Risk factors analysis of longitudinal viral dynamics and immunological responses among HIV-discordant couples

Xiaoyue Wang

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Reading Committee:

Ying Qing Chen, Chair

Ying Huang

Program Authorized to Offer Degree:

Biostatistics

University of Washington

Abstract

Risk factors analysis of longitudinal viral dynamics and immunological responses among HIV-discordant couples

Xiaoyue Wang

Chair of the Supervisory Committee:
Ying Qing Chen
Department of Biostatistics

In their 2011 paper "Prevention of HIV-1 infection with Early Antiretroviral Therapy" (Cohen, et al, *New England Journal of Medicine,* 2011), an interim analysis was performed for the HPTN 052 Study to address both personal and public health benefits from early antiretroviral therapy (ART) among HIV-infected patients and their HIV-uninfected partners. It was found that at least 96% of genetically linked infections caused by human immunodeficiency virus type I (HIV-1) in serodiscordant couples were prevented by ART [1].

Furthermore, as described in their 2016 paper "Antiretroviral Therapy for the Prevention of HIV-1 Transmission" ((Cohen, et al, *New England Journal of Medicine,* 2016), offering the antiretroviral therapy to all patients with HIV-1 infection including 5 years of follow-up proves the durability of ART for the prevention of HIV-1 transmission comparing early and delayed ART. As a result, they conclude that the early initiation of ART results in meaningful decrease in genetically linked HIV-1 infections in sexual partners [2].

In this thesis, we intend to perform risk factors analyses to identify key prognostic factors that are associated with the longitudinal progression of viral dynamics and immunological response, reflected by viral loads and CD4+ counts, respectively, via repeated measurements analysis with Linear Mixed-Effects Models and Generalized Estimating Equations (GEE), for the HIV-infected patients receiving early and delayed ART for the HPTN 052 Study. Such analyses provide more insights into how HIV-1 viral dynamics and patient's immunological response are progressing over time, when the patients are experiencing or yet to experience HIV disease symptoms with ART [2].

After performing all the analyses, it was found that it is better to use Antiretroviral therapy as early as possible based on multivariate analyses for both immediate arm and delayed arm. In addition, we also found that baseline CD4 count, baseline viral load and region are the key risk factors when predicting how the CD4 count and viral load change over time in either immediate arm or delayed arm.

TABLE OF CONTENTS

LIST OF PICTURES

LIST OF TABLES

Chapter 1 INTRODUCTION AND BACKGROUND

1.1 SCIENTIFIC BACKGROUND

The global HIV-1 epidemic is mainly driven by the way of sexual transmission. Thus, durable and effective HIV-1 prevention strategies are highly needed. Several observational studies have shown a decreased acquisition of HIV-1 by sexual partners of patients who receiving antiretroviral therapy.

There are seven main stages of the HIV life cycle. They are binding, fusion, reverse transcription, integration, replication, assembly and budding. HIV damages people's immune system by targeting the CD4 cells. The virus grabs onto the surface of CD4 cell, gets inside of it and then becomes a part of it. As infected CD4 cells multiply, more copies of HIV are made. Those new viruses will find and take over more CD4 cells, and the HIV-free CD4 cells will become fewer and fewer as the cycle continues. Furthermore, bacteria, viruses and any other invading germs can access people's body easily and result in illness (Opportunistic infections) since people's body is lack of defense. Antiretroviral therapy is the use of HIV medicines in order to treat HIV infection and HIV medicines protect people's immune system by blocking HIV at different stages of HIV life cycle [6].

The CD4 also known as T cells are white blood cells that move throughout people's whole body to fight against infection and destroy bacteria, viruses, and other invading germs. CD4 count is a test that measures the number of CD4 cells people have in their blood and can serve as a snapshot of how well one's immune system is functioning, namely if you have more CD4 cells you have better health condition. CD4 count from 500 to 1400 cells per cubic millimeter of blood can be considered as normal. The number of CD4 cells drops as

the HIV infection progresses and a person will be diagnosed with AIDS when his/her CD4 count drops under 200 cells per cubic millimeter. The CD4 count can be increased when effective antiretroviral treatment is used [4].

Viral load is of important interest in the treatment process for people with HIV infection. Viral load is a numerical expression that measures the quantity of virus in a given volume of blood and reflects the amount of HIV in patient's blood, namely the higher the viral load the more severe the active viral infection is. Viral load test results are described as the number of copies of HIV RNA in a milliliter of blood. It is usually undetectable if the viral load drops below a threshold. But, it doesn't mean HIV disappeared completely from patient's body. The more HIV RNA there is in patient's blood, the faster the CD4 count will drop and the greater the risk is for getting ill because of HIV infection [4].

HPTN 052 is a phase III, two arms, randomized, controlled, multi-center trial. It aims to determine if antiretroviral therapy can prevent sexual transmission of HIV-1 among HIV-1 serodiscordant couples. An serodiscordant relationship means one partner is HIV positive while the other is HIV negative. Patients with HIV positive are considered as index patients while patients with HIV negative are classified as partners. Total of 1,763 couples at 13 sites in nine countries (Malawi, Zimbabwe, South Africa, Botswana, Kenya, Thailand, India, Brazil, and the United States) were enrolled into study and most of the couples in which one partner was HIV-1 positive and the other was HIV-negative were heterosexual, accounting for 97% of the whole participants [1].

The 052 Data with repeated time measurements of CD4 counts and viral load for HIV-1 infection participants were collected. Infected participants were randomly assigned to receive antiretroviral therapy either immediately or after a decline in CD4 count or the onset of HIV-1 related symptoms in a 1:1 ratio [1]. On one hand, patients who are in immediate arm of the study means they start ART treatment right after enrollment with CD4+ between

350 and 550 cells per cubic millimeter. On the other hand, patients assigned in the delayed arm of the study start ART treatment after their two consecutive CD4+ counts dropped below 250 cells per cubic millimeter or an illness indicative of the acquired immunodeficiency syndrome was developed instead of providing ART treatment right after enrollment. In addition to this, all index patients who were not already receiving ART treatment were provided ART no matter how much their CD4+ count is after May 12, 2011 [2]. Follow up visits were performed monthly for the first 3 months after initial enrollment and then conducted quarterly.

In their 2011 paper "Prevention of HIV-1 infection with Early Antiretroviral Therapy" (Cohen, et al, *New England Journal of Medicine,* 2011), an interim analysis was performed in order to show both personal and public health benefits from Early Antiretroviral Therapy [1]. The primary clinical endpoint was the first occurrence of pulmonary tuberculosis, severe bacterial infection, a World Health Organization stage 4 event, or death while the primary prevention endpoint was linked HIV-1 transmission in HIV-1-negative partners. A Lan-DeMets implementation of an O'Brien-Fleming monitoring boundary was used with an attempt to evaluate the interim data and Kaplan-Meier method was used in order to calculate the survival probabilities. Based on the results of the interim analysis, more than 96% of genetically linked infections caused by human immunodeficiency virus type I in serodiscordant couples were prevented by antiretroviral therapy. They concluded that the early initiation of antiretroviral therapy has both personal and public health benefits with reduced rates of sexual transmission of HIV-1 and clinical events as well as improved survival of infected individuals.

Furthermore, as described in their 2016 paper "Antiretroviral Therapy for the Prevention of HIV-1 Transmission" (Cohen, et al, *New England Journal of Medicine,* 2016), the antiretroviral therapy was offered to all patients with HIV-1 infection with 5 years of

follow-up after May 12, 2011. It is intended to assess the durability of ART treatment for the prevention of HIV-1 transmission in either early or delayed ART treatment arm [2]. The diagnosis of genetically linked HIV-1 infection of previously HIV-1-negative partner in an intention to treat analysis was the primary study endpoint. Kaplan-Meier method and Cox regression models were used to calculate cumulative event probabilities and estimate relative risk for the treatment effect of immediate arm versus delayed arm, with or without key baseline covariates respectively. A 93% reduction of HIV transmission within couples was found when comparing patients in delayed arm and in immediate arm. As a result, they conclude that the early initiation of ART results in meaningful decrease in genetically linked HIV-1 infections in sexual partners. The benefits and importance of early initiation of ART were emphasized.

With an attempt to figure out how CD4 count and viral load change over time in either treatment arm and detect what variables are important when analyzing patients' performance in terms of CD4 count and viral load, this paper is organized by involving 6 main chapters. Chapter 1 provides readers the background of the reference papers in order to understand the choices of analysis and regression model. Chapter 2 contains the exploratory analysis based on the data we use in this study. Chapter 3 focuses on univariate analysis for viral load and CD4 counts respectively. Chapter 4 provides the multivariate analysis for viral load and CD4 counts respectively. Chapter 5 gives the reader deviance analysis in order to perform the variables selection. Chapter 6 concludes with explanation of the choices of modeling, question discussion and presents the findings throughout the paper.

1.2 STATISTICAL QUESTIONS

The main scientific question of interest is to compare patients' performance of Antiretroviral Therapy between Immediate and delayed arm with respect to CD4 counts and viral load, to

figure out how the changes of the Viral Load and CD4 count differ among different characteristics in general and in different treatment arms and to investigate what factors matters most when analyzing the change of viral load and CD4 count over clinical visits.

1.3 STATISTICAL METHODS

For the data analysis, I used the simulation studies, descriptive analysis, simple linear regression, linear mixed effects model, random effect model, GEE and deviance analysis. Linear mixed-effects models and GEE are appropriate regression models to make inferences about changes in the outcomes over time between groups. Deviance analysis was used to figure out what risk factors matters most with an attempt to predict changes of CD4+ counts and viral load.

1.3.1 SIMULATION STUDIES

Simulation studies can be used to evaluate the proposed estimator's properties. Parametric regression methods for analyzing longitudinal data require parametric specification for the baseline mean function of the response variable over time. In order to avoid this difficulty, semiparametric regression model was proposed. The response variables were generated by the following random-effects model:

$$Y(t) = \alpha_0(t) + \beta_{01}X_1 + \beta_{02}X_2 + \epsilon(t) \quad (1.1)$$

Where $\alpha_0(t) = \alpha_0 + \alpha_1 t^{1/2}$ or $\alpha_0 + \alpha_1 \text{sint}$, $X_1$is Bernoulli with success probability is 0.5, X_2 is independent standard normal, $\epsilon(t)$ is normal with mean ϕ and variance σ_ϵ^2 for all t where ϕ follows normal distribution with mean zero and variance σ_ϕ^2. In addition, the observation times was a random-effects Poisson process with rate $\eta e^{\gamma_{01}X_1+\gamma_{02}X_2}$ where η is an independent gamma random variable with mean 1 and variance σ_η^2. The censoring time was

a variable with an independent distribution of Uniform $(0, \tau)$. Furthermore, τ is chosen to yield a desired number of uncensored observations [5].

For the simulation studies stated in the paper, the results were generated by setting α_0=0, α_1=β_{01}=β_{02}=1, σ_ϕ=0.2, and σ_η=0.5. Independent observation times were generated by setting $\gamma_{01} = \gamma_{02} = 0$ while the dependent observation times were generated by setting γ_{01}=-0.25 and γ_{02}=0.5. τ=4 and 20 yielded on average 4 and 10 observations per subject [5].

The reason for doing simulation studies is to gain insights of the real world and to evaluate if the random effects model we conducted is reasonable. Plots of how the values change over observation times for different error and τ values were displayed below with number of observations being 100.

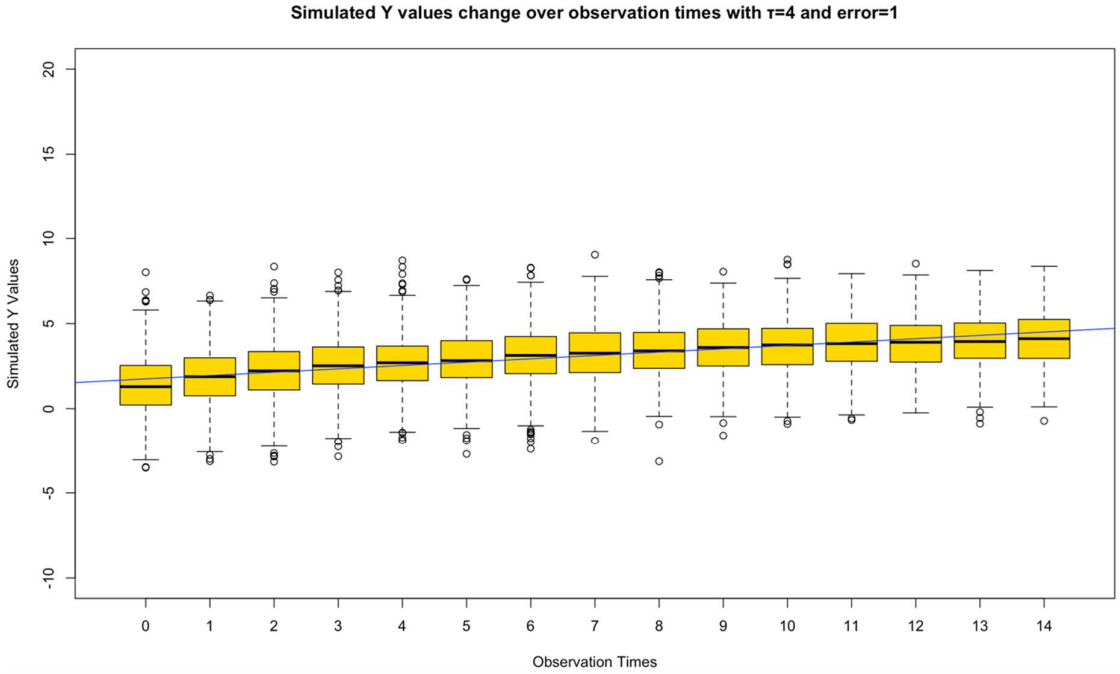

Figure 1.1: Simulated Y values change over observation times with τ=4 and error=1

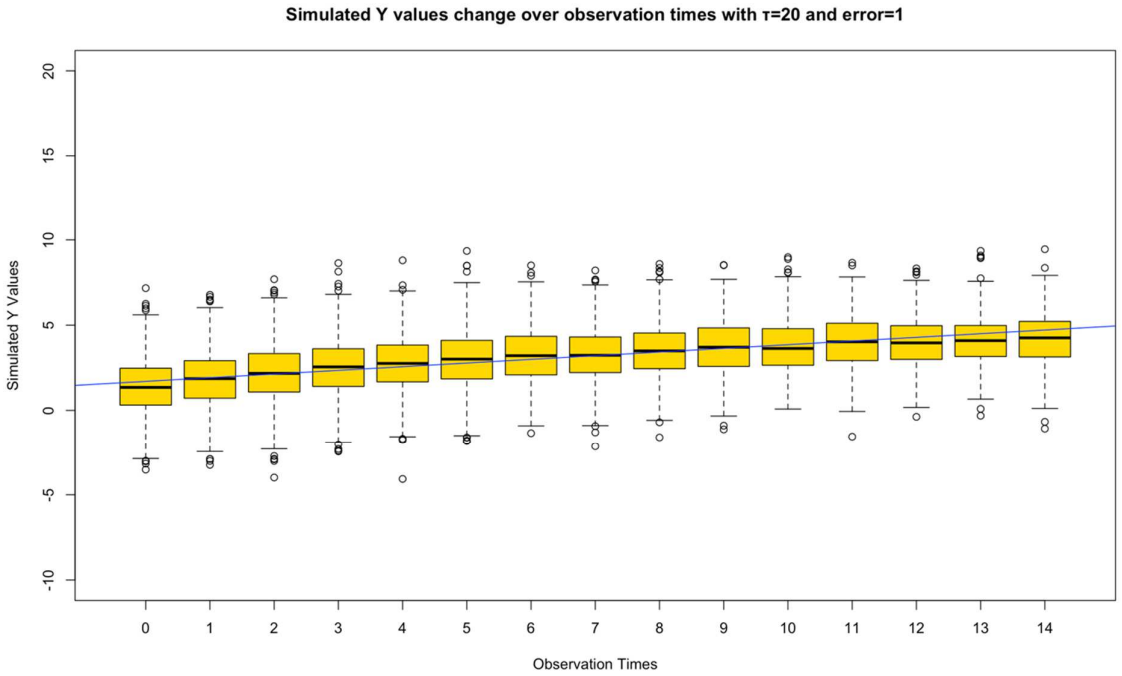

Figure 1.2: Simulated Y values change over observation times with τ=20 and error=1

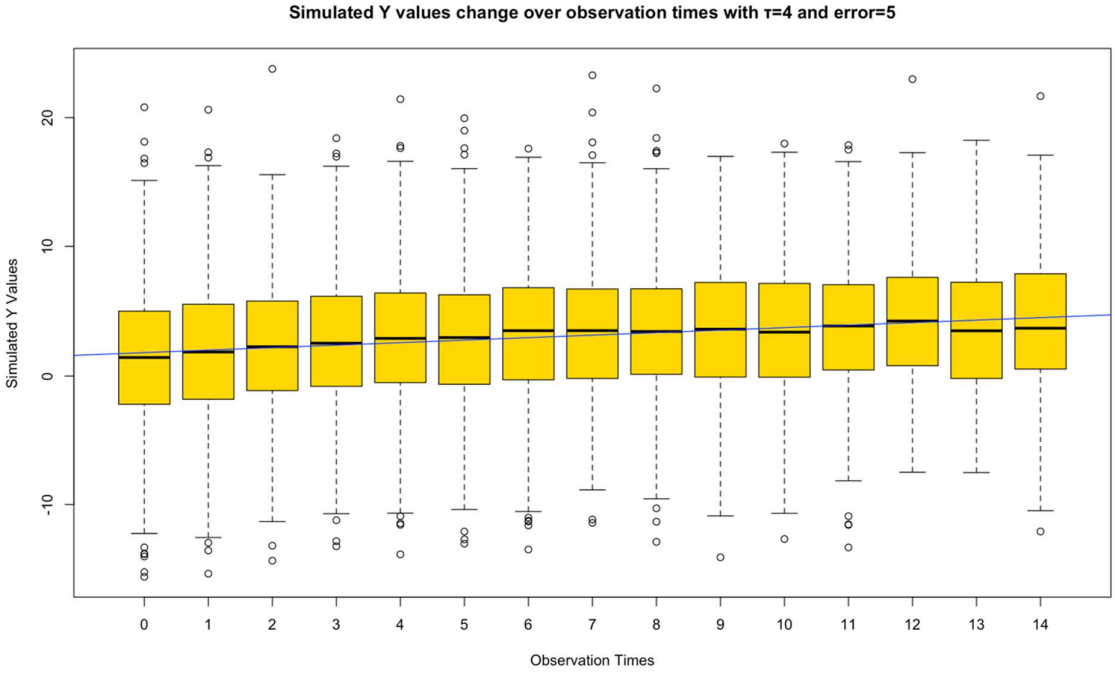

Figure 1.3: Simulated Y values change over observation times with τ=4 and error=5

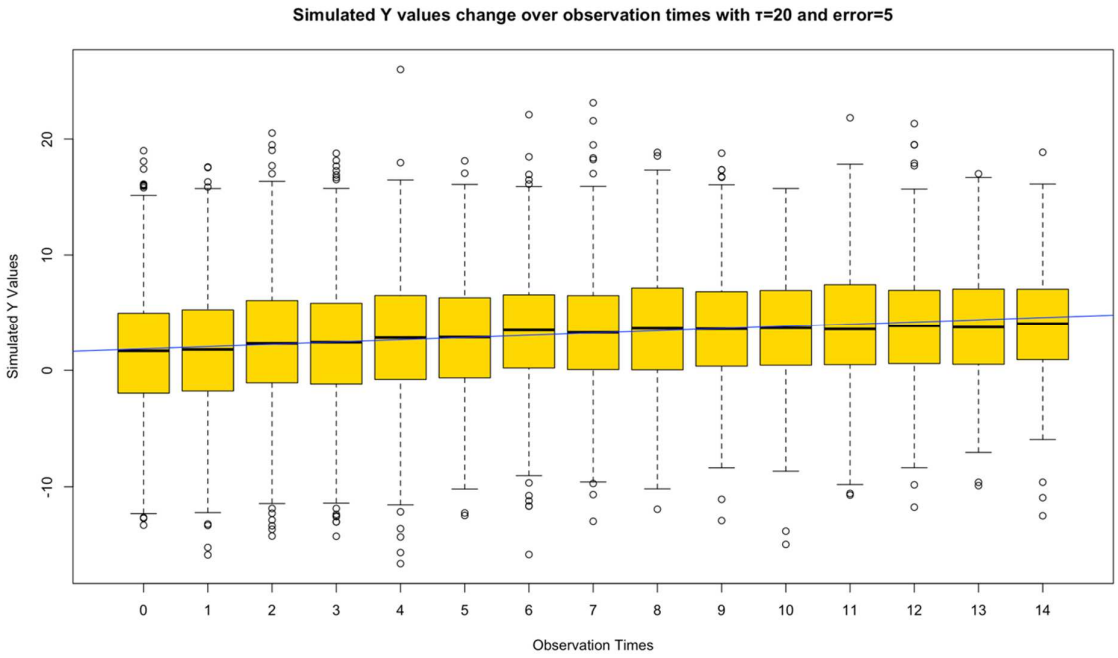

Figure 1.4: Simulated Y values change over observation times with τ=20 and error=5

According to the figures above, we concluded that the change in values of τ will not affect the trend a lot while lower values of the error term tend to have an increasing trend of values over observation times. As stated in paper "Semiparametric and Nonparametric Regression Analysis of Longitudinal Data" (Lin, D., & Ying, Z., *Journal Of The American Statistical Association,* 2001), bias, standard error of the estimates and error sum of squares of simulated Y values were compared among all the combinations of different values of τ and variance of the error for both independent observation times and dependent observation times. It was found that the discrepancies are fairly small [5]. The semiparametric regression model where $\alpha_0(\cdot)$ is unspecified, looks reasonable for dealing with longitudinal data depend on the simulation studies presented by the four plots above.

1.3.2 SIMPLE LINEAR REGRESSION

Linear regression often used to study relationships between a continuous response variable Y and one or more predictor variables X. Linear regression models are commonly fitted by minimizing the least squares criterion. It is very useful when assessing the accuracy of

coefficient estimates, estimate relationship between *Y* and *X*, and make prediction for *Y* based on new value of *X*.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2) \quad (1.2)$$

Typical assumptions of simple linear regression are:

i) Normality: This assumption can be tested with a Q-Q-Plot or by conducting a histogram. Log transformation can also be used when the data are not normally distributed or skewed

ii) Linearity: This assumption can be checked by performing scatterplots to check for outliers or non-linear shape

iii) No collinearity or little collinearity: Collinearity means two or more independent variables are highly correlated with each other. This assumption can be tested by calculating a correlation matrix

iv) Constant variance: This assumption can be tested by plotting residuals versus fitted values. Errors have constant variance when residuals scattered randomly around zero.

1.3.3 LINEAR MIXED-EFFECTS MODEL

The linear mixed-effects model is very useful especially in the scenarios to deal with problems in the longitudinal setting with repeated measurements on the same subject by accounting for correlation of measurements from the same individual. In addition, it has an advantage of handling the missing data. Linear mixed-effects models give unbiased results in the presence of missing data as long as the missing data are missing at random. Missing at random means the propensity for the data point to be missing is related to some of the observed data, but not related to the missing data [3].

We assume the response variable is linearly related to the covariates of interest in an additive way by using linear regression. It assumes that each individual is independent. However, inter-dependence may exist between factors of interest. Thus, we choose to use the linear mixed-effects model since it can have random effects in the model that allow us to assume a different baseline value of the response for each factor. A mixed-effects model has both random and fixed effects while the standard linear regression model only has the fixed effects. In our analyses later, we impose random effects on intercepts which means intercept is an unspecified estimator in our linear mixed-effects model.

Typical assumptions of linear mixed-effects model are:

i) Explanatory variables are linearly related to the response variable: This assumption can be tested by performing scatterplots to check for non-linear shape

ii) Errors have constant variance: This assumption can be tested by plotting residuals versus fitted values. Errors have constant variance when residuals scattered randomly around zero

iii) Errors are normally distributed: This assumption can best be checked with a histogram or a Q-Q-Plot

iv) Random effects are independent of the errors

1.3.4 GENERALIZED ESTIMATING EQUATION (GEE)

Generalized Estimating Equation can be used to estimate the parameters of the generalized linear model. GEE average across the population and only account for one source of clustering. It is a semi-parametric approach to analyze longitudinal data with categorical response. It uses quasi-likelihood estimation instead of maximum likelihood estimation (MLE) or ordinary least squares (OLS) to estimate the parameters [3]. In this paper, GEE was only used to check if the results from linear mixed-effects model is similar with the results by using GEE and not exactly using GEE for analysis.

Typical assumptions of GEE are:

i) The responses variables are correlated or clustered

ii) There is a linear relationship between the covariates and transformation of the response variable

iii) Absence of multicollinearity

iv) Within-subject covariance has some structure

- Independence: Observations over time are independent to each other
- Exchangeable: All observations over time have the same correlation
- AR (1): Correlation decrease as a power of the number of time points apart two observations are
- Unstructured: Correlation between al time points may be different

1.3.5 DEVIANCE ANALYSIS

Deviance is a measure of the discrepancy between the full model and the current model. The full model is the model that contains all the variables of interests and it can give a point of comparison for any model has parameters fewer than the full model. Scaled deviance was used for the comparison between current model to the full model.

$$D^* = 2[ln\ \left(\frac{L_f}{L_c}\right)] \tag{1.3}$$

The difference in deviance between the current model and the full model asymptotically follows a chi-squared distribution with degrees of freedom d where d is the difference between the number of parameters in full model and the current model [3]. ANOVA is used to analyze the differences among group means and it is useful for comparing three or more means for statistical significance.

Typical assumptions of deviance analysis are:

i) Independence of errors

ii) Normality: residuals follow normal distribution

iii) Constant error variance (Homoscedasticity)

Chapter 2 EXPLORATORY DATA ANALYSIS

It is important to perform comprehensive exploratory analysis for demographic data and variables of interests, including both continuous and categorical variables. The variables of interests of this study include age group (categorical:18-20,21-25, 26-30, 31-40, 41-50 and 50+), region (categorical: America, Africa and Asia), gender(categorical), marital status (categorical), education (categorical), household income (continuous), personal income (continuous), country (categorical), baseline CD4 counts (continuous), baseline viral load (continuous), prior use of ART for perinatal purpose (categorical), biomedical intervention (categorical), number of sex partners (continuous), number of sex acts (continuous), uses of condoms (continuous), disease (categorical) and treatment arm (categorical).

Prior use of ART for perinatal purpose only applies to female. For male, missing information of prior use of ART for perinatal purpose is categorized into the category of no prior use of ART for perinatal purpose. Education and income information combined can be used to reflect and address patients' socioeconomic status (SES).

Exploratory analyses were conducted for partner patients at baseline (Table 2.1), index patients at baseline (Table 2.2), partner patients at baseline by site (Table 2.3) and index patients at baseline by site (Table 2.4).

Table 2.1: Partner patients at baseline

Variables (N=1793)	Mean	Median	SD	Min	Max
Age	33.82	32	9.555	18	75
Personal Income	1215867	1600	3241266	0	9999999
Household Income	1287231	4500	3323412	0	9999999
# of sex	3.043	3	0.5488	1	9

partners						
Region	Africa	America	India	Thailand	Asia	
	N=967	N=294	N=425	N=107	N=532	
Gender	Female	Male				
	N=864	N=929				
Marital Status	Married/living with partner	Single	Widowed/Separated /Divorced			
	N=1680	N=100	N=13			
Education	No schooling	Post-secondary	Primary schooling	Secondary schooling		
	N=189	N=187	668	748		
Treatment Arm	Delayed Arm	Immediate Arm				
	N=890	N=903				
Age Group	18-20	21-25	26-30	31-40	41-50	50+
	N=56	N=273	N=464	N=613	N=277	N=110

There are total of 1793 sexual partner of patients enrolled into this study. According to the Table 2.1, the extremely high household income and personal income for partners are due to currency inflation in Africa. The number of personal income and household income are all displayed based on their own local currency. Most of the partners have primary schooling and secondary schooling. Many of the partners are from Africa, secondly Asia and the next is America. The partners of patients average 32 years of age while the oldest partner is 74 years old and the youngest partner is 18 years old. Most of the partners are married.

Table 2.2: Index patients at baseline

N=1763	Mean	Median	SD	Min	Max	
age	33.74	33	8.558	18	67	
Household Income (Missing:19)	1319084	4200	3360634	0	9999999	
Personal Income	1225527	2000	3255179	0	9999999	
Education	No schooling	Post-secondary	Primary schooling	Secondary		
	N=170	N=152	N=706	N=734		
Baseline CD4	213.9	203	107.3	1	471	
Baseline Viral Load	79836.29	26542.5	184178	400	5006049	
Use of condoms	2.194	2	2.334	0	56	
# of sex acts	2.349	2	2.348	0	56	
Age Group	18-20	21-25	26-30	31-40	41-50	50+
	N=45	N=261	N=413	N=690	N=275	N=79
Region	Africa	America	India	Thailand	Asia	
	N=954	N=278	N=425	N=106	N=531	
Gender	Female	Male				
	N=873	N=890				

Marital Status	Married/living with partner	Single	Widowed/Separated/ Divorced		
	N=1667	N=86	N=10		
Prior use of ART for perinatal purposes	Yes	No	Missing		
	N=192	N=494	N=1077		
Treatment Arm	Delayed Arm	Immediate Arm			
	N=877	N=886			

Biomedical intervention									
None	(AZT/3TC)/(LPV/RTV)	(AZT/3TC)/ATV	(AZT/3TC)/ATV/RTV	(AZT/3TC)/ATV/RTV-generic	(AZT/3TC)/EFV	(AZT/3TC)/NVP	(FTC/TDF)/ATV	(FTC/TDF)/ATV/RTV	(FTC/TDF)/EFV
92	70	137	8	1	1152	39	1	9	213
(LPV/RTV)/(FTC/TDF)	3TC/(LPV/RTV)/d4T	3TC/ATV/d4T	3TC/EFV/d4T	3TC/TDF/ATV/RTV	3TC/TDF/EFV	Any supplemental non-study ART	NVP/(FTC/TDF)	NVP/3TC/d4T	
3	1	1	3	2	14	1	9	7	

There are total of 1763 index patients enrolled into this study. According to Table 2.2, the extremely high household income and personal income for index patients are due to currency inflation in Africa. For education, index patients who have primary schooling and secondary schooling take the largest proportion. In addition, most of the index patients are from Africa, secondly Asia and the next is America. The index patients' age is around 33 while the oldest index patient is as high as 67 and as low as 18. Married patients account for the largest proportion in the index patients.

From the distribution of biomedical intervention above, we can see there are total of 19 categories and the distribution is very sparse with most of the categories only have 1 patient. In general, most of the patients are in the biomedical intervention category of (AZT/3TC)/EFV.

Table 2.3: Demographic, by Site-Partner

	Africa N=967			America N=294			Asia N=532		
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Age	34.31	9.8	32	35.78	10.48	34	31.85	8.143	30
Gender	Female	Male		Female	Male		Female	Male	
	407	560		105	189		352	180	
Marital	Married	Single	Divorced	Married	Single	Divorced	Married	Single	Divorced

Status	911		56		0		246		37		11		523		7		2	
Education	None		Post - secondary		Primary	2nd	None		Post - secondary		Primary	2nd	None		Post-secondary	Primary	secondary	
	73		84		372	437	3		42		118	131	113		61	178	180	
Treat Arm	Delayed		Immediate				Delayed		Immediate				Delayed		Immediate			
	483		484				143		151				264		268			
Age Group	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
	21	147	238	341	152	68	10	32	71	91	60	30	25	94	155	181	65	12

Table 2.3 demonstrates demographic information grouped by different site for partners. Age distribution of partners for each site is different with a mean age around 34.31 for Africa, around 35.78 for America and 31.85 for Asia. Asia has the lowest mean age of partner participants. Moreover, female partners account for more in Asia while the gender is relatively proportionally balanced for the other two sites. Most of the partners for each site are married and with an education status in the category of either primary or secondary. In Asia, the partners with no education also account for a big proportion in total. Most partners concentrated on the age group 3 and age group 4 for each site.

Table 2.4: Demographic, by Site-Index

	Africa (N=954)						America (N=278)						Asia (N=531)					
	Mean		SD		Median		Mean		SD		Median		Mean		SD		Median	
Age	33.75		9.209		32		33.71		9.026		32		33.73		6.959		33	
Gender	Female		Male				Female		Male				Female		Male			
	550 (57.7%)		404 (42.3%)				150 (53.9%)		128 (46.1%)				173 (32.6%)		358 (67.4%)			
Marital Status	Married		Single		Divorced		Married		Single		Divorced		Married		Single		Divorced	
	906 (95%)		47 (4.9%)		1 (0.1%)		238		34		6		523		5		3	
Education	No school		Post - secondary		Primary	secondary	No school		Post - secondary		Primary	secondary	No school		Post-secondary	Primary	secondary	
	72		60		401	420	0		25		117	136	98		67	188	178	
Treat Arm	Delayed		Immediate				Delayed		Immediate				Delayed		Immediate			
	477 (50%)		477 (50%)				136		142				264		267			
Age Group	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
	26	163	223	342	146	54	13	37	68	96	47	17	6	61	122	252	82	8

Table 2.4 illustrates demographic information grouped by different site for index patients. Age distribution of index patients for each site is relatively similar with a mean age around 33. Male index patients account for more in Asia while the gender is relatively proportionally balanced for the other two sites. Most of the index patients for each site are married and with an education status in the category of primary and secondary. Index patients concentrated most on the age group 3 and age group 4 for each site.

Plots of CD4 counts and viral load are all only based on index patients since only index have measurements of CD4 counts and viral load.

We also plot the graphs of how the mean values change over observation times in both immediate arm and delayed arm for both log10(RNA) and CD4 counts separately.

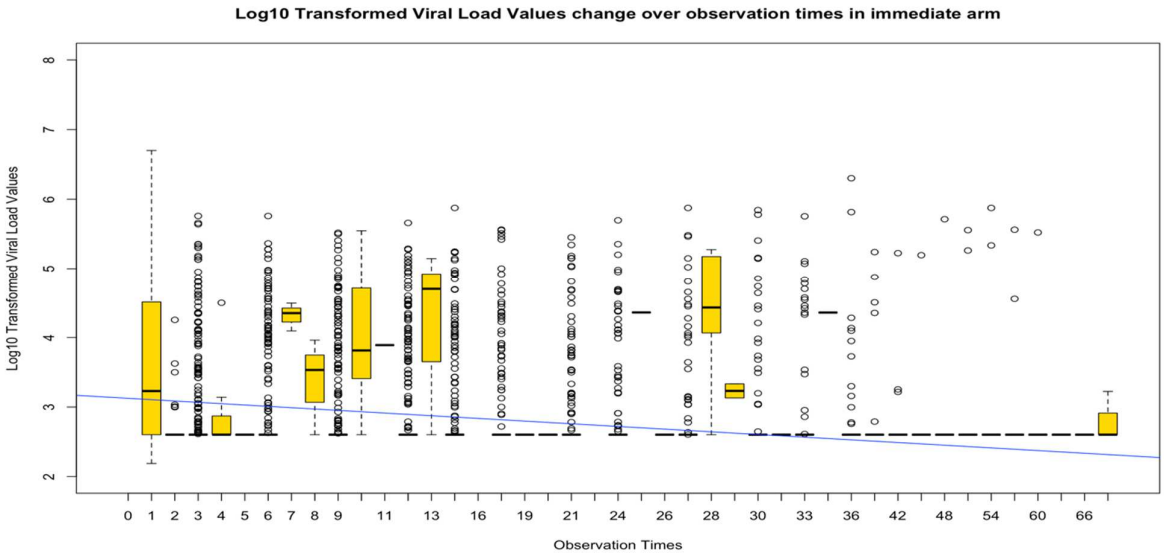

Figure 2.1: log10 Viral load change over observation times in immediate arm

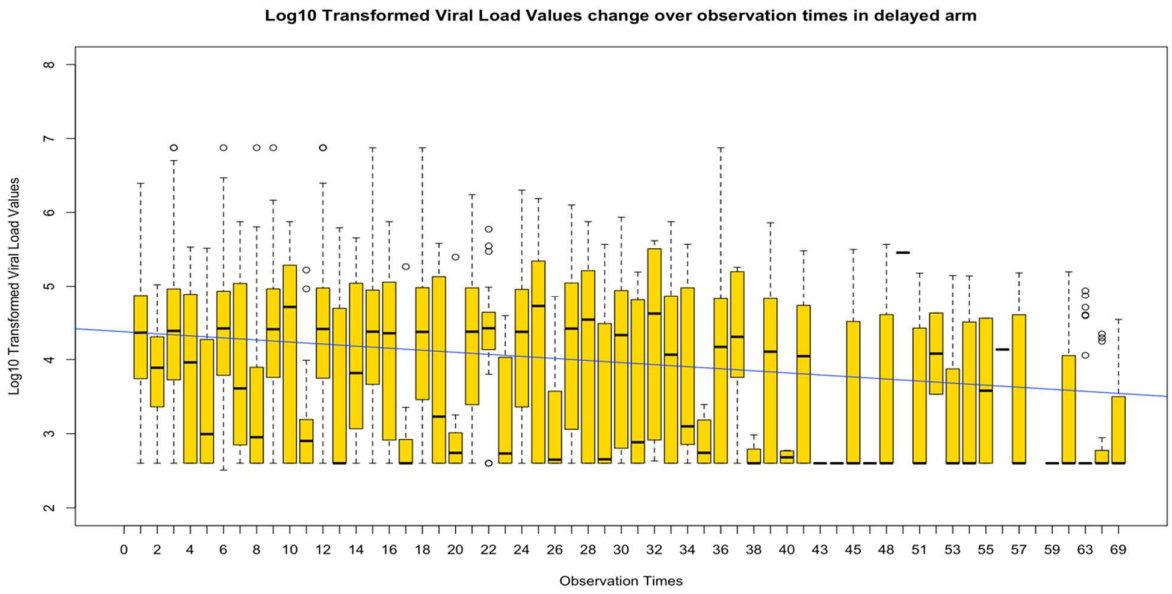

Figure 2.2: log10 Viral load change over observation times in delayed arm

Figure 2.1 and Figure 2.2 above all show a decreasing trend of the estimated mean viral load values over time for both delayed and immediate arms. However, the estimated mean viral load value from the delayed arm is always higher than from the immediate arm over time.

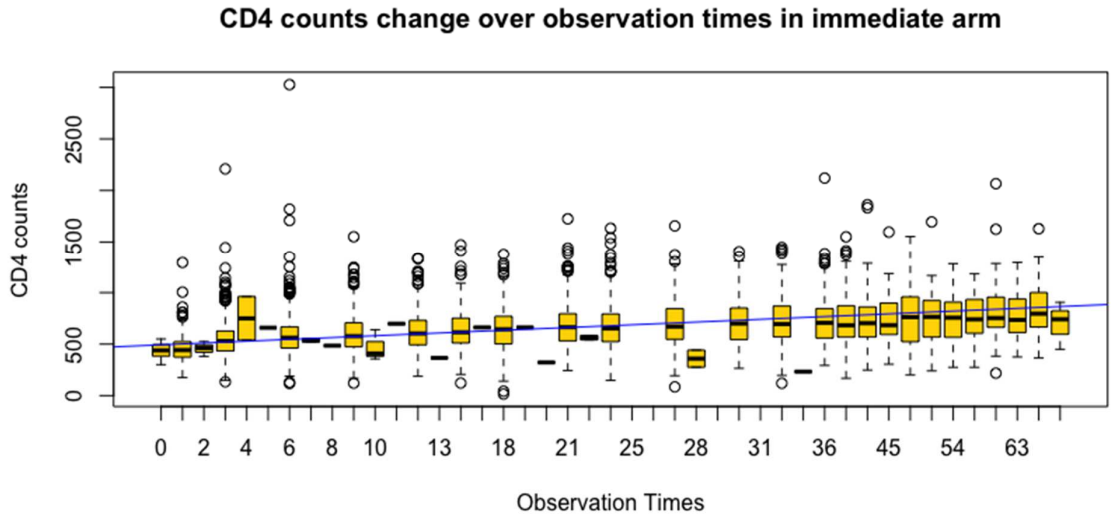

Figure 2.3: Repeated CD4 counts change over observation times in immediate arm

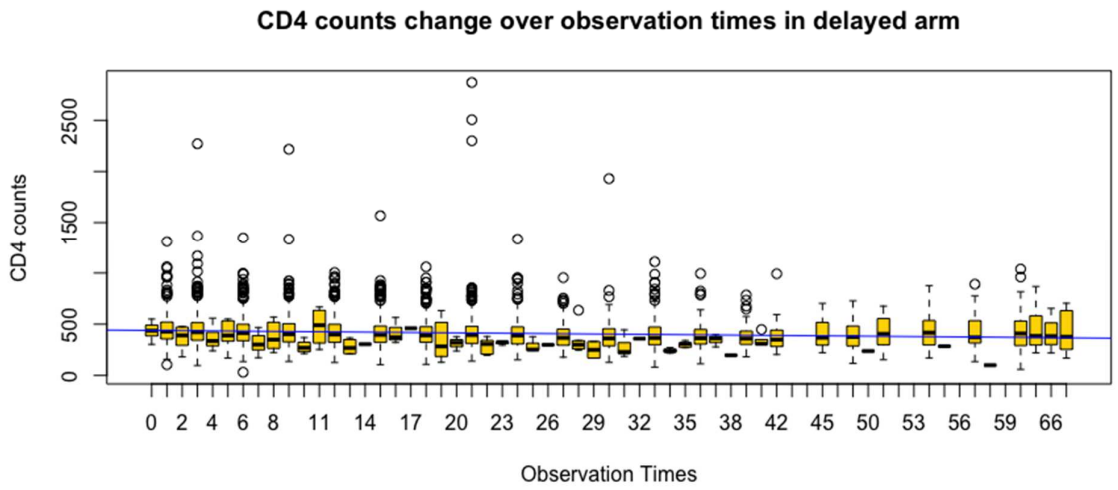

Figure 2.4 :Repeated CD4 counts change over observation times in delayed arm

Figure 2.3 and Figure 2.4 show the estimated mean CD4 count in immediate arm which indicates an increasing trend over time. In contrast, estimated CD4 count in delayed arm shows a slightly decreasing trend over time. The estimated mean CD4 counts from the delayed arm are always below the estimated mean CD4 counts in the immediate arm.

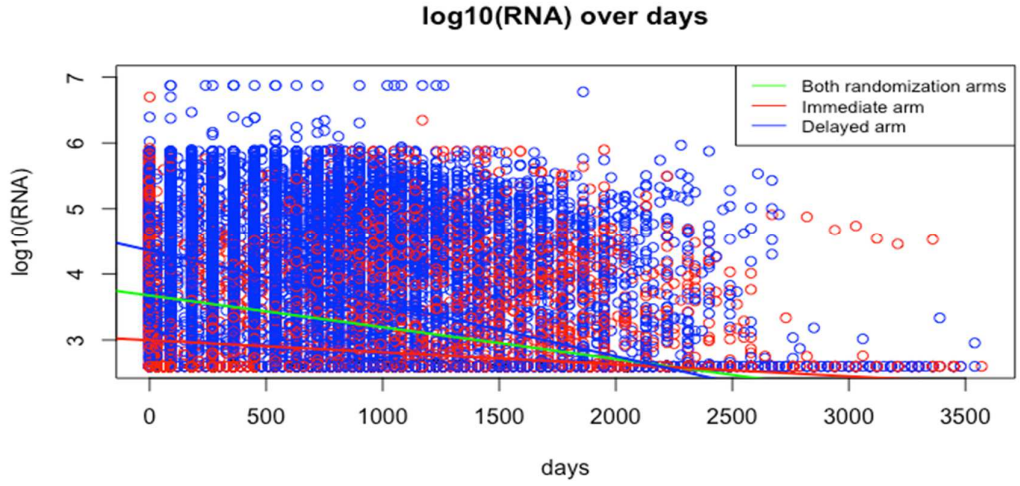

Figure 2.5: Repeated measurements of log10(Viral Load) change over days for all patients, patients in delayed arm and patients in immediate arm

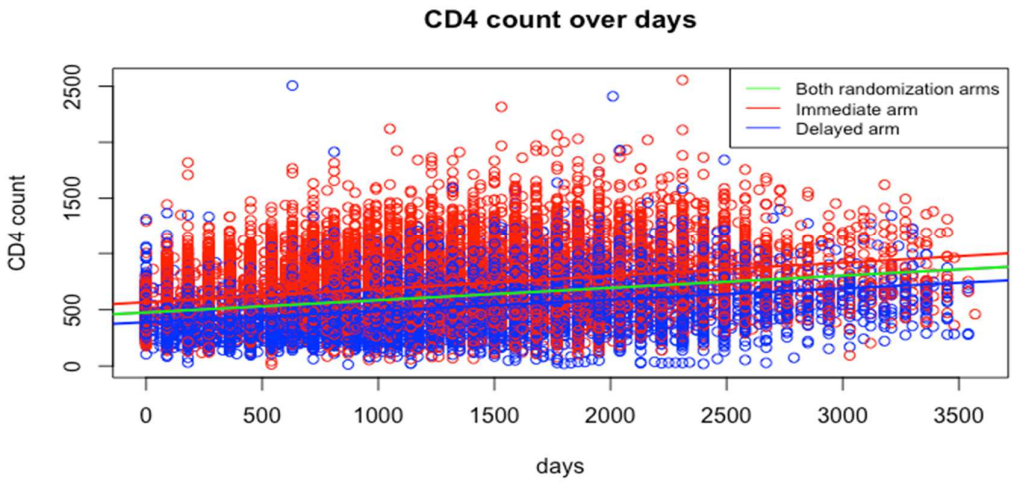

Figure 2.6: Repeated measurements of CD4 count change over days for all patients, patients in delayed arm and patients in immediate arm

For reference, blue dots represent the measurements of log10(RNA) (Figure 9) or CD4 count (Figure 10) in delayed arm while red dots indicate the measurements of log10(RNA) or CD4 counts in immediate arm. Each figure consists of trend lines that were also fitted for immediate arm, delayed arm and for all index patients with a different color. We can see that the estimated average CD4 count over days in immediate arm are always above it in delayed arm. Unlike CD4 count, the estimated average log10(RNA) over days in immediate arm are always below the value in delayed arm. However, in the end the estimated average viral load over days in immediate arm are starting to go below the trend line in delayed arm and it may be due to the lack of data points in the end. Many participants were lost to follow up at the end of study.

Chapter 3 UNIVARIATE ANALYSIS

We are interested in how the viral load and CD4 counts change over time for each variable of interest separately between delayed treatment arm and immediate treatment arm (Table 4.1 and Table 4.2). Only index patients have the measurements information of CD4 counts and

viral loads. Thus, we restricted our univariate analysis to index patients only. Within immediate arm and delayed arm, linear mixed effects models were fitted for each variable of interest with a response variable of either viral load or CD4 count. In addition, we also performed univariate analyses for all the index patients regardless of treatment arm. Linear mixed effects models were used to perform all the univariate analyses.

Before conducting univariate analyses, we changed personal income, household income, condom use and number of sex actions in the past week from continuous variables to categorical variables. For personal income and household income, we divided personal income and household income into 4 categories based on the quantiles for each region due to the amount of income varies a lot in different region. On one hand, we divided the number of condom use during past week into ≥ 4, $0-4$ and any responses not fit into each category were classified into a third category. On the other hand, the number of sex actions in the past week was also distributed into ≥ 4, $0-4$ and any responses not fit into each category were classified into a third category. Baseline RNA was also changed from raw RNA value to log scale RNA.

3.1 UNIVARIATE ANALYSES OF LOG(RNA) IN IMMEDIATE ARM

According to Table 4.1, age group 18-20 has the largest rate of change of the estimated log(RNA) per visit compared to all the other age groups. The rate of change of log(RNA) per visit is lowest in age group 41-50 compared to other age groups. In addition, the estimated log(RNA) decreases 0.011 copies/ml per visit when age increases 1 year. The estimated log(RNA) is 0.071 copies/ml higher per visit for male than female. For marital status, the estimated log(RNA) is 0.14 copies/ml higher per visit among single than married and the estimated log(RNA) is 0.2 copies/ml higher per visit among divorced than married.

The rate of change of estimated log(RNA) is the lowest for people with post-secondary education while it is the highest for the people with no education. People with prior ART use for perinatal purposes tend to have lower rate of change than people without ART use for perinatal purposes. For region, the rate of change of estimated log(RNA) is 0.95 copies/ml higher per visit in America versus Africa and increases 0.06 copies/ml per visit in Asia versus Africa. The estimated log(RNA) decreases 0.098 copies/ml, 0.058 copies/ml, 0.000072 copies/ml per visit according to one unit increase of the household income, personal income and baseline CD4 respectively while increases 0.13 copies/ml per visit for one unit increase of baseline viral load. As regards to use of condoms, the estimated log(RNA) decreases 0.047 copies/ml per visit with 1 unit increase of the use of condoms. In the opposite, the estimated log(RNA) of number of sex acts increases 0.015 per visit with 1 more increase in the number of sex acts.

3.2 UNIVARIATE ANALYSES OF CD4 IN IMMEDIATE ARM

As regards to Table 4.2, the estimated CD4 count has a lower rate of change per visit compared all the age groups to reference age group (18-20) and the estimated rate of change of CD4 count is the lowest in age group 50+. In addition, the estimated CD4 decreases $2.11/mm^3$ per visit when age increases 1 unit. As for gender, the estimated CD4 is $64.11/mm^3$ lower per visit for male than female. For marital status, the rate of change of estimated CD4 decreases $20.61/mm^3$ per visit for single versus married and it increases $104.18/mm^3$ per visit compared divorced to married.

For education, the rate of change of estimated CD4 decreases quickest for people with primary education. People with prior ART use for perinatal purposes tend to have lower change rate per visit than people without ART use for perinatal purposes. According to region, the estimated CD4 is $94.37/mm^3$ higher per visit in America than Africa and is 59.72

higher per visit in Asia than Africa. The estimated CD4 decreases 3.40/mm^3 and 19.50/ mm^3 per visit respectively according to one unit increase of the household income and personal income. The estimated CD4 increases 0.56 /mm^3 and 14.043 /mm^3 per visit respectively with respect to one unit increase in baseline CD4 and log (Baseline RNA). The estimated CD4 increases 0.81/mm^3 per visit for 1 unit increase in the use of condoms while the estimated CD4 decreases 12.073/mm^3 per visit with 1 unit increase in the number of sex acts.

3.3 UNIVARIATE ANALYSES OF CD4 IN DELAYED ARM

As regards to Table 4.3, the estimated CD4 count decreases compared all the age groups to reference group (18-20) and the estimated rate of change of CD4 count is the lowest in age group 50+. In addition, the estimated CD4 decreases 1.88/mm^3 per visit with 1 year increase in age. As for gender, the estimated CD4 is 42.99/mm^3 lower per visit for males to females. For marital status, the estimated CD4 is 5.72/mm^3 lower per visit for single to married and the estimated CD4 is 6.56/mm^3 higher for divorced than married per visit.

The rate of change of estimated CD4 decreases quickest in the category of secondary education. People with prior ART use for perinatal purposes tend to have higher rate per visit than people without ART use for perinatal purposes. For region, the estimated CD4 changes 17.74/mm^3 lower per visit compared America to Africa and 37.12/mm^3 lower per visit compared Asia to Africa. The estimated CD4 decreases 1.88/mm^3 and 9.62/mm^3 per visit respectively according to one unit increase of the household income and personal income. The estimated CD4 increases 0.45/mm^3 per visit with respect to one unit increase in baseline CD4 while decreases 3.18/mm^3 per visit with one unit increase in baseline RNA in log scale. The estimated CD4 increases 5.35/mm^3 per visit with 1 unit increase in the number of

condoms used in the past week while the estimated CD4 increases $2.41/mm^3$ per visit with 1 unit increase in the number of sex acts.

3.4 UNIVARIATE ANALYSES OF LOG(RNA) IN DELAYED ARM

According to Table 4.4, the estimated rate of change of log(RNA) per visit is higher for age group 21-25 to age group 18-20 while the estimated log(RNA) decreases per visit for all the other age groups compared to age group 18-20. The rate of change of log(RNA) per visit is the lowest in age group 41-50 compared to all the other age groups. In addition, the estimated log(RNA) decreases 0.0054 copies/ml per visit with 1 year increase in age. The estimated log(RNA) increases 0.36 copies/ml per visit for males to females. For marital status, the estimated log(RNA) decreases 0.48 copies/ml per visit for single to married and the estimated log(RNA) increases 0.57 copies/ml per visit for divorced to married.

The rate of change of estimated log(RNA) is the lowest for people with post-secondary education while it is the highest for the people with primary education. People with prior ART use for perinatal purposes tend to have lower rate of change than people without ART use for perinatal purposes. For region, the estimated log(RNA) increases 0.97 copies/ml per visit compared America to Africa and increases 0.85 copies/ml per visit in Asia versus Africa. The estimated log(RNA) decreases 0.01 copies/ml and 0.000033 copies/ml per visit according to one unit increase of the household income and baseline CD4 respectively while increases 0.0073 copies/ml and 0.32 copies/ml per visit for one unit increase of personal income and baseline viral load. As regards to use of condoms, the estimated log(RNA) increases 0.001 copies/ml per visit with 1 unit increase of the use of condoms. In the opposite, the estimated log(RNA) of number of sex acts decreases 0.0022 copies/ml per visit with 1 more increase of sex acts.

3.5 UNIVARIATE ANALYSES OF LOG(RNA)

According to Table 4.5, the estimated log(RNA) decreases per visit for all the other age groups compared to reference age group (18-20). The rate of change of RNA per visit is the lowest in age group 41-50. In addition, the estimated log(RNA) decreases 0.011 copies/ml per visit when age increases 1 year. The estimated log(RNA) increases 0.18 copies/ml per visit for males to females. For marital status, the estimated log(RNA) decreases 0.21 copies/ml per visit for single to married and the estimated log(RNA) increases 0.73 copies/ml per visit compared divorced to married.

The rate of change of estimated log(RNA) is the lowest in post-secondary education while it is the highest for the people with primary education. People with prior ART use for perinatal purposes tend to have lower rate of change than people without ART use for perinatal purposes. For region, the estimated log(RNA) increases 1.047 copies/ml per visit in America versus Africa and increases 0.44 copies/ml per visit in Asia versus Africa. The estimated log(RNA) decreases 0.01 copies/ml, 0.0296 copies/ml and 0.00017 copies/ml per visit according to one unit increase of the household income, personal income and baseline CD4 respectively while increases 0.24 copies/ml per visit for one unit increase of baseline viral load. As regards to use of condoms, the estimated log(RNA) increases 0.019 copies/ml per visit with 1 unit increase of the use of condoms. In the opposite, the estimated log(RNA) of number of sex acts decreases 0.017 copies/ml per visit with 1 more increase of sex acts. The estimated rate of change of log(RNA) is 1.69 copies/ml lower in immediate arm than it in delayed arm.

3.6 UNIVARIATE ANALYSES OF CD4

As regards to Table 4.6, the estimated CD4 count decreases compared all the age groups to reference group (18-20) and the estimated rate of change of CD4 count is the lowest in age

group 50+. In addition, the estimated CD4 decreases $1.72/mm^3$ per visit when age increases 1 year. As for gender, the estimated CD4 decreases $50.52/mm^3$ per visit for males to females. For marital status, the estimated CD4 decreases $3.45/mm^3$ per visit for single to married and the estimated CD4 increases $27.58/mm^3$ per visit compared divorced to married.

The rate of change of estimated CD4 is the lowest in post-secondary education. People with prior ART use for perinatal purposes tend to have higher rate per visit than people without ART use for perinatal purposes. For region, the estimated CD4 increases $41.23/mm^3$ per visit in America versus Africa and increases $11.42/mm^3$ per visit in Asia versus Africa. The estimated CD4 decreases $2.65\ /mm^3$ and $14.13\ /mm^3$ per visit respectively according to one unit increase of the household income and personal income. The estimated CD4 increases $0.505/mm^3$ and $4.68/mm^3$ per visit respectively with respect to one unit increases in baseline CD4 and log (Baseline RNA). The estimated CD4 decreases $0.402/mm^3$ per visit for 1 unit increase in the use of condoms while the estimated CD4 decreases $6.083/mm^3$ per visit with 1 unit increase in the number of sex acts. The estimated rate of change in CD4 is $182.65/mm^3$ higher in immediate arm than it in delayed arm.

Chapter 4 MULTIVARIATE ANALYSIS

We also did multivariate linear mixed effects models aiming to detect how the viral load and CD4 counts change over time for all the variables of interest separately between delayed treatment arm and immediate treatment arm. Only index patients have the measurements information of CD4 counts and viral loads. Thus, we restricted our multivariate analysis to index patients only. One linear mixed-effects model was fitted with log-transformed viral load as the response variable and independent variables containing age group, gender, region, marital status, education, personal income, household income, and prior use of ART for perinatal purposes (Table 4.5). Another linear mixed-effects model was fitted with CD4

count as the response variable and independent variables containing age group, gender, region, marital status, education, personal income, household income, and prior use of ART for perinatal purposes (Table 4.6). Biomedical intervention is not included in the multivariate model for the reason that the data are too sparse and too small in some categories of the biomedical intervention (Table 2.2). We also did a multivariate analysis for all the patients with no regards to the treatment arm.

4.1 MULTIVARIATE ANALYSES OF LOG(RNA) IN IMMEDIATE ARM

According to Table 4.1, the estimated log(RNA) decreases per visit for all the age groups compared to reference age group (18-20). The rate of change of log(RNA) per visit is the lowest in age group 31-40 compared to other age groups. In addition, the estimated log(RNA) decreases 0.0089 copies/ml per visit when age increases 1 unit. The estimated log(RNA) increases 0.08 copies/ml per visit for males versus females. For marital status, the estimated log(RNA) increases 0.14 copies/ml per visit compared single to married and the estimated log(RNA) increases 0.20 copies/ml per visit compared divorced to married.

The rate of change of estimated log(RNA) is the lowest for people with post-secondary education while it is the highest for the people with no education. People with prior ART use for perinatal purposes tend to have higher rate of change than people without ART use for perinatal purposes. For region, the estimated log(RNA) increases 0.911 copies/ml per visit compared America to Africa and decreases 0.02 copies/ml per visit compared Asia to Africa. The estimated log(RNA) decreases 0.103 copies/ml per visit according to one unit increase of the household income while increases 0.065 copies/ml, 0.00028 copies/ml and 0.13 copies/ml per visit for one unit increase of personal income, baseline CD4 and baseline log viral load. As regards to use of condoms, the estimated log(RNA) increases 0.14 copies/ml per visit with 1 unit increase of the use of condoms. In

the opposite, the estimated log(RNA) of number of sex acts increases 0.071 copies/ml per visit with 1 more increase in the number of sex acts.

4.2 MULTIVARIATE ANALYSES OF CD4 IN IMMEDIATE ARM

As regards to Table 4.2, the estimated CD4 count increases compared all age groups to reference group (18-20) and the estimated rate of change of CD4 count is the lowest in age group 18-20. In addition, the estimated CD4 decreases $0.14/mm^3$ per visit when age increases 1 year. As for gender, the estimated CD4 decreases $77.074/mm^3$ per visit for males to females. For marital status, the estimated CD4 decreases $31.035/mm^3$ per visit for single to married and the estimated CD4 increases $117.05/mm^3$ per visit for divorced versus married.

The rate of change of estimated CD4 is the lowest for people with primary education. People with prior ART use for perinatal purposes tend to have lower rate per visit than people without ART use for perinatal purposes. As for region, the estimated CD4 increases $92.45/mm^3$ per visit in America versus Africa and increases $73.97/mm^3$ per visit in Asia versus Africa. The estimated CD4 decreases $8.67/mm^3$ per visit according to one unit increase in personal income. The estimated CD4 increases $1.79/mm^3$, $8.67/mm^3$ and $0.62/mm^3$ per visit respectively with respect to one unit increase in household income, Baseline CD4 and log Baseline RNA. The estimated CD4 is $17.95/mm^3$ higher per visit for 1 unit increase in the use of condoms while the estimated CD4 decreases $13.11/mm^3$ per visit with 1 unit increase in the number of sex acts.

4.3 MULTIVARIATE ANALYSES OF CD4 IN DELAYED ARM

As regards to Table 4.3, the estimated CD4 count decreases compared all the other age groups to reference group (18-20) and the estimated rate of change of CD4 count is the

lowest in age group 50+. In addition, the estimated CD4 increases $1.67/mm^3$ per visit when age increases 1 year. For gender, the estimated CD4 decreases $17.43/mm^3$ per visit for males to females. For marital status, the estimated CD4 decreases $3.015/mm^3$ per visit compared single to married and the estimated CD4 increases $48.074/mm^3$ per visit compared divorce to married.

The rate of change of estimated CD4 is the lowest for people with post-secondary education. People with prior ART use for perinatal purposes tend to have higher rate per visit than people without ART use for perinatal purposes. According to region, the estimated CD4 decreases $24.40/mm^3$ per visit in America versus Africa and decreases 29.026 per visit in Asia versus Africa. The estimated CD4 decreases $10.613/mm^3$per visit according to one unit increase of the personal income. The estimated CD4 increases $5.27/mm^3$, $0.44/mm^3$and $3.74/mm^3$ per visit respectively with respect to one unit increase in household income, baseline CD4 and log (Baseline RNA). The estimated CD4 increases $5.663/mm^3$ per visit for 1 unit increase in the use of condoms while the estimated CD4 increases $1.056/mm^3$ per visit with 1 unit increase in the number of sex acts.

4.4 MULTIVARIATE ANALYSES OF LOG(RNA) IN DELAYED ARM

According to Table 4.4, the estimated log(RNA) increases per visit for all the other age groups compared to reference age group (18-20). The rate of change of log(RNA) per visit is the lowest in age group 18-20 compared to all the other age groups. In addition, the estimated log(RNA) decreases 0.038 copies/ml per visit with 1 year increase in age. The estimated log(RNA) increases 0.15 copies/ml per visit for males to females. For marital status, the estimated log(RNA) decreases 0.43 copies/ml per visit compared single to married and the estimated log(RNA) increases 0.202 copies/ml per visit compared divorced to married.

The rate of change of estimated log(RNA) is the lowest for people with post-secondary education while it is the highest for the people with secondary education. People with prior ART use for perinatal purposes tend to have lower rate of change than people without ART use for perinatal purposes. For region, the estimated log(RNA) increases 1.013 copies/ml per visit in America versus Africa and increases 0.64 copies/ml per visit in Asia versus Africa. The estimated log(RNA) increases 0.0038 copies/ml, 0.0109 copies/ml, 0.00082 copies/ml and 0.305 copies/ml respectively per visit for one unit increase of household, personal income, baseline CD4 and baseline log viral load. As regards to the use of condoms, the estimated log(RNA) increases 0.001 copies/ml per visit with 1 unit increase of the use of condoms. In the opposite, the estimated log(RNA) of number of sex acts decreases 0.0022 copies/ml per visit with 1 more increase of sex acts.

4.5 MULTIVARIATE ANALYSES OF LOG(RNA)

According to Table 4.5, the estimated log(RNA) decreases per visit for age groups 21-25 and 26-30 compared to age group 18-20. The rate of change of RNA per visit is the highest in age group 50+ compared to all the other age groups. In addition, the estimated log(RNA) decreases 0.027 copies/ml per visit with one year increase in age. The estimated log(RNA) increases 0.079 copies/ml per visit for males to females. For marital status, the estimated log(RNA) decreases 0.27 copies/ml per visit for single to married and the estimated log(RNA) increases 0.17 copies/ml per visit for divorced to married.

The rate of change of estimated log(RNA) is the lowest in for people with post-secondary education while it is the highest for the people with no education. People with prior ART use for perinatal purposes tend to have lower rate of change than people without ART use for perinatal purposes. For region, the estimated log(RNA) increases 0.95 copies/ml per visit in America versus Africa and increases 0.33 copies/ml per visit in Asia

versus Africa. The estimated log(RNA) decreases 0.046 copies/ml per visit accompanied with one unit increase of the household income while increases 0.039 copies/ml, 0.00045 copies/ml and 0.22 copies/ml per visit for one unit increase of personal income, baseline CD4 and baseline log viral load. As regards to use of condoms, the estimated log(RNA) decreases 0.075 copies/ml per visit with 1 unit increase of the use of condoms. In the opposite, the estimated log(RNA) of number of sex acts increases 0.0342 copies/ml per visit with 1 more increase of sex acts. The rate of change estimated log(RNA) is 1.66 times lower in immediate arm than in delayed arm.

4.6 MULTIVARIATE ANALYSES OF CD4

As regards to Table 4.6, the estimated CD4 count increases compared age group 21-25 and 26-30 to reference group (18-20) and the estimated rate of change of CD4 count is the lowest in age group 50+. In addition, the estimated CD4 increases $0.64/mm^3$ per visit when age increases 1 year. As for gender, the estimated CD4 decreases $44.53/mm^3$ per visit for males to females. For marital status, the estimated CD4 decreases $17.3143/mm^3$ per visit for single to married and the estimated CD4 increases $56.79/mm^3$ per visit for divorced to married.

The rate of change of estimated CD4 is the lowest for people with no education. People with prior ART use for perinatal purposes tend to have lower rate per visit than people without ART use for perinatal purposes. For region, the estimated CD4 increases $32.93/mm^3$ per visit in America versus Africa and increases $20.41/mm^3$ per visit in Asia versus Africa. The estimated CD4 decreases $10.7/mm^3$ per visit according to one unit increase of the personal income. The estimated CD4 increases $3.18/mm^3$, $0.52/mm^3$ and $14.25/mm^3$ per visit respectively with respect to one unit increase in household income, baseline CD4 and log (Baseline RNA). The estimated CD4 increases $9.01/mm^3$ per visit for 1 unit increase in the use of condoms while the estimated CD4 decreases $6.41/mm^3$ per visit with 1 unit

increase in the number of sex acts. The estimated rate of change of CD4 is 184.04/mm^3 higher in immediate arm than in delayed arm.

Table 4.1: Univariate and Multivariate analysis of log(RNA) in Immediate Arm

Variable (Baseline, Index)		N	Univariate (Immediate)		Multivariate (Immediate)		Final Model	
			Est	95% CI	Est	95% CI	Est	95% CI
Age group								
	18-20	23	REF					
	21-25	122	-0.61	(-1.068, -0.14)	-0.44	(-0.89, 0.0089)	-0.47	(-0.91, -0.033)
	26-30	193	-0.64	(-1.091, -0.19)	-0.47	(-0.96, 0.016)	-0.54	(-0.96, -0.11)
	31-40	363	-0.79	(-1.23 -0.35)	-0.59	(-1.18, 0.0021)	-0.70	(-1.123, -0.28)
	41-50	148	-0.8	-1.256, -0.34)	-0.59	(-1.38, 0.203)	-0.78	(-1.22, -0.34)
	50+	37	-0.78	(-1.31, -0.24)	-0.49	(-1.56, 0.58)	-0.77	(-1.28, -0.27)
Age		886	-0.011	(-0.019, -0.0034)	-0.0089	(-0.035, 0.018)		
Gender								
	Female	432	REF					
	Male	454	0.071	(-0.062, 0.205)	0.07998	(-0.086, 0.25)		
Marital								
	Married	834	REF					
	Single	48	0.14	(-0.155, 0.45)	-0.073	(-0.37, 0.22)		
	Divorced	4	0.2	(-0.87, 1.27)	-0.17	(-1.17, 0.83)		
Education								
	None	101	REF					
	Post-secondary	79	-0.299	(-0.593, -0.004)	-0.37	(-0.67, -0.082)	-0.37	(-0.65, -0.077)
	Primary	359	-0.025	(-0.25, 0.20)	-0.14	(-0.35, 0.071)	-0.13	(-0.34, 0.078)
	Secondary	347	-0.149	(-0.37, 0.07)	-0.28	(-0.50, -0.053)	-0.26	(-0.48, -0.04)
Prior ART								
	Yes	93	-0.048	(-0.27, 0.17)	0.08	(-0.14, 0.30)		
	No	793	REF					
Region								
	Africa	477	REF					
	America	142	0.95	(0.76, 1.14)	0.911	(0.72, 1.11)	0.91	(0.72, 1.104)
	Asia	267	0.06	(-0.084, 0.20)	-0.02	(-0.17, 0.13)	-0.002	(-0.15, 0.14)

Household Income		886	-0.098	(-0.16, -0.04)	-0.103	(-0.18, -0.023)	-0.11	(-0.19, -0.036)
Personal Income		886	-0.058	(-0.115, -0.001)	0.065	(-0.021, 0.15)	0.077	(0.0016, 0.15)
Baseline CD4		886	-0.00007	(-0.00066,0.0005)	0.00028	(-0.00028,0.0008)	0.00028	(-0.0003,0.0008)
Baseline RNA		886	0.134	（0.099，0.17）	0.131	(0.096, 0.165)	0.13	(0.098, 0.17)
Use of condoms		886	-0.047	(-0.16, 0.07)	-0.14	(-0.29, 0.0039)	-0.15	(-0.29, 0.0012)
# of Sex acts		886	0.015	(-0.061, 0.091)	0.071	(-0.025, 0.17)	0.074	(-0.022, 0.17)

Based on Table 4.1 results of log(RNA) in immediate arm, we can visualize that only age group, education, region, household income, personal income and baseline viral load are statistically significant based on the p value and 95% confidence interval in univariate analyses and multivariate analyses of final reduced models. In addition, only education, region, household income and baseline viral load remain statistically significant in multivariate analysis where all the variables of interests are included.

Table 4.2: Univariate and Multivariate results of CD4 in Immediate Arm

Variable (Baseline, Index)		N	Univariate (CD4, Immediate)		Multivariate (CD4, Immediate)		Final Model	
			Est	95% CI	Est	95% CI	Est	95%CI
Age group								
	18-20	23	REF					
	21-25	122	-0.24	(-93.21, 47.55)	65.48	(-5.96, 136.91)	57.004	(-12.39, 126.40)
	26-30	193	-30.95	(-75.11, 61.26)	35.61	(-41.99, 113.22)	25.97	(-41.61, 93.54)
	31-40	363	-46.15	(-57.44, 75.47)	45.95	(-48.71, 140.61)	35.17	(-31.72, 102.05)
	41-50	148	-48.23	(-50.57, 87.92)	64.31	(-63.26, 191.88)	48.87	(-21.37, 119.12)
	50+	37	-93.96	(-65.30, 97.65)	29.97	(-143.21, 203.16)	14.67	(-67.50, 96.84)
Age		886	-2.11	(-3.54, -0.66)	-0.14	(-4.44, 4.17)		
Gender								
	Female	432	REF					
	Male	454	-64.11	(-87.28, -40.94)	-77.1	(-103.94, -50.21)	-79.86	(-102.73, -57.00)
Marital								
	Married	834	REF					
	Single	48	-20.61	(-72.29, 31.074)	-31.04	(-76.84, 14.77)		
	Divorced	4	104.18	(-69.75, 278.11)	117.05	(-32.90, 267.001)		

Education								
	None	101	REF					
	Post-secondary	79	3.67	(-48.88, 56.22)	37.85	(-10.03, 85.72)		
	Primary	359	-18.25	(-57.84, 21.34)	-2.32	(-37.24, 32.61)		
	Secondary	347	3.86	(-35.85, 43.57)	32.07	(-4.72, 68.86)		
Prior ART use for perinatal								
	Yes	93	17.99	(-20.17, 56.16)	-15.04	(-50.30, 20.21)		
	No	250						
	Missing	543	REF					
Region								
	Africa	477	REF					
	America	142	94.37	(61.81, 126.93)	92.45	(63.024, 121.87)	97.56	(68.75, 126.38)
	Asia	267	59.72	(33.51, 85.95)	73.97	(49.05, 98.89)	75.23	(51.45, 99.019)
Household Income		886	-3.398	(-13.65, 6.85)	1.79	(-11.17, 14.75)		
Personal Income		886	-19.499	(-29.41, -9.58)	-8.67	(-22.72, 5.38)		
Baseline CD4		886	0.56	(0.463, 0.657)	0.62	(0.533, 0.72)	0.63	(0.54, 0.72)
Baseline RNA		886	14.04	(7.82, 20.25)	25.598	(19.979, 31.22)	25.33	(19.68, 30.97)
Condom Use		886	0.81	(-19.62, 21.25)	17.95	(-5.778, 41.69)		
# of Sex acts		886	-12.07	(-25.43, 1.29)	-13.11	(-28.589, 2.37)		

As for Table 4.2 results for CD4 count in immediate arm, we detected that only age (continuous), gender, region, personal income, baseline CD4 count and baseline viral load are statistically significant based on the p value in univariate analysis while gender, region, baseline CD4 and baseline viral load are statistically significant in the final reduced model. Moreover, only gender, region, baseline CD4 count and baseline viral load remain statistically significant from the multivariate analysis where all the variables of interests are included.

Table 4.3: Univariate and Multivariate results of CD4 in Delayed Arm

Variable (Baseline, Index)	N	Univariate (CD4, Delay)		Multivariate (CD4, Delay)		Final Model	
		Est	95% CI	Est	95% CI	Est	95% CI

Age group								
	18-20	22	REF					
	21-25	139	-15.45	(-35.33, 62.28)	-19.37	(-71.162, 32.43)	-14.12	(-64.15, 35.92)
	26-30	220	-27.43	(-33.23, 61.96)	-24.496	(-80.78, 31.79)	-15.50	(-64.34, 33.34)
	31-40	327	-37.55	(-40.46, 53.36)	-33.61	(-103.21, 35.98)	-12.97	(-61.37, 35.42)
	41-50	127	-40.497	(-57.22, 41.01)	-52.15	(-146.54, 42.23)	-15.84	(-66.72, 35.044)
	50+	42	-110.62	(-116.17, -4.20)	-126.37	(-253.76, 1.014)	-70.74	(-128.76, -12.73)
Age		877	-1.88	(-2.83, -0.93)	1.67	(-1.5189, 4.853)		
Gender								
	Female	441	REF					
	Male	436	-42.99	(-59.28, -26.70)	-17.40	(-36.14, 1.34)	-27.66	(-43.88, -11.44)
Marital								
	Married	833	REF					
	Single	38	-5.72	(-46.05, 34.61)	-3.015	(-39.798, 33.77)		
	Divorced	6	6.56	(-93.24, 106.36)	48.074	(-40.59, 136.74)		
Education								
	None	69	REF					
	Post-secondary	73	-11.49	(-52.49, 29.52)	-1.43	(-40.59, 37.73)		
	Primary	347	20.72	(-11.61, 53.04)	9.59	(-20.098, 39.28)		
	Secondary	387	25.45	(-6.58, 57.49)	11.514	(-19.44, 42.46)		
Prior ART use for perinatal								
	Yes	99	48.89	(23.01, 74.77)	8.32	(-16.86, 33.50)		
	No	776	REF					
Region								
	Africa	477	REF					
	America	136	-17.74	(-41.24, 5.77)	-24.398	(-45.56, -3.238)	-20.51	(-41.33, 0.31)
	Asia	264	-37.12	(-55.73, -18.50)	-29.026	(-46.93, -11.13)	-27.83	(-45.044, -10.61)
Household Income		877	-1.88	(-9.07, 5.30)	5.266	(-4.018, 14.55)		
Personal Income		877	-9.62	(-16.58, -2.65)	-10.605	(-20.17, -1.04)		
Baseline CD4		877	0.45	(0.39, 0.505)	0.44	(0.379, 0.496)	0.44	(0.38, 0.495)
Baseline RNA		877	-3.18	(-7.63, 1.28)	3.743	(-0.305, 7.79)	3.699	(-0.31, 7.71)
Condom use		877	5.35	(-9.00, 19.69)	5.66	(-11.011, 22.33)		

Sex acts		877	2.41	(-7.18, 12.00)	1.056	(-9.89, 12.00)		

As for Table 4.3 results for CD4 count in delayed arm, we detected that only age group, age (continuous), prior use of ART, gender, region, personal income and baseline CD4 count are statistically significant based on the p value in univariate analysis while age group, gender, region and baseline CD4 are statistically significant in the final reduced model. Moreover, only region, personal income and baseline CD4 count remain statistically significant in multivariate analysis where all the variables of interests are included.

Table 4.4: Univariate and Multivariate results of log(RNA) in Delayed Arm

Variable (Baseline, Index)		N	Univariate (RNA, Delay)		Multivariate (RNA, Delay)		Final Model	
			Est	95% CI	Est	95% CI	Est	95% CI
Age group								
	18-20	22	REF					
	21-25	139	0.114	(-0.47, 0.69)	0.34	(-0.15, 0.83)		
	26-30	220	-0.048	(-0.61, 0.52)	0.32	(-0.21, 0.86)		
	31-40	327	0.066	(-0.49, 0.62)	0.55	(-0.11, 1.21)		
	41-50	127	-0.007	(-0.59, 0.58)	0.74	(-0.16, 1.63)		
	50+	42	-0.13	(-0.79, 0.54)	1.19	(-0.022, 2.39)		
Age		877	-0.0054	(-0.015, 0.0043)	-0.038	(-0.068, -0.0077)	-0.011	(-0.019, -0.0026)
Gender								
	Female	441	REF					
	Male	436	0.36	(0.20, 0.53)	0.15	(-0.025, 0.33)		
Marital								
	Married	833	REF					
	Single	38	-0.48	(-0.89, -0.068)	-0.434	(-0.785, -0.082)	-0.4497	(-0.79, -0.11)
	Divorced	6	0.57	(-0.45, 1.58)	0.202	(-0.65, 1.052)	0.138	(-0.70, 0.98)
Education								
	None	69	REF					
	Post-secondary	73	-0.048	(-0.47, 0.37)	-0.064	(-0.44, 0.31)		
	Primary	347	0.12	(-0.21, 0.45)	0.122	(-0.16, 0.403)		

	Secondary	387	-0.025	(-0.35, 0.30)	0.14	(-0.15, 0.43)		
Prior ART use for perinatal								
	Yes	99	-0.55	(-0.81, -0.29)	-0.29	(-0.53, -0.05)	-0.34	(-0.56, -0.11)
	No	776	REF					
Region								
	Africa	477	REF					
	America	136	0.97	(0.74, 1.19)	1.013	(0.81, 1.22)	0.996	(0.79, 1.20)
	Asia	264	0.85	(0.68, 1.031)	0.64	(0.47, 0.81)	0.66	(0.50, 0.82)
Household Income		877	-0.01	(-0.083, 0.062)	0.0038	(-0.084, 0.092)		
Personal Income		877	0.0073	(-0.064, 0.078)	0.011	(-0.08, 0.102)		
Baseline CD4		877	-0.000033	(-0.0007, 0.0006)	0.00082	(0.0003, 0.0014)	0.00084	(0.00028, 0.0014)
Baseline RNA		877	0.32	(0.29, 0.36)	0.31	(0.27, 0.34)	0.31	(0.27, 0.35)
Condom use		877	0.001	(-0.14,0.15)	-0.073	(-0.23, 0.085)		
Sex acts		877	-0.0022	(-0.10, 0.095)	0.031	(-0.073, 0.14)		

According to Table 4.4 results for log(RNA) in delayed arm, we detected that only gender, marital status, prior use of ART, region and baseline viral load are statistically significant based on the p value in univariate analysis while age (continuous), marital status, prior use of ART, region, baseline viral load and baseline CD4 are statistically significant in the final reduced model. Moreover, only age (continuous), marital status, prior use of ART, region, baseline viral load and baseline CD4 count remain statistically significant in multivariate analysis where all the variables of interests are included.

Table 4.5: Univariate and Multivariate results of log(RNA) for all index patients

Variable (Baseline, Index)		N	Univariate (RNA)		Multivariate (RNA)		Final Model	
			Est	95% CI	Est	95% CI	Est	95% CI
Age group								
	18-20	45	REF		REF			
	21-25	261	-0.206	(-0.664, 0.25)	-0.011	(-0.343, 0.322)	-0.13	(-0.45, 0.195)
	26-30	413	-0.302	(-0.75, 0.14)	-0.033	(-0.393, 0.328)	-0.27	(-0.583, 0.049)
	31-40	690	-0.4	(-0.84, 0.036)	0.083	(-0.357, 0.524)	-0.33	(-0.64, -0.014)

	41-50	275	-0.45	(-0.91, 0.0029)	0.216	(-0.38, 0.81)	-0.44	(-0.77, -0.11)
	50+	79	-0.395	(-0.92, 0.13)	0.574	(-0.23, 1.37)	0.57	(-0.75, 0.0079)
Age		1763	-0.011	(-0.019, -0.003)	-0.027	(-0.047, -0.0076)	-0.37	
Gender						(-0.0414, 0.199)		
	Female	873	REF		REF			
	Male	890	0.18	(0.047, 0.31)	0.079	(-0.058, 0.064)	0.11	(0.0028, 0.22)
Marital								
	Married	1667	REF					
	Single	86	-0.21	(-0.52, 0.101)	-0.27	(-0.49, -0.041)		
	Divorced	10	0.73	(-0.18, 1.63)	0.17	(-0.47, 0.81)		
Education								
	None	170	REF		REF			
	Post-secondary	152	-0.012	(-0.321, 0.297)	-0.21	(-0.44, 0.018)	-0.26	(-0.48, -0.0299)
	Primary	706	0.197	(-0.04, 0.44)	-0.0075	(-0.18, 0.16)	-0.0082	(-0.18, 0.16)
	Secondary	734	0.124	(-0.112, 0.36)	-0.068	(-0.25, 0.11)	-0.1	(-0.28, 0.075)
Prior ART use								
	Yes	192	-0.27	(-0.038, -0.036)	-0.16	(-0.32, 0.00099)		
	No	1569	REF					
Region								
	Africa	954	REF		REF			
	America	278	1.047	(0.86, 1.24)	0.95	(0.81, 1.09)	0.93	(0.79, 1.075)
	Asia	531	0.44	(0.30, 0.59)	0.33	(0.21, 0.44)	0.35	(0.24, 0.46)
Household Income		1763	-0.058	(-0.12, -0.00025)	-0.046	(-0.105, 0.013)		
Personal Income		1763	-0.0296	(-0.086, 0.027)	0.039	(-0.023, 0.10)	0.012	(-0.034, 0.057)
Baseline CD4		1763	-0.00017	(-0.00073, 0.0004)	0.0005	(0.00011, 0.00089)	0.0005	(0.0001,0.0009)
Baseline RNA		1763	0.241	(0.21, 0.28)	0.22	(0.19, 0.25)	0.22	(0.196, 0.25)
Condom use		1763	0.019	(-0.096, 0.134)	-0.075	(-0.18, 0.033)		
# of Sex acts		1763	0.017	(-0.059, 0.093)	0.034	(-0.036, 0.10)		
Treat arm								
	Immediate	886	-1.69	(-1.80, -1.59)	-1.66	(-1.75, -1.57)	-1.66	(-1.75, -1.57)
	Delayed	877	REF					

According to Table 4.5 results for log(RNA) for all index patients, we found that only age (continuous), gender, prior use of ART, region, household income, randomization arm and baseline viral load are statistically significant based on the p value in univariate analysis while age group, gender, education, region, randomization arm, baseline viral load and baseline CD4 are statistically significant in the final reduced model. Moreover, only age (continuous), marital status, region, randomization arm, baseline viral load and baseline CD4 count remain statistically significant in multivariate analysis where all the variables of interests are included.

Table 4.6: Univariate and Multivariate results of CD4 counts for all index patients

Variable (Baseline, Index)		N	Univariate (CD4)		Multivariate (CD4)		Final Model	
			Est	95% CI	Est	95% CI	Est	95% CI
Age group								
	18-20	45	REF		REF			
	21-25	261	-17.11	(-73.97, 39.75)	12.64	(-32.93, 58.22)	-29.385	(-29.39, 58.87）
	26-30	413	-37.73	(-93.03, 17.57)	0.031	(-49.50, 49.56)	-39.449	(-39.45, 46.91）
	31-40	690	-39.63	(-93.85, 14.58)	-1.79	(-62.67, 59.10)	-35.711	(-35.71, 49.88）
	41-50	275	-39.89	(-96.53, 16.75)	-2.028	(-84.36, 80.31)	-30.58	(-30.58, 59.51）
	50+	79	-109.58	(-175.2, -43.96)	-51.66	(-163.1, 59.79)	-79.90	(-79.9, 24.35）
Age		1763	-1.72	(-2.69, -0.74)	0.64	(-2.14, 3.41)		
Gender								
	Female	873	REF		REF			
	Male	890	-50.52	(-67.07, -33.97)	-44.53	(-61.34, -27.73)	-62.34	(-62.34, -31.55）
Marital								
	Married	1667	REF					
	Single	86	-3.45	(-42.07, 35.17)	-17.31	(-48.04, 13.42)		
	Divorced	10	27.58	(-83.1, 138.26)	56.79	(-28.63, 142.20)		
Education								
	None	170	REF		REF			

	Post-secondary	152	-18.75	(-57.92, 20.43)	18.31	(-13.92, 50.55)	-13.48	(-13.483, 49.81)
	Primary	706	-16.64	(-46.73, 13.44)	0.99	(-22.98, 24.96)	-22.59	(-22.585, 25.19)
	Secondary	734	-9.54	(-39.50, 20.42)	20.88	(-4.21, 45.97)	-4.197	(-4.197, 45.04)
Prior ART use for perinatal								
	Yes	192	30.67	(3.97, 57.37)	-5.87	(-28.15,16.41)		
	No	494						
	Missing	1075	REF					
Region								
	Africa	954	REF		REF			
	America	278	41.23	(17.48, 64.99)	32.927	(14.25, 51.61)	15.237	(15.24, 52.16）
	Asia	531	11.42	(-7.53, 30.37)	20.41	(4.56, 36.26)	5.798	(5.80， 37.05）
Household Income		1763	-2.65	(-9.92, 4.61)	3.175	(-4.996, 11.35)		
Personal Income		1763	-14.13	(-21.18, -7.08)	-10.7	(-19.39, -2.02)	-13.729	(-13.73, -0.99)
Baseline CD4		1763	0.505	(0.44, 0.57)	0.521	(0.466, 0.576)	0.466	(0.466,0.58)
Baseline RNA		1763	4.68	(0.21, 9.16)	14.247	(10.67, 17.824)	10.779	(10.78, 17.93)
Use of condoms		1763	-0.402	(-14.90, 14.10)	9.008	(-5.872, 23.89)		
# of Sex acts		1763	-6.08	(-15.67, 3.51)	-6.413	(-16.148, 3.32)		
Treat arm								
	Immediate	886	182.65	(167.94, 197.36)	184.04	(171.14,196.94)	170.68	(170.68, 196.49）
	Delayed	877	REF					

According to Table 4.6 results for CD4 count for all index patients, we examined that only age (continuous), age group, gender, prior use of ART, region, personal income, randomization arm, baseline CD4 count and baseline viral load are statistically significant based on the p value in univariate analysis while gender, region, personal income, randomization arm, baseline viral load and baseline CD4 are statistically significant in the final reduced model. Moreover, only gender, region, personal income, randomization arm, baseline viral load and baseline CD4 count remain statistically significant in multivariate analysis where all the variables of interests are included.

Chapter 5 VARIABLE SELECTION

Deviance analysis approach was used to investigate the final reduced model of each full model for both log(RNA) (Table 4.1, 4.2) and CD4 count in either delayed arm (Table 4.3) or immediate arm (Table 4.4) as well as log(RNA) (Table 4.5) and CD4 count (Table 4.6) for all index patients regardless of treatment arm. Compare the log likelihood difference between current model and reduced model to chi-squared distribution with degree of freedom d where d is the parameter difference between current model and reduced model [3]. Multivariate analyses by using linear mixed effects models were performed for each final reduced model of with response variable is either log(RNA) or CD4 count in each treatment arm. This process is summarized in statistical method 1.3.5.

5.1 REDUCED MODEL OF LOG(RNA) IN DELAYED ARM

Regarding to Table 5.1, the reduced model for log(RNA) in delayed arm is:

$$log(RNA) = \beta_0 + \beta_1 * Age + \beta_2 * Marital\ Status + \beta_3 * Prior\ ART + \beta_4 * Region + \beta_5 * BaselineCD4 + \beta_6 * Baseline\ RNA \quad (5.1)$$

Based on Table 5.1, the rate of change of estimated log(RNA) decreases 0.011 copies/ml per visit with 1 year increase in age with all the other variables of interests fixed. The rate of change of estimated log(RNA) is 0.4497 copies/ml lower for single people than married people per visit while it is 0.138 higher for divorced people than married people per visit. According to prior ART use of perinatal purposes, people with prior ART use for perinatal purposes tend to have 0.335 copies/ml lower rate of change per visit than people without prior ART use of perinatal purposes. The rate of change of estimated log(RNA) increases 0.996 copies/ml per visit in America versus Africa while it increases 0.66 copies/ml per visit in Asia versus Africa. The rate of change of estimated log(RNA) increases 0.000835

copies/ml and 0.308 copies/ml per visit respectively with 1 unit increase in baseline CD4 and baseline viral load with all the other variables fixed.

5.2 REDUCED MODEL OF LOG(RNA) IN IMMEDIATE ARM

Regarding to Table 5.2, the reduced model for log(RNA) in immediate arm is:

$$log(RNA) = \beta_0 + \beta_1 * Agegroup + \beta_2 * Education + \beta_3 * Region + \beta_4 * Income_h + \beta_5 * Income_p + \beta_6 * BaselineCD4 + \beta_7 * BaseineRNA + \beta_8 * CondomUse + \beta_9 * Sex\ acts \quad (5.2)$$

The rate of change of estimated log(RNA) is the highest in age group 18-20 among all other age groups with all the other variables fixed. As for education, the rate of change of estimated log(RNA) is the highest for people with no education and is the lowest for people with post-secondary education with all the other variables fixed. The rate of change of estimated log(RNA) increases 0.913 copies/ml per visit in America to Africa while it decreases 0.00179 copies/ml visit in Asia versus Africa. The rate of change of estimated log(RNA) decreases 0.111 copies/ml per visit with 1 unit increase in household income while it increases 0.0772 copies/ml per visit with 1 unit increase of personal income with all the variables fixed. The rate of change of estimated log(RNA) increases 0.000276 copies/ml and 0.132 copies/ml respectively per visit with 1 unit increase in baseline CD4 and baseline viral load with all the other variables fixed.

The estimated log(RNA) decreases 0.146 copies/ml per visit for 1 unit increase in the use of condoms while the estimated log(RNA) increases 0.0738 copies/ml per visit with 1 unit increase in the number of sex acts with all the other variables fixed.

5.3 REDUCED MODEL OF CD4 IN DELAYED ARM

According to Table 5.3, the reduced model for CD4 count in delayed arm is:

$$CD4 = \beta_0 + \beta_1 * Agegroup + \beta_2 * Gender + \beta_3 * Region + \beta_4 * BaselineCD4 + \beta_5 * BaselineRNA \quad (5.3)$$

The rate of change of estimated CD4 is the highest in age group 18-20 among other age groups with all the other variables fixed. With regard to gender, the rate of change of estimated CD4 of males is $27.66/mm^3$ lower than females per visit. The rate of change of estimated CD4 decreases $20.5/mm^3$ per visit in America to Africa while it decreases $27.83/mm^3$ per visit in Asia to Africa. The rate of change of estimated CD4 increases $0.436/mm^3$ and $3.699/mm^3$ per visit with 1 unit increase in baseline CD4 and baseline viral load respectively with all the other variables fixed.

5.4 REDUCED MODEL OF CD4 IN IMMEDIATE ARM

With regards to Table 5.4, the reduced model for CD4 count in immediate arm is:

$$CD4 = \beta_0 + \beta_1 * Agegroup + \beta_2 * Gender + \beta_3 * Region + \beta_4 * BaslineCD4 + \beta_5 * BaselineRNA \quad (5.4)$$

With regards to Table 5.4, the rate of change of estimated CD4 is the highest in age group 21-25 among all the other age groups with other variables fixed. With regard to gender, the rate of change of estimated CD4 of males is $79.86/mm^3$ lower than females per visit. The rate of change of estimated CD4 increases $97.56/mm^3$ per visit in America versus Africa while it increases $75.23/mm^3$ per visit compared Asia to Africa. The rate of change of estimated CD4 increases $0.628/mm^3$ and $25.33/mm^3$ per visit with 1 unit increase in baseline CD4 and baseline viral load respectively with all the other variables fixed.

5.5 REDUCED MODEL OF LOG(RNA)

As for Table 5.5, the reduced model of log(RNA) count is:

$$log(RNA) = \beta_0 + \beta_1 * Agegroup + \beta_2 * Gender + \beta_3 * Region + \beta_4 * Education + \beta_5 * Income_p + \beta_6 * BaselineCD4 + \beta_7 * BaselineRNA + \beta_8 * TreatmentArm \quad (5.5)$$

The rate of change of estimated log(RNA) is the highest in age group 50+ among other age groups with all the other variables fixed. With regard to gender, the rate of change of estimated log(RNA) of males is 0.113 copies/ml higher than females per visit. As for education, the rate of change of estimated log(RNA) per visit is the highest for people with no education and is the lowest for people with post-secondary education with all the other variables fixed. The rate of change of estimated log(RNA) increases 0.934 copies/ml per visit compared America to Africa while it increases 0.352 copies/ml per visit compared Asia to Africa. The rate of change of estimated log(RNA) decreases 0.118 copies/ml per visit with 1 unit increase in personal income with all the variables fixed. The rate of change of estimated log(RNA) increases 0.00051 copies/ml and 0.221 copies/ml per visit with 1 unit increase in baseline CD4 and baseline viral load respectively with all the other variables fixed. The estimated log(RNA) is 1.66 copies/ml lower per visit in immediate arm than in delayed arm with all the other variables fixed.

5.6 REDUCED MODEL OF CD4

According to Table 5.6, the reduced model of CD4 count is:

$$CD4 = \beta_0 + \beta_1 * Agegroup + \beta_2 * Gender + \beta_3 * Region + \beta_4 * Education + \beta_5 * Income_p + \beta_6 * BaselineCD4 + \beta_7 * BaselineRNA + \beta_8 * TreatmentArm \quad (5.6)$$

The rate of change of estimated CD4 is the highest in age group 18-20 among all the other age groups with all the other variables fixed. With regard to gender, the rate of change of estimated CD4 of males is $62.336/mm^3$ lower than females per visit. As for education, the rate of change of estimated CD4 is the highest for people with no education and lowest in for people with primary education with all the other variables fixed. The rate of change of

estimated CD4 increases $15.24/mm^3$ per visit compared America to Africa while it increases $5.798/mm^3$ per visit compared Asia to Africa. The rate of change of estimated CD4 decreases $13.73/mm^3$ per visit with 1 unit increase in personal income with all the variables fixed. The rate of change of estimated CD4 increases $0.466/mm^3$ and $10.78/mm^3$ per visit with 1 unit increase in baseline CD4 and baseline viral load respectively with all the other variables fixed. The estimated log(RNA) is $170.68/mm^3$ higher per visit in immediate arm than in delayed arm other variables fixed.

Table 5.1: Deviance analysis for log(RNA) in delayed arm

log(RNA) in Delayed Arm					
Full Model: (LogLik: -39540.03) log(RNA)~month+agegroup+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna					
Model	Deviance	Delta	DF	P value	AIC
log(RNA)~month+agegroup+age+gender+region+education+marstat+prsnlinc+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna	2*(-39540.03-(-39540.03))	0	1	0.9	79128.06
log(RNA)~month+agegroup+age+gender+region+education+marstat+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna	2*(-39540.1 -(-39540.03))	-0.14	1	0.7	79126.21
log(RNA)~month+agegroup+age+gender+region+education+marstat+IPRart+tot_con+BLifvcd4cl+logBLifvrna	2*(-39540.26 -(-39540.1))	-0.18	1	0.5	79124.52
log(RNA)~month+agegroup+age+gender+region+education+marstat+IPRart+BLifvcd4cl+logBLifvrna	2*(-39540.48-(-39540.26))	-0.26	1	0.5	79122.96
log(RNA)~month+agegroup+age+gender+region+marstat+IPRart+BLifvcd4cl+logBLifvrna	2*(-39541.92-(-39540.48))	-2.62	1	0.3	79119.84
log(RNA)~month+age+gender+region+marstat+IPRart+BLifvcd4cl+logBLifvrna	2*(-39544.39-(-39541.92))	-2.32	1	0.07	79114.78
log(RNA)~month+age+region+marstat+IPRart+BLifvdd4l+logBLifvrna	2*(-39546.16-(-39544.39))	-1.22	1	0.05	79116.32

Table 5.2: Deviance analysis of log(RNA) in immediate arm

log(RNA) in Immediate Arm

Full Model: （logLik: -28385.14） log(RNA)~month+agegroup+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+BLifvcd4cl+logBLifvrna+tot_sex+tot_con					
Model	Deviance	Delta	DF	P value	AIC
log(RNA)~month+agegroup+age+gender+region+education+prsnlinc+hshldinc+IPRart+BLifvcd4cl+logBLifvrna+tot_sex+tot_con	2*(-28385.32-(-28385.14))	-0.36	1	0.6	56816.63
log(RNA)~month+agegroup+gender+region+education+prsnlinc+hshldinc+IPRart+BLifvcd4cl+logBLifvrna+tot_sex+tot_con	2*(-28385.52 -(-28385.32))	-0.04	1	0.5	56815.04
log(RNA)~month+agegroup+gender+region+education+prsnlinc+hshldinc+BLifvcd4cl+logBLifvrna+tot_sex+tot_con	2*(-28385.82 -(-28385.52))	-0.56	1	0.4	56813.64
log(RNA)~month+agegroup+region+education+prsnlinc+hshldinc+BLifvcd4cl+logBLifvrna+tot_sex+tot_con	2*(-28386.08 -(-28385.82))	0.04	1	0.4	56812.16

Table 5.3: Deviance analysis for CD4 in delayed arm

CD4 in Delayed Arm					
Full Model: (LogLik: -135089) CD4~month+agegroup+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna					
Model	Deviance	Delta	DF	P value	AIC
CD4~month+agegroup+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+tot_con+BLifvcd4cl+logBLifvrna	2*(-135089 -(-135089))	0	1	0.8	270228
CD4~month+agegroup+age+gender+region+education+marstat+prsnlinc+hshldinc+tot_con+BLifvcd4cl+logBLifvrna	2*(-135089.2 -(-135089))	-0.4	1	0.5	270224.5
CD4~month+agegroup+age+gender+region+marstat+prsnlinc+hshldinc+tot_con+BLifvcd4cl+logBLifvrna	2*(-135089.9 -(-135089.2))	-1	1	0.4	270219.9
CD4~month+agegroup+gender+region+prsnlinc+hshldinc+tot_con+BLifvcd4cl+logBLifvrna	2*(-135091 -(-135089.9))	-1.2	2	0.3	270215.9
CD4~month+agegroup+gender+region+prsnlinc+BLifvcd4cl+logBLifvrna	2*(-135092.3 -(-135091))	-1.4	2	0.2	270214.5
CD4~month+agegroup+gender+region+BLifvcd4cl+logBLifvrna	2*(-135093.9 -(-135092.3))	-1.8	1	0.06	270215.8

Table 5.4: Deviance analysis of CD4 count in immediate arm

CD4 in immediate Arm					
Full Model: (Loglik: -133272.5) CD4~month+agegroup+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna					
Model	Deviance	Delta	DF	P value	AIC
CD4~month+agegroup+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna	2*(-133272.5 -(-133272.5))	0	1	0.9	266593.1
CD4~month+agegroup+gender+region+education+marstat+prsnlinc+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna	2*(-133272.6-(-133272.5))	-0.2	1	0.7	266591.1
CD4~month+agegroup+gender+region+education+marstat+prsnlinc+tot_sex+tot_con+BLifvcd4cl+logBLifvrna	2*(-133272.9-(-133272.6))	-0.4	1	0.4	266589.8
CD4~month+agegroup+gender+region+education+tot_sex+BLifvcd4cl+logBLifvrna	2*(-133276.7-(-133272.9))	-7.2	3	0.1	266589.4
CD4~month+agegroup+gender+region+BLifvcd4cl+logBLifvrna	2*(-133280.9-(-133276.7))	-1.2	2	0.1	266589.8

Table 5.5: Deviance analysis of log(RNA) for all patients

General model for log(RNA)					
Full Model: (LogLik: -72102.18) log(RNA)~month+agegroup+randarm+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna					
Model	Deviance	Delta	DF	P value	AIC
log(RNA)~month+agegroup+randarm+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+tot_con+BLifvcd4cl+logBLifvrna	2*(-72102.64 -(-72102.18))	-0.92	1	0.3	144255.3
log(RNA)~month+agegroup+randarm+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+BLifvcd4cl+logBLifvrna	2*(-72103.12 -(-72102.64))	-0.04	1	0.3	144254.2
log(RNA)~month+agegroup+randarm+age+gender+region+education+marstat+hshldinc+IPRart+BLifvcd4cl+logBLifvrna	2*(-72103.79 -(-72103.12))	-1.3	1	0.2	144253.6
log(RNA)~month+agegroup+randarm+age+gender+region+education+marstat+IPRart+BLifvcd4cl+logBLifvrna	2*(-72104.37-(-72103.79))	0.14	1	0.2	144252.7

Table 5.6: Deviance analysis of CD4 count for all patients

General model for CD4 count					
Full Model: (LogLik: -268658.2) CD4~month+agegroup+randarm+age+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+totsex+totcon+BLifvcd4cl+logBLifvrna					
Model	Deviance	Delta	DF	P value	AIC
CD4~month+agegroup+randarm+gender+region+education+marstat+prsnlinc+hshldinc+IPRart+tot_sex+tot_con+BLifvcd4cl+logBLifvrna	2*(-268658.3-(-268658.2))	-0.2	1	0.6	537368.5
CD4~month+agegroup+randarm+gender+region+education+marstat+prsnlinc+tot_sex+tot_con+BLifvcd4cl+logBLifvrna	2*(-268658.8-(-268658.3))	-0.8	1	0.4	537363.6
CD4~month+agegroup+randarm+gender+region+education+marstat+prsnlinc+tot_sex+BLifvcd4cl+logBLifvrna	2*(-268659.5-(-268658.8))	-0.6	2	0.2	537363
CD4~month+agegroup+randarm+gender+region+education+marstat+prsnlinc+BLifvcd4cl+logBLifvrna	2*(-268659.8 -(-268659.5))	0	1	0.2	537361.5
CD4~month+agegroup+randarm+gender+region+education+prsnlinc+BLifvcd4cl+logBLifvrna	2*(-268661.4 -(-268659.8))	-3.2	1	0.1	537360.8

To sum up, age group, gender, region, baseline CD4 and baseline RNA are the five main risk factors that associated with the repeated measurements of CD4 over time in either delayed arm or immediate arm. Age, region, marital status, prior ART use for perinatal purposes, baseline CD4 and baseline viral load are the main risk factors associated with the repeated measurements of viral load in delayed arm while age group, region, education, personal income, household income, baseline CD4, baseline RNA, total condoms use in the past week and total sex acts in the past week are the main risk factors that can be used to predict the log(RNA) change over time. Regardless of randomization arm, age group, randomization arm, age, gender, region, education, marital status, prior ART use for perinatal purposes, baseline CD4 and baseline viral load are the main risk factors associated with repeated measurements of log(RNA) over time while age group, randomization arm, gender, region, education, personal income, baseline CD4 and baseline RNA are the eight main risk factors that associated with the repeated measurements for CD4 count over time. In general,

it turns out that baseline CD4, baseline RNA and region are the common risk factors that exist in every model.

By comparing the linear mixed effects models' results of final reduced model, univariate analyses and multivariate analyses performed above, the results with response variable log(RNA) are all consistent while the results with response variable CD4 count have some discrepancies.

Chapter 6 SUMMARY AND DISCUSSION

In conclusion, HIV infection can cause AIDS and it destroys people's immune system allowing life-threatening opportunistic infections and cancers to thrive. The mean survival time after HIV infection is estimated to be 9 to 11 years without treatment. As stated in this paper, early initiation of antiretroviral therapy has obvious benefits in reducing rates of sexual transmission of HIV-1 and clinical events and improving the survival of infected individuals.

According to the univariate analysis and multivariate analysis performed above in chapter 3 and chapter 4, we conclude that gender is a statistically significant risk factor for CD4 count among immediate arm, delayed arm, and for all index patients. Moreover, age and age group are also a statistically significant risk factor for log(RNA) among delayed arm, immediate arm, and all index patients. Thus, it is very useful to consider gender difference when predicting the change in CD4 count over time and age difference when predicting the change of log(RNA) over time. CD4 counts of males decrease quicker than females and viral load increase quicker for males than females regardless of treatment arm. Older patients decrease quicker than younger patients in CD4 counts and increase quicker in the viral load regardless of treatment arm.

According to the risk factors analysis, we detected that baseline CD4 count, baseline viral load and region information play an important role in predicting the trend of CD4 count

and viral load over time. Accordingly, we can take more specific steps in advance based on patient's baseline CD4 count and viral load in the future. There are also other variables that are important for analyzing and predicting in either immediate arm or delayed arm, depending on CD4 count and viral load.

Furthermore, there are still some limitations exist in the analyses. For people in the delayed arm, they were not given the treatment at the same date while they were given the treatment when their CD4 count drops below a specific threshold or after May 12, 2011. Unfortunately, some of patients didn't even start treatment after May 12, 2011. However, people in the immediate arm were started treatment right after the enrollment date. In our discussion thus far, the analysis results we got for delayed arm only reflect an averaged effect from the results of the analyses in this setting since we didn't take individual ART initiation into account. Therefore, a better strategy is to take antiretroviral therapy initiation date into consideration for further analyses with an attempt to improve the predictions in delayed arm for both CD4 counts and viral load measurements.

BIBLIOGRAPHY

[1] Cohen, M., Chen, Y., McCauley, M., Gamble, T., Hosseinipour, M., & Kumarasamy, N. et al. (2011). Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *New England Journal Of Medicine*, *365*(6), 493-505. doi: 10.1056/nejmoa1105243

[2] Cohen, M., Chen, Y., McCauley, M., Gamble, T., Hosseinipour, M., & Kumarasamy, N. et al. (2016). Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *New England Journal Of Medicine*, *375*(9), 830-839. doi: 10.1056/nejmoa1600693

[3] Diggle, P., Liang, K., & Zeger, S. (2000). *Analysis of longitudinal data*. Oxford: Clarendon Press.

[4] Home. (n.d.). Retrieved from http://www.aidsmap.com/

[5] Lin, D., & Ying, Z. (2001). Semiparametric and Nonparametric Regression Analysis of Longitudinal Data. *Journal Of The American Statistical Association*, *96*(453), 103-126. doi: 10.1198/016214501750333018

[6] The HIV Life Cycle Understanding HIV/AIDS. (2018, July 27). Retrieved from https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/19/73/the-hiv-life-cycle