covariance was chosen for the covariance structure of the random effects, $\boldsymbol{\Sigma}$, as it resulted the lowest AIC (AIC = 7810.63) compared to other simpler structures as shown in Table 8. After determining the appropriate covariance structure of random effects the next step was to test whether serial correlation should be included in the model. The result is summarized in Table 9 (also in Appendix). The inclusion of serial correlation did not considerably improve the model because the increment in the REML based log-likelihood values was less than 10. This increment may be important statistically but the gain is not worthwhile because it leads to adding one more parameter that increases the complexity of the model. Regarding to the covariates used in the analysis, WHO baseline clinical stages were not found having a significant effects in describing the evolution as well as the interaction effect between covariates, as a result these insignificant factors were dropped from the model.

$$\begin{aligned} \textbf{log CD4}_i(t) &= \beta_0 + \beta_1 * NRTI_i + \beta_2 * NNRTI_i + \beta_3 * NRTI_i * NNRTI_i \\ &+ \beta_4 * Gender_i + \beta_5 * Baseline\_Age_i + \beta_6 * WHOII_i + \beta_7 * WHOIII_i \\ &+ \sum_{l=1}^{v} \gamma_l f_l(t) + b_{0i} + b_{1i} * t + \epsilon_i(t) \end{aligned} \tag{3}$$

Where $\textbf{log CD4}_i(t)$ is log transformed CD4 cell counts of the i^{th} subject measured at the t^{th} time point. $NRTI_i$=1 for TDF and 0 for AZT. $NNRTI_i$=1 for EFV and 0 for NVP. $Gender_i$=1 for Male patient and 0 for Female patient. $BaselineAge_i$ is the i^{th} patient's age at the start of ART respectively. The other components are as defined in Section 3.3.2. $WHOI_i = 1$ for WHO clinical stage II, 0 otherwise. $WHOIII_i = 1$ for WHO clinical stage III or IV, 0 otherwise. The estimated smoothing parameters $\hat{\lambda}$ (std.error) were 0.1535 (0.0617), 0.1763 (0.0291), 0.11412 (0.0408), and 0.4359 (0.2196) for each combination of NRTI and NNRTI levels where all of them were significantly different from 0.

The result displayed in Table 4 and particular comparisons shown in Table 3 show that the interaction effect between NNRTI and NRTI was significant (p-value=0.0070) implying that the effect of NNRTI depends on the type of backbone with which it was combined. That is, the mean difference between the evolution of log CD4 counts subject to different types of NNRTIs (NVP or EFV) also varied between the type of backbone (AZT or TDF). At a particular baseline age, baseline WHO clinical stage, and gender, the evolution level of log CD4 counts from NVP was significantly lower than EFV if the backbone is AZT + 3TC (effect difference (EFV - NVP) = $\hat{\beta}_2 = 0.0906$, p-value=0.03146). In other words, the mean evolution of log CD4 cell counts among patients on first line ART of AZT/3TC/NVP was lower than those on first line ART of AZT/3TC/EFV. Again holding the effect of baseline age, gender, and WHO clinical stages the effect of NVP was not significantly different from EFV if the backbone is TDF + 3TC (effect difference = $\hat{\beta}_2 + \hat{\beta}_3$ = -0.1531, p-value=0.0591). On the other way around we can compare the backbones at a given level of NNRTI as it can be seen also in Table 3. These findings also showed that the average trajectory of log CD4 cell count was significantly different between the backbones depending on of the type of NNRTI they were given with. The mean evolution was larger for AZT + 3TC compared to TDF + 3TC if the type of NNRTI is NVP (effect difference = $\hat{\beta}_1 = 0.2067$, p-value = 0.0006). Also for patients who were given ART that had EFV type of NNRTI, the mean evolution was not significantly different between those who were given AZT and TDF (effect difference = $\hat{\beta}_1 + \hat{\beta}_3$ = -0.0370, p-value = 0.5840). These results confirmed what we have seen in the exploratory data analysis that showed some gap between the smoothed mean evolution of log CD4 cell counts at different regimens visually (see Figure 3 (a)). In order to make the results more imaginable, these findings are graphically explained in Figure 6 to 9.

Table 3: Comparisons between ART regimens

Comparison	**Estimate**	**Std. Error**	**p-value**
NVP vs. EFV given NRTI = AZT	0.0906	0.04205	0.0313
NVP vs. EFV given NRTI = TDF	-0.1531	0.0811	0.0591
AZT vs. TDF given NNRTI = NVP	0.2067	0.0602	0.0006
AZT vs. TDF given NNRTI = EFV	-0.0370	0.0675	0.5840

The confounding factors gender, baseline age, and WHO clinical stages were also found having a statistically significant effect on the evolution of log CD4 cell count (see Table 4). Regarding gender, on average women showed larger log CD4 cell count than men patients (effect difference (male - female) = $\hat{\beta}_4$ = -0.1950, p-value<0.0001) at a fixed level of all other factors included in the model and this was in line with the result from exploratory data analysis as shown in Figure 3 (b). Similarly, baseline age was also found one of the significant factors (p-value = 0.0002) with negative estimate of the coefficient ($\hat{\beta}_5$ = -0.0063) implying the higher the baseline age the lower log CD4 cell count. In other words, the increase in CD4 cell count in response to ART initiation was slower for elderly HIV-1 patients than younger HIV-1 patients. Also, baseline WHO clinical stages had undeniable effect on the trajectory, that is, those patient who started ART at late clinical stages showed lower evolution of CD4 counts compared to those who started ART at earlier stages. However, WHO baseline clinical stages I and II did not show statistically different level of the trajectory (p-value=0.0865) where as the last two stages (III and IV) were found pulling down the evolution compared to Stage I (p-value=0.0016).

Figure 6 ((a) to (d)) shows the thin plate smoothed mean evolution of log CD4 cell count stratified by NRTI and then by NNRTI (also first by NNRTIs and then by NRTIs) and Figure 7 ((a) to (d)) shows the pairwise difference between the smoothed evolution accompanied with 95% confidence band. The plots are particularly for women patents at a median baseline age 33 as well as on the first baseline WHO stage (the same smoothed evolution applies to the other categories of the covariates except that it shifts up or down depending on their effect, see the assumption stated in Section 3.3.2). The average evolution of log CD4 cell count from those patients taking NVP combined with AZT + 3TC was substantially lower compared to those who took EFV combined with AZT + 3TC (see Figure 6 (a)). The difference was more pronounced at long run (after 18 months from ART initiation) than earlier time and this difference was statistically significant as discussed earlier (p-value = 0.0006) and also from the figure it can be seen that the 95% confidence bands almost did not overlap each other (slight overlap was seen because we used approximate confidence interval using Normal distribution assumption).

Table 4: Parameter estimates of the final model for CD4 cell count

Effects	**Estimate**	**Std.Error**	**p-value**
Intercept (β_0)	6.0896	0.0617	<0.0001
NRTI			
TDF (β_1)	0.2067	0.0602	0.0006
AZT*	0.0000	.	.
NNRTI			
EFV (β_2)	0.0906	0.0420	0.0314
NVP*	0.0000	.	.
NRT*NNRTI (Interaction Effect)			
TDF*EFV (β_3)	-0.2437	0.0903	0.0070
Gender			
Male (β_4)	-0.1950	0.0342	<0.0001
Female*	0.0000	.	.
Baseline Age (β_5)	-0.0063	0.0017	0.0002
Baseline WHO stages			
Stage II (β_6)	-0.0628	0.0366	0.0865
Stage III and IV (β_7)	-0.4467	0.1416	0.0016
Stage I*	0.0000	.	.
Variance Components			
Std.Dev($\mathbf{b}_{0i}$)	0.5400		
Std.Dev($\mathbf{b}_{1i}$)	0.0003		
Covariance($\mathbf{b}_{0i}$, $\mathbf{b}_{1i}$)	-0.0002	($\hat{\rho}_{\mathbf{b}_{0i},\mathbf{b}_{1i}} = -0.5540$)	
Std. dev($\epsilon_i(t)$)	0.4048		

**Reference category*

Std.Dev = Standard Deviation

Figure 7 (a) also supports this finding because the over time mean difference (indicated with blue solid line) and its confidence band (light blue shade) between NVP and EFV was lied below the reference line 0 after the 18th month providing an evidence to conclude a difference between EFV and NVP when the backbone is AZT + 3TC. This may also lead to a conclusion that the performance of NVP fallen down after 18 months from ART initiation when the backbone was AZT. Similarly Figure 6 (b) shows that NVP seems to have larger effect than EFV in the first 36 months when they are combined with TDF + 3TC backbone, but since the two confidence bands overlapped each other widely, the superiority of NVP over EFV was by chance. This was also discussed earlier with p-value = 0.0591 indicating

insignificant difference. The mean difference over time shown in Figure 7(b) is also supporting this fact because the solid line and its confidence region lied on the reference line. The width of the confidence band for the evolution of mean difference was wider than the one shown in Figure 7 (a) which is probably due to lower sample size for TDF + 3TC backbone (see Table 1).

Figure 6 (c) and (d) also show the smoothed trajectory of mean log CD4 counts subject to AZT and TDF given the NNRTI was NVP (c) and EFV (d). As it can be seen in Figure 6 (c) patients who received ART with backbone AZT+ 3TC showed lower level of the evolution throughout the study period than those who received ART that contains TDF + 3TC backbone. The difference was even statistically more meaningful in the first 12 months and it continued with a reduced gap. The confidence bands were overlapped slightly until the 36th month. However both showed indistinguishable outcome at the end period which was probably because of the high uncertainty after the 36th month as indicated with wider confidence band that was caused probably because of lower sample size at the end period of the study. This result was again reproduced in Figure 7 (c) with the smoothed mean difference through time where the solid blue line and its confidence region were below 0 up to the 36th month indicating again superiority of TDF over AZT. This result was actually in line with what we discussed earlier with p-value 0.0006 (see Table 3). In similar ways, on the other hand, Figure 6(d) and Figure 7(d) show the absence of evidence to conclude that AZT and TDF resulted different mean evolution when they are combined with EFV.

The previous paragraph focused on describing the difference in the evolution of log CD4 counts with respect to the factors of interest. Now we discuss the rate of changes in log CD4 counts over time in response to the factors included in the study. The rate of change was described by the first order derivative of model (3) with respect to time and the plots are shown in Figure 8 and 9. From the property of first order derivative of a function, an increasing function will have a decreasing first order derivative function and if there is no change in the function the first order derivative will show a constant line over 0. Keeping these properties in mind we discussed its meaning in the subsequent paragraphs.

When we start from Figure 8(a) we can see that the rate of change in log CD4 counts was higher in the first about 6 to 7 months and it continued stable after that point. Specifically, patients who received NVP and EFV along with AZT + 3TC backbone showed equal rate of change, which was increasing in log CD4 counts up to the 12th month except a small difference up to the 4th month. And then they showed almost stable level (no change) for the rest of the time as the solid lines (representing the estimated first order derivative) and their corresponding confidence bands (indicated with shaded area) were overlapped each other and lied over the 0 reference line after the 12th month indicating statistically insignificant different rate of change as well as no change in log CD4 counts. The tiny difference observed in the first few months was not significant because their confidence bands were crossed each other widely. The overtime difference in the rate of change was also shown in Figure 9(a) and it confirms that no difference in the rate of change as the solid blue line indicating the difference and its 95% confidence band lied over

the reference line (0) and the tiny bump at the beginning was not also statistically acceptable as the band lied over the reference line. This insignificant difference in the rate of change for log CD4 counts between NVP and EFV when AZT + 3TC is the backbone was also held true when the backbone is TDF + 3TC except that the width of the confidence band was quite larger indicating relatively high uncertainty.

On the other hand, from Figure 8(c) we can see that TDF was associated to relatively fast growth in log CD4 counts in the first about 7 months compared to AZT when the NNRTI was NVP. From the 7th to 36th month both did not show any change as the first order derivative in this period was almost 0. After the 36th month TDF showed again an increase in log CD4 counts but not AZT. However, at the end period the confidence band was so wide and hence any justification of a change in log CD4 counts during the last period is masked by relatively high uncertainty. Figure 9(c) shows the difference through time and it showed that the rate of change was not statistically different between AZT and TDF when the NNRTI was NVP. When the NNRTI is EFV the rate of change was observed much faster again for TDF than AZT in the first 6 months and almost no change was seen after this point as it can be seen in Figure 8(d) and the time through difference shown in Figure 9(d) also verifies that AZT showed slow change in log CD4 counts compared to TDF.

The estimated standard deviation of the random intercepts and slopes were 0.5400 and 0.0003 respectively. The correlation between them was also estimated to be -0.5540, which indicates moderate but indirect linear association between them. The estimated residual standard deviation was also 0.4048 indicating lower variability of residuals compared to the random intercepts. The scatter plot diagram between the random intercepts and slopes shown in Figure 21(a) showed that there were no severe outliers except that few subjects showed an estimate of subject specific intercept ($\beta_0 + b_{0i}$) which were relatively deviated from the average intercept to the left. On the other hand, the normal quantile plot shown in Figure 21(b) shows that there was no too much deviation from the normality assumption for residuals except that the distribution showed extreme tails in both sides.

In order to check the validity of the fitted model, a comparison between observed and predicted evolution of log CD4 cell count was done for randomly selected 9 patients as shown in Figure 17 (in Appendix) and also at a randomly selected fixed time points as shown in 19 (also in Appendix). The plots show that the fitted model fitted the observed data almost accurately. Also, the comparison between the predicted and observed log CD4 counts at a particular time points showed strong and direct association because the points were lied almost on the 45^o line that represents a perfect positive association. This strong and direct association between the predicted and observed data can be depicted to the meaning that the model captured the real trend of log CD4 counts data with little error so that the model can be used to make inference in comparing the evolution of CD4 cell count between different ART regimens.

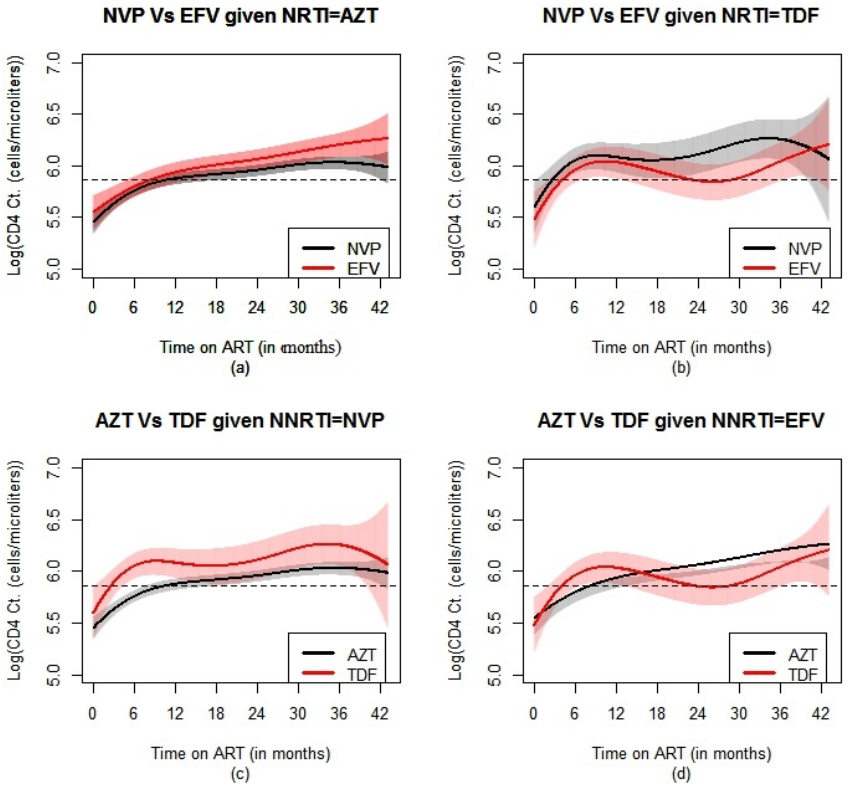


Figure 6: Fitted smoothing splines for the evolution of log CD4 cell count (a) NVP versus EFV given NRTI is AZT + 3TC, (b) NVP versus EFV given NRTI is TDF + 3TC, (c) AZT versus TDF given NNRTI is NVP, and (d) AZT versus TDF given NNRTI is EFV. The shaded region is the 95% confidence band of the curve (pink shade is for the red line and grey shade is for the black line)

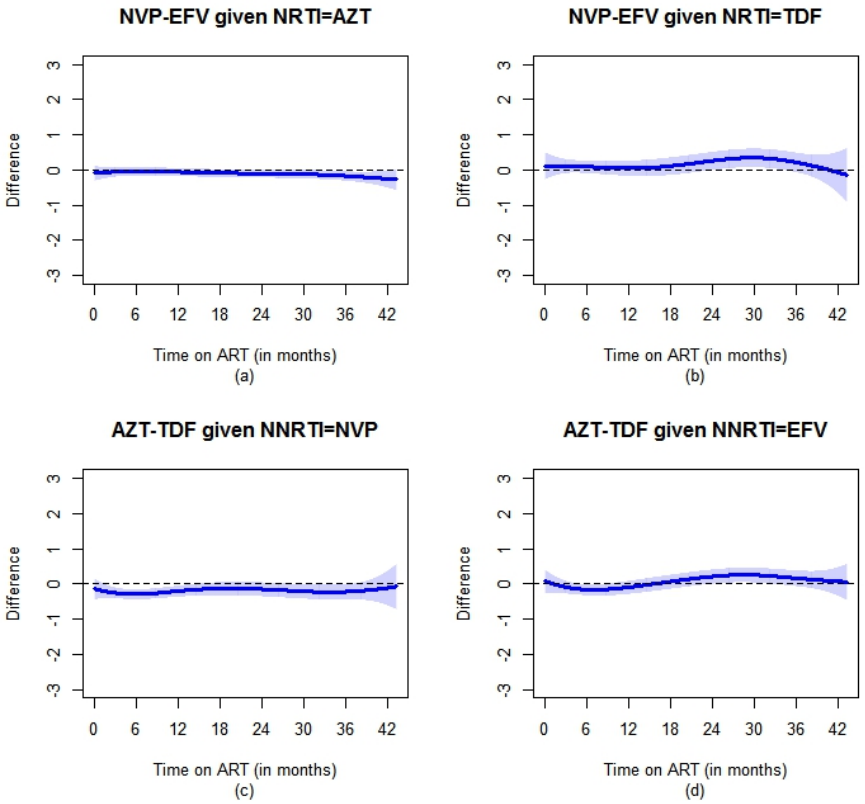


Figure 7: The difference in the smoothed evolution of log CD4 counts (a) NVP versus EFV given NRTI is AZT + 3TC, (b) NVP versus EFV given NRTI is TDF + 3TC, (c) AZT versus TDF given NNRTI is NVP, and (d) AZT versus TDF given NNRTI is EFV. The shaded region (light blue) is the 95% confidence band of the curve (blue line)

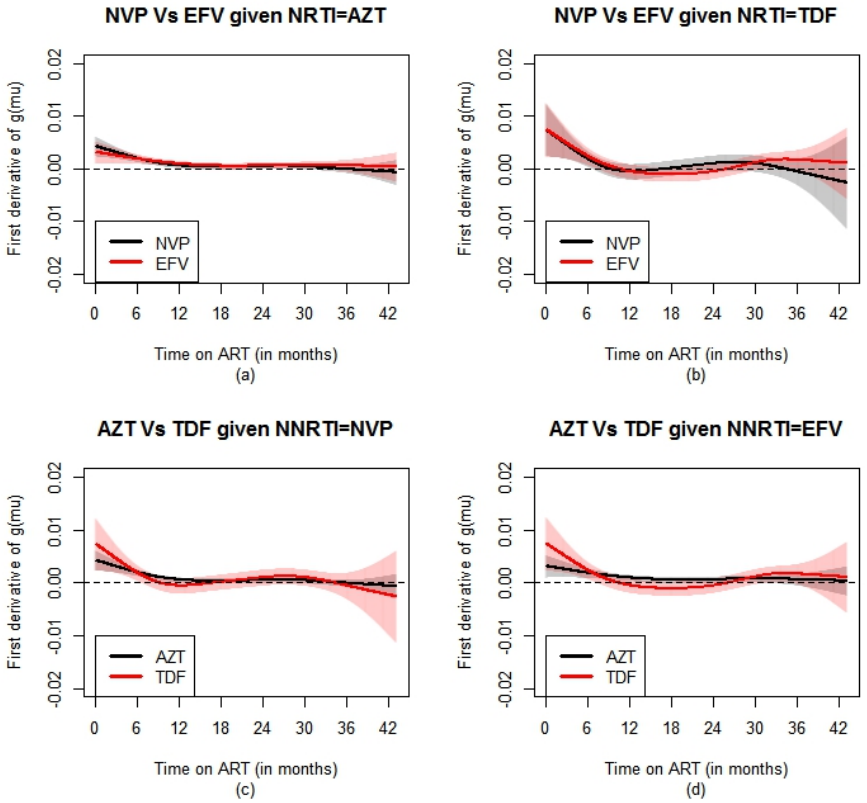


Figure 8: The smoothed first order derivative of the evolution of log CD4 cell count in response to (a) NVP versus EFV given AZT + 3TC backbone, (b) NVP versus EFV given TDF + 3TC backbone, (c) AZT versus TDF given that the NNRTI is NVP, and (d) AZT versus TDF given that the NNRTI is EFV. The shaded region is the 95% confidence band of the curve (pink shade is for the red line and grey shade is for the black line)

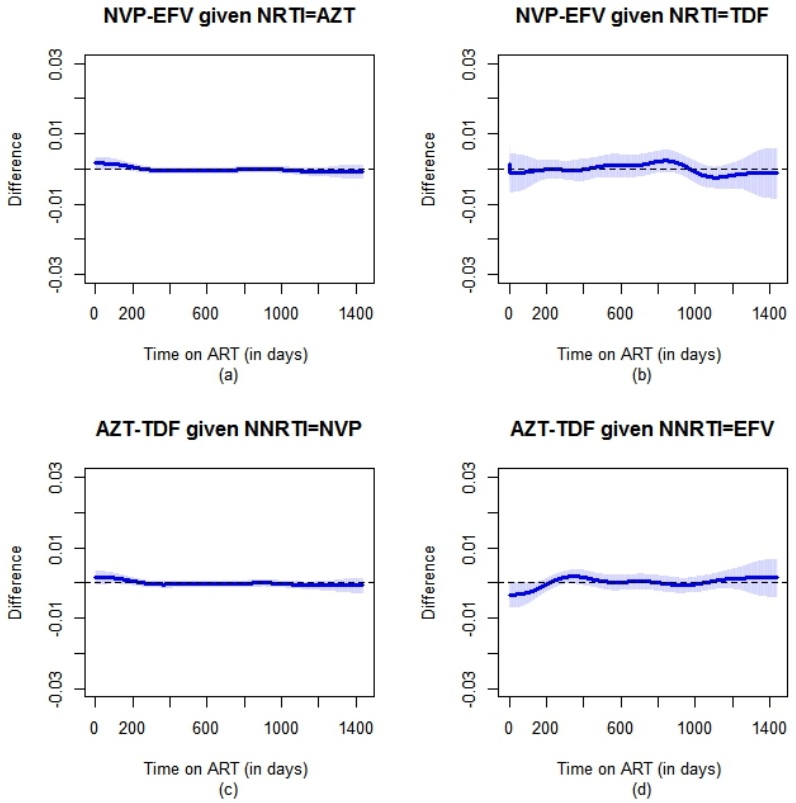


Figure 9: The difference in the smoothed first order derivative of the evolution of log CD4 counts (a) NVP versus EFV given NRTI is AZT + 3TC, (b) NVP versus EFV given NRTI is TDF + 3TC, (c) AZT versus TDF given NNRTI is NVP, and (d) AZT versus TDF given NNRTI is EFV. The shaded region (light blue) is the 95% confidence band of the curve (blue line)

4.3 Penalized Thin Plate Regression Splines for Hemoglobin Concentration level

Procedures used to arrive on the final model for log CD4 counts data were also used for modelling hemoglobin data. Different mean structures were also compared using log-likelihood ratio test and AIC then the final one that resulted the smallest AIC is mathematically expressed in model (4) and its parameter estimates are shown in Table 5. Again as it is shown in Table 7 (in Appendix) both random intercept $\boldsymbol{b}_{0i}$ and random slope for linear time effect $\boldsymbol{b}_{1i}$ were found sufficient (p-value <0.0001) for the random part of model 2 to account the correlation between repeatedly measured Hemoglobin concentration levels from a single patient. As shown in Table 8 (in Appendix), Unstructured and Diagonal covariance structures $\boldsymbol{\Sigma}$ resulted almost equal AIC which suggests no clear win between these two structures. For generalizability purpose we preferred Unstructured covariance assumption for the random effects as it adds up only one parameter. Also, the need for serial correlation was not supported by the data as shown in Table 9 (also in Appendix) because no improvement in the REML log-liklihood was observed from modelling serial correlation. Interaction effect between NRTI and NNRTI was not significant, as well as WHO clinical stages did not show a significant difference in explaining the evolution of Hemoglobin level, as a result they were dropped from the model to reduce model complexity. Unlike to model (3), the interaction effect between gender and NRTI was found significant and it is included to the model.

$$\begin{aligned} \boldsymbol{Hemoglobin}_i(t) &= \beta_0 + \beta_1 * NRTI_i + \beta_2 * NNRTI_i + \beta_3 * Gender_i \\ &+ \beta_4 * NRTI_i * Gender_i + \beta_5 * Baseline\_Age_i + \beta_6 * Baseline\_Age_i^2 \\ &+ \sum_{l=1}^{v} \gamma_l f_l(t) + \boldsymbol{b}_{0i} + \boldsymbol{b}_{1i} * t + \epsilon_i(t) \end{aligned} \tag{4}$$

Where $\boldsymbol{Hemoglobin}_i(t)$ is Hemoglobin concentration (in g/dL) of the i^{th} subject measured at the t^{th} time point. $NRTI_i$=1 for TDF and 0 for AZT. $NNRTI_i$=1 for EFV and 0 for NVP. $Gender_i$=1 for Male patient and 0 for Female patient. $Baseline\_Age_i$ is the i^{th} patient's age at the start of ART. The other components are as defined in Section 3.3.2.

The result displayed in Table 5 shows that the effect of NRTI in describing the evolution of Hemoglobin level depends on the sex of the patients as the interaction effect was significant (p-value <0.0001). This implies that at a particular baseline age, and NNRTI, female HIV-1 patients who took ART that had AZT + 3TC backbone had higher mean evolution of Hemoglobin level than those patients with ART given with TDF + 3TC backbone (effect difference (AZT - TDF) = -$\hat{\beta}_1$ = 0.4085, p-value=0.0027). Similarly, if the patient is male, the smoothed mean evolution was lower among patients who were given ART with backbone AZT + 3TC than those who took ART with backbone TDF + 3TC (effect difference (AZT - TDF)= $\hat{\beta}_1 + \hat{\beta}_4$ = -0.4100, p-value=0.0026). Regarding NNRTI, there was no significant difference in the mean evolution of Hemoglobin level between EFV and NVP (effect

difference = $\hat{\beta}_2 = 0.0682$, p-value = 0.4594) and this indifference did not depend on either the type of backbone with which it is given or other factors included in the study as the interaction was statistically worthless. Similarly, baseline age was also found having a significant linear and quadratic effect in predicting the mean evolution with negative leading coefficient (p-value <0.0001) implying quadratic (slightly downward parabolic as $\hat{\beta}_6 < 0$) relationship with the mean evolution of hemoglobin level. This suggests that for children and elderly HIV-1 patients relatively lower Hemoglobin level is observed and for middle age slightly higher level is expected because the coefficient that determines the non-linear (quadratic) relationship $\hat{\beta}_6$ was considerably near to 0 but different from 0 (p-value <0.0001). The standard deviation of the random intercept and slope were 1.1815 and 0.0007 respectively with covariance -0.0002 and this resulted -0.2080 correlation coefficient which indicates indirect but weak linear association between them. The residual standard deviation was also estimated to be 1.2716. The estimated smoothing parameters $\hat{\lambda}$ (std.error) were 0.5481 (0.3202), 0.5359 (0.2258), 1.0750 (0.7668), and 2.8220 (0.9864) for each combination of NRTI and NNRTI levels.

Figure 10 ((a) to (d)) shows the thin plate smoothed mean evolution of Hemoglobin level stratified by NRTI and then by NNRTI (also first by gender and then by NRTIs) and Figure 11 ((a) to (d)) show the pairwise difference in the smoothed mean evolution accompanied with 95% confidence band. In order to avoid large number of plots, the presented plots are particularly for patients at a median baseline age 33. Figure 10 (a) and (b) show that the average evolution of Hemoglobin level from those patients taking either NVP or EFV combined with any of the backbones (AZT/TDF + 3TC) was not different as the two smoothed lines overlap each other (even their confidence band). This difference was not statistically significant as discussed earlier and also the interaction between NNRTI and NRTI. Figure 11 (a) and (b) also support this finding because the over time mean difference between NVP and EFV (indicated with blue solid line) and its confidence band (light blue shade) lied over the 0 reference line. Figure 10 (c) and (d) also show the smoothed mean trajectory of Hemoglobin level subject to AZT and TDF given the NNRTI is NVP for female patients (c) and male patients (d). As it can be seen in Figure 10 (c) AZT was associated to higher concentration level than TDF for female HIV-1 patients. The confidence bands were almost do not overlap especially after the 10th month from ART initiation and in Figure 10 (d) we can see that AZT became associated to lower evolution than TDF specially in the first 12 months but they showed equal performance after the 13th month. This was also discussed earlier that TDF performs inferior to AZT for female patients regardless of the type of NNRTI with p-value <0.0027 and TDF performs superior to AZT for male patients with p-value 0.0026. Also, Figure 11 (c) explains the same finding that the mean difference between AZT and TDF and its confidence band were almost above 0 for female patients and below the 0 reference line for male patients. The difference was larger after about 12 months for females and lower for male patients.

So far we have seen the smoothed mean trajectory of Hemoglobin concentration level in response to different ARV combinations that build up ART regimen taking into account gender and baseline age of

Table 5: Parameter estimates of the final model for Hemoglobin concentration level data

Effects	**Estimate**	**Std.Error**	**p-value**
Intercept (β_0)	10.3715	0.1274	<0.0001
NRTI			
TDF (β_1)	-0.4085	0.1361	0.0027
AZT*	0.0000	.	.
NNRTI			
EFV (β_2)	0.0682	0.0922	0.4594
NVP*	0.0000	.	.
Gender			
Male (β_3)	1.2136	0.0843	<0.0001
Female*	0.0000	.	.
NRTI*Gender			
TDF*Male (β_4)	0.9800	0.2282	<0.0001
Baseline Age (β_5)	0.1170	0.0085	<0.0001
Baseline Age^2 (β_6)	-0.0015	0.0001	<0.0001
Variability components			
Std. dev($\mathbf{b}_{0i}$)	1.1815		
Std. dev($\mathbf{b}_{1i}$)	0.0007		
Covariance($\mathbf{b}_{0i}$, $\mathbf{b}_{1i}$)	-0.0002	($\hat{\rho} = -0.2080$)	
Std. dev($\epsilon_i(t)$)	1.2716		

**Reference category*

Std.Dev = Standard Deviation

patients. Now we focus on the rate of change in the concentration level over time. The rate of change was described by the first order derivative of model (4) with respect to time and the plots are shown in Figure 8 and 9. When we start with Figure 8(a) we can see that the rate of change in Hemoglobin level was higher in the first about 7 to 8 months (200 to 240 days), which was a little bit longer than the time used to stabilize CD4 counts, and then it continued stable with almost no change. Particularly, NVP showed higher change (increasing) compared to EFV at the beginning but both were showing almost equal rate (stable evolution) after about 7 months from ART initiation when the backbone is AZT + 3TC. The difference in the rate of change was also shown in Figure 9(a) and it confirms again that higher change in earlier time for NVP was observed compared to EFV and then it became non differentiable as the blue line (indicating the difference) lied over the 0 reference line until about the 40th month but after this time point increase in the concentration level was observed among patients who took ART that contains NVP as a part of NNRTI compared to EFV. When the backbone is TDF + 3TC, see Figure 8

(b), quite high and almost equal rate of change in Hemoglobin level was observed from both NVP and EFV up to 12 months and then almost stagnant evolution of Hemoglobin level was observed. However, as shown in Figure 8 (b) the difference in the rate of change was insignificant as the confidence band of the blue line embraced the reference line.

Similarly, Figure 8(c) and (d) show similar rate of change in the evolution of Hemoglobin level because the interaction effect between gender and NRTI was significant but not with the type of NNRTI and except that male patients had higher Hemoglobin level compared to female patients as shown in Figure 10 (c) and (d) the evolution had the same trajectory movement which resulted identical type of first order derivative trend. These plots (Figure 12 (a) and (b)) show that slow rate of change was observed in the first about 12 months and also at the end periods. The difference in this rate as shown in Figure 13 (a) and (b) was statistically meaningful on the first few months and the last few months as the blue line and its confidence band were almost above (for the first few months) and below (for the last few months) the reference line.

The scatter plot diagram between estimated random intercepts and slops shown in Figure 22(a) shows that there were few outlying subjects who had relatively higher estimate of random intercept and also slope. On the other hand, the normal quantile plot shown in Figure 21(b) shows that there was no too much deviation from the normality assumption for residuals except that the distribution showed extreme tails in both sides. Again, in order to validate the fitted model, a comparison between observed and predicted evolution of Hemoglobin level was done for randomly selected 9 patients as shown in Figure 18 (in Appendix) and at a randomly selected 4 fixed time points as shown in 20 (also in Appendix). The plots show that the fitted model 'described' the observed data almost accurately. Also, the comparison between the predicted and observed Hemoglobin level at a particular time points showed strong and direct association because the points were lied almost on the 45^o line that represents a perfect positive association. This strong and direct association between the predicted and observed data can be translated to the meaning that the model had captured the real trend of Hemoglobin level data well so that the model can be used to make inference in comparing the evolution between different ART regimens.

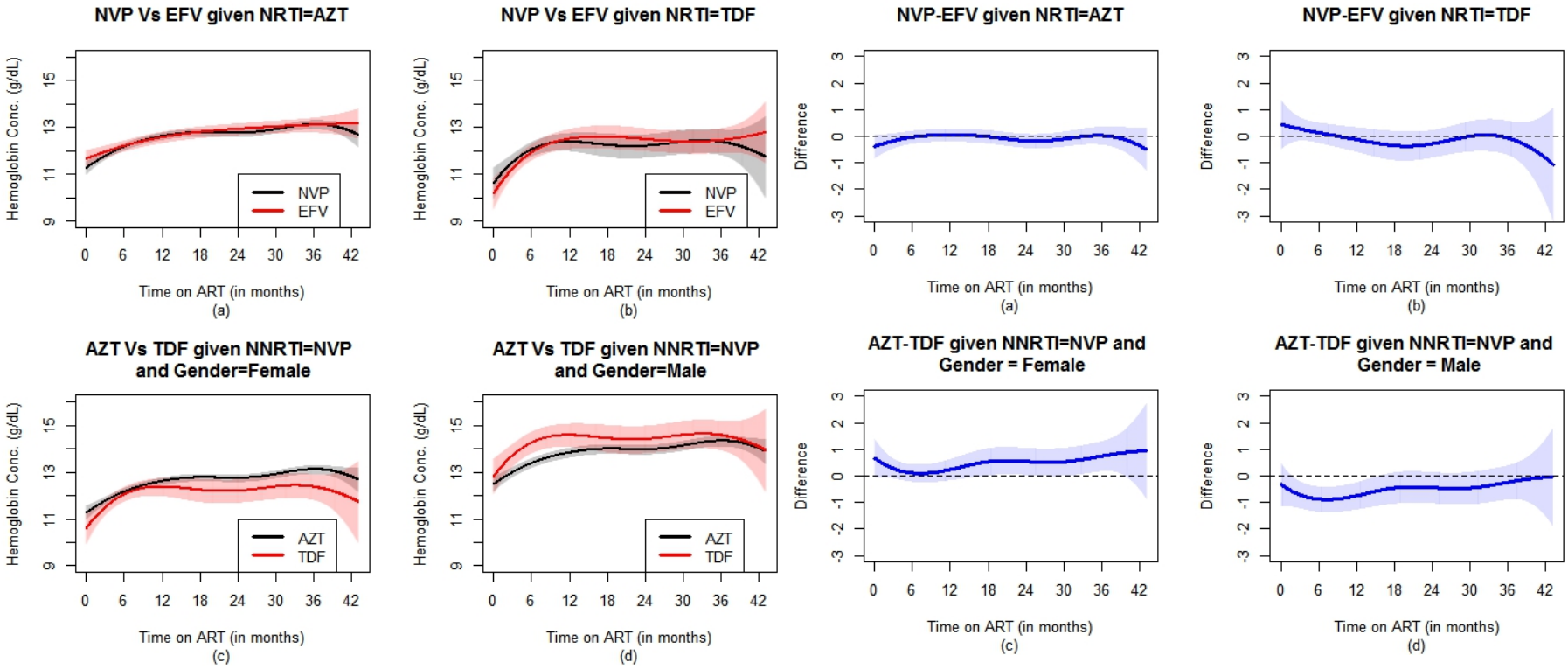
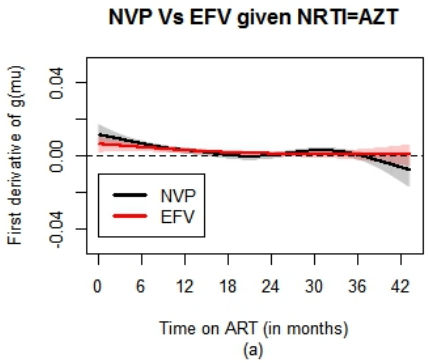
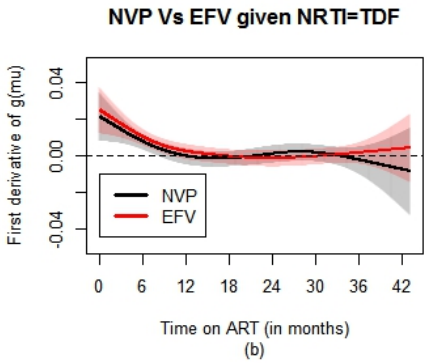
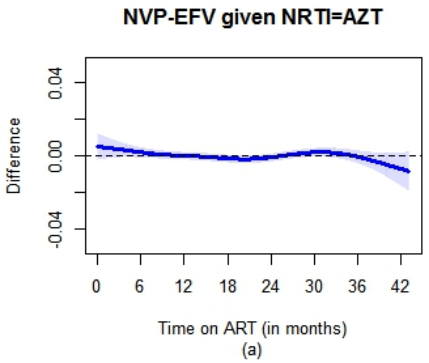


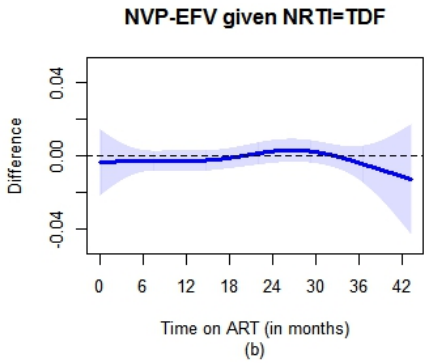
Figure 10: Fitted smoothing splines for the mean evolution of Hemoglobin concentration level (a) NVP versus EFV given NRTI is AZT + 3TC, (b) NVP versus EFV given NRTI is TDF + 3TC, (c) AZT versus TDF given NNRTI is NVP for female patients, and (d) AZT versus TDF given NNRTI is NVP for male patients. The shaded region is the 95% confidence band of the curve (pink shade is for the red line and grey shade is for the black line)

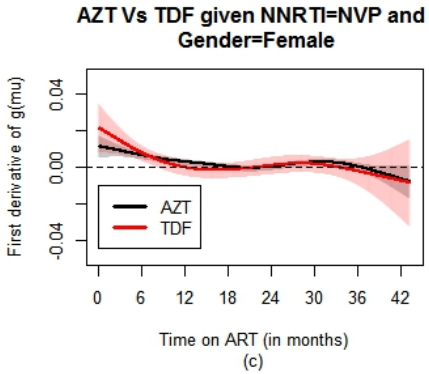
Figure 11: The difference in the smoothed mean evolution of Hemoglobin concentration level (a) NVP versus EFV given NRTI is AZT + 3TC, (b) NVP versus EFV given NRTI is TDF + 3TC, (c) AZT versus TDF given NNRTI is NVP for female patients, and (d) AZT versus TDF given NNRTI is NVP for male patients. The shaded region (light blue) is the 95% confidence band of the curve (blue line)

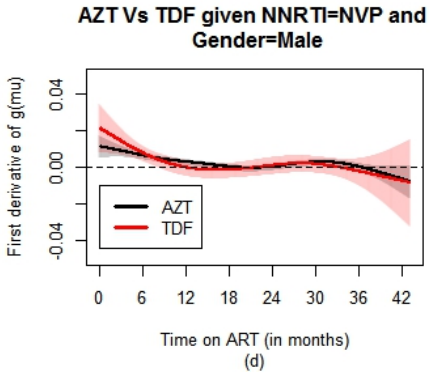


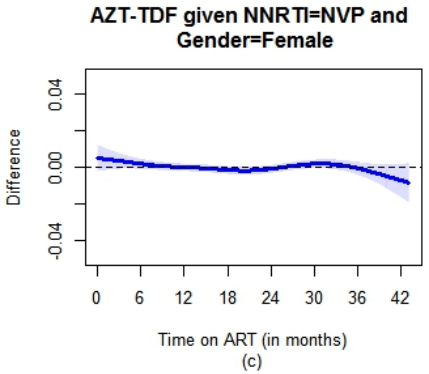












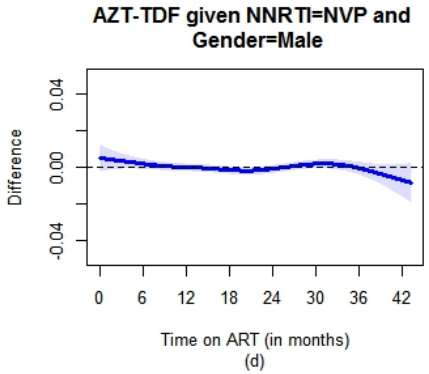


Figure 12: The smoothed first order derivative of the mean evolution of Hemoglobin concentration level in response to (a) NVP versus EFV given AZT + 3TC backbone, (b) NVP versus EFV given TDF + 3TC backbone, (c) AZT versus TDF given NNRTI is NVP for female patients, and (d) AZT versus TDF given NNRTI is NVP for male patients. The shaded region is the 95% confidence band of the curve (pink shade is for the red line and grey shade is for the black line)

Figure 13: The difference in the smoothed first order derivative of the mean evolution of Hemoglobin level (a) NVP versus EFV given AZT + 3TC backbone, (b) NVP versus EFV given TDF + 3TC backbone, (c) AZT versus TDF given NNRTI is NVP for female patients, and (d) AZT versus TDF given NNRTI is NVP for male patients. The shaded region (light blue) is the 95% confidence band of the curve (blue line)

4.4 Joint Modelling of Multivariate Longitudinal Data Using Mixed Effect Models

The model that best fits the joint evolution of log CD4+ cell count and Hemoglobin level is expressed mathematically in eq. (5) and the parameter estimates are shown in Table 6.

$$\begin{aligned} Y_{mi}(t) &= \beta_{10} + \beta_{11} * NRTI_{1i} + \beta_{12} * NNRTI_{1i} + \beta_{13} * NRTI_{1i} * NRTI_{1i} + \beta_{14} * Gender_{1i} \\ &+ \beta_{15} * Age_{1i} + \beta_{16} * WHOII_{1i} + \beta_{17} * WHOIII_{1i} \\ &+ \beta_{18} * t_1 + \beta_{19} * t_1^2 + \beta_{1,10} * t_1^3 + b_{10i} + b_{11i} * t_1 + \epsilon_{1i}(t_1) \\ &+ \beta_{20} + \beta_{21} * NRTI_{2i} + \beta_{22} * NNRTI_{2i} + \beta_{23} * Gender_{2i} + \beta_{24} * NRTI_{2i} * Gender_{2i} \\ &+ \beta_{25} * Age_{2i} + \beta_{26} * Age_{2i}^2 + \beta_{27} * t_2 + \beta_{28} * t_2^2 + \beta_{29} * t_2^3 + b_{20i} + b_{21i} * t_2 + \epsilon_{2i}(t_2) \quad (5) \end{aligned}$$

Where $Y_{mi}(t)$ is the value of the mth (m=1, 2) outcome from subject i (i=1, 2, ..., N) measured at time point t (m=1 for log CD4 counts and m=2 for Hemoglobin level). Let X be one of the covariates such as NRTI or Gender etc, then X_{im} is the value of the covariate for the i^{th} subject if he/she was measured for the m^{th} outcome. Based on AIC and log-likelihood ratio test this model had the best mean structure (AIC=32118.7) and it was equivalent to the independent models in terms of the fixed effect covariates included in the models so that it is easy to make comparison more efficiently. From the marginal models of each outcome random intercept and slope for the linear time effect were sufficient to account the correlated nature of the repeated measurements (see Table 7 in Appendix) and this result was also used in the joint modelling, that is, random intercept and slope were included for each response and Unstructured covariance was also assumed for the variability measurement of the random effects. The non linear evolution of the responses was captured with a three degree polynomial time effect, this was also the case in [1, 20] except that they modelled the response using Univariate mixed effect model. Another interesting part of the joint analysis was the parameter estimates and the standard errors were close to the corresponding results obtained from the independent analysis using thin-plate regression splines except that a small increase in the standard errors were observed in the estimates of joint analysis which was probably because the standard errors were adjusted for the correlation between the responses [10]. Baseline WHO clinical stage was also insignificant factor in influencing the evolution of Hemoglobin and hence partially dropped from the final model.

With respect to the effect of the factors, the results almost resemble to the results obtained from univariate analysis of each response. As displayed in Table 6, holding constant all other factors, NRTI showed significant effect on the evolution of both outcomes and had also significant interaction effect with NNRTI on the evolution of log CD4 counts (p-value = 0.0130) and with Gender on the evolution of Hemoglobin level (p-value = <0.0001). This was also true for each separate analysis except that now the main effect of NNRTI on the evolution of log CD4 counts was not significant (p-value = 0.2894). Given AZT + 3TC backbone, NVP and EFV showed no significant difference (effect difference = $\hat{\beta}_{12}$ = 0.0445, p-value = 0.2894). But the result from the univariate analysis for CD4 counts outcome (see Table

3) showed significant difference. Also given TDF + 3TC backbone, the evolution did not show again significant variation between NVP and EFV (effect difference (EFV - NVP) = $\hat{\beta}_{12} + \hat{\beta}_{13}$ = -0.1809, p-value = 0.0622). Similarly, given NVP, there was significant gap between AZT and TDF (effect difference = $\hat{\beta}_{11}$ = 0.2260, p-value = 0.0002), where AZT was associated to lower CD4 counts. On the other hand, given EFV there was no significant difference between AZT and TDF (effect difference (EFV - NVP) = $\hat{\beta}_{11} + \hat{\beta}_{13}$ = 0.0006, p-value = 0.4985). Regarding to Hemoglobin concentration outcome, again the effect of NNRTI was insignificant (p-value = 0.7474) and it had no even interaction effect with any of the other factors. Among female HIV-1 patients, subjects with ART that had AZT backbone showed higher level of Hemoglobin than those with TDF (effect difference (TDF - AZT) = $\hat{\beta}_{21}$ = -0.3402, p-value = 0.0045), where as for male patients AZT was significantly associated to lower level than TDF (effect difference (TDF - AZT) = $\hat{\beta}_{21} + \hat{\beta}_{24}$ = 0.6510, p-value = 0.0304). This result was actually in line with the univariate analysis for hemoglobin level data.

Similarly, baseline WHO clinical stages and age showed significant effect, that is, as it was discussed in the univariate analysis, late baseline WHO clinical stages were directly related to lower CD4 counts where as reverse linear association between baseline age and log CD4 counts was observed, also in the multivariate analysis it was noted again that the higher baseline age the lower log CD4 counts. Additionally, baseline age again showed downward parabolic association with Hemoglobin level suggesting lower Hemoglobin concentration among children and elderly HIV-1 patients compared to adult patients. In order to visualize this result the fitted average evolution of the two outcomes are shown visually in Figure 14 particularly for patients with median baseline age 33 and at baseline WHO clinical stage I. Each combination of NRTI, NNRTI, and gender represented with different colors. From Figure 14 (a) we can see that the evolution of CD4+ cell count had quite gap between female and male patients where the line associated to females was above the line associated to males. The predicted mean evolution of log CD4 counts showed a small gap between NVP and EFV when the backbone is AZT, where as relatively large gap was shown when the backbone was EFV for all genders but this was not significant. The gap was wider probably because the standard error of the difference was large. Similarly from Figure 14(b) we can notice that for female patients the gap between the evolution of Hemoglobin subject to AZT and TDF was smaller than the gap between AZT and TDF for male patients, where AZT was over TDF in the former one and TDF was over AZT in the later one, this was actually because of the interaction effect between gender and backbone of the ART regimen. Also we can see that the gap between NVP and EFV at each level of NRTI and gender was indistinguishable indicating no significant difference.

The other interesting result obtained from the joint analysis was the association structure between the two outcomes. The correlation between the evolution of the two outcomes was estimated to be 0.0471 (i.e $r_E = 0.0002/\sqrt{0.0129 \times 0.0014} = 0.0471$). This result reveals the weak linear association between the evolution of the two outcomes but the association may be non-linear. The marginal linear association between these two outcome at a given time point is also summarized in Figure 15. As it can

be seen from this plot, there was a positive but weak linear relationship in the first about 5 to 6 months and showed negative and very weak association as the red line crossed the threshold line (0) after about the seventh month.

Table 6: Parameter estimates for the joint mixed effect model.

log CD4+ cell count outcome				**Hemoglobin level outcome**			
Effect	**Estimate**	**Std. Er.**	**P-value**	**Effect**	**Estimate**	**Std. Er.**	**P-value**
Intercept 1	5.7154	0.0837	<.0001	**Intercept 2**	9.2613	0.1402	<.0001
NRTI				**NRTI**			
TDF	0.2260	0.0609	0.0002	**TDF**	-0.3402	0.1362	0.0045
AZT	0.0000			**AZT**	0.0000		
NNRTI				**NNRTI**			
EFV	0.0445	0.0420	0.2894	**EFV**	0.0284	0.0941	0.7474
NVP	0.0000			**NVP**	0.0000		
NRTI*NNRTI				**Gender**			
TDF*EFV	-0.2254	0.0907	0.013	**Male**	1.2280	0.0843	<.0001
Gender				**Female**	0.0000		
Male	-0.2197	0.0345	<.0001	**NRTI*Gender**			
Female	0.0000			**TDF*Male**	0.9912	0.2301	<.0001
Baseline Age	-0.0048	0.0017	0.0048	**Baseline Age**	0.1152	0.0085	<.0001
WHO Clinical Stages				**Baseline Age^2**	-0.0015	0.0001	<.0001
Stage II	-0.0747	0.0365	0.0407	**Time**	0.1417	0.0113	<.0001
Stage III and IV	-0.2868	0.1402	0.0409	**Time^2**	-0.0048	0.0006	<.0001
Stage I	0.0000			**Time^3**	0.00001	0.00001	<.0001
Time	0.0390	0.0107	0.0003				
Time^2	-0.0014	0.0005	0.0103				
Time^3	0.00002	0.00001	0.0257				
Variance components							
Var(ϵ_{1i})	0.8724	Var(b_{10i})	0.0457	Cov(b_{10i}, b_{11i})	-0.0555	Cov(b_{11i}, b_{21i})	0.0002
Var(ϵ_{2i})	1.2405	Var(b_{11i})	0.0129	Cov(b_{10i}, b_{20i})	0.0954	Cov(b_{20i}, b_{21i})	-0.0277
Cov(ϵ_{1i}, ϵ_{2i})	0.0291	Var(b_{20i})	2.2720	Cov(b_{10i}, b_{21i})	-0.0050		
		Var(b_{21i})	0.0014	Cov(b_{11i}, b_{20i})	0.0018		

Var=Variance *Cov=Covariance* *Std.Er=Standard Error*

When we compare the univariate analysis with the multivariate one, slight change in the effect of some factors was occurred such as NNRTI, that is, the main effect of NNRTI was significant in the univariate analysis for CD4 counts outcome model but in the case of multivariate analysis it was insignificant for the same outcome and hence this difference led us to make opposite conclusions about the difference between the evolution of CD4 counts subject to NVP and EFV at a particular type of backbone. To be specific, the significant difference between NVP and EFV observed in the univariate analysis for CD4 counts outcome was not significant in the multivariate analysis probably because the increase in standard error of the parameter estimates in multivariate setting. Also, the two approaches resulted quite different AIC values. The AIC from the univariate analysis for log CD4 counts and

hemoglobin level outcomes were 7810.63 and 20541.11 respectively, where as the multivariate analysis resulted AIC = 32118.7 which was larger compared the AIC from the univariate analysis but we can not make comparison, however, with the univariate analysis using AIC because the response variables in the two approaches had different structures. The residual variability was larger in joint analysis (0.7380) compared to the univariate analysis for CD4 counts outcome (0.4048) and smaller than the univariate analysis for hemoglobin level outcome (1.2716).

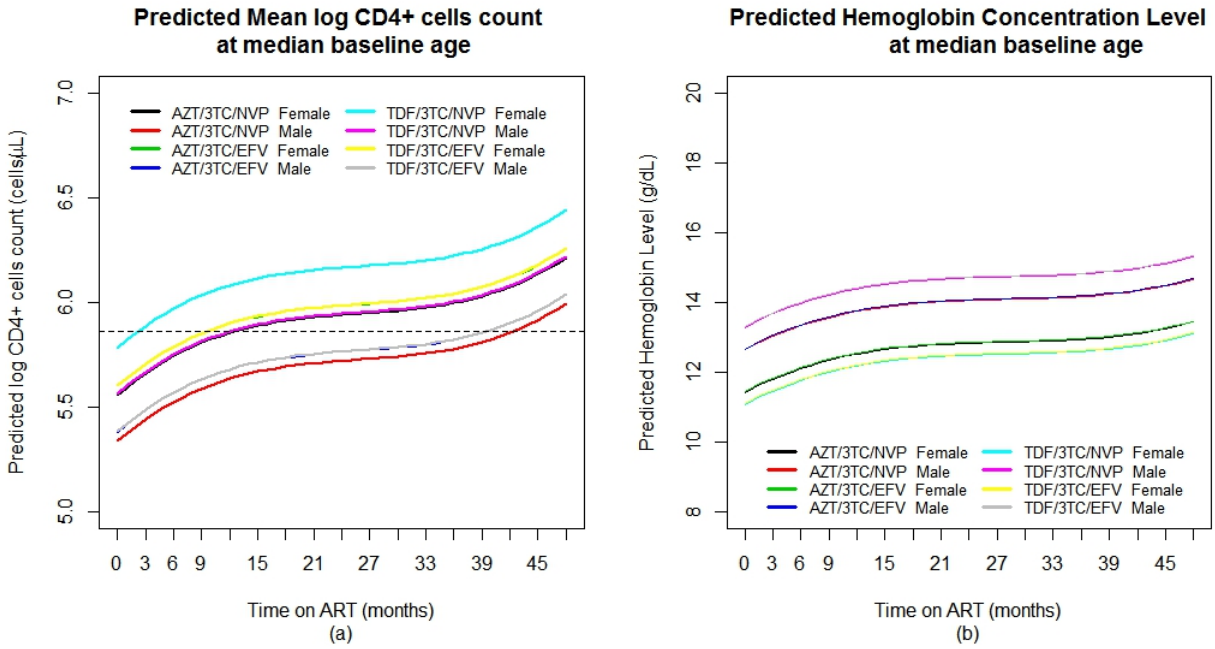


Figure 14: The predicted average evolution of CD4+ cell count (a) and Hemoglobin level (b) from joint modelling for patients with median baseline age 33 years and baseline CD4+ cell count 350 cells/μL at each combination of NRTI, NNRTI, and gender. Note that in (a) the yellow and green lines are overlapped, the purple and black lines are overlapped, also the grey and blue lines overlapped. For (b) the purple and grey lines are overlapped, the arctic and yellow lines are overlapped, the blue and red colors are overlapped, also the black and green line are overlapped.

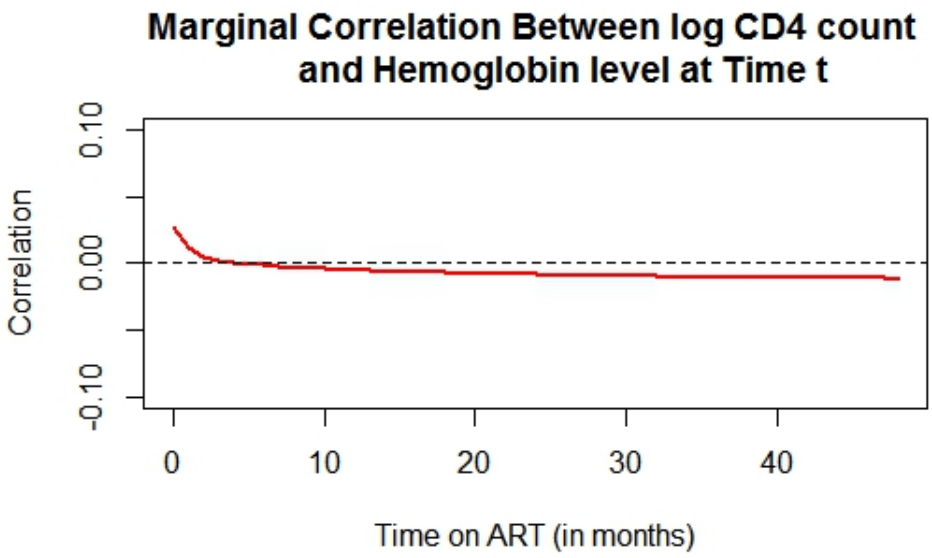


Figure 15: The marginal linear association between CD4+ cell count and Hemoglobin concentration level at each time point. The correlation is indicated with red solid line, where as the black dashed line is the reference line, which is 0 that indicates an absolute independence.

5 Discussion and Conclusion

One main goal of this study was to describe the evolution of CD4 cell count and Hemoglobin concentration level in HIV-1 patients on first line ART in Mildmay Uganda. Two approaches were implemented: semi-parametric mixed effect model for each outcome independently, and joint mixed effect modelling of the two outcomes together. Both approaches ended up with similar results except that the joint analysis added up another information about the association between the two outcomes through time.

From a general view, in the first few months (0 to 7 months) from ART initiation date, an increase in CD4 counts was observed and then it was also noted that from 6 to 9 months till the end of the study the evolution of CD4 cell count showed on average stable level almost above 350 cells /μL, which was the threshold during the study period. Also, Hemoglobin concentration level showed increasing trend in the first 7 to 14 months and then showed stable level around 12.5 g/dL, which is within the normal range of hemoglobin level for adults [30]. Patients who were taking ART that had AZT + 3TC backbone showed relatively longer period to reach at the threshold level of CD4 counts and Hemoglobin level compared to those who were taking ART which had TDF + 3TC backbone.

From both approaches we were able to get evidence on the difference between the evolution of CD4 counts in HIV-1 patients on first line ART subject to different regimens. Based on the univariate analysis the average evolution of CD4 counts (the analysis was based on logarithmic transformation) varied between NVP and EFV when the backbone is AZT + 3TC, where as no significant difference was observed when they are given with TDF + 3TC backbone. In this situation EFV was associated to higher evolution level of CD4 counts compared to NVP. This result was partially in line with the result obtained in [20], which discussed that there was interaction effect between NNRTI and time that indicates that the difference between the evolution of CD4 counts subject to NVP and EFV varies through time, but in our case time was modelled non-parametrically using thin-plate regression splines and from the smoothed mean evolution of CD4 counts subject to NVP and EFV an interaction with time was not seen, which could be shown with significant crossed predicted lines. However, in contrast to the univariate analysis and [20], the multivariate analysis showed no difference between NVP and EFV in describing the trajectory of CD4 cell counts. Comparing to the univariate analysis for Hemoglobin outcome, the result from the multivariate setting also confirmed the absence of variation in the evolution of Hemoglobin level between NVP and EFV. The difference between the backbone levels of ART in describing the evolution of both outcomes obtained in the univariate analysis was also supported by the joint analysis. Since the joint analysis was based on discretized time points which created artificial time scales, we conclude about the result from the joint analysis with caution.

CD4 cell count was also found evolving differently between women and men patients based on the result from both approaches. The evolution level was higher for female patients compared to males. This result also conforms to the result obtained in [1, 21], which discussed that men had the lowest mean

CD4 counts, but in [20] it was discussed that there was no difference. Also, consistent with findings in [21], that older men had fewer CD4+ cell per microliter on average when compared with the younger men, our study showed that treatment was less effective for elderly patients: This is expected, because it is well known from literature that immune function declines with age.

In addition, there was also statistical evidence that proved the difference between the evolution of Hemoglobin level subject to AZT and TDF which depends on the sex of patients. Women patients taking ART with AZT + 3TC backbone showed higher level, where as men patients taking ART with TDF + 3TC backbone showed higher level. Age had also showed an effect in changing the trend of Hemoglobin level. Age effect in particular showed that children and elderly patients had relatively lower level than middle age patients. When we compare this result with what discussed in [15], it was consistent in some aspects, that is, it was discussed that Hemoglobin level varies between sex, where women most of the time have lower level than men, as well as age but it was also mentioned that the effect of gender depends on age. In contrast to this, however, in our result, age and gender had no interaction effect in describing the trajectory of Hemoglobin concentration.

The multivariate modelling, on the other hand, added up an information about how CD4 cell counts and Hemoglobin concentration level evolve jointly in response to ART regimens taking into account baseline patient characteristics. The result was particularly very helpful to discover the time through association between the two outcomes of interest. Our result suggested that there was a weak linear association between the evolution of CD4 counts and Hemoglobin level which was consistent with the result obtained in [4] that discussed no significant association between CD4 counts and Hemoglobin level. The marginal correlation at a particular time point obtained in our result showed also that a weak positive association in the first 5 to 6 months and showed weak negative association between them after the 7th month from ART initiation day and this result was actually more close to the result obtained in [11].

Many different studies on hematology data for HIV-1 patients have used different predictor variables and some of them arrived on different results. Therefore we recommended that further studies on this topic include other important covariates that were not included in this study which could have a potential confounding effect. Such covariates include: viral load results, treatment failure, opportunistic infections and many others.

References

[1] Adams, M., Luguterah, A. (2013) Longitudinal analysis of change in CD4+ cell counts of HIV-1 patients on antiretroviral therapy (ART) in the Builsa district hospital. *Eouropean Scientific Journal* **9**(33) 1857-7881.

[2] Adewumi, AA., et al. (2014) Prevalence of HIV-related autoimmune haemolytic anaemia in Lagos, Nigeria. *Niger Med J.* Jan; **55**(1):63-6. DOI: 10.4103/0300-1652.128175.

[3] AIDSmeds. Retrieved July 7, 2015, from http://www.aidsmeds.com.

[4] Alavi, S.M., et al. (2007) Correlation between Total Lymphocyte Count, Hemoglobin, Hematocrit and CD4 Count in HIV/AIDS Patients. *Acta Medica Iranica*. **47**(1): 1-4.

[5] Barry O., Powell J., et al. (2013) Effectiveness of first-line antiretroviral therapy and correlates of longitudinal changes in CD4 and viral load among HIV-infected children in Ghana. *BMC Infectious Diseases*. **13**:476

[6] Centers for Disease Control and Prevention (CDC). 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR Recomm Rep. 1992 Dec 18; **41**(RR-17):1-19. Accessed August 2, 2015.

[7] Cleveland, W. S. (1993), *Visualizing Data*, Summit, NJ: Hobart Press.

[8] Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2002) *Analysis of Longitudinal Data*. New York: Oxford University Press, Inc.

[9] Fieuws, S., and Verbeke, G. (2004). Joint modeling of multivariate longitudinal profiles: pitfalls of the random effects approach. *Statististics in Medicine*, **23**: 3093-3104.

[10] Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2009). *Longitudinal Data Analysis*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Boca Raton: Chapman & Hall/CRC

[11] Fuchs, D., Shearer, GM., Boswell, RN., Lucey, DR., Clerici, M., Reibnegger, G., Werner, ER., Zajac, RA., Wachter, H. (1991) Negative correlation between blood cell counts and serum neopterin concentration in patients with HIV-1 infection. *Aids*; **5**(2):20912. DOI: 10.1097/00002030-199102000-00012.

[12] Funk, A., Kanters, S., Nansubuga, M., Mwehire, D., Featherstone, A., et al. (2012) Cohort Profile: The MUg Observational Cohort. *Int J Epidemiol*. **41**(16): 1594-1594f.

[13] Global Fact Sheet, Joint United Nations Programme on HIV and AIDS, 20 November 2012.

[14] Hammer, SM., Saag, MS., Schechter, M., Montaner, JS., Schooley, RT., Jacobsen, DM., Thompson, MA., Carpenter, CC., Fischl, MA., Gazzard, BG., et al. (2006) Treatment for adult HIV infection: 2006 recommendations of the international AIDS society-USA panel. *Jama*, **296**(7):827843.

[15] Hawkins, WW., Speck, E., Leonard, VG. (1954) Variation of the hemoglobin level with age and sex. *Blood 9*: 9991007.

[16] HIV AND AIDS UGANDA COUNTRY PROGRESS REPORT; 2013, 31 March, 2014.

[17] HIV Diagnostic Tests (HIV/AIDS Treatment Information). (n.d.). Retrieved July 7, 2015, from http://cig.salk.edu/extra_html/etc_hiv_diagonstics.htm

[18] HIV/AIDS Research. (n.d.). Retrieved July 7, 2015, from http://www.scripps.edu/discover/hiv-aids.html

[19] Ledergerber, B., Telenti, A., Effer, M. (1999) Risk of HIV related Kaposis sarcoma and non-Hodgkins lymphoma with potent antiretroviral therapy: Prospective cohort study. *BMJ.* **319**:23-24.

[20] Lubyayi, L., Shkedy, Z., Kajungu, D., Mwehire, D., Kirungi, C., et al. (2015) Evolution of CD4 Cell Counts over Time for HIV/AIDS Patients on Antiretroviral Therapy (ART) In Mildmay Uganda. *Biom Biostat Int J* **2**(2): 00025. DOI: 10.15406/bbij.2015.02.00025

[21] Mair, C., Hawes, S. E., Agne, H. D., Sow P. S., N'doye, I., Manhart, L. E., Gottlieb, G.S and Kiviat, N. B (2008). Factors associated with CD4+ lymphocyte counts in HIV - negative Senegalese individuals. *Clin Exp Immunol.* Vol **151**(3): 432-440.

[22] Palella, FJ. Jr, Delaney, KM., Moorman, AC., et al. (1998) Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection.*N Engl J Med.* **338**:853-860.

[23] Pinheiro, J.C., and Bates, D.M. (2000) *Mixed-Effects Models in S and S-PLUS*, Springer.

[24] Pushpike J. et al. (2011) The Use of Semi-parametric Mixed Models to Analyze PamChip Peptide Array Data: an Application to an Oncology Experiment. *doi:10.1093/bioinformatics/btr475.* Vol. **27** no. 20, pages 28592865

[25] Ruppert, D., Wand, M., and Carrol, R., editors (2003). *Semiparametric regression.* Cambridge University Press.

[26] Spaulding, A., Rutherford, GW., Siegfried, N., (2010) Tenofovir or zidovudine in threedrug combination therapy with one nucleoside reverse transcriptase inhibitor and one nonnucleoside reverse transcriptase inhibitor for initial treatment of HIV infection in antiretroviralnave individuals. *Cochrane Database of Systematic Reviews.* Issue 10. Art. No.: CD008740. DOI: 10.1002/14651858.CD008740.

[27] Treatment and care. (n.d.). Retrieved July 7, 2015, from http://www.who.int/hiv/topics/treatment/en/

[28] UNAIDS (2013). Access to Antiretroviral Therapy in Africa: Status Report on Progress Towards the 2015 Targets.

[29] Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag, Inc.

[30] Volberding, PA., Bake,r KR., Levine, AM. (2003) Human immunodeficiency virus hematology. *Hematology Am Soc Hematol Educ Program*. 294-313.

[31] Watts, S.A., Daly, B., et al. (2001) The effect of age, gender, risk level and glycosylated hemoglobin in predicting foot amputation in HMO patients with diabetes. *J Am Acad Nurse Pract.*; **13**(5):230-5.

[32] WHO (2013, June) The Use of Antiretroviral Drugs for Treating and Preventing HIV infection.

[33] Wood S.N. (2006) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press.

[34] World Population by continents and countries - Nations Online Project. Retrieved March 18, 2015, from http://www.nationsonline.org/oneworld/world_population.htm

Appendix

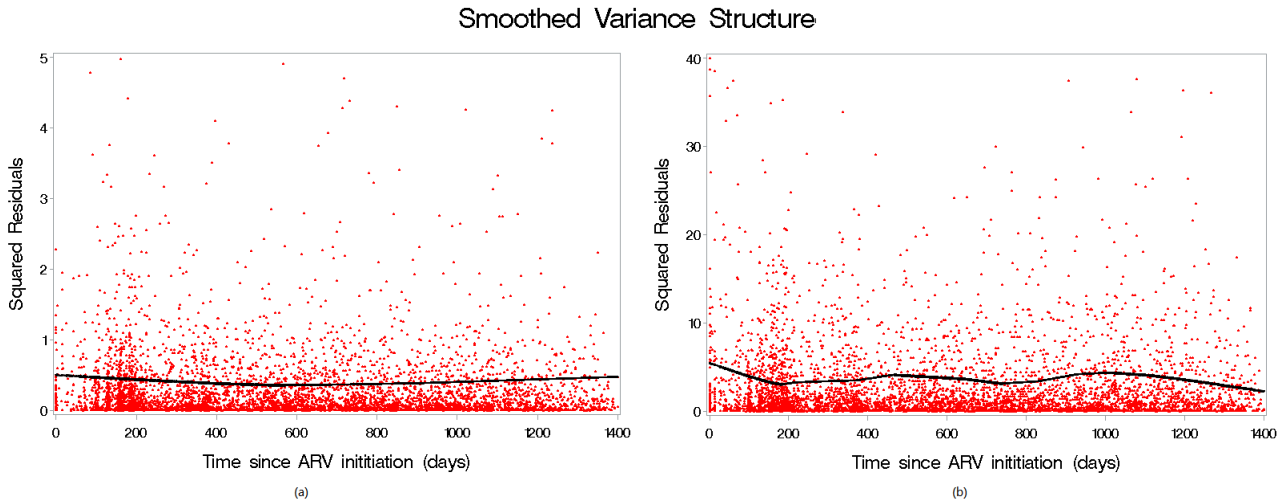


Figure 16: Smoothed variance structure for (a) Log CD4 cell count, and (b) Hemoglobin level

Table 7: Testing the need for random effects

		Model for log CD4 Count		Model for Hemoglobin	
	Random Effect	Test-Statitic	Pvalue	Test-Statitic	Pvalue
1	bo	1814.49	<0.0001	1279.42	<0.0001
2	**bo + b1*Time**	53.06	<0.0001	14.53	<0.0001
3	bo + b1*Time + b2*Time^2	1.64	0.5454	4.00	0.2000

Table 8: Comparing different covariance structures of the random effects

	Model for log CD4 Count				Model for Hemoglobin			
Covariance structure	df	AIC	BIC	ML-logLik	df	AIC	BIC	ML-logLik
Diagonal/Simple	17.00	7847.81	7959.02	-3906.91	17	20542.33	20654.66	-10254.16
Unstructured/CS	18.00	7810.63	7928.38	-3887.32	18	20541.11	20660.05	-10252.55

Table 9: Comparing different serial correlation structures

		Model for log CD4 Count	Model for Hemoglobin
	Seria_Correlation	REML_LogLik	REML_LogLik
1	**No Ser. Corr**	-3912.45	-10086.34
2	Sp. Exponential	-3904.77	-10086.34
3	Sp. Gaussian	-3905.23	-10086.34

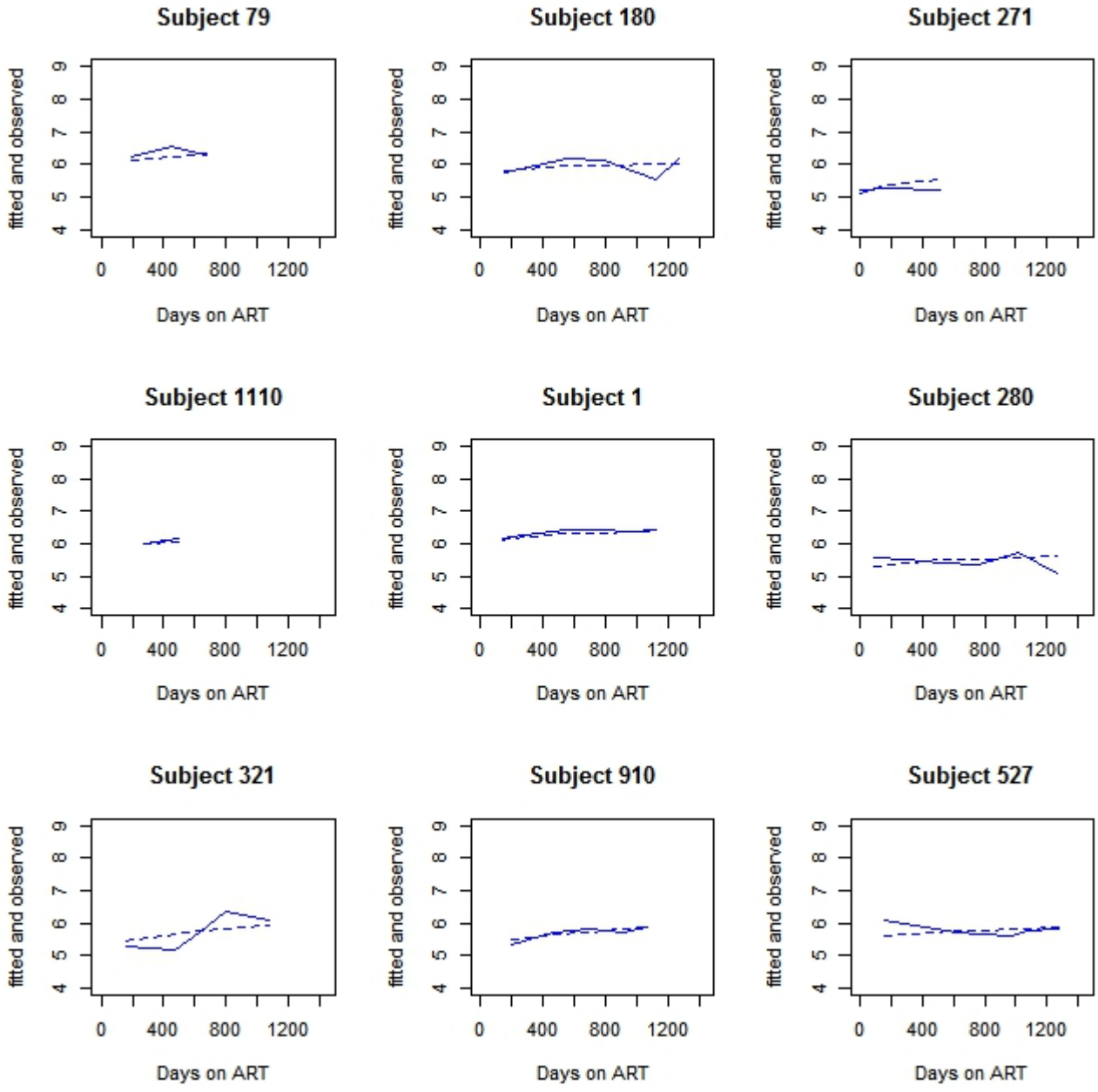


Figure 17: Fitted versus observed trajectory of log CD4 cell count for randomly selected 9 patients. The solid blue line represents the observed log CD4 counts and the dashed blue line is for the predicted trajectory

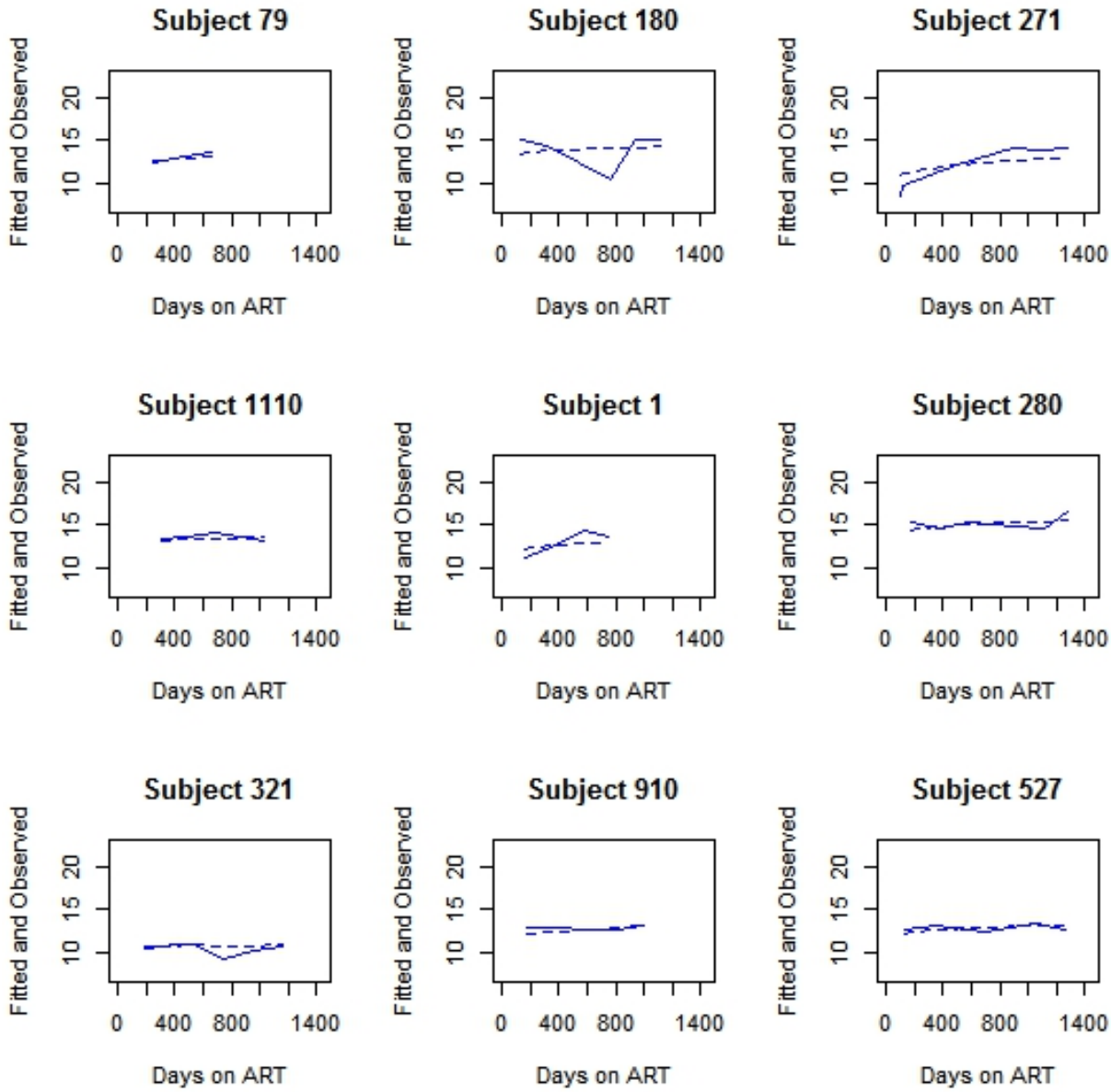


Figure 18: Fitted versus observed trajectory of Hemoglobin concentration level for randomly selected 9 patients. The solid blue line represents the observed Hemoglobin level and the dashed blue line is for the predicted trajectory

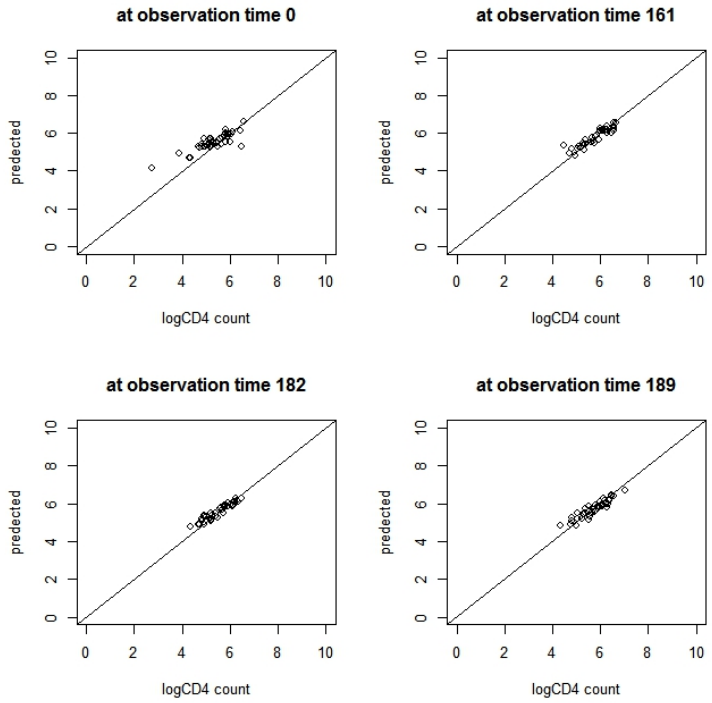


Figure 19: Fitted versus observed trajectory of log CD4 cell count at randomly selected 4 time points.

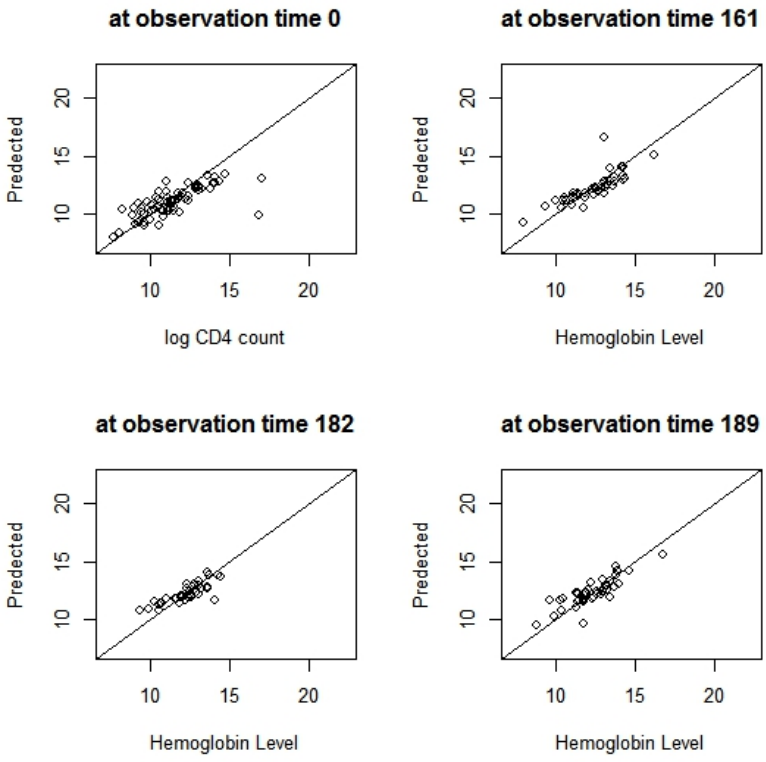


Figure 20: Fitted versus observed trajectory of Hemoglobin level at randomly selected 4 time points.

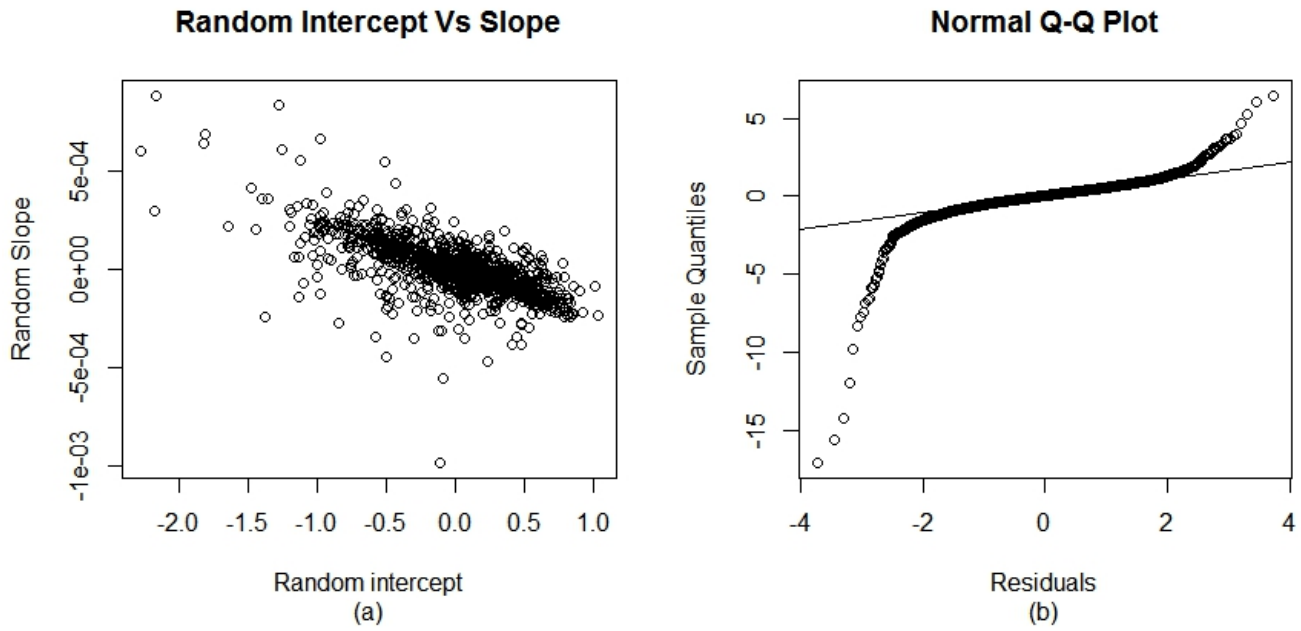


Figure 21: (a) Scatter plot between random intercept and slope estimates for CD4 counts data model, (b) A normal quantile plot of the standardized LS level-1 residuals. The large deviations from the reference line indicates that we are dealing with a very heavy-tailed distribution.

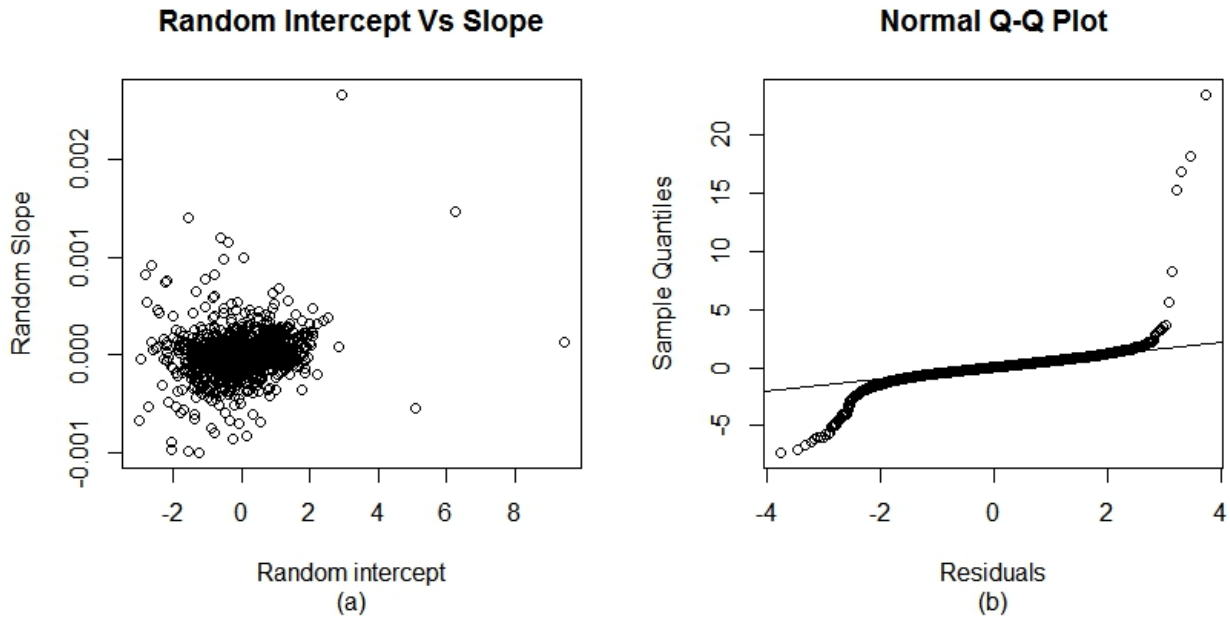


Figure 22: (a) Scatter plot between random intercept and slope estimates for Hemoglobin concentration level data model, (b) A normal quantile plot of the standardized LS level-1 residuals. The large deviations from the reference line indicates that we are dealing with a very heavy-tailed distribution.

Codes Used in the Analysis

```
#-------------------------- CD4 Cells count ----------------------------#
#-----------------------------------------------------------------------#
CD4=read.table(choose.files(), header=T, sep="&")

#Creating new ID variable with sequence of numbers
CD4$ID=NULL
CD4$ID[1]=1
for(i in 2:dim(CD4)[1]){
if(CD4$PTIDNO[i]==CD4$PTIDNO[i-1]) CD4$ID[i] = CD4$ID[i-1]
else if(CD4$PTIDNO[i]!=CD4$PTIDNO[i-1]) CD4$ID[i] =CD4$ID[i-1] +1
}
CD4$time2=CD4$time*CD4$time

#Creating new WHO stages
WHOs=NULL
for(i in 1:dim(CD4)[1]){
if(CD4$ARTWHOStage[i]=="T1") WHOs[i]="T1"
else if(CD4$ARTWHOStage[i]=="T2") WHOs[i]="T2"
else if(CD4$ARTWHOStage[i]=="T3" |
CD4$ARTWHOStage[i]=="T4") WHOs[i]="T34"
}
CD4$WHOs=factor(WHOs)

#ART Regimens
R=NULL
for(i in 1:dim(CD4)[1]){
if(CD4$NRTI[i]=="AZT" & CD4$NNRTI[i]=="NVP") R[i]=1
else if(CD4$NRTI[i]=="AZT" & CD4$NNRTI[i]=="EFV") R[i]=2
else if(CD4$NRTI[i]=="TDF" & CD4$NNRTI[i]=="NVP") R[i]=3
else if(CD4$NRTI[i]=="TDF" & CD4$NNRTI[i]=="EFV") R[i]=4
}
CD4$ART=factor(R)

#-------------------------EDA, Profile Plots ----------------------#
#Overall Mean profile
win.graph()
par(mfrow=c(1,2))
plot(CD4$time, CD4$logCD4, type="n", xlab="Time since ARV start date (in days)",
ylab="Log CD4 Count (cells/microlitre)", sub="(a)",
main="Trajectory of CD4 Cells Count in HIV 1 Patients")
for(i in 1:dim(CD4)[1]){
lines(CD4$time[CD4$ID==i],
CD4$logCD4[CD4$ID==i], col="grey")
}

#*************** SAS Program Smoothed mean profile *****************#
proc sort data=thesis.final3_CD4 out=sorted;
by CurrentRegimen;
run;
proc loess data=sorted PLOTS(MAXPOINTS=100000);
by CurrentRegimen;
ods output scoreresults=out1;
model logCD4=time;
score data=sorted;
run;
proc sort data=out1;
by time CurrentRegimen;
run;

proc sort data=thesis.final3_CD4 out=sorted2;
by Gender;
run;
proc loess data=sorted2 PLOTS(MAXPOINTS=100000);
by Gender;
ods output scoreresults=out1b;
model logCD4=time;
score data=sorted2;
run;
proc sort data=out1b;
by time Gender;
run;
```

```
#*****************************************************************#

#Importing PROC LOESS result from SAS program
LWsmoothCD4=read.table(choose.files(), header=T, sep="&")
head(LWsmoothCD4)
lines(LWsmoothCD4$time, LWsmoothCD4$p_logCD4, lwd=3, col="blue")
legend(200,-2,c("Loess-Smooth mean of Log CD4 count"), lty=1, lwd=3, col="blue")


#------------------Semi-Parametric Mixed Model Analysis -----------#
library(mgcv)

#Assigning reference groups
CD4$Gender=relevel(CD4$Gender, ref="Fema")
CD4$NRTI=relevel(CD4$NRTI, ref="AZT")
CD4$NNRTI=relevel(CD4$NNRTI, ref="NVP")
CD4$ARTWHOStage=relevel(CD4$ARTWHOStage, ref="T1")
CD4$WHOs=relevel(CD4$WHOs, ref="T1")

#Comparing Different Mean structures
fit1=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI + Gender+ Age_at_ARVstart + WHOs,
data=CD4, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))

fit2=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI + Gender+ Gender*NRTI + Gender*NNRTI +
Age_at_ARVstart+ WHOs,
data=CD4, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))

fit3=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI + Gender+ Gender*NRTI + Gender*NNRTI +
Age_at_ARVstart+Age_at_ARVstart*NRTI + Age_at_ARVstart*NNRTI +
WHOs, data=CD4, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))

fit4=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI + Gender+ Gender*NRTI + Gender*NNRTI +
Age_at_ARVstart+Age_at_ARVstart*NRTI + Age_at_ARVstart*NNRTI +
WHOs + WHOs*NRTI + WHOs*NNRTI,
data=CD4, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))

#..... #Other models are omitted in the report

anova(fit1$lme, fit2$lme, fit3$lme, fit4$lme)
anova(fit1$lme, fit2$lme)
anova(fit1$lme, fit3$lme)
anova(fit1$lme, fit4$lme)

#The need for Random Effect Test (mixture chi-square test)
fit1o=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=CD4, method="REML", correlation=NULL, random=NULL)

fit1a=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=CD4, method="REML", correlation=NULL,
random=list(ID=pdSymm(~1)))
chi1=-2*(fit1o$lme$logLik-fit1a$lme$logLik)
pval1=0.5*pchisq(chi1,0, lower.tail=F)+0.5*pchisq(chi1,1, lower.tail=F)

fit1b=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=CD4, method="REML", correlation=NULL,
random=list(ID=pdSymm(~time)))

chi2=-2*(fit1a$lme$logLik-fit1b$lme$logLik)
pval2=0.5*pchisq(chi2,1, lower.tail=F)+0.5*pchisq(chi2,2, lower.tail=F)

fit1c=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=CD4, method="REML", correlation=NULL,
```

```
random=list(ID=pdDiag(~time+time2)))

chi3=-2*(fit1b$lme$logLik-fit1c$lme$logLik)
chi3=0 #b/s the LL difference was -ve
pval3=0.5*pchisq(chi3,2, lower.tail=F)+0.5*pchisq(chi3,3, lower.tail=F)
table1=data.frame(Test=c("bo", "bo + b1*Time", "bo + b1*Time + b2*Time^2"),
Statitic=c(chi1, chi2, chi3), Pvalue=c(pval1, pval2, pval3))
table1
#Reducing Covariance Structure based on AIC (method=ML)
fit2a=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=CD4, method="ML", correlation=NULL,
random=list(ID=pdDiag(~time))) #Simple structure

fit2b=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=CD4, method="ML", correlation=NULL,
random=list(ID=pdSymm(~time))) #Unstructured

table2=data.frame(anova(fit2a$lme, fit2b$lme, test=F))
table2

#Testing Serial Correlation based on REML-likelihood
fit3a=fit1b #No serial correlation
fit3b=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=CD4, method="REML", correlation=corExp(),
random=list(ID=pdSymm(~time))) #Spatial Exponential erial correlation
fit3c=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=CD4, method="REML", correlation=corGaus(),
random=list(ID=pdSymm(~time))) #Spatial gaussian serial correlation

table3=data.frame(Seria_Correlation=c("No Ser. Corr", "Sp. Exponential", "Sp. Gaussian"),
REML_LogLik=c(fit3a$lme$logLik, fit3b$lme$logLik, fit3c$lme$logLik))
table3
#####

Final=gamm(logCD4~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI +
Gender+
Age_at_ARVstart+
WHOs,
data=CD4, method="REML", correlation=NULL,
random=list(ID=pdSymm(~time)))

summary(Final$gam)
summary(Final$lme)

########### Contrast ################
library(multcomp)
library(mvtnorm)

f1=Final$lme
f2=Final$gam

length(coef(f2))
names(coef(f2))

contrast.matrix1 <- rbind("NVP vs. EFV given NRTI= AZT" =  c(0, 0, 1, rep(0, times=45)))
contrast.matrix2 <- rbind("NVP vs. EFV given NRTI= TDF" =  c(0, 0, 1, 0, 0, 0, 0, 1, rep(0, times=40)))
contrast.matrix3 <- rbind("AZT vs. TDF given NNRTI=NVP" =  c(0, 1,  rep(0, times=46)))
contrast.matrix4 <- rbind("AZT vs. TDF given NNRTI=EFV" =  c(0, 1, 0, 0, 0, 0, 0, 1, rep(0, times=40)))

summary(glht(f2, contrast.matrix1))
summary(glht(f2, contrast.matrix2))
summary(glht(f2, contrast.matrix3))
summary(glht(f2, contrast.matrix4))



#------------------------ Smoothed Mean Plots-----------------------------#
GammObj1<-Final
summary(GammObj1$lme)
```

```
summary(GammObj1$gam)

###-------------------------predict-----------------------------------------------
yfit1<-predict(GammObj1$lme, asList=F, level = 0:1) #BLUP
#names(yfit1)

t=seq(0,1439, 1)
time=rep(t, times=24)
v=length(t)
age=rep(33, length(time))
#blc=rep(median(CD4$blcd4, na.rm=T), length(time))
#who=rep(rep(c("T1", "T2", "T3", "T4"), each=v), times=8)
who=rep(rep(c("T1", "T2", "T34"), each=v), times=8)
sex=rep(rep(c("Fema", "Male"), each=v*3), times=4)
nnrti=rep(rep(c("NVP", "EFV"), each=6*v), times=2)
nrti=rep(rep(c("AZT", "TDF"), each=12*v), times=1)
hlp1<-data.frame(time=time, NRTI=nrti, NNRTI=nnrti, Gender=sex, WHOs=who,
Age_at_ARVstart=age)# sequence for time and ranges from 0 to 60
R=NULL
for(i in 1:dim(hlp1)[1]){
if(hlp1$NRTI[i]=="AZT" & hlp1$NNRTI[i]=="NVP") R[i]=1
else if(hlp1$NRTI[i]=="AZT" & hlp1$NNRTI[i]=="EFV") R[i]=2
else if(hlp1$NRTI[i]=="TDF" & hlp1$NNRTI[i]=="NVP") R[i]=3
else if(hlp1$NRTI[i]=="TDF" & hlp1$NNRTI[i]=="EFV") R[i]=4
}
hlp1$ART=factor(R)

#ylogfit<-CD4average$ylogm# observed mean at each time point

yfitfixef1<-predict(GammObj1$gam,hlp1, se.fit=T)# is it the mean at each time point?

hlp1a=data.frame(hlp1, yfitfixed=as.data.frame(yfitfixef1))
hlp1a=data.frame(hlp1a, lower=hlp1a$yfitfixed.fit-1.96*(hlp1a$yfitfixed.se.fit)^1.0,
upper=hlp1a$yfitfixed.fit+1.96*(hlp1a$yfitfixed.se.fit)^1.0)
##############################################################################
addTrans <- function(color,trans)
{
# This function adds transparancy to a color.
# Define transparancy with an integer between 0 and 255
# 0 being fully transparant and 255 being fully visable
# Works with either color and trans a vector of equal length,
# or one of the two of length 1.

if (length(color)!=length(trans)&!any(c(length(color),length(trans))==1)) stop("Vector lengths not correct")
if (length(color)==1 & length(trans)>1) color <- rep(color,length(trans))
if (length(trans)==1 & length(color)>1) trans <- rep(trans,length(color))

num2hex <- function(x)
{
hex <- unlist(strsplit("0123456789ABCDEF",split=""))
return(paste(hex[(x-x%%16)/16+1],hex[x%%16+1],sep=""))
}
rgb <- rbind(col2rgb(color),trans)
res <- paste("#",apply(apply(rgb,2,num2hex),2,paste,collapse=""),sep="")
return(res)
}
##############################################################################
win.graph()
par(mfrow=c(2,2))
plot(c(0:1439), seq(min(hlp1a$lower), max(hlp1a$upper), length=1440), type="n", sub="(a)",
main="NVP Vs EFV given NRTI=AZT",ylab="Log(CD4 Ct. (cells/microliters))",
xlim=c(0,1440), ylim=c(5, 7),  xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" & hlp1a$WHOs=="T1"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" & hlp1a$WHOs=="T1"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" & hlp1a$WHOs=="T1"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" & hlp1a$WHOs=="T1"],
lwd=2, col=addTrans("black",5))
segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" & hlp1a$WHOs=="T1"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" & hlp1a$WHOs=="T1"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" & hlp1a$WHOs=="T1"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" & hlp1a$WHOs=="T1"],
lwd=2, col=addTrans("red",5))
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" & hlp1a$WHOs=="T1"],
```

```
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" & hlp1a$WHOs=="T1"],
lwd=2, col=1)
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" & hlp1a$WHOs=="T1"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV"& hlp1a$WHOs=="T1"],
lwd=2, col=2)
abline(h=5.86, lty=2)
legend(950, 5.4, c("NVP", "EFV"), lty=1, col=c(1,2),lwd=3)

plot(c(0:1439), seq(min(hlp1a$lower), max(hlp1a$upper), length=1440), type="n", sub="(b)",
main="NVP Vs EFV given NRTI=TDF",ylab="Log(CD4 Ct. (cells/microliters))",
xlim=c(0,1440), ylim=c(5, 7),   xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))

segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
lwd=2, col=addTrans("black",5))

segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
lwd=2, col=addTrans("red",5))
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
lwd=2, col=1)
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
lwd=2, col=2)
abline(h=5.86, lty=2)
legend(950, 5.4, c("NVP", "EFV"), lty=1, col=c(1,2),lwd=3)

plot(c(0:1439), seq(min(hlp1a$lower), max(hlp1a$upper), length=1440), type="n", sub="(c)",
main="AZT Vs TDF given NNRTI=NVP",ylab="Log(CD4 Ct. (cells/microliters))",
xlim=c(0,1440), ylim=c(5, 7),   xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
lwd=2, col=addTrans("black",5))
segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
lwd=2, col=addTrans("red",5))
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
lwd=2, col=1)
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
lwd=2, col=2)
abline(h=5.86, lty=2)
legend(950, 5.4, c("AZT", "TDF"), lty=1, col=c(1,2),lwd=3)

plot(c(0:1439), seq(min(hlp1a$lower), max(hlp1a$upper), length=1440), type="n", sub="(d)",
main="AZT Vs TDF given NNRTI=EFV",ylab="Log(CD4 Ct. (cells/microliters))",
xlim=c(0,1440), ylim=c(5, 7),   xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP" &  hlp1a$WHOs=="T1"],
lwd=2, col=addTrans("black",5))
segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
lwd=2, col=addTrans("red",5))
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
lwd=2, col=1)
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
```

```
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV" &  hlp1a$WHOs=="T1"],
lwd=2, col=2)
abline(h=5.86, lty=2)
legend(950, 5.4, c("AZT", "TDF"), lty=1, col=c(1,2),lwd=3)



#-------------------------------------Model 1: first order derivatives ------------------------#
t.mesh<-seq(0,1439,1)
delta<-1e-5
hlp2a=hlp1
hlp2a$time<-hlp1$time-delta
X01a<-predict(GammObj1$gam,hlp2a,type="lpmatrix")#?
hlp2a$time<-hlp1$time+delta
X11a<-predict(GammObj1$gam,hlp2a,type="lpmatrix")#?
Xp1a<-(X11a-X01a)/delta
v1a<-Xp1a%*%GammObj1$gam$coef
v.sd1a<-rowSums(Xp1a%*%GammObj1$gam$Vp*Xp1a)^.5
#XDer1<-cbind(0,1,2*t.mesh)
#DerRan1<-GammObj$lme$coef$random$ID%*%t(XDer1)# the deriative of the random effect

hlp3a=data.frame(hlp1, Deriv=v1a, lowerDer=v1a-1.96*v.sd1a, upperDer=v1a+1.96*v.sd1a)
head(hlp3a)

win.graph()
par(mfrow=c(2,2))
plot(c(0:1439), seq(min(hlp3a$lowerDer), max(hlp3a$upperDer), length=1440), type="n", sub="(a)",
xlim=c(0,1440), ylab="First derivative of g(mu)", main="NVP Vs EFV given NRTI=AZT",
ylim=c(-0.02, 0.02), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
lwd=2, col=addTrans("black", 5))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
lwd=2, col=addTrans("red", 5))
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
lwd=2, col=1)
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
lwd=2, col=2)
abline(h=0, lty=2)
legend(0, -0.01, c("NVP", "EFV"), lty=1, col=c(1,2),lwd=3)


plot(c(0:1439), seq(min(hlp3a$lowerDer), max(hlp3a$upperDer), length=1440), type="n", sub="(b)",
xlim=c(0,1440), ylab="First derivative of g(mu)",
main="NVP Vs EFV given NRTI=TDF", ylim=c(-0.02, 0.02), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
lwd=2, col=addTrans("black", 5))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
lwd=2, col=addTrans("red", 5))
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
lwd=2, col=1)
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
lwd=2, col=2)
abline(h=0, lty=2)
legend(0, -0.01, c("NVP", "EFV"), lty=1, col=c(1,2),lwd=3)

plot(c(0:1439), seq(min(hlp3a$lowerDer), max(hlp3a$upperDer), length=1440), type="n", sub="(c)",
```

```
xlim=c(0,1440), ylab="First derivative of g(mu)",
main="AZT Vs TDF given NNRTI=NVP", ylim=c(-0.02, 0.02), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))

segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
lwd=2, col=addTrans("black", 5))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
lwd=2, col=addTrans("red", 5))
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
lwd=2, col=1)
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP" & hlp3a$WHOs=="T1"],
lwd=2, col=2)
abline(h=0, lty=2)
legend(0, -0.01, c("AZT", "TDF"), lty=1, col=c(1,2),lwd=3)


plot(c(0:1439), seq(min(hlp3a$lowerDer), max(hlp3a$upperDer), length=1440), type="n", sub="(d)",
xlim=c(0,1440), ylab="First derivative of g(mu)", main="AZT Vs TDF given NNRTI=EFV",
ylim=c(-0.02, 0.02), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
lwd=2, col=addTrans("black", 5))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
lwd=2, col=addTrans("red", 5))
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
lwd=2, col=1)
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV" & hlp3a$WHOs=="T1"],
lwd=2, col=2)
abline(h=0, lty=2)
legend(0, -0.01, c("AZT", "TDF"), lty=1, col=c(1,2),lwd=3)



#------------------------- Difference in CD4 Profiles ---------------------------------#

t=seq(0,1439, 2)
time=rep(t, times=24)
v=length(t)
age=rep(33, length(time))
#blc=rep(median(blc, na.rm=T), length(time))
#who=rep(rep(c("T1", "T2", "T3", "T4"), each=v), times=8)
who=rep(rep(c("T1", "T2", "T34"), each=v), times=8)
sex=rep(rep(c("Fema", "Male"), each=v*3), times=4)
nnrti=rep(rep(c("NVP", "EFV"), each=6*v), times=2)
nrti=rep(rep(c("AZT", "TDF"), each=12*v), times=1)
hlp1z<-data.frame(time=time, NRTI=nrti, NNRTI=nnrti, Gender=sex,
Age_at_ARVstart=age, WHOs=who)# sequence for time and ranges from 0 to 60
Rz=NULL
for(i in 1:dim(hlp1z)[1]){
if(hlp1z$NRTI[i]=="AZT" & hlp1z$NNRTI[i]=="NVP") Rz[i]=1
else if(hlp1z$NRTI[i]=="AZT" & hlp1z$NNRTI[i]=="EFV") Rz[i]=2
else if(hlp1z$NRTI[i]=="TDF" & hlp1z$NNRTI[i]=="NVP") Rz[i]=3
else if(hlp1z$NRTI[i]=="TDF" & hlp1z$NNRTI[i]=="EFV") Rz[i]=4
}
hlp1z$ART=factor(Rz)
############

Xp02<-predict(GammObj1$gam,hlp1z,type="lpmatrix")#?
```

```
fv3=Xp02%*%GammObj1$gam$coef
#fv3=as.vector(predict(GammObj1$gam, hlp1z))
Vall3=Xp02%*%GammObj1$gam$Vp%*%t(Xp02)

win.graph()
par(mfrow=c(2,2))

#NVP - EFV
def2=NULL
sddef2=NULL
tp=1    # you can change depending on the group you want to test
L=c(1,-1)
for(i in 1:(length(t)*12)){
ind.extract <- c(tp, (6*length(t)+tp))
VarVini2 <- Vall3[ind.extract, ind.extract]
VintE <- c(fv3[ind.extract[1]], fv3[ind.extract[2]])
def2[i]=L %*% VintE
sddef2[i]=sqrt(t(L) %*% VarVini2 %*%L)
if(i==length(t)*6) tp=tp+(length(t)*6+1)
else tp=tp+1
}


NVP_EFV=data.frame(Time=rep(t, times=12),
NRTI=hlp1z$NRTI[c(1:(length(t)*6), (length(t)*12+1):(length(t)*18)) ],
Gender=hlp1z$Gender[c(1:(length(t)*6), (length(t)*12+1):(length(t)*18)) ],
WHOs=hlp1z$WHOs[c(1:(length(t)*6), (length(t)*12+1):(length(t)*18)) ],
Dif=def2, SD.Dif=sddef2, lb=def2-1.96*sddef2, ub=def2+1.96*sddef2)

plot(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT" & NVP_EFV$WHOs=="T1"],
NVP_EFV$Dif[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT" & NVP_EFV$WHOs=="T1"],
type="n",  ylab="Difference", main="NVP-EFV given NRTI=AZT", sub="(a)",
ylim=c(-3,3), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT" & NVP_EFV$WHOs=="T1"],
NVP_EFV$lb[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT" & NVP_EFV$WHOs=="T1"],
NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT" & NVP_EFV$WHOs=="T1"],
NVP_EFV$ub[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT" & NVP_EFV$WHOs=="T1"],
col=addTrans("blue",16))
lines(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT" & NVP_EFV$WHOs=="T1"],
NVP_EFV$Dif[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT" & NVP_EFV$WHOs=="T1"],
col="blue",lwd=3)
abline(h=0, lty=2)

plot(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF" & NVP_EFV$WHOs=="T1"],
NVP_EFV$Dif[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF" & NVP_EFV$WHOs=="T1"],
type="n",ylab="Difference", sub="(b)", main="NVP-EFV given NRTI=TDF",
ylim=c(-3,3), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF" & NVP_EFV$WHOs=="T1"],
NVP_EFV$lb[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF" & NVP_EFV$WHOs=="T1"],
NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF" & NVP_EFV$WHOs=="T1"],
NVP_EFV$ub[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF" & NVP_EFV$WHOs=="T1"],
col=addTrans("blue",16))
lines(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF" & NVP_EFV$WHOs=="T1"],
NVP_EFV$Dif[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF" & NVP_EFV$WHOs=="T1"],
col="blue",lwd=3)
abline(h=0, lty=2)


#AZT - TDF
def=NULL
sddef=NULL
tp=1    # you can change depending on the group you want to test
L=c(1,-1)
for(i in 1:(length(t)*12)){
ind.extract <- c(tp, ((length(t)*12)+tp))
VarVini2 <- Vall3[ind.extract, ind.extract]
VintE <- c(fv3[ind.extract[1]], fv3[ind.extract[2]])
def[i]=L %*% VintE
sddef[i]=sqrt(t(L) %*% VarVini2 %*%L)
tp=tp+1
}
```

```
AZT_TDF=data.frame(Time=rep(t, times=12), NNRTI=hlp1z$NNRTI[1:(length(t)*12)],
Gender=hlp1z$Gender[1:(length(t)*12)], WHOs = hlp1z$WHOs[1:(length(t)*12)],
Dif=def, SD.Dif=sddef, lb=def-1.96*sddef, ub=def+1.96*sddef)

plot(AZT_TDF$Time[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="NVP" & AZT_TDF$WHOs=="T1"],
AZT_TDF$Dif[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="NVP" & AZT_TDF$WHOs=="T1"],
type="n", ylab="Difference", main="AZT-TDF given NNRTI=NVP", sub="(c)" ,
ylim=c(-3,3), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(AZT_TDF$Time[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="NVP" & AZT_TDF$WHOs=="T1"],
AZT_TDF$lb[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="NVP" & AZT_TDF$WHOs=="T1"],
AZT_TDF$Time[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="NVP" & AZT_TDF$WHOs=="T1"],
AZT_TDF$ub[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="NVP" & AZT_TDF$WHOs=="T1"],
col=addTrans("blue",16))
lines(AZT_TDF$Time[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="NVP" & AZT_TDF$WHOs=="T1"],
AZT_TDF$Dif[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="NVP" & AZT_TDF$WHOs=="T1"],
col="blue",lwd=3)
abline(h=0, lty=2)

plot(AZT_TDF$Time[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="EFV" & AZT_TDF$WHOs=="T1"],
AZT_TDF$Dif[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="EFV" & AZT_TDF$WHOs=="T1"],
type="n", ylab="Difference", sub="(d)",
main="AZT-TDF given NNRTI=EFV",  ylim=c(-3,3), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))

segments(AZT_TDF$Time[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="EFV" & AZT_TDF$WHOs=="T1"],
AZT_TDF$lb[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="EFV" & AZT_TDF$WHOs=="T1"],
AZT_TDF$Time[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="EFV" & AZT_TDF$WHOs=="T1"],
AZT_TDF$ub[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="EFV" & AZT_TDF$WHOs=="T1"],
col=addTrans("blue",16))
lines(AZT_TDF$Time[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="EFV" & AZT_TDF$WHOs=="T1"],
AZT_TDF$Dif[AZT_TDF$Gender=="Fema"  & AZT_TDF$NNRTI=="EFV" & AZT_TDF$WHOs=="T1"],
col="blue",lwd=3)
abline(h=0, lty=2)



#----------------------Difference of CD4 First Derivative over time------------------
t=seq(0,1439, 2)
time=rep(t, times=24)
v=length(t)
age=rep(33, length(time))
#blc=rep(median(CD4$blcd4, na.rm=T), length(time))
#who=rep(rep(c("T1", "T2", "T3", "T4"), each=v), times=8)
who=rep(rep(c("T1", "T2", "T34"), each=v), times=8)
sex=rep(rep(c("Fema", "Male"), each=v*3), times=4)
nnrti=rep(rep(c("NVP", "EFV"), each=6*v), times=2)
nrti=rep(rep(c("AZT", "TDF"), each=12*v), times=1)
hlp11<-data.frame(time=time, NRTI=nrti, NNRTI=nnrti, Gender=sex, WHOs=who,
Age_at_ARVstart=age)# sequence for time and ranges from 0 to 60
R=NULL
for(i in 1:dim(hlp11)[1]){
if(hlp1$NRTI[i]=="AZT" & hlp1$NNRTI[i]=="NVP") R[i]=1
else if(hlp1$NRTI[i]=="AZT" & hlp1$NNRTI[i]=="EFV") R[i]=2
else if(hlp1$NRTI[i]=="TDF" & hlp1$NNRTI[i]=="NVP") R[i]=3
else if(hlp1$NRTI[i]=="TDF" & hlp1$NNRTI[i]=="EFV") R[i]=4
}
hlp11$ART=factor(R)
yfitfixef11<-predict(GammObj1$gam,hlp11, se.fit=T)# is it the mean at each time point?

hlp11a=data.frame(hlp11, yfitfixed=as.data.frame(yfitfixef11))
hlp11a=data.frame(hlp11a, lower=hlp11a$yfitfixed.fit-1.96*(hlp11a$yfitfixed.se.fit)^1.0,
upper=hlp11a$yfitfixed.fit+1.96*(hlp11a$yfitfixed.se.fit)^1.0)
head(hlp11a)

memory.limit(size = 8072)

delta<-1e-2
hlp2z=hlp11a
hlp2z$time<-hlp11a$time-delta
X01a<-predict(GammObj1$gam,hlp2z,type="lpmatrix")#?
hlp2z$time<-hlp11a$time+delta
X11a<-predict(GammObj1$gam,hlp2z,type="lpmatrix")#?
```

```
Xp12<-(X11a-X01a)/delta
v12<-Xp12%*%GammObj1$gam$coef
v.sd1a<-rowSums(Xp12%*%GammObj1$gam$Vp*Xp12)^.5
Xp02<-predict(GammObj1$gam,hlp2z,type="lpmatrix")#?
fv3=Xp12%*%GammObj1$gam$coef
Vall <- Xp12%*% GammObj1$gam$Vp %*% t(Xp12)

#NVP - EFV
derdef4=NULL
sdderdef4=NULL
tp=1    # you can change depending on the group you want to test
L=c(1,-1)
for(i in 1:(length(t)*12)){
ind.extract <- c(tp, (6*length(t)+tp))
VarVini2 <- Vall[ind.extract, ind.extract]
VintE <- c(fv3[ind.extract[1]], fv3[ind.extract[2]])
derdef4[i]=L%*%VintE
sdderdef4[i]=sqrt(t(L) %*% VarVini2 %*%L)
if(i==length(t)*6) tp=tp+length(t)*6+1
else tp=tp+1
}

dNVP_EFV=data.frame(Time=rep(t, times=12), NRTI=hlp2z$NRTI[c(1:(length(t)*6),
(length(t)*12+1):(length(t)*18))], Gender=hlp2z$Gender[c(1:(length(t)*6),
(length(t)*12+1):(length(t)*18))], WHOs=hlp2z$WHOs[c(1:(length(t)*6),
(length(t)*12+1):(length(t)*18))],Dif=derdef4, SD.Dif=sdderdef4,
lb=derdef4-2*sdderdef4^1, ub=derdef4+2*sdderdef4^1)

win.graph()
par(mfrow=c(2,2))
plot(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$Dif[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
type="n" , ylab="Difference", main="NVP-EFV given NRTI=AZT", sub="(a)",
ylim=c(-0.0001,0.0001), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$lb[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$ub[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
col=addTrans("blue",12))
lines(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$Dif[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
col="blue",lwd=3)
abline(h=0, lty=2)

plot(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$Dif[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT" & dNVP_EFV$WHOs=="T1"],
type="n" , ylab="Difference", main="NVP-EFV given NRTI=AZT", sub="(a)",
ylim=c(-0.0001,0.0001), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1440, 200)), labels=c(seq(0, 42, 6)))
segments(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$lb[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$ub[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF" & dNVP_EFV$WHOs=="T1"],
col=addTrans("blue",12))
lines(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF" & dNVP_EFV$WHOs=="T1"],
dNVP_EFV$Dif[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF" & dNVP_EFV$WHOs=="T1"],
col="blue",lwd=3)
abline(h=0, lty=2)


#AZT - TDF
derdef3=NULL
sdderdef3=NULL
tp=1    # you can change depending on the group you want to test
L=c(1,-1)
for(i in 1:(length(t)*12)){
ind.extract <- c(tp, (12*length(t)+tp))
VarVini2 <- Vall[ind.extract, ind.extract]
VintE <- c(fv3[ind.extract[1]], v12[ind.extract[2]])
derdef3[i]=L %*% VintE
sdderdef3[i]= sqrt(t(L) %*% VarVini2 %*%L)
tp=tp+1
}
```

```
dAZT_TDF=data.frame(Time=rep(t, times=12), NNRTI=hlp11a$NNRTI[1:(length(t)*12)],
Gender=hlp11a$Gender[1:(length(t)*12)],
WHOs=hlp11a$WHOs[1:(length(t)*12)],
Dif=derdef3, SD.Dif=sdderdef3,
lb=derdef3-2*sdderdef3^1, ub=derdef3+2*sdderdef3^1)
plot(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP" & dAZT_TDF$WHOs=="T1"],
dAZT_TDF$Dif[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP" & dAZT_TDF$WHOs=="T1"],
type="n",  ylab="Difference", main="AZT-TDF given NNRTI=NVP", sub="(c)",
ylim=c(-0.001,0.001), xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))
segments(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$lb[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$ub[hlp1z$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
col=addTrans("blue",12))
lines(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$Dif[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
col="blue",lwd=3)
abline(h=0, lty=2)

plot(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="EFV"],
dAZT_TDF$Dif[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="EFV"],
type="n", xlab="Time on ART (in days) ", ylim=c(-0.008,0.008), ylab="Difference",
main="AZT-TDF given NNRTI=EFV", sub="(d)")
segments(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="EFV"],
dAZT_TDF$lb[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="EFV"],
dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="EFV"],
dAZT_TDF$ub[hlp1z$Gender=="Fema"  & dAZT_TDF$NNRTI=="EFV"],
col=addTrans("blue",12))
lines(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="EFV"],
dAZT_TDF$Dif[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="EFV"],
col="blue",lwd=3)
abline(h=0, lty=2)
par(mfrow=c(3,3))


#------------------------------- Residual Analysis -------------------------#
rand=as.data.frame(ranef(GammObj1$lme))
head(rand.effect)
re=rand[,c(1dim(rand)[2]-1, dim(rand)[2])]
head(re)
names(re)=c("b0", "b1")
r=residuals(GammObj1$lme, type = "normalized")
win.graph()
par(mfrow=c(1,2))
plot(re$b0, re$b1, xlab="Random intercept",
ylab="Random Slope", main="Random Intercept Vs Slope", sub="(a)")
qqnorm(r, xlab="Residuals", sub="(b)")
qqline(r)
qqnorm(re$b0 , xlab="Random Intercepts", sub="(c)")
qqline(re$b0)
qqnorm(re$b1, xlab="Random Slopes", sub="(d)")
qqline(re$b1)

#---------------------model check-------------------------------------------------
CD4b=CD4[complete.cases(CD4),]
CD4b$pp=predict(GammObj1$lme)
win.graph()
par(mfrow=c(2,2))

plot(CD4b$logCD4[CD4b$time==0],CD4b$pp[CD4b$time==0],xlim=c(0,10),
ylim=c(0,10),ylab="predected",xlab="log CD4 count",main="at observation time 0")
abline(0,1)

plot(CD4b$logCD4[CD4b$time==161],CD4b$pp[CD4b$time==161],xlim=c(0,10),
ylim=c(0,10),ylab="predected",xlab="logCD4 count",main="at observation time 161")
abline(0,1)

plot(CD4b$logCD4[CD4b$time==182],CD4b$pp[CD4b$time==182],xlim=c(0,10),
ylim=c(0,10),ylab="predected",xlab="logCD4 count",main="at observation time 182")
abline(0,1)

plot(CD4b$logCD4[CD4b$time==189],CD4b$pp[CD4b$time==189],xlim=c(0,10),ylim=c(0,10),
```

```
ylab="predected",xlab="logCD4 count",main="at observation time 189")
abline(0,1)


par(mfrow=c(3,3))
plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed",main="Subject 79")
lines(CD4b$time[CD4b$ID==79],CD4b$logCD4[CD4b$ID==79],col="blue")
lines(CD4b$time[CD4b$ID==79],CD4b$pp[CD4b$ID==79],lty="dashed",col="blue")

plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed", main="Subject 180")
lines(CD4b$time[CD4b$ID==180],CD4b$logCD4[CD4b$ID==180],col="blue")
lines(CD4b$time[CD4b$ID==180],CD4b$pp[CD4b$ID==180],col="blue",lty="dashed")

plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed",main="Subject 271")
lines(CD4b$time[CD4b$ID==271],CD4b$logCD4[CD4b$ID==271],col="blue")
lines(CD4b$time[CD4b$ID==271],CD4b$pp[CD4b$ID==271],col="blue",lty="dashed")

plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed",main="Subject 1110")
lines(CD4b$time[CD4b$ID==1110],CD4b$logCD4[CD4b$ID==1110],col="blue")
lines(CD4b$time[CD4b$ID==1110],CD4b$pp[CD4b$ID==1110],col="blue",lty="dashed")

plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed",main="Subject 1")
lines(CD4b$time[CD4b$ID==1],CD4b$logCD4[CD4b$ID==1],col="blue")
lines(CD4b$time[CD4b$ID==1],CD4b$pp[CD4b$ID==1],lty="dashed",col="blue")

plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed",main="Subject 280")
lines(CD4b$time[CD4b$ID==280],CD4b$logCD4[CD4b$ID==280],col="blue")
lines(CD4b$time[CD4b$ID==280],CD4b$pp[CD4b$ID==280],col="blue",lty="dashed")

plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed",main="Subject 321")
lines(CD4b$time[CD4b$ID==321],CD4b$logCD4[CD4b$ID==321],col="blue")
lines(CD4b$time[CD4b$ID==321],CD4b$pp[CD4b$ID==321],col="blue",lty="dashed")

plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed",main="Subject 910")
lines(CD4b$time[CD4b$ID==910],CD4b$logCD4[CD4b$ID==910],col="blue")
lines(CD4b$time[CD4b$ID==910],CD4b$pp[CD4b$ID==910],col="blue",lty="dashed")

plot(c(0,1439),c(4,9),type="n",xlab="Days on ART", ylab="fitted and observed",main="Subject 527")
lines(CD4b$time[CD4b$ID==527],CD4b$logCD4[CD4b$ID==527],col="blue")
lines(CD4b$time[CD4b$ID==527],CD4b$pp[CD4b$ID==527],col="blue",lty="dashed")

##############################################################################################
##############################################################################################
##############################################################################################

#---------------------Hemoglobin Concentration Level -------------------#
#----------------------------------------------------------------------#
Hemog=read.table(choose.files(), header=T, sep="&")
head(Hemog)
Hemog$hemog=Hemog$LABRESULT
Hemog$ID=NULL
Hemog$ID[1]=1
for(i in 2:dim(Hemog)[1]){
if(Hemog$PTIDNO[i]==Hemog$PTIDNO[i-1]) Hemog$ID[i] = Hemog$ID[i-1]
else if(Hemog$PTIDNO[i]!=Hemog$PTIDNO[i-1]) Hemog$ID[i] =Hemog$ID[i-1] +1
}
Hemog$time2=Hemog$time*Hemog$time
Hemog$time3=Hemog$time*Hemog$time*Hemog$time
WHOs=NULL
for(i in 1:dim(Hemog)[1]){
if(Hemog$ARTWHOStage[i]=="T1") WHOs[i]="T1"
else if(Hemog$ARTWHOStage[i]=="T2") WHOs[i]="T2"
else if(Hemog$ARTWHOStage[i]=="T3" |
Hemog$ARTWHOStage[i]=="T4") WHOs[i]="T34"
}
Hemog$WHOs=factor(WHOs)

R=NULL
for(i in 1:dim(Hemog)[1]){
if(Hemog$NRTI[i]=="AZT" & Hemog$NNRTI[i]=="NVP") R[i]=1
else if(Hemog$NRTI[i]=="AZT" & Hemog$NNRTI[i]=="EFV") R[i]=2
else if(Hemog$NRTI[i]=="TDF" & Hemog$NNRTI[i]=="NVP") R[i]=3
else if(Hemog$NRTI[i]=="TDF" & Hemog$NNRTI[i]=="EFV") R[i]=4
}
```

```
Hemog$ART=factor(R)
head(Hemog)

###################
plot(Hemog$time, Hemog$hemog, type="n", xlab="Time since ARV start date (in days)",
ylab="Hemoglobin Level (grams/deciliter)", ylim=c(0,25), sub="(b)",
main="Trajectory of Hemoglobin Level in HIV 1 Patients")
for(i in 1:(dim(Hemog)[1]-50)){
lines(Hemog$time[Hemog$ID==i & Hemog$hemog<=25],
Hemog$hemog[Hemog$ID==i & Hemog$hemog<=25], col="grey")
}

LWsmoothHemog=read.table(choose.files(), header=T, sep="&")
head(LWsmoothHemog)
lines(LWsmoothHemog$time, LWsmoothHemog$p_LABRESULT, lwd=3, col="blue")
legend(200,1,c("Loess-Smooth mean of Hemoglobin Level"), lty=1, lwd=3,
col="blue")


# #-------------------Semi-Parametric Mixed Model Analysis -----------#
library(mgcv)
Hemog$Gender=relevel(Hemog$Gender, ref="Fema")
Hemog$NRTI=relevel(Hemog$NRTI, ref="AZT")
Hemog$NNRTI=relevel(Hemog$NNRTI, ref="NVP")
Hemog$ARTWHOStage=relevel(Hemog$ARTWHOStage, ref="T1")
Hemog$WHOs=relevel(Hemog$WHOs, ref="T1")

#Comparing Different Mean structures
fit1=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI +
Gender+
Age_at_ARVstart+
WHOs,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))

fit2=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI +
Gender+ Gender*NRTI + Gender*NNRTI +
Age_at_ARVstart+
WHOs,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))

fit3=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI +
Gender+ Gender*NRTI + Gender*NNRTI +
Age_at_ARVstart+Age_at_ARVstart*NRTI + Age_at_ARVstart*NNRTI +
WHOs,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))

fit4=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI +
Gender+ Gender*NRTI + Gender*NNRTI +
Age_at_ARVstart+Age_at_ARVstart*NRTI + Age_at_ARVstart*NNRTI +
WHOs + WHOs*NRTI + WHOs*NNRTI,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))


anova(fit1$lme, fit2$lme, fit3$lme, fit4$lme)
anova(fit1$lme, fit2$lme)
anova(fit1$lme, fit3$lme)
anova(fit1$lme, fit4$lme)

#Random Effect Test (mixture chi-square test)
fit1o=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart +
I(Age_at_ARVstart^2),
```

```
data=Hemog, method="REML", correlation=NULL, random=NULL)

fit1a=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart +
I(Age_at_ARVstart^2),
data=Hemog, method="REML", correlation=NULL,
random=list(ID=pdSymm(~1)))
chi1=-2*(fit1o$lme$logLik-fit1a$lme$logLik)
pval1=0.5*pchisq(chi1,0, lower.tail=F)+0.5*pchisq(chi1,1, lower.tail=F)

fit1b=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart +
I(Age_at_ARVstart^2),
data=Hemog, method="REML", correlation=NULL,
random=list(ID=pdSymm(~time)))

chi2=-2*(fit1a$lme$logLik-fit1b$lme$logLik)
pval2=0.5*pchisq(chi2,1, lower.tail=F)+0.5*pchisq(chi2,2, lower.tail=F)

fit1c=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart +
I(Age_at_ARVstart^2),
data=Hemog, method="REML", correlation=NULL,
random=list(ID=pdDiag(~time+time2)))

chi3=-2*(fit1c$lme$logLik-fit1b$lme$logLik)
chi3=0 #b/s the LL difference was -ve
pval3=0.5*pchisq(chi3,2, lower.tail=F)+0.5*pchisq(chi3,3, lower.tail=F)
table1=data.frame(Test=c("bo", "bo + b1*Time", "bo + b1*Time + b2*Time^2"),
Statitic=c(chi1, chi2, chi3), Pvalue=c(pval1, pval2, pval3))
table1


#Reducing Covariance Structure based on AIC (method=ML)
fit2a=gamm(hemog~s(time, bs="tp", by=ART,m=2)+NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdDiag(~time))) #Simple structure

fit2b=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~time))) #Unstructured

table2=data.frame(anova(fit2a$lme, fit2b$lme, test=F))
table2

#Testing Serial Correlation based on REML-likelihood
fit3a=fit1b #No serial correlation
fit3b=gamm(hemog~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
data=Hemog, method="REML", correlation=corExp(),
random=list(ID=pdSymm(~time))) #Spatial Exponential erial correlation
fit3c=gamm(hemog~s(time, bs="tp", by=ART,m=2)+NRTI*NNRTI +
Gender+Age_at_ARVstart+WHOs,
```

```
data=Hemog, method="REML", correlation=corGaus(),
random=list(ID=pdSymm(~time))) #Spatial gaussian serial correlation

table3=data.frame(Seria_Correlation=c("No Ser. Corr", "Sp. Exponential", "Sp. Gaussian"),
REML_LogLik=c(fit3a$lme$logLik, fit3b$lme$logLik, fit3c$lme$logLik))
table3

########################################
fit3=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI +
Gender+ Gender*NRTI + Gender*NNRTI +
Age_at_ARVstart+Age_at_ARVstart*NRTI + Age_at_ARVstart*NNRTI +
WHOs,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~1)))

Final_Model1=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI*NNRTI +
Gender +
Age_at_ARVstart +
WHOs,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~time)))
Final_Model2=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Gender*NNRTI +
Age_at_ARVstart,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~time)))

Final_Model3=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart,
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~time)))
Final_Model8=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart +
I(Age_at_ARVstart^2),
data=Hemog, method="ML", correlation=NULL,
random=list(ID=pdSymm(~time)))

summary(Final_Model4$gam)
AIC(Final_Model4$lme)
AIC(Final_Model1$lme, Final_Model2$lme, Final_Model3$lme)
BIC(Final_Model1$lme, Final_Model2$lme, Final_Model3$lme)


Final_Model=gamm(hemog~s(time, bs="tp", by=ART,m=2)+
NRTI +
NNRTI +
Gender +
Gender*NRTI +
Age_at_ARVstart +
I(Age_at_ARVstart^2),
data=Hemog, method="REML", correlation=NULL,
random=list(ID=pdSymm(~time)))

summary(Final_Model$gam)
summary(Final_Model$lme)
anova(Final_Model$lme)

GammObj1=Final_Model

########## Contrast ################
```

```
library(multcomp)
library(mvtnorm)

f1=Final_Model$lme
f2=Final_Model$gam

length(coef(f2))
names(coef(f2))

contrast.matrix1 <- rbind("AZT vs. TDF given Gender= Female" =  c(0, 1, 0, rep(0, times=40)))
contrast.matrix2 <- rbind("AZT vs. TDF given Gender= Male" =  c(0, 1, 0, 0, 0, 1, rep(0, times=37)))

summary(glht(f2, contrast.matrix1))
summary(glht(f2, contrast.matrix2))

#------------------------ Smoothed Mean Plots----------------------------#
###-------------------------predict-----------------------------------------------
yfit1<-predict(GammObj1$lme, asList=F, level = 0:1)

#names(yfit1)
t=seq(0,1439, 1)
time=rep(t, times=8)
v=length(t)
age=rep(33, length(time))
#who=rep(rep(c("T1", "T2", "T3", "T4"), each=v), times=8)
#who=rep(rep(c("T1", "T2", "T34"), each=v), times=8)
sex=rep(rep(c("Fema", "Male"), each=v), times=4)
nnrti=rep(rep(c("NVP", "EFV"), each=2*v), times=2)
nrti=rep(rep(c("AZT", "TDF"), each=4*v), times=1)
hlp1<-data.frame(time=time, NRTI=nrti, NNRTI=nnrti,
Gender=sex,Age_at_ARVstart=age)
R=NULL
for(i in 1:dim(hlp1)[1]){
if(hlp1$NRTI[i]=="AZT" & hlp1$NNRTI[i]=="NVP") R[i]=1
else if(hlp1$NRTI[i]=="AZT" & hlp1$NNRTI[i]=="EFV") R[i]=2
else if(hlp1$NRTI[i]=="TDF" & hlp1$NNRTI[i]=="NVP") R[i]=3
else if(hlp1$NRTI[i]=="TDF" & hlp1$NNRTI[i]=="EFV") R[i]=4
}
hlp1$ART=factor(R)

yfitfixef1<-predict(GammObj1$gam,hlp1, se.fit=T)# is it the mean at each time point?
hlp1a=data.frame(hlp1, yfitfixed=as.data.frame(yfitfixef1))
hlp1a=data.frame(hlp1a, lower=hlp1a$yfitfixed.fit-2*(hlp1a$yfitfixed.se.fit)^1.0,
upper=hlp1a$yfitfixed.fit+2*(hlp1a$yfitfixed.se.fit)^1.0)


#Function to create Transparent Color
addTrans <- function(color,trans)
{
# This function adds transparancy to a color.
# Define transparancy with an integer between 0 and 255
# 0 being fully transparant and 255 being fully visable
# Works with either color and trans a vector of equal length,
# or one of the two of length 1.

if (length(color)!=length(trans)&!any(c(length(color),length(trans))==1)) stop("Vector lengths not correct")
if (length(color)==1 & length(trans)>1) color <- rep(color,length(trans))
if (length(trans)==1 & length(color)>1) trans <- rep(trans,length(color))

num2hex <- function(x)
{
hex <- unlist(strsplit("0123456789ABCDEF",split=""))
return(paste(hex[(x-x%%16)/16+1],hex[x%%16+1],sep=""))
}
rgb <- rbind(col2rgb(color),trans)
res <- paste("#",apply(apply(rgb,2,num2hex),2,paste,collapse=""),sep="")
return(res)
}

win.graph()
par(mfrow=c(2,2))
plot(c(0:1439), seq(min(hlp1a$lower), max(hlp1a$upper), length=1440), type="n", sub="(a)",
main="NVP Vs EFV given NRTI=AZT",ylab="Hemoglobin Conc. (g/dL)",
xlim=c(0,1440), ylim=c(9, 16),   xlab="Time on ART (in months)", xaxt="n")
```

```
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))

segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
lwd=2, col=addTrans("black",5))
segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV"],
lwd=2, col=addTrans("red",5))
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
lwd=2, col=1)
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="EFV"],
lwd=2, col=2)
abline(h=5.86, lty=2)
legend(800, 11, c("NVP", "EFV"), lty=1, col=c(1,2),lwd=3)

plot(c(0:1439), seq(min(hlp1a$lower), max(hlp1a$upper), length=1440), type="n", sub="(b)",
main="NVP Vs EFV given NRTI=TDF",ylab="Hemoglobin Conc. (g/dL)",
xlim=c(0,1440), ylim=c(9, 16),   xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))


segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
lwd=2, col=addTrans("black",5))

segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV"],
lwd=2, col=addTrans("red",5))
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
lwd=2, col=1)
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="EFV"],
lwd=2, col=2)
abline(h=5.86, lty=2)
legend(800, 11, c("NVP", "EFV"), lty=1, col=c(1,2),lwd=3)



plot(c(0:1439), seq(min(hlp1a$lower), max(hlp1a$upper), length=1440), type="n", sub="(c)",
main="AZT Vs TDF given NNRTI=NVP \n and Gender=Female",ylab="Hemoglobin Conc. (g/dL)",
xlim=c(0,1440), ylim=c(9, 16),   xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))

segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
lwd=2, col=addTrans("black",5))


segments(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$lower[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$upper[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
lwd=2, col=addTrans("red",5))
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
lwd=2, col=1)
lines(hlp1a$time[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Fema" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
lwd=2, col=2)
abline(h=5.86, lty=2)
legend(800, 11, c("AZT", "TDF"), lty=1, col=c(1,2),lwd=3)
```

```
plot(c(0:1439), seq(min(hlp1a$lower), max(hlp1a$upper), length=1440), type="n", sub="(d)",
main="AZT Vs TDF given NNRTI=NVP \n and Gender=Male",ylab="Hemoglobin Conc. (g/dL)",
xlim=c(0,1440), ylim=c(9, 16),   xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))

segments(hlp1a$time[hlp1a$Gender=="Male" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$lower[hlp1a$Gender=="Male" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$time[hlp1a$Gender=="Male" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$upper[hlp1a$Gender=="Male" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
lwd=2, col=addTrans("black",5))


segments(hlp1a$time[hlp1a$Gender=="Male" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$lower[hlp1a$Gender=="Male" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$time[hlp1a$Gender=="Male" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$upper[hlp1a$Gender=="Male" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
lwd=2, col=addTrans("red",5))
lines(hlp1a$time[hlp1a$Gender=="Male" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Male" & hlp1a$NRTI=="AZT" & hlp1a$NNRTI=="NVP"],
lwd=2, col=1)
lines(hlp1a$time[hlp1a$Gender=="Male" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
hlp1a$yfitfixed.fit[hlp1a$Gender=="Male" & hlp1a$NRTI=="TDF" & hlp1a$NNRTI=="NVP"],
lwd=2, col=2)
abline(h=5.86, lty=2)
legend(800, 11, c("AZT", "TDF"), lty=1, col=c(1,2),lwd=3)


##############################################################################################
################################### The difference in the Profile ############################
t=seq(0,1439, 2)
time=rep(t, times=8)
v=length(t)
age=rep(33, length(time))
#who=rep(rep(c("T1", "T2", "T3", "T4"), each=v), times=8)
#who=rep(rep(c("T1", "T2", "T34"), each=v), times=8)
sex=rep(rep(c("Fema", "Male"), each=v), times=4)
nnrti=rep(rep(c("NVP", "EFV"), each=2*v), times=2)
nrti=rep(rep(c("AZT", "TDF"), each=4*v), times=1)
hlp1z<-data.frame(time=time, NRTI=nrti, NNRTI=nnrti, Gender=sex,
Age_at_ARVstart=age)# sequence for time and ranges from 0 to 60
Rz=NULL
for(i in 1:dim(hlp1z)[1]){
if(hlp1z$NRTI[i]=="AZT" & hlp1z$NNRTI[i]=="NVP") Rz[i]=1
else if(hlp1z$NRTI[i]=="AZT" & hlp1z$NNRTI[i]=="EFV") Rz[i]=2
else if(hlp1z$NRTI[i]=="TDF" & hlp1z$NNRTI[i]=="NVP") Rz[i]=3
else if(hlp1z$NRTI[i]=="TDF" & hlp1z$NNRTI[i]=="EFV") Rz[i]=4
}
hlp1z$ART=factor(Rz)
############

memory.limit()
memory.limit(size=8000)

Xp02<-predict(GammObj1$gam,hlp1z,type="lpmatrix")#?
fv3=Xp02%*%GammObj1$gam$coef
#fv3=as.vector(predict(GammObj1$gam, hlp1z))
Vall3=Xp02%*%GammObj1$gam$Vp%*%t(Xp02)


win.graph()
par(mfrow=c(2,2))

#NVP - EFV
def2=NULL
sddef2=NULL
tp=1    # you can change depending on the group you want to test
L=c(1,-1)
for(i in 1:(length(t)*4)){
ind.extract <- c(tp, (2*length(t)+tp))
VarVini2 <- Vall3[ind.extract, ind.extract]
VintE <- c(fv3[ind.extract[1]], fv3[ind.extract[2]])
def2[i]=L %*% VintE
```

```
sddef2[i]=sqrt(t(L) %*% VarVini2 %*%L)
if(i==(length(t)*2)) tp=(4*length(t)+1)
else tp=tp+1
}

NVP_EFV=data.frame(Time=rep(t, times=4)
, NRTI=hlp1z$NRTI[c(1:(length(t)*2), (length(t)*4+1):(length(t)*6))],
Gender=hlp1z$Gender[c(1:(length(t)*2),  (length(t)*4+1):(length(t)*6))],
Dif=def2, SD.Dif=sddef2, lb=def2-2*sddef2, ub=def2+2*sddef2)

plot(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT"],
NVP_EFV$Dif[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT"],
type="n", ylim=c(-3,3), ylab="Difference",
main="NVP-EFV given NRTI=AZT", sub="(a)", xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))

segments(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT"],
NVP_EFV$lb[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT"],
NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT"],
NVP_EFV$ub[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT"],
col=addTrans("blue",12))
lines(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT"],
NVP_EFV$Dif[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="AZT"],
col="blue",lwd=3)
abline(h=0, lty=2)


plot(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF"],
NVP_EFV$Dif[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF"],
type="n",  ylab="Difference", sub="(b)",
main="NVP-EFV given NRTI=TDF", ylim=c(-3,3),xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))

segments(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF"],
NVP_EFV$lb[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF"],
NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF"],
NVP_EFV$ub[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF"],
col=addTrans("blue",12))
lines(NVP_EFV$Time[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF"],
NVP_EFV$Dif[NVP_EFV$Gender=="Fema"  & NVP_EFV$NRTI=="TDF"],
col="blue",lwd=3)
abline(h=0, lty=2)

#AZT - TDF
def=NULL
sddef=NULL
tp=1    # you can change depending on the group you want to test
L=c(1,-1)
for(i in 1:(length(t)*4)){

ind.extract <- c(tp, ((length(t)*4)+tp))
VarVini2 <- Vall3[ind.extract, ind.extract]
VintE <- c(fv3[ind.extract[1]], fv3[ind.extract[2]])
def[i]=L %*% VintE
sddef[i]=sqrt(t(L) %*% VarVini2 %*%L)
tp=tp+1
}

AZT_TDF=data.frame(Time=rep(t, times=4), NNRTI=hlp1z$NNRTI[1:(length(t)*4)],
Gender=hlp1z$Gender[1:(length(t)*4)],
Dif=def, SD.Dif=sddef,
lb=def-2*sddef, ub=def+2*sddef)
plot(AZT_TDF$Time[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="NVP"],
AZT_TDF$Dif[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="NVP"],
type="n",  ylim=c(-3,3), ylab="Difference",
main="AZT-TDF given NNRTI=NVP and \n Gender = Female", sub="(c)",
xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))
segments(AZT_TDF$Time[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="NVP"],
AZT_TDF$lb[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="NVP"],
AZT_TDF$Time[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="NVP"],
AZT_TDF$ub[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="NVP"],
col=addTrans("blue",12))
lines(AZT_TDF$Time[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="NVP"],
```

```
AZT_TDF$Dif[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="NVP"],
col="blue",lwd=3)
abline(h=0, lty=2)


plot(AZT_TDF$Time[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="EFV"],
AZT_TDF$def[hlp1z$Gender=="Fema"  & hlp1z$NNRTI=="EFV"],
type="n", ylab="Difference", sub="(d)",
main="AZT-TDF given NNRTI=NVP and \n Gender = Male", ylim=c(-3,3),
xlim=c(0, 1440),
xlab="Time on ART (in months)", xaxt="n")
axis(1, at=c(seq(0, 1437, 200)), labels=c(seq(0, 42, 6)))

segments(AZT_TDF$Time[hlp1z$Gender=="Male"  & hlp1z$NNRTI=="NVP"],
AZT_TDF$lb[hlp1z$Gender=="Male"  & hlp1z$NNRTI=="NVP"],
AZT_TDF$Time[hlp1z$Gender=="Male"  & hlp1z$NNRTI=="NVP"],
AZT_TDF$ub[hlp1z$Gender=="Male"  & hlp1z$NNRTI=="NVP"],
col=addTrans("blue",12))
lines(AZT_TDF$Time[hlp1z$Gender=="Male"  & hlp1z$NNRTI=="NVP"],
def[hlp1z$Gender=="Male"  & hlp1z$NNRTI=="NVP"],
col="blue",lwd=3)
abline(h=0, lty=2)


################################################################################
#------------------------------Model 1: first order derivatives ------------------------
t.mesh<-seq(0,1439,1)
delta<-1e-5
hlp2a=hlp1
hlp2a$time<-hlp1$time-delta
X01a<-predict(GammObj1$gam,hlp2a,type="lpmatrix")#?
hlp2a$time<-hlp1$time+delta
X11a<-predict(GammObj1$gam,hlp2a,type="lpmatrix")#?
Xp1a<-(X11a-X01a)/delta
v1a<-Xp1a%*%GammObj1$gam$coef
v.sd1a<-rowSums(Xp1a%*%GammObj1$gam$Vp*Xp1a)^.5
#XDer1<-cbind(0,1,2*t.mesh)
#DerRan1<-GammObj$lme$coef$random$ID%*%t(XDer1)# the deriative of the random effect

hlp3a=data.frame(hlp1, Deriv=v1a, lowerDer=v1a-2*v.sd1a, upperDer=v1a+2*v.sd1a)

win.graph()
par(mfrow=c(2,2))
plot(c(0:1439), seq(min(hlp3a$lowerDer), max(hlp3a$upperDer), length=1440), type="n", sub="(a)",
xlim=c(0,1440), ylim=c(-0.05, 0.05), ylab="First derivative of g(mu)",
xlab="Time on ART (in days)", main="NVP Vs EFV given NRTI=AZT")
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
lwd=2, col=addTrans("black", 5))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV"],
lwd=2, col=addTrans("red", 5))
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
lwd=2, col=1)
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="EFV"],
lwd=2, col=2)
abline(h=0, lty=2)
legend(0, -0.01, c("NVP", "EFV"), lty=1, col=c(1,2),lwd=3)


plot(c(0:1439), seq(min(hlp3a$lowerDer), max(hlp3a$upperDer), length=1440), type="n", sub="(b)",
xlim=c(0,1440), ylim=c(-0.05, 0.05), ylab="First derivative of g(mu)",
xlab="Time on ART (in days)", main="NVP Vs EFV given NRTI=TDF")

segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
```

```
lwd=2, col=addTrans("black", 5))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV"],
lwd=2, col=addTrans("red", 5))
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
lwd=2, col=1)
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="EFV"],
lwd=2, col=2)
abline(h=0, lty=2)
legend(0, -0.01, c("NVP", "EFV"), lty=1, col=c(1,2),lwd=3)

plot(c(0:1439), seq(min(hlp3a$lowerDer), max(hlp3a$upperDer), length=1440), type="n", sub="(c)",
xlim=c(0,1440), ylim=c(-0.05, 0.05), ylab="First derivative of g(mu)",
xlab="Time on ART (in days)", main="AZT Vs TDF given NNRTI=NVP and \n Gender=Female")
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
lwd=2, col=addTrans("black", 5))
segments(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$lowerDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$upperDer[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
lwd=2, col=addTrans("red", 5))
lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
lwd=2, col=1)

lines(hlp3a$time[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$Deriv[hlp3a$Gender=="Fema" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
lwd=2, col=2)
abline(h=0, lty=2)
legend(0, -0.01, c("AZT", "TDF"), lty=1, col=c(1,2),lwd=3)

plot(c(0:1439), seq(min(hlp3a$lowerDer), max(hlp3a$upperDer), length=1440), type="n", sub="(d)",
xlim=c(0,1440), ylim=c(-0.05, 0.05), ylab="First derivative of g(mu)",
xlab="Time on ART (in days)", main="AZT Vs TDF given NNRTI=NVP and \n Gender=Male")
segments(hlp3a$time[hlp3a$Gender=="Male" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$lowerDer[hlp3a$Gender=="Male" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$time[hlp3a$Gender=="Male" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$upperDer[hlp3a$Gender=="Male" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
lwd=2, col=addTrans("black", 5))
segments(hlp3a$time[hlp3a$Gender=="Male" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$lowerDer[hlp3a$Gender=="Male" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$time[hlp3a$Gender=="Male" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$upperDer[hlp3a$Gender=="Male" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
lwd=2, col=addTrans("red", 5))
lines(hlp3a$time[hlp3a$Gender=="Male" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
hlp3a$Deriv[hlp3a$Gender=="Male" & hlp3a$NRTI=="AZT" & hlp3a$NNRTI=="NVP"],
lwd=2, col=1)

lines(hlp3a$time[hlp3a$Gender=="Male" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
hlp3a$Deriv[hlp3a$Gender=="Male" & hlp3a$NRTI=="TDF" & hlp3a$NNRTI=="NVP"],
lwd=2, col=2)
abline(h=0, lty=2)
legend(0, -0.01, c("AZT", "TDF"), lty=1, col=c(1,2),lwd=3)



#----------------------difference of CD4 First Derivative  over time for NRTI(AZT-TDF)-----------------
######################plot of difference in rate of change for NRTI(AZT-TDF) ######-----------------
t.mesh<-seq(0,1439,1)
delta<-1e-5
hlp2z=hlp1z
hlp2z$time<-hlp1z$time-delta
X01a<-predict(GammObj1$gam,hlp2z,type="lpmatrix")#?
hlp2z$time<-hlp1z$time+delta
X11a<-predict(GammObj1$gam,hlp2z,type="lpmatrix")#?
Xp12<-(X11a-X01a)/delta
v12<-Xp12%*%GammObj1$gam$coef
```

```
v.sd1a<-rowSums(Xp12%*%GammObj1$gam$Vp*Xp12)^.5

Xp02<-predict(GammObj1$gam,hlp1z,type="lpmatrix")#?
fv3=Xp12%*%GammObj1$gam$coef
Vall <- Xp12%*% GammObj1$gam$Vp %*% t(Xp12)

#NVP - EFV
derdef4=NULL
sdderdef4=NULL
tp=1    # you can change depending on the group you want to test
L=c(1,-1)
for(i in 1:(length(t)*4)){

ind.extract <- c(tp, (2*length(t)+tp))
VarVini2 <- Vall[ind.extract, ind.extract]
VintE <- c(fv3[ind.extract[1]], fv3[ind.extract[2]])
derdef4[i]=L %*% VintE
sdderdef4[i]=sqrt(t(L) %*% VarVini2 %*%L)
if(i==(length(t)*2)) tp=(length(t)*4 +1)
else tp=tp+1
}

dNVP_EFV=data.frame(Time=rep(t, times=4), NRTI=hlp1z$NRTI[c(1:(length(t)*2),(length(t)*4+1):(length(t)*6)) ],
Gender=hlp1z$Gender[c(1:(length(t)*2),(length(t)*4+1):(length(t)*6)) ],
Dif=derdef4, SD.Dif=sdderdef4, lb=derdef4-2*sdderdef4^1, ub=derdef4+2*sdderdef4^1)

win.graph()
par(mfrow=c(2,2))
plot(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
dNVP_EFV$Dif[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
type="n", xlab="Time on ART (in days) ", ylim=c(-0.03,0.03), ylab="Difference",
main="NVP-EFV given NRTI=AZT", sub="(a)")
segments(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
dNVP_EFV$lb[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
dNVP_EFV$ub[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
col=addTrans("blue",12))
lines(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
dNVP_EFV$Dif[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
col="blue",lwd=3)
abline(h=0, lty=2)


plot(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
dNVP_EFV$Dif[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="AZT"],
type="n", xlab="Time on ART (in days) ", ylim=c(-0.03,0.03), ylab="Difference",
main="NVP-EFV given NRTI=TDF", sub="(b)")
segments(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF"],
dNVP_EFV$lb[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF"],
dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF"],
dNVP_EFV$ub[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF"],
col=addTrans("blue",12))
lines(dNVP_EFV$Time[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF"],
dNVP_EFV$Dif[dNVP_EFV$Gender=="Fema"  & dNVP_EFV$NRTI=="TDF"],
col="blue",lwd=3)
abline(h=0, lty=2)


#AZT - TDF
derdef3=NULL
sdderdef3=NULL
tp=1    # you can change depending on the group you want to test
L=c(1,-1)
for(i in 1:(length(t)*4)){
ind.extract <- c(tp, (2*length(t)+tp))
VarVini2 <- Vall[ind.extract, ind.extract]
VintE <- c(fv3[ind.extract[1]], v12[ind.extract[2]])
derdef3[i]=L %*% VintE
sdderdef3[i]= sqrt(t(L) %*% VarVini2 %*%L)
tp=tp+1
}


dAZT_TDF=data.frame(Time=rep(t, times=4), NNRTI=hlp1z$NNRTI[1:(length(t)*4)],
```

```
Gender=hlp1z$Gender[1:(length(t)*4)],
Dif=derdef3, SD.Dif=sdderdef3,
lb=derdef3-2*sdderdef3^1, ub=derdef3+2*sdderdef3^1)
plot(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$Dif[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
type="n", xlab="Time on ART (in days) ", ylim=c(-0.03,0.03), ylab="Difference",
main="AZT-TDF given NNRTI=NVP and \n Gender=Female", sub="(c)")
segments(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$lb[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$ub[hlp1z$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
col=addTrans("blue",12))
lines(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$Dif[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
col="blue",lwd=3)
abline(h=0, lty=2)


plot(dAZT_TDF$Time[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$Dif[dAZT_TDF$Gender=="Fema"  & dAZT_TDF$NNRTI=="NVP"],
type="n", xlab="Time on ART (in days) ", ylim=c(-0.03,0.03), ylab="Difference",
main="AZT-TDF given NNRTI=NVP and \n Gender=Male", sub="(d)")
segments(dAZT_TDF$Time[dAZT_TDF$Gender=="Male"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$lb[dAZT_TDF$Gender=="Male"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$Time[dAZT_TDF$Gender=="Male"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$ub[hlp1z$Gender=="Male"  & dAZT_TDF$NNRTI=="NVP"],
col=addTrans("blue",12))
lines(dAZT_TDF$Time[dAZT_TDF$Gender=="Male"  & dAZT_TDF$NNRTI=="NVP"],
dAZT_TDF$Dif[dAZT_TDF$Gender=="Male"  & dAZT_TDF$NNRTI=="NVP"],
col="blue",lwd=3)
abline(h=0, lty=2)


#---------------------Residual Analysis and Model check-----------------------------------------------
#Estimate of random effects and residual Anlaysis

rand=as.data.frame(ranef(GammObj1$lme))
head(rand)
re=rand[,c(dim(rand)[2]-1, dim(rand)[2])]
head(re)
names(re)=c("b0", "b1")
r=residuals(GammObj1$lme, type = "normalized")

win.graph()
par(mfrow=c(2,2))
plot(re$b0[re$b0], re$b1[re$b0], xlab="Random intercept",
ylab="Random Slope", main="Random Intercept Vs Slope", sub="(a)")
qqnorm(r, xlab="Residuals", sub="(b)")
qqline(r)
qqnorm(re$b0 , xlab="Random Intercepts", sub="(c)")
qqline(re$b0)
qqnorm(re$b1, xlab="Random Slopes", sub="(d)")
qqline(re$b1)


Hemb=Hemog
Hemb$pp=predict(GammObj1$lme)
attach(Hemb)
win.graph()
par(mfrow=c(2,2))

plot(Hemb$hemog[Hemb$time==0],Hemb$pp[Hemb$time==0],xlim=c(min(pp),max(pp)),
ylim=c(min(pp),max(pp)),ylab="Predected",xlab="log CD4 count",main="at observation time 0")
abline(0,1)

plot(Hemb$hemog[Hemb$time==161],Hemb$pp[Hemb$time==161],xlim=c(min(pp), max(pp)),
ylim=c(min(pp),max(pp)),ylab="Predected",xlab="Hemoglobin Level",main="at observation time 161")
abline(0,1)

plot(Hemb$hemog[Hemb$time==182],Hemb$pp[Hemb$time==182],xlim=c(min(pp), max(pp)),
ylim=c(min(pp),max(pp)),ylab="Predected",xlab="Hemoglobin Level",main="at observation time 182")
abline(0,1)

plot(Hemb$hemog[Hemb$time==189],Hemb$pp[Hemb$time==189],xlim=c(min(pp), max(pp)), ylim=c(min(pp),
```

```
max(pp)), ylab="Predected",xlab="Hemoglobin Level",main="at observation time 189")
abline(0,1)


par(mfrow=c(3,3))
plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 79")
lines(Hemb$time[Hemb$ID==79],Hemb$hemog[Hemb$ID==79],col="blue")
lines(Hemb$time[Hemb$ID==79],Hemb$pp[Hemb$ID==79],lty="dashed",col="blue")

plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 180")
lines(Hemb$time[Hemb$ID==180],Hemb$hemog[Hemb$ID==180],col="blue")
lines(Hemb$time[Hemb$ID==180],Hemb$pp[Hemb$ID==180],col="blue",lty="dashed")

plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 271")
lines(Hemb$time[Hemb$ID==271],Hemb$hemog[Hemb$ID==271],col="blue")
lines(Hemb$time[Hemb$ID==271],Hemb$pp[Hemb$ID==271],col="blue",lty="dashed")

plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 1110")
lines(Hemb$time[Hemb$ID==1110],Hemb$hemog[Hemb$ID==1110],col="blue")
lines(Hemb$time[Hemb$ID==1110],Hemb$pp[Hemb$ID==1110],col="blue",lty="dashed")

plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 1")
lines(Hemb$time[Hemb$ID==1],Hemb$hemog[Hemb$ID==1],col="blue")
lines(Hemb$time[Hemb$ID==1],Hemb$pp[Hemb$ID==1],lty="dashed",col="blue")

plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 280")
lines(Hemb$time[Hemb$ID==280],Hemb$hemog[Hemb$ID==280],col="blue")
lines(Hemb$time[Hemb$ID==280],Hemb$pp[Hemb$ID==280],col="blue",lty="dashed")

plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 321")
lines(Hemb$time[Hemb$ID==321],Hemb$hemog[Hemb$ID==321],col="blue")
lines(Hemb$time[Hemb$ID==321],Hemb$pp[Hemb$ID==321],col="blue",lty="dashed")

plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 910")
lines(Hemb$time[Hemb$ID==910],Hemb$hemog[Hemb$ID==910],col="blue")
lines(Hemb$time[Hemb$ID==910],Hemb$pp[Hemb$ID==910],col="blue",lty="dashed")

plot(c(0,1439),c(min(pp), max(pp)),type="n",xlab="Days on ART", ylab="Fitted and Observed",
main="Subject 527")
lines(Hemb$time[Hemb$ID==527],Hemb$hemog[Hemb$ID==527],col="blue")
lines(Hemb$time[Hemb$ID==527],Hemb$pp[Hemb$ID==527],col="blue",lty="dashed")


###############################################################################################
###############################################################################################
###############################################################################################

*-------------------------- Joint Analysis ------------------------
libname thesis "C:\Users\Alemu\Desktop\Thesis2";

*Preparing the data for joint analysis
PROC IMPORT OUT=thesis.aa
DATAFILE="C:\Users\Alemu\Desktop\Thesis2\aa.csv"
REPLACE;
getnames=yes;
RUN;
proc freq data=thesis.aa;
table WHOs;
run;

PROC IMPORT OUT=thesis.bb
DATAFILE="C:\Users\Alemu\Desktop\Thesis2\bb.csv"
REPLACE;
getnames=yes;
RUN;
proc sort data=thesis.aa;
```

```
by PTIDNO tdisc;
run;
proc sort data=thesis.bb;
by PTIDNO tdisc;
run;


proc sort data=b out=sorted;
by PTIDNO Gender Age_at_ARVstart Agegroup_at_ARVstart NRTI NNRTI WHOs ART blcd4 blhemog tdisc;
proc transpose data=sorted out=b1;
by PTIDNO Gender Age_at_ARVstart Agegroup_at_ARVstart NRTI NNRTI WHOs ART blcd4 blhemog tdisc;
var logCD4 hemog;run;

proc means data=b1 n;
var col1 col2;
run;

data b2;
set b1;
rename _NAME_=outcome;
rename Age_at_ARVstart=Age;
drop col2 Agegroup_at_ARVstart ART blhemog;
run;

data thesis.joint2;
set b2;
rename Col1=Y;
*if Col1 ^=.;
tclass=tdisc;
run;

proc export data=thesis.joint2
outfile='C:\Users\Alemu\Desktop\j.txt'
dbms=dlm;
delimiter='&';
run;

*Some Analysis were undertaken in Excel file (not reported)
PROC IMPORT OUT=thesis.joint3
DATAFILE="C:\Users\Alemu\Desktop\j.csv"
REPLACE;
getnames=yes;
RUN;

data b;
set thesis.joint3;

if outcome="logCD4" then int_1=1;
else int_1=0;
if outcome="hemog" then int_2=1;
else int_2=0;

if outcome="logCD4" then T_1=tdisc;
else T_1=0;
if outcome="hemog" then T_2=tdisc;
else T_2=0;
if outcome="logCD4" then T2_1=tdisc*tdisc;
else T2_1=0;
if outcome="hemog" then T2_2=tdisc*tdisc;
else T2_2=0;

if outcome="logCD4" then T3_1=tdisc*tdisc*tdisc;
else T3_1=0;
if outcome="hemog" then T3_2=tdisc*tdisc*tdisc;
else T3_2=0;

if outcome="logCD4" then x1_1=NRTI;
else x1_1=0;
if outcome="hemog" then x1_2=NRTI;
else x1_2=0;

if outcome="logCD4" then x2_1=NNRTI;
else x2_1=0;
if outcome="hemog" then x2_2=NNRTI;
```

```
else x2_2=0;

if outcome="logCD4" then x12_1=NNRTI*NRTI;
else x12_1=0;
if outcome="hemog" then x12_2=NNRTI*NRTI;
else x12_2=0;

if outcome="logCD4" then x1t_1=NRTI*tdisc;
else x1t_1=0;
if outcome="hemog" then x1t_2=NRTI*tdisc;
else x1t_2=0;

if outcome="logCD4" then x1t2_1=NRTI*tdisc*tdisc;
else x1t2_1=0;
if outcome="hemog" then x1t2_2=NRTI*tdisc*tdisc;
else x1t2_2=0;

if outcome="logCD4" then x1t3_1=NRTI*tdisc*tdisc*tdisc;
else x1t3_1=0;
if outcome="hemog" then x1t3_2=NRTI*tdisc*tdisc*tdisc;
else x1t3_2=0;

if outcome="logCD4" then x2t_1=NNRTI*tdisc;
else x2t_1=0;
if outcome="hemog" then x2t_2=NNRTI*tdisc;
else x2t_2=0;

if outcome="logCD4" then x2t2_1=NNRTI*tdisc*tdisc;
else x2t2_1=0;
if outcome="hemog" then x2t2_2=NNRTI*tdisc*tdisc;
else x2t2_2=0;

if outcome="logCD4" then x2t3_1=NNRTI*tdisc*tdisc*tdisc;
else x2t3_1=0;
if outcome="hemog" then x2t3_2=NNRTI*tdisc*tdisc*tdisc;
else x2t3_2=0;

if outcome="logCD4" then x3_1=Gender;
else x3_1=0;
if outcome="hemog" then x3_2=Gender;
else x3_2=0;

if outcome="logCD4" then x3t_1=Gender*tdisc;
else x3t_1=0;
if outcome="hemog" then x3t_2=Gender*tdisc;
else x3t_2=0;

if outcome="logCD4" then x3t2_1=Gender*tdisc*tdisc;
else x3t2_1=0;
if outcome="hemog" then x3t2_2=Gender*tdisc*tdisc;
else x3t2_2=0;

if outcome="logCD4" then x3t3_1=Gender*tdisc*tdisc*tdisc;
else x3t3_1=0;
if outcome="hemog" then x3t3_2=Gender*tdisc*tdisc*tdisc;
else x3t3_2=0;

if outcome="logCD4" then x4_1=Age;
else x4_1=0;
if outcome="hemog" then x4_2=Age;
else x4_2=0;

if outcome="logCD4" then x4t_1=Age*tdisc;
else x4t_1=0;
if outcome="hemog" then x4t_2=Age*tdisc;
else x4t_2=0;

if outcome="logCD4" then x4t2_1=Age*tdisc*tdisc;
else x4t2_1=0;
if outcome="hemog" then x4t2_2=Age*tdisc*tdisc;
else x4t2_2=0;

if outcome="logCD4" then x4t3_1=Age*tdisc*tdisc*tdisc;
else x4t3_1=0;
```

```
if outcome="hemog" then x4t3_2=Age*tdisc*tdisc*tdisc;
else x4t3_2=0;

if outcome="logCD4" then x5_1=WHOs;
else x5_1=0;
if outcome="hemog" then x5_2=WHOs;
else x5_2=0;
run;

data thesis.c;
set b;
if outcome="logCD4" then outc=1;
else if outcome="hemog" then outc=2;
drop outcome PTIDNO;
run;

* Fitting Joint mixed model
proc mixed data=thesis.C covtest noclprint PLOTS(MAXPOINTS=20000) method=REML;
class ID tdisc outc
X1_1 (ref="0")  X1_2 (ref="0") X2_1 (ref="0") X2_2 (ref="0") X12_1 (ref="0")
X12_2 (ref="0") X3_1 (ref="0") X3_2 (ref="0") X5_1 (ref="1");
*X1t_1 X1t_2 X2t_1 X2t_2   X3t_1 X3t_2;
*X1t2_1 X1t2_2 X2t2_1 X2t2_2   X3t2_1 X3t2_2
X1t3_1 X1t3_2 X2t3_1 X2t3_2   X3t3_1 X3t3_2;

model Y=int_1 int_2 X1_1  X2_1   X12_1 X3_1  X4_1  X5_1  T_1   T2_1  T3_1
X1_2  X2_2  X3_2 X1_2*X3_2 X4_2  X4_2*X4_2       T_2   T2_2  T3_2
/noint solution outpm=resid_1;
random int_1  T_1
int_2  T_2/subject=ID type=UN g gcorr;
repeated outc/subject=ID*tdisc type=CS;
ods output covparms=cov_1 solutionF=Fixed_1;
run;


*Calculating Correlaions
proc iml;
D = {
0.03135 0.000803 0.1886 -0.00284,
0.000803 0.00001 -0.00385 0.000109,
0.1886 -0.00385 2.0928 -0.02693,
-0.00284 0.000109 -0.02693 0.001361};

D_slope=
{2.0928 -0.02693,
-0.02693 0.001361};

R={0.8081 0.8237,
0.8237 0.8081};

T = {1 2}; /*Here 2 is used to calculate marginal correlation at time 2*/
D_marg=T*D*t(D) + R;
/*Association between slopes*/
corr_bet_evol=j(nrow(D_slope), nrow(D_slope), 0);
do i=1 to nrow(D_slope);
do j=1 to ncol(D_slope);
corr_bet_evol[i,j] = D_slope[i,j]/sqrt(D_slope[i,i]*D_slope[j,j]);
end;
end;
print corr_bet_evol;
/*Marginal correlation at time 2*/
Marg_corr=j(nrow(D_marg), nrow(D_marg), 0);
do i=1 to nrow(D_marg);
do j=1 to ncol(D_marg);
Marg_corr[i,j]=D_marg[i,j]/sqrt(D_marg[i,i]*D_marg[j,j]);
end;
end;
print Marg_corr;
run;
quit;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Modelling the Evolution of CD4+ Cell Counts and Hemoglobin Concentration Level for HIV-1 Patients on Antiretroviral Therapy (ART) in Mildmay Uganda

Richting: **Master of Statistics-Biostatistics**
Jaar: **2015**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Assefa, Alemu Takele

Datum: **4/09/2015**