Nonlinear Mixed Effect Models for Modeling Initial Viral Decay Rates in an HIV Study

Su-Jin Kang

University of St Andrews

MSc Applied statistics and datamining

2008

Supervisor

Dr Carl Donovan

ABSTRACT

The Nonlinear Mixed Effect Viral Dynamic Model can easily handle unbalanced repeated and continuous measures data for individuals and is also popular in many other research areas such as biology and pharmacokinetics. Wu *et al.* (2004) [13] described a Nonlinear Mixed Effects Biphasic Model to estimate short-term population and individual viral decay rates in their study. Perelson *et al.* (1999) [43] and Ding *et al.* (1999) [15] reported that initial viral decay estimated for viral decay models would be good markers of the potency of antiretroviral regimens. The aim of this study was to model viral decay rates, and check the validity of the model for the set of data provided and investigate whether the relationships found with baseline covariates and long-term response are consistent with Wu *et al.*'s (2004) findings [13].

The Nonlinear Mixed Effect Single and Biphasic Viral Dynamic Models were fitted, and their respective initial viral decay rates were derived. In this study, analyses and reports are focused on the first-phase viral decay rates of the models. The study found that the actual treatment groups were more potent than the control group. It was found that actual treatment effect and the number of multi-PI mutations at baseline had impacts on the initial viral decay rates for both models. Besides, baseline HIV-1 RNA levels had an impact on the initial viral decay rates for the biphasic model. There were no significant differences in the initial viral decay rates for different ages, ethnicities, and gender groups.

The study also shows that the initial viral decay rates were somewhat negatively correlated with the baseline HIV-1 RNA levels. A strong correlation between the initial viral decay rates and week 1 virus load reduction from baseline was observed. It was also observed that individuals with the higher initial viral decay rates were more likely to have

suppressed virus load at week 24. Also, individuals with higher week 1 virus load reduction, i.e. early viral dynamics, were more likely to have suppressed virus load at week 24. These findings suggest that the antiviral potency or the initial viral decay rates are predictive of long-term viral load response [13].

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Special thanks go to Dr Carl Donovan for guidance throughout this project.

Thanks also to the GLAXOSMITHKLINE (GSK) Research and Development team who suggested the project and provided dataset. Thanks are also due to the department of statistics at the School of Mathematics and Statistics for facilitating the project with GSK.

Thank you to Hulin Wu, Adam A Ding and colleagues and other authors for their tremendous works on HIV clinical research which provided the guidance for the study.

Finally, I would like to thank my lovely family, my Father, Mother and Brother who supported my study in the UK.

SECTION 1: INTRODUCTION

1.1 Study Background

In HIV clinical trials, one important measure of the effectiveness of an antiviral treatment is the extent to which it suppresses the viral load in the patient's plasma. Once therapy has been initiated, achieving and maintaining an undetectable HIV viral load is an important treatment goal. Most current tri-therapy regimens achieve reasonably good viral suppression over the short-term. An outstanding challenge is to find regimens that sustain viral load suppression over the long-term, which is made difficult by the mutating aspects of the virus [30].

The viral load will drop within a few weeks and remain low long-term when patients are taking an efficacious therapy and their virus is sensitive to this therapy. The development of resistance from the virus to the drugs (through mutations) means that the viral load might eventually rebound [30].

The data is often summarized by calculating the change from baseline in viral load at each visit and also, by categorizing subjects as responders or non-responders at an acceptable long-term time-point (e.g. 48 weeks) according to a Time to Loss of Virological Response (TLOVR). From this, subjects will be classified as non-responders if they discontinue their treatment, or their viral load never falls below a certain threshold (either 400 or 50 copies/mL), or their viral load rebounds above the threshold [30].

Wu *et al.* (2004) [13] described a nonlinear mixed effects biphasic model to estimate short-term population and individual viral decay rates. They then investigated the

relationship between the individual early viral decay rates and subjects' characteristics, and also investigate whether the early viral decay rate could predict the subjects' long-term responses [30]. Perelson *et al.* (1999) [43] and Ding *et al.* (1999) reported that initial viral decay estimated for viral decay models would be good markers of the potency of antiretroviral regimens.

In Wu *et al.*'s (2004) study [13], viral dynamics in HIV-1–infected individuals aged 12–22 years were similar to those of HIV-1–infected adults over 22 years and infants. Also, the 3TC/ZDV/EFV regimen may be more potent than 3TC/ZDV/NFV or other regimens. In addition, early viral dynamics or week 1 virus load reduction measurements may be useful in evaluating the potency of antiretroviral regimens. The first-phase viral decay rates were positively correlated with baseline RNA levels and week 1 virus load reductions [13].

1.2 Initial Issues

Assumptions with regards to independence, constant variance and normality are crucial for modeling to obtain valid results. Most standard statistical techniques such as the unpaired t-test, linear regression and the chi-square test for association, assume that each of the primary observations are independent of all of the others [1-3].

If repeated observations are taken within subjects [1-7], the independence assumption can be unsuitable because observations within an individual tend to be correlated with one another. If we take two observations at random from the same individual, they are likely to be more similar or correlated in value than two random observations from two different individuals which cause the correlations in the errors [1, 4, 6-9]. This means that each

repeated observation in an individual may provide less additional information than a new observation in a new individual [1] and can be affect inferences.

In this case, assuming the correlation structure is crucial rather than assuming that the errors are independent of one another incorrectly. If the independence is not to be satisfied, the estimates from the result will not be reasonable. In specific, the test statistic and test of significance will be incorrect which can impact in a model selection process or any inferences.

There are alternative modeling approaches or estimation procedures which offer the possibility of analysing non-independent error structure such as Mixed Models, Generalized Estimating Equations (GEEs). Also, Generalized Additive Models (GAMs) or many other regression methods are when it is used in a Mixed Models or GEE estimation. Of the aforementioned, GEEs approach is only going to work for a linear problem, (i.e. linear in its parameters); however, an appropriate linearizing transformation is not always possible.

The Nonlinear mixed effect model is an alternative methodology that deals with this problem. Observations are assumed to be independent across all the individuals in the model, but it allows for the existence of within-subject covariance thanks to repeated data gathering on the same subject. It is suitable for repeat measuring data.

1.3 Objectives

The main objective of this research is to model viral decay rates and check the validity of the model for the set of data provided and investigate if the relationships found with baseline covariates and long-term response are consistent with Wu *et al.*'s (2004) findings [13]. This will be done as per what is outlined below:

1. Deriving initial viral decay rates for each subject.
2. Identifying baseline characteristics which are correlated with the viral decay rates.
3. Examining whether the initial viral decay rates predict long-term response.
4. Examining whether the relationship of the initial viral decay rates with baseline covariates and long-term response are consistent with Wu *et al.*'s (2004) findings.
5. Examining other methods that could be used for analysis.
6. Discussion and suggestions.

SECTION 2: LTERATURE REVIEW

2.1 Nonlinear Models

How do we determine whether we should fit our model to our data using linear or nonlinear regression? If the relationship between the response variable, y, and the explanatory variable, x, appears to be roughly linear then linear regression may be a fairly reasonable thing to do. However, even if the plotted relationship appears to be distinctly nonlinear, this does not necessarily mean that a linear regression model cannot be used [12].

Even if the plotted relationship does not appear to be linear, with a careful choice of the form of the model, we may wish to fit to the data if it is still possible to use ordinary linear least squares regression, also, it is much easier than using nonlinear least squares which use an iterative search method in general. If we can express the relationship between the response variable Y, the explanatory variables x_i and the parameters in the form $Y = X\beta + \varepsilon$ (linear in their parameters), we can fit the model to the data using a least ordinary squares fitting approach. Sometimes it may be necessary to perform simple transformations on the variables to allow the model to be expressed in a linear form [12].

Nevertheless, this is not always the most appropriate solution because once we transform the variables, important assumptions about the errors associated with the data which are normally distributed may no longer hold[1] and the inference on confidence and prediction intervals should be treated cautiously [12].

[1] A basic assumption such as normality is crucial for obtaining valid results.

Then, we can use a Generalized Linear Model (GLM) as an alternative method. Models that are based on a particular subset of nonlinear relationships can be fitted using the GLM framework. The subset of models that are allowed are those that have a linearizing transformation. Within this GLM framework, any class of non-normal error distributions[2] can be expressed as a member of the exponential family can be specified [12].

However, if the model does not possess an appropriate linearizing transformation, and also if we are fairly confident that the error distribution consists of normal errors with zero mean and constant variance, then nonlinear regression using nonlinear least squares is a viable alternative [12].

In summary, we are interested in specifying an appropriate model for the relationship between the response variable and the explanatory variable. If a linear regression model using linear least squares cannot explain the relationships very well, i.e. if we can't transform the form to be linear in its parameters, then, nonlinear regression using nonlinear least squares is one of the alternatives to fit the proper model which can explain/predict our data well.

2.2 Nonlinear Mixed Effect Models

Nonlinear mixed effects models, also referred to as hierarchical nonlinear models[3], are considered as a popular form for analysis for the repeated and continuous measures data on each of the individuals when interest focuses on individual-specific characteristics [18].

[2] GLM is an extension of standard linear models and we can relax the linearity, non-normality and even constant variance assumptions.

[3] This structure of the nonlinear mixed effects model makes it a natural candidate for Bayesian inference [18].

Mixed effects models have some advantages for example, they can easily handle unbalanced repeated measures data that occur in many areas such as pharmacokinetics, economics, biology, and many others, and the flexible variance-covariance structures of the response vector which allows for the nonconstant correlation among observations [17]. It is intuitively appealing because the notion that individuals' responses all follow a similar functional form with parameters that vary among individuals seems to be appropriate in many situations [16].

Nonlinear mixed effects models are mixed effects models in which the intrasubject model relating to the response variable to covariate (time typically) is nonlinear in the parameters [17] and may involve both fixed effects and random effects. Model building for nonlinear mixed effects models is considered as the process of determining the characteristics of both the fixed and the random effects so as to give an adequate but parsimonious model [19].

Several different nonlinear mixed effects models have been proposed by many scholars such as Sheiner and Beal (1980) [38], Mallet, Mentre, Steimer and Lokiek (1988) [39], author and author[4] [17]. In this dissertation, a slightly modified form of the model proposed in Lindstrom and Bates (1990) [35] was considered. This model can be presented as a two-stage hierarchy model which in some ways generalizes both the linear mixed effects model of Laird and Ware (1982) [36] and the usual nonlinear model for independent data of Bates and Watts (1988) [37] [16]. In the first stage the i th observation on the j th individual is modelled as in the following formula:

[4] Lindstrom and Bates (1990) [35], Vonesh and Carter (1992) [46], Davidian and Gallant (1992) [34], Wakefield, Smith, Racine-Poon and Gelfand (1994) [47].

$$y_{ij} = f(x_{ij}, \phi_{ij}) + \varepsilon_{ij}, \qquad i = 1, \cdots, M, \qquad j = 1, \cdots, n_i \tag{1}$$

where f is a nonlinear function governing within-individual behaviour of an individual-specific parameter vector ϕ_{ij} and y_{ij} is j th response on the i th individual. x_{ij} is the predictor vector, ε_{ij} is a normally distributed noise term, M is the total number of individuals, and n_i is the number of observations on the i th individual [19].

In the second stage, the individual-specific parameter vector ϕ_{ij} is modelled as:

$$\phi_{ij} = A_{ij}\beta + B_{ij}b_i, \qquad b_i \sim N(0, \sigma^2 D) \tag{2}$$

where A_{ij} and B_{ij} are design matrices for the fixed and random effects respectively and β is a p - dimensional vector of fixed population parameters, b_i is a q - dimensional random effects vector associated with the i th individual (not varying with j). $\sigma^2 D$ is a (general) variance-covariance matrix. Also, it is further assumed that observations made on different subjects are independent and that the ε_{ij} are iid $N(0, \sigma^2 D)$ and independent of the b_i [16, 18].

Different methods can be used to estimate the parameters in model (1). In this study, maximum likelihood and restricted maximum likelihood estimation was considered. Maximum likelihood estimation in (1) is based on the marginal density of y:

$$p(y \mid \beta, D, \sigma^2) = \int p(y \mid b, \beta, D, \sigma^2)\, p(b) db \qquad (3)$$

In general, this integral does not have a closed-form expression when the model function f is nonlinear in b_i, thus different approximations have been proposed for estimating it. Some of these methods consist of taking a first-order Taylor expansion of the model function f around the expected value of the random effects in Sheiner *et al.* [32] and Vonesh *et al.* [33]. Others have proposed the use of Gaussian quadrature rules (in Davidian and Gallant (1992) [34]) [17].

Nonlinear mixed effects models can be operated using the NLMIXED procedure in SAS. In this dissertation, the NLMIXED procedure will be applied to fit the specified nonlinear mixed model by maximizing an approximation to the likelihood integrated over random effects[5] .

The NLMIXED procedure assumes that we have an observed data vector y_i for each of i subjects, $i = 1, \cdots, M$. The y_i are assumed to be independent across *i*, but it allows the existence of within-subject covariance since y_i , each of the elements, are measured on the same subject. As a statistical mechanism for modeling this within-subject covariance, assume that there exist latent random-effect vectors of small dimensions which are also independent across i [11].

[5] PROC NLMIXED only implements maximum likelihood, whereas PROC MIXED can perform both maximum likelihood and restricted maximum likelihood (REML) estimation [11].

PROC NLMIXED fits nonlinear mixed models by maximizing an approximation to the likelihood integrated over the random effects. Different integral approximations are available such as the principal ones being adaptive Gaussian quadrature and a first-order Taylor series approximation. However, a variety of alternative optimization techniques are also available to carry out the maximization and the default is a dual quasi-Newton algorithm [11].

We are able to use the estimated model to construct predictions of arbitrary functions by using the parameter estimates and the empirical Bayes estimates[6] of the random effects in PROC NLMIXED [24].

2.3 Nonlinear Mixed Effect Multi-phase Viral Dynamic Models

Wu and Ding (1999) [14] introduced an application of hierarchical nonlinear mixed effect models to HIV dynamics. They illustrated various phases of HIV-1 dynamics (Figure 1).

[6] Estimates of the individual parameters by modes (or mean) of their posterior distributions given the data, i.e. estimates of the unobservable covariance parameters instead of specifying a full prior distribution [27].

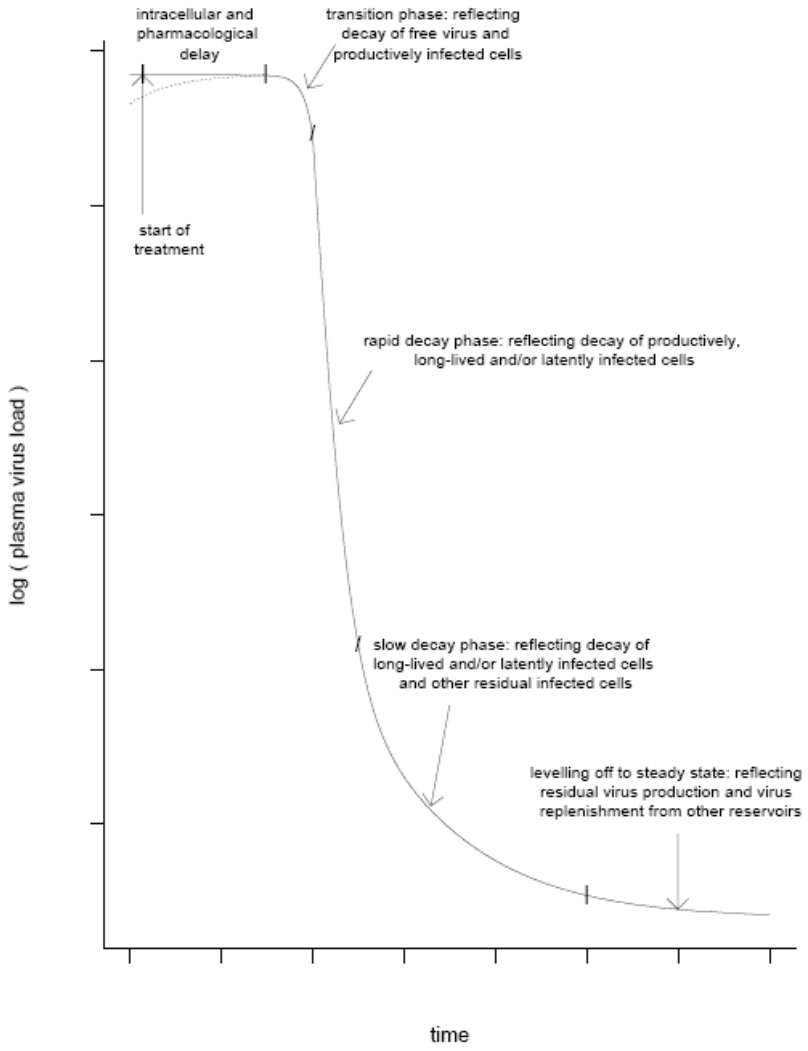

Figure 1: Illustration of the different phases of plasma viral dynamics following antiviral drug treatment in Wu *et al.* (1999) [14]. In the phase of intracellular and pharmacological delay, the dotted line denotes non-steady-state case before treatment [14].

According to the illustration, if the data on the transition phase and rapid decay phase are available, the model $V(t) = P_0 + P_1 e^{-\delta_p t} + (P_2 + P_3 t)e^{-ct}$ is suggested. If the data on the rapid decay phase and slow decay phase are available, the model $V(t) = P_0 + P_1 e^{-\delta_p t} + P_2 e^{-\lambda t}$ is suggested. Also, if the data on the slow decay phase and leveling off phase are available, the suggested model is $V(t) = P_0 + P_2 e^{-\lambda t}$, where Parameter P_i represents the initial viral production rate, and parameter λ is a possibly confounded clearance rate of long-lived and latently infected cells [21, 14]. Also, $\delta_p = \lambda_1$, $c = \lambda_5$ and t is time [14].

They also have developed a nonlinear mixed effects biphasic viral decay model to estimate population and individual viral decay rates [14, 26] where the antiretroviral drugs are not assumed to be perfect [27]. Before that, Perelson *et al.* [25] developed a two-phase plasma viral decay model which assumes that there are two major HIV-infected cell compartments which are productively infected cells and long-lived infected cells.

According to their model, the viral dynamic model after initiation of antiviral therapy, i.e. including treatment effects, can be written as [15]:

$$\begin{aligned}
\frac{d}{dt}T_1 &= (1-\gamma_1)\alpha_1 kTV_1 - \delta_1 T_1 \\
\frac{d}{dt}T_2 &= (1-\gamma_2)\alpha_2 kTV_1 - \delta_2 T_2 \\
\frac{d}{dt}V_I &= (1-\eta_0)[(1-\eta_2)N_2\delta_2 T_2 + (1-\eta_1)N_1\delta_1 T_1] - cV_I \\
\frac{d}{dt}V_{NI} &= (\eta_0 + (1-\eta_0)\eta_2)N_2\delta_2 T_2 + (\eta_0 + (1-\eta_0)\eta_1)N_1\delta_1 T_1 - cV_{NI}.
\end{aligned} \tag{4}$$

where V_I and V_{NI} denote the concentration of infectious virions and non-infectious virions respectively, and T_1 and T_2 denote the concentration of two infected cell compartments, productively infected cells and long-lived/latently infected cells, respectively [27]. Parameters γ_1 and γ_2 represent the protease inhibitor drug efficacy in the two infected cell compartments and η_1 and η_2 represent the protease inhibitor drug efficacy in the two corresponding compartments. Thus, the overall combination treatment potency in the two infected cell compartments can be defined by $e_1 = 1-(1-\eta_1)(1-\gamma_1)$ and $e_2 = 1-(1-\eta_2)(1-\gamma_2)$, respectively [27].

Wu and Ding (1999) [15] have also shown that the total virus observation, $V(t) = V_I(t) + V_{NI}(t)$, in this model can be approximated by:

$$V(t) = P_1 e^{-d_1 t} + P_2 e^{-d_2 t}, \quad t \geq t_c, \tag{5}$$

where t_c is the time that the "shoulder" disappears (usually 2 or 3 days) [15]. Parameters, P_1 and P_2, are reparametrized parameters from the solution of (4). Figure 2 shows the observed data and model-fitting result using model (5) for four selected patients from the above application [27].

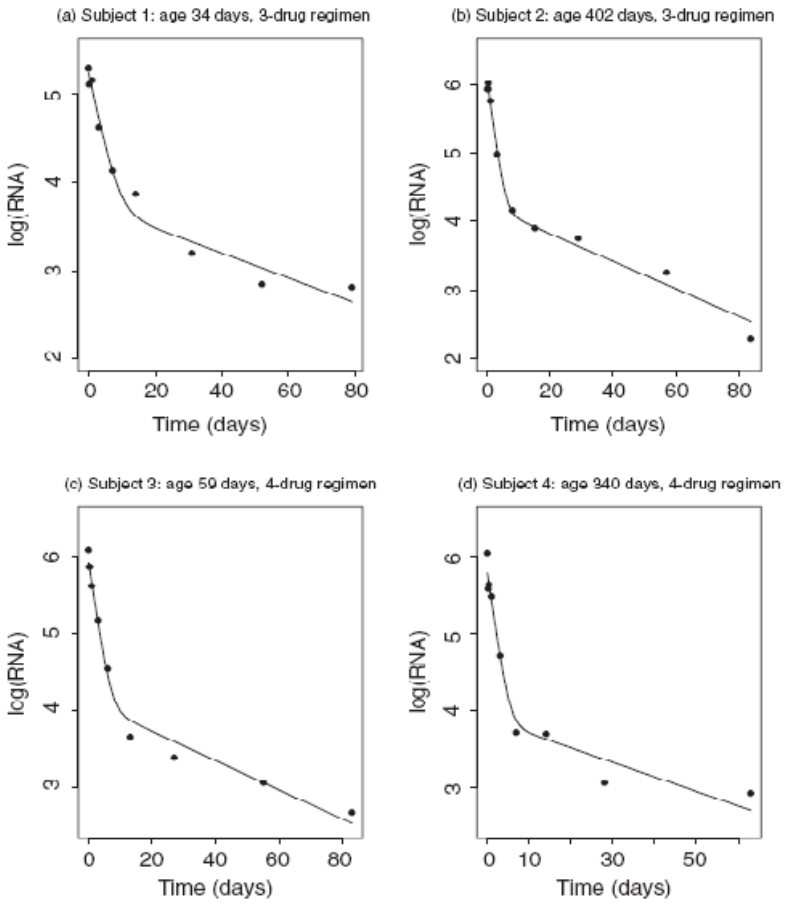

Figure 2: The Population nonlinear mixed effect model of model (5) fitted individual curves from four patients (one from each cohort) in Wu and Ding (1999). The dots are observed viral loads [27].

In addition, they showed that parameter d_1 and d_2 are the decay rates of the two phases of plasma virus and can be approximated by:

$$d_1 = |1 - R_1(1 - e_1)|\delta_1$$
$$d_2 = \left(1 - R_2 \frac{1 - e_2}{e_1}\right)\delta_2 \qquad (6)$$

where $R_1 = (1-\eta_0)N_1\alpha_1 kT/c$ and $R_2 = (1-\eta_0)N_2\alpha_2 kT/c$ are the baseline reproduction/ clearance ratios of the virus from the two infected cell compartments. Three factors, loss rates of infected cell (δ_1 and δ_2), baseline reproduction/clearance ratios (R_1 and R_2) and treatment effects (e_1 and e_2), determine the decay rates. Hence we can use viral decay rates to compare the potencies of antiviral therapies if other factors (R_1, R_2, δ_1 and δ_2) are homogeneous between treatment arms (ideally using a randomized design) [27].

Based on model (5), to estimate population and individual viral decay rates, Wu *et al.* (2004) [13] used a nonlinear mixed effect biphasic viral dynamic model as follows [13,14, 27]:

$$\log_{10}[V_i(t)] = \log_{10}[\exp(P_{1i})\exp(-d_{1i}t) + \exp(P_{2i})\exp(-d_{2i}t)] + \varepsilon_i(t), \qquad (7)$$

where $\varepsilon_i(t)$ is a HIV-1 RNA measurement error (on the $\log_{10}$ scale) with mean zero, and $V_i(t)$ is HIV-1 RNA copies/mL plasma at treatment time t for the i th subject. The viral decay rates for the i th subject are $d_{1i} = d_1 + b_{1i}$ and $d_{2i} = d_2 + b_{2i}$, where the fixed-effect parameters, i.e. d_1 and d_2, are the population decay rates for the two viral decay phases. b_{1i} and b_{2i} are random effect parameters assumed to be *iid* $N(0, \sigma_{b1}^2)$ and *iid* $N(0, \sigma_{b2}^2)$

respectively, that quantify the between-subject variation of viral decay rates. Parameters $P_{1i} = P_1 + b_{3i}$ and $P_{2i} = P_2 + b_{4i}$ are "macroparameters", with $\exp(P_{1i}) + \exp(P_{2i})$ being baseline virus load at time $t = 0$ (the day of starting treatment) [13, 14, 20, 26, 27].

The fitted virus load trajectories from 6 selected subjects using the model (7) are shown in figure 3 in Wu *et al.* (2004) [13].

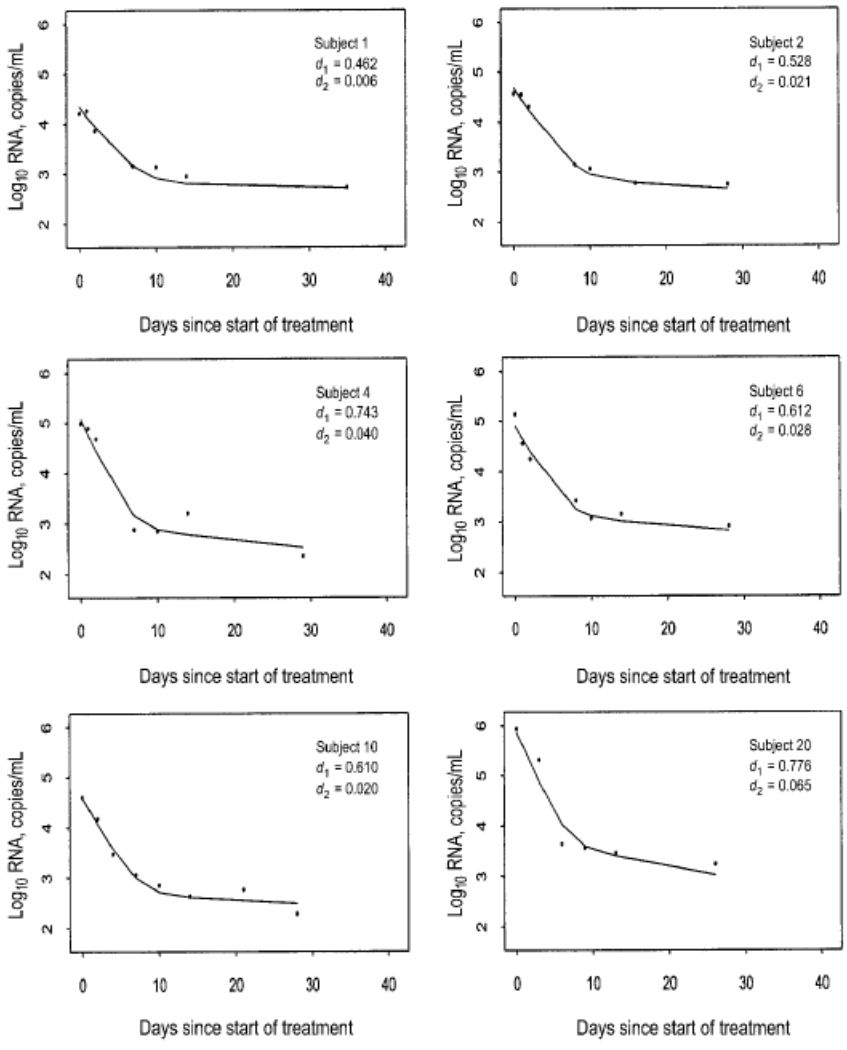

Figure 3: HIV-1 RNA data (dots) from 6 selected individuals and corresponding fitted trajectories using the NLME modeling approach in Wu *et al.* (2004) [13].

Wu *et al.* (2004) [13] only included HIV-1 RNA data from day 0 to week 4, i.e. with a 2-week window, during treatment to fit the biphasic viral dynamic model. This decision was made for several reasons to eliminate the possibility of falling below the lower limit of detection, i.e. 400 copies/mL, or viral rebound may occur after 4 weeks in the data and also, the viral dynamic model is valid only for the early stage of treatment [13, 14, 28].

They also excluded the rebounded data, an increase from the previous virus load measurement within 4 weeks for a patient [13]. If the HIV-1 RNA level fell below the limit of detection within 4 weeks, they only included the first limit of detection value to prevent an artificial effect. Their analysis and report focused on the first-phase viral decay rates; the second-phase viral decay rates for individual subjects may not be reliable because 60% of the total number of subjects did not have the data on the second-phase (no viral decay data after weeks 2) in the study [13].

In another study by Wu *et al.* (2003) [20], to fit the biphasic viral dynamic model, the HIV-1 RNA data from 0 to week 8 on treatment was only included because weeks 8 was the time when subjects might change their treatment. They also excluded the rebounded data if there was a viral rebound within 8 weeks and if the HIV-1 RNA level fell below the limit of detection within 8 weeks; they only included the first limit of detection value.

Anthony *et al.* [23] suggested a simple and flexible nonlinear mixed effects model for the trajectory of HIV-1 RNA until rebound. They were interested in the relationship between the lowest level of plasma HIV attained after initiation of therapy and the time until rebound [23]. In the study, they modelled the initial 2-week follow-up. The first-phase was modelled by a linear slope in the $\log_{10}$ HIV-1 RNA as follows:

$$f(\theta_i;t) = \beta_{0i} - \beta_{1i}(t - t_0), \qquad t = 0 \text{ (2 days}^7\text{)} < t < t_0 \qquad (9)$$

where the parameter β_{1i} is the rate of decline during this phase, and $\beta_{0i} = f(\theta_i, t_0)$, the expected $\log_{10}$ viral load for subject i at time t_0 [23].

After the initial phase, viral load is assumed for $t > t_0$ to be the sum of two distinct components which is the one that declines in response to therapy, and the other of which may either decline or increase:

$$f(\theta_i;t) = \log_{10}(A_i \exp\{-\beta_{2i}(t - t_0)\} + B_i \exp\{-\beta_{3i}(t - t_0)\}, \qquad t_0 < t \qquad (10)$$

where A_i and B_i are the levels of the two components of RNA at the start of the second phase (at $t = t_0$); $\beta_{2i} > 0$ and β_{3i} are the rate of decay of first component and the rate of change either growth or decay of the second component respectively. The first exponential term in (10) implies that a component of the HIV-1 RNA continues to fall log-linearly. For some subjects (with $\beta_{3i} > 0$), the second component may increase log-linearly, perhaps reflecting resistance; in others ($\beta_{3i} < 0$), it will continue to decline. This model guarantees a smooth transition from RNA decline to RNA increase. The two phases (9) and (10) are combined by imposing the constraint that the two mean equations agree for $t = t_0$:

$$\log_{10}(A_i + B_i) = \beta_{0i} \qquad (11)$$

[7] They do not include the day 0 values in the analysis and start instead with day 2 since no measurements are taken during the first two days.

For subjects with viral rebound; $\beta_{3i} > 0$ a turning point T_i can be defined, as the time from baseline to reaching the minimum expected viral load. This is given by:

$$T_i = \max\{t_0; t_0 + \{\ln(A_i / B_i) + \ln(\beta_{2i} / \beta_{3i})\}/(\beta_{2i} + \beta_{3i})\} \quad (12)$$

For subjects whose HIV-1 RNA continues to decline throughout the second phase; $\beta_{3i} < 0$, the turning point is not defined [23].

They fit the equations (10) and (11) jointly using a nonlinear mixed effect model, where the subject-specific parameters are treated as random effects. For the subject *i*, the response is $y_{ij} = y_i(t_{ij})$, $j = 1, \cdots n_i$. The values $y_{ij} < 2 = \log_{10}(100)$ are censored. And they assume independent normal errors: $y_{ij} = f(\theta_i, t_{ij}) + \varepsilon_{ij}$, ε_{ij} are iid $N(0, \sigma^2 D)$ [23]. Also, to impose the condition $\beta_{2i} > 0$, they reparameterize $\beta_{2i} = e^{\phi_i}$. The condition (11) and the restrictions A_i, $B_i > 0$ are modelled by taking $A_i = 10^{\beta_{0i}}/(1 + e^{\tau_i})$, $B_i = 10^{\beta_{0i}} e^{\tau_i}/(1 + e^{\tau_i})$. And the random effects vectors $\theta_i = (\beta_{0i}, \beta_{1i}, \phi_i, \beta_{3i}, \tau_i)^{\mathrm{T}}$ are assumed independent multivariate normal with unspecified covariance matrix are θ_i are iid $N(0, G)$, where G is an arbitrary positive-definite matrix [23].

Although this model is not intended to describe the long-term behaviour of HIV-1 RNA in response, it seems to appropriate for an analysis that only considers progression up to the first reading after reaching the nadir value [23].

SECTION 3: METHODS

3.1 Data Available for Analysis

3.1.1 Data Description

The HIV dataset was provided by GLAXOSMITHKLINE (GSK) Research and Development for the proposed project. Dataset 1 contains the main information for the study such as demographic information, information related to treatment and assessment, and viral load results. Dataset 2 contains subjects' characteristics at baseline and Intention–To–Treat[8] (ITT) exposed population information, and dataset 3 contains classification of type of failure or success information at week 24.

After selecting variables for analysis, these datasets were combined into one. The original data sets given are as follows:

[8] "Intention to treat" is a strategy for the analysis of randomised controlled trials that compares patients in the groups to which they were originally randomly assigned. This is generally interpreted as including all patients, regardless of whether they actually satisfied the entry criteria, the treatment actually received, and subsequent withdrawal or deviation from the protocol [49].

Table 1.1: List of variables and Attributes of dataset 1:

Categories	Variable Name	Description
Subject and Demographic information (PI resistant is classifying subject according to whether their virus is resistant to Protease Inhibitor (PI) which is the class that the study drug belongs to):	SUBJID	Subject ID
	AGE	Age in years
	SEX	Sex
	RACECD	Race code
	RACE	Race
	ATRTCD	Actual treatment code
	atrtgrp	Actual treatment group
Visit name and code, number of days since treatment start date and date for each assessment	AVISIT	Actual visit description
	AVISNUM	Actual visit sequence number
	LBACTDY	Actual study day of collection
	LBDT	Actual date of collection
Whether the assessment prior to treatment start date or post treatment stop date	ATTYPECD	Time in relation to treatment - code
	ATTYPE	Time in relation to treatment
Viral Load results (with units, logs and change from BL and assay information)	LBORUNIT	Original unit
	LBORRES	Original text result
	LBORRESN	Original numeric result
	LBORCHBL	Change from baseline - original unites
	LBORRLG	Original numeric result - $\log_{10}$
	LBORRLGC	Change from baseline - $\log_{10}$
	ASSAYV	Assay version
	ASSAYVCD	Assay version code
	LLOD	Lower level of detection
	ULOD	Upper level of detection
	KEEPLOD	Retained <or> LOD valid value flag
Viral load results in the situation where there are several values for the same visit window with the appropriate flag	LBEVFLG	Evaluable flag
	LBORRSNW	Windowed original numeric result
	LBORRLGW	Windowed $\log_{10}$ original numeric result
	LBORLGCW	Windowed change from baseline - $\log_{10}$

Table 1.2: List of variables and Attributes of dataset 2:

Attribute	Variable Name	Label
Subject characteristics at baseline and ITT exposed population	SUBJID	Subject ID
	ACT20GCD	Actual introduction of a dose of T20 (yes/no) code
	ACT20GRP	Actual introduction of a dose of T20 (yes/no)
	BLVLGCD	Baseline viral load subgroup code
	BLVLGRP	Baseline viral load subgroup
	CD4CG2CD	CD4+ cells/cu mm group 2 code
	CD4CGRP2	CD4+ cells/cu mm group 2
	MULPICD2	Number of multi-PI mutations code
	MULPIG2	Number of multi-PI mutations group
	REGION	Region of recruitment
	PNITTE	Intent-to-Treat Exposed population
	CD4__BLC	Baseline CD4+ cell count
	CD4__BLQ	Baseline CD4+ cell percentage
	CD8__BLC	Baseline CD8+ cell count
	CD8__BLQ	Baseline CD8+ cell percentage

Table 1.3: List of variables and Attributes of dataset 3:

Attribute	Variable Name	Label
Subject number and visit name and code	SUBJID	Subject ID
	AVISIT	Actual visit description
	AVISNUM	Actual visit sequence number
Response at week 24 (has the subject managed to get below 400 copies/mL) according to 2 algorithms: "Observed" and "Time to Loss of Virological Response"	P400_OBS	<400 copies/mL
	P400_TLO	<400 copies/mL, TLOVR
Detailed classification of type of failure at week 24	r400_tlc	Reason for failure code
	r400_tl	Reason for failure

3.1.2 Data Processing

Some values such as negative study days, i.e. AVISIT= "Screening", and the repeated first day measurement, i.e. AVISIT= "Day 1", and also re-test measurement on the same day, i.e. LBEVFLG[9] = "0" and KEEPLOD[10] = "0", were eliminated for presenting summary statistics which means it allows us to select one observation per subject per visit. In addition, all the missing cases of LBEVFLG and KEEPLOD are also excluded. Also, to model viral decay, using ATTYPE, all the measurements where the value does not equal "Treatment" were excluded, e.g. drop pre-treatment measurements.

Some patient's Plasma HIV-1 RNA was repeatedly quantified beyond 24 weeks, i.e. by 32 or 40 weeks, and the values by 24 weeks are only used for the analysis.

[9] This flag allows our group to select 1 observation per subject per visit for presenting summary statistics. In particular, if there are 2 assessments for the same visit "window", they will take the closest to the target date or if they are equidistant, the average value. When selecting observations based on this flag, i.e. LBEVFLG, for summary statistics, the "windowed type variables", i.e. LBORLGCW, LBORRLGW, and LBORRSNW, need to be used because these variables include the correct observation or the averaged value when applicable. However, for the model part of this project, all assessments need to be used i.e. if several tests were carried out for the same visit window, so this aspect can be omitted [31].

[10] This flag selects a re-test (on the same day) if the first test reached the limit of detection of the assay, this aspect needs to be used for this project (both for summaries and the model part). (Sometimes a test reached the limit of detection e.g. <50 and can not be re-tested so this value needs to be kept, hence the reason why selection can not be done on LBORRES only). These flags only apply to viral load [31].

Original viral load results, such as $\log_{10}$ (original numeric result), i.e. LBORRLG or original numeric result, i.e. LBORRESN, are considered as response variables. Also, variables for the baseline characteristics, such as variable number of mutations at baseline, i.e. MULPICD2; baseline CD4+ cell count, i.e. CD4_BLC; and introduction of T20 (enfuvirtide, background HIV medication), i.e. ACT20GCD, were considered as covariates. Because they are considered as an important part of the description of the study population, and ones that are likely to have an impact on the effect of the drug [31].

Also, to investigate whether the early viral decay rates had an impact on long-term response (week 24), the binary response variable indicating viral load measurements below the 400 copies/mL threshold at week 24 according to TLOVR, i.e. P400_TLO, was used.

3.2 Statistical Methodologies

Using the NLMIXED procedure in SAS, nonlinear mixed effect models (which have both fixed and random effects) were fitted.

Spearman's rank tests (using the CORR procedure in SAS) performed for the correlation of estimated the first-phase viral decay rates[11], i.e. d_1, with '$\log_{10}$ (baseline RNA)' and the correlation of d_1 with respective 'week 1, week 20 and week 24 $\log_{10}$ (RNA) change from baseline'. The correlation of d_1 with 'baseline CD4+ cell count' and the correlations of 'week 1 $\log_{10}$ (RNA) change from baseline', i.e. early viral dynamics or week 1 virus load reduction, with 'weeks 20 and 24 $\log_{10}$ (RNA) change from baseline', i.e. week 20 and 24 virus load reduction, were performed respectively.

[11] In this study, analysis and report focused on the first-phase decay rates for initial viral decay rates.

The ANOVA procedure in SAS was performed for age, ethnicity, gender, and the actual treatment group effects on the first-phase viral decay rates, respectively. Tukey's Studentized Range (HSD) test in GLM was also performed for the actual treatment group effects on estimated d_1 viral decay rates.

The Wilcoxon rank sum and Kruskal-Wallis tests (using the NPAR1WAY procedure in SAS) was performed for examining properties of actual treatment group on the first-phase viral decay rates.

In addition, Univariate Regression Models (GLM) were fitted (using the GLM procedure in SAS) to identify baseline characteristics which are correlated with the estimated first-phase viral decay rates.

The Wilcoxon rank sum and Kruskal-Wallis tests was applied to examine if the early viral decay rates can predict the long-term response (24 weeks) and Univariate logistic regression analyses (using the CATMOD procedure in SAS) was used to examine if the actual treatment group is a significant predictor for the long-term response.

3.3 Software used

All analyses were conducted in SAS v9.1 (SAS Institute Inc., 2003). The NLMIXED procedure and relevant command statements were conducted in SAS, and statistical software-R v2.6.2 and SPSS v14.0 (SPSS Institute Inc., 1989-2005) were also used to produce graphs and tables.

SECTION 4: RESULTS

4.1 Complete Dataset Results

4.1.1 Data Summary

A total of 116 subjects' Plasma HIV-1 RNAs were repeatedly quantified and their demographic information was provided for the study (dataset 1). This number came from an original dataset 2 of 288 subjects. There were 17 (14.7%) females and 99 (85.3%) males. Their median age was 43 years (range: 16-65 years). The median baseline CD4 cell count was 152 (cells/mm^3) and median $\log_{10}$ (pre-treatment Plasma HIV-1 RNA) was 4.55. 87 HIV infected patients (75%) were treated with three potent antiviral drugs; does 1 (30; 25.9%), dose 2 (28; 24.1%) and dose 3 (29; 25.0%) and 29 patients (25%) were in the control group.

Plasma HIV-1 RNA was repeatedly quantified on days 1, 2, 3, 8, 10, 15, and weeks 4, 8, 12, 16, 20, 24, 32 and 40 after initiation of treatment. However, for various reasons such as Plasma HIV-1 RNA rebound, never achieved VL suppression by weeks 24, 40 patients (34.5%) completed their treatments before 24 weeks as follows:

- 1 patient (0.9%) on 10 days
- 3 patients (2.6%) on 4 weeks
- 2 patients (1.7%) on 8 weeks
- 6 patients (5.2%) on 12 weeks
- 19 patients (16.4%) on 16 weeks
- 9 patients (7.8%) on 20 weeks

76 patients (65.5%) completed their treatments beyond 24 weeks:

- 41 patients (35.3%) on 24 weeks
- 29 patients (25.0%) on 32 weeks
- 6 patients (5.2%) on 40 weeks

Of the 116 subjects, 107 were used for the viral dynamic analysis, including the model fitting; because 9 subjects had no initial viral decline at all after starting treatment, they are ineligible for viral dynamic analysis.

4.1.2 *Subjects' Viral Decay Patterns*

Before fitting a model, subjects' viral decay patterns were examined through plotting the data. Primarily, some randomly selected subjects' viral decay patterns through time after they started their treatments were examined. The next fitted plots (figure 4), for "windowed" $\log_{10}$ (original numeric result) through day after starting treatment for each randomly selected 7 people show diverse viral decay or rebound patterns, respectively. Overall, after the early rapid decay term before approximately 10 days ($t = 7$, week 1), some subjects' plasma virus seems to have increased/rebounded (and maintained in that way or slowly declined again) and some of them declined slowly.

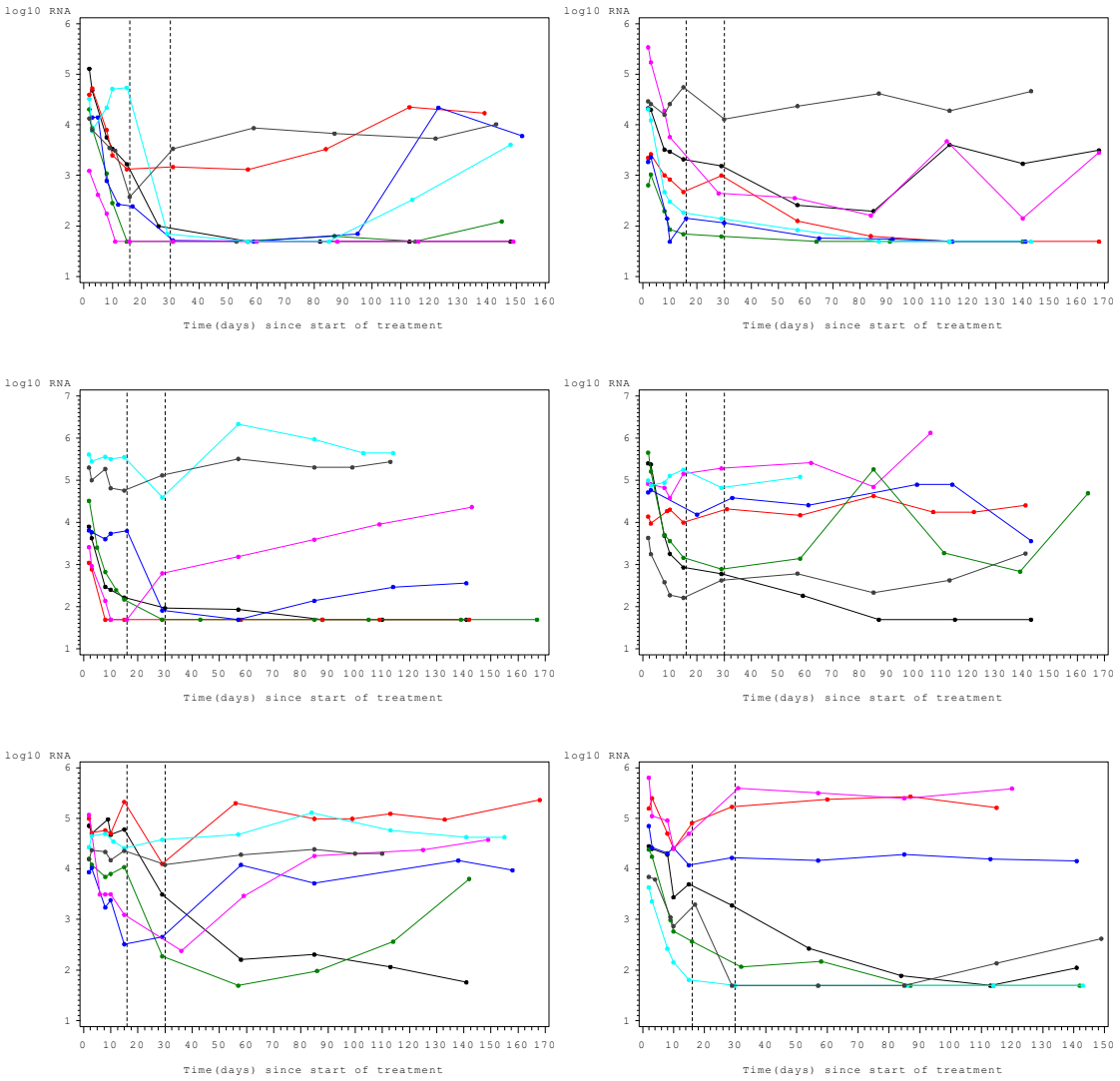

Figure 4: The fitted curves of each 7 randomly selected patients. The first reference vertical dotted line indicates 2 weeks (16 days) and the second reference vertical dotted line indicates 4 weeks (30 days) which is the end of period Wu *et al.* (2004) [13] derived viral decay rates.

The subjects were classified as responders or non-responders at a long-term time point (24 weeks) according to a Time to Loss of Virological Response (P400_TLO) algorithm. According to this rule, subjects who discontinued their treatments or viral load never fell below 400 copies/mL, i.e. never achieved VL suppression by week 24, or viral load rebounded above the threshold, i.e. Plasma HIV-1 RNA rebounded, were classified as non-responders. Some subjects of insufficient viral load response, or Protocol mandated switch from 150mg BCV/r were also classified as non-responders.

Specifically, 34 subjects (29.3%) never achieved VL suppression by week 24, 21 subjects' (18.1%) Plasma HIV-1 RNA rebounded, 15 subjects (12.9%) had insufficient viral load responses and 18 subjects (15.6%) failed to get below 400 copies/mL at week 24. 10 subjects' (8.6%) Protocol mandated switch from 150mg Brecanavir[12]/ritonavir[13] (BCV/r) and only 18 subjects (15.5%) were classified as responders at visit and censored thereafter [Appendix 1, Chart 2.1-2.3].

[12] Brecanavir, a novel protease inhibitor (PI), has sub-nM in vitro antiviral activity against multi-PI-resistant HIV-1 and in vitro is >100-fold more potent than previously marketed PIs and approx. 10-fold more potent than the recently marketed PI, darunavir [29].

[13] Ritonavir, also known as Norvir, is a type of medicine called a protease inhibitor (PI). PIs act by blocking protease, a protein that HIV needs to make more copies of itself [48].

The following plots (figure 5) present classification types of response at week 24.

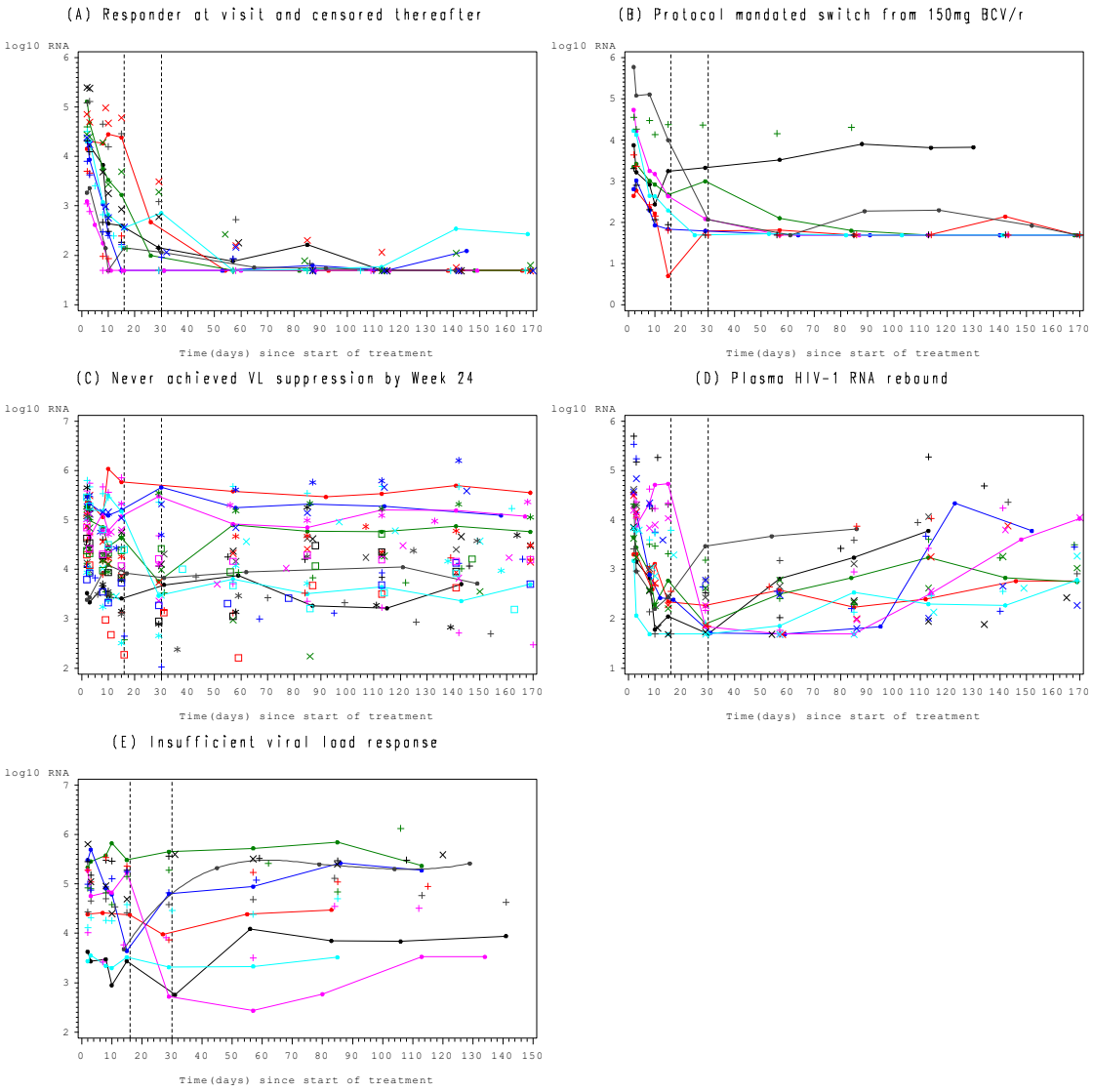

Figure 5: Plots of patients demonstrating (A) responder at visit and censored thereafter, (B) Protocol mandated switch from 150mg BCV/r, (C) never achieved VL suppression by week 24, (D) Plasma HIV-1 RNA rebounded and (E) insufficient viral load response, clockwise from top left. The first reference vertical dotted line indicates 2 weeks (16 days) and the second reference vertical dotted line indicates 4 weeks (30 days) which is the end of period Wu *et al.* (2004) [13] derived viral decay rates.

Detailed classification types of each individual plot are presented in figure 6.1- 6.5.

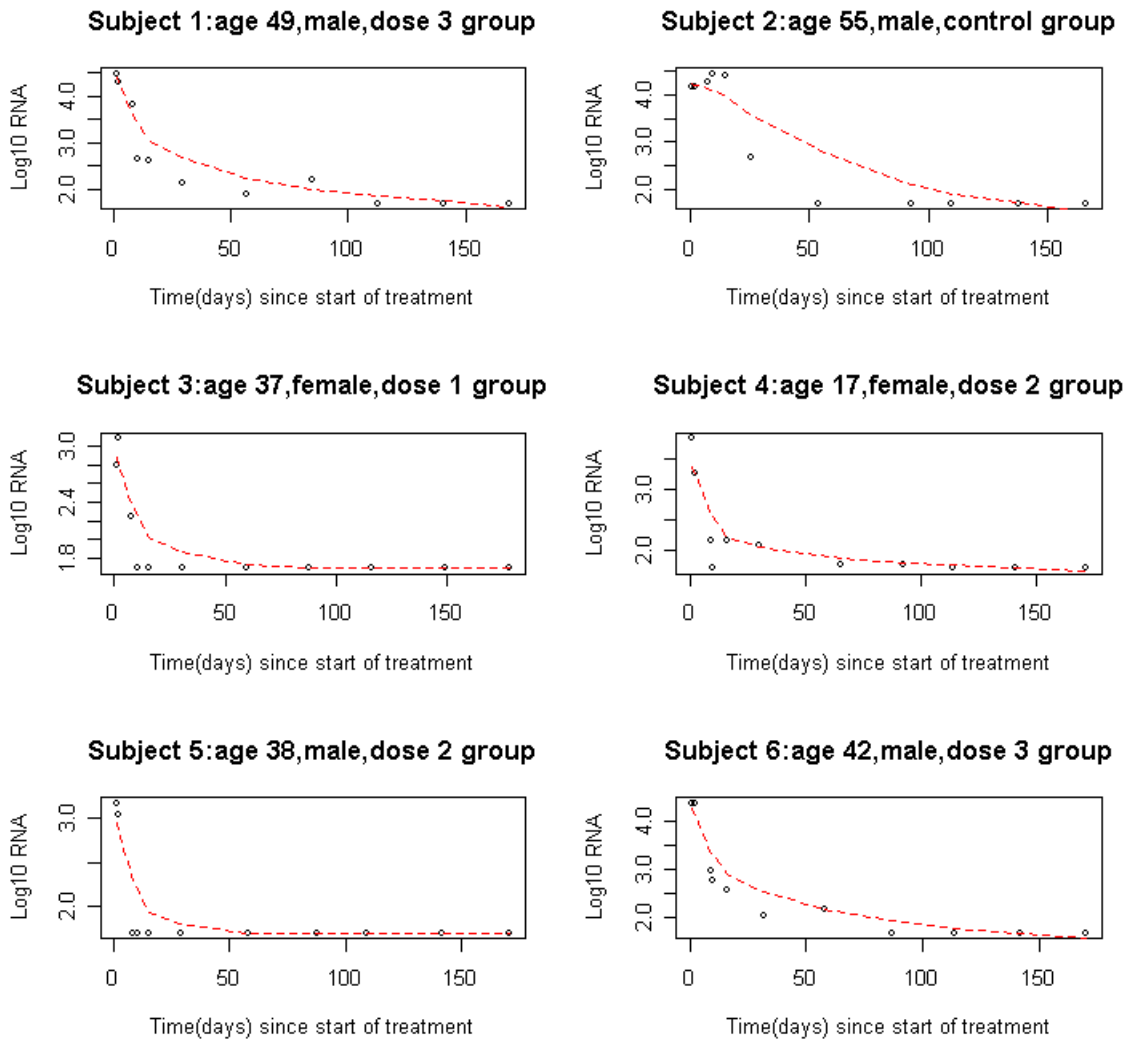

Figure 6.1: The fitted curves from 6 randomly selected patients who were responder at visit and censored thereafter. The dots are the observations, and the solid lines are Friedman's super smoother.

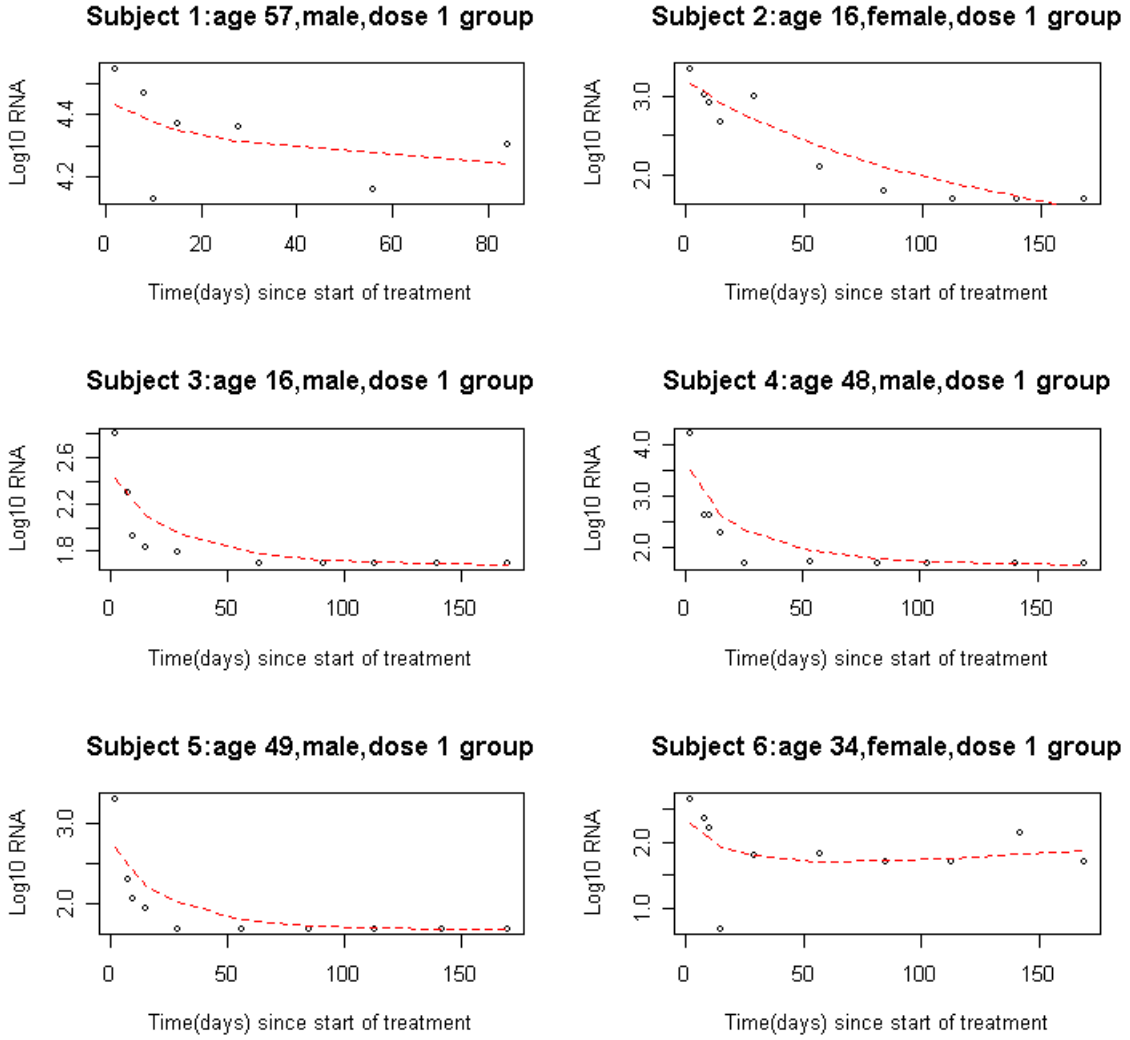

Figure 6.2: The fitted curves from 6 randomly selected patients who were Protocol mandated switch from 150mg BCV/r. The dots are the observations, and the solid lines are Friedman's super smoother.

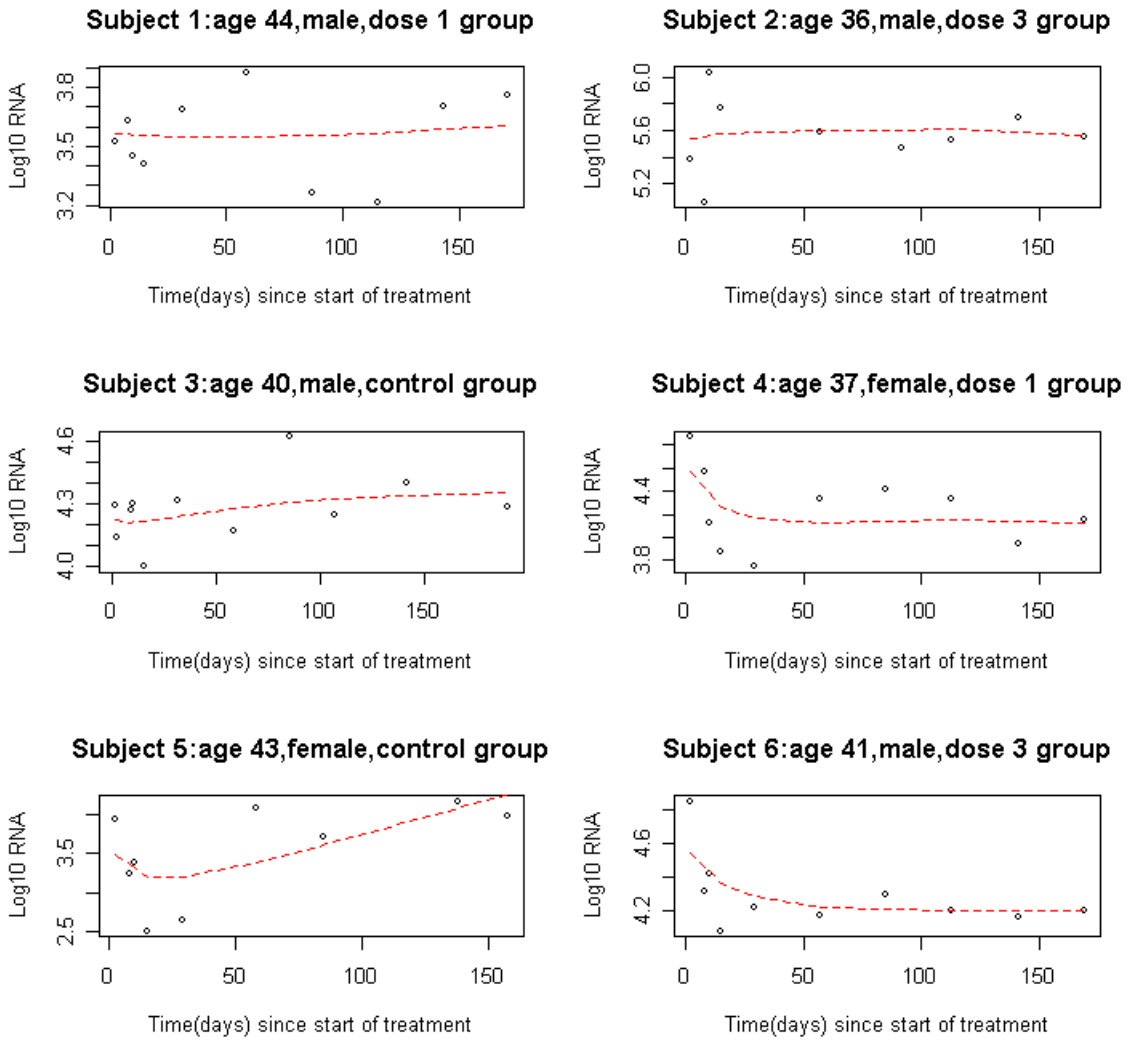

Figure 6.3: The fitted curves from 6 randomly selected patients who never achieved VL suppression by week 24. The dots are the observations, and the solid lines are Friedman's super smoother.

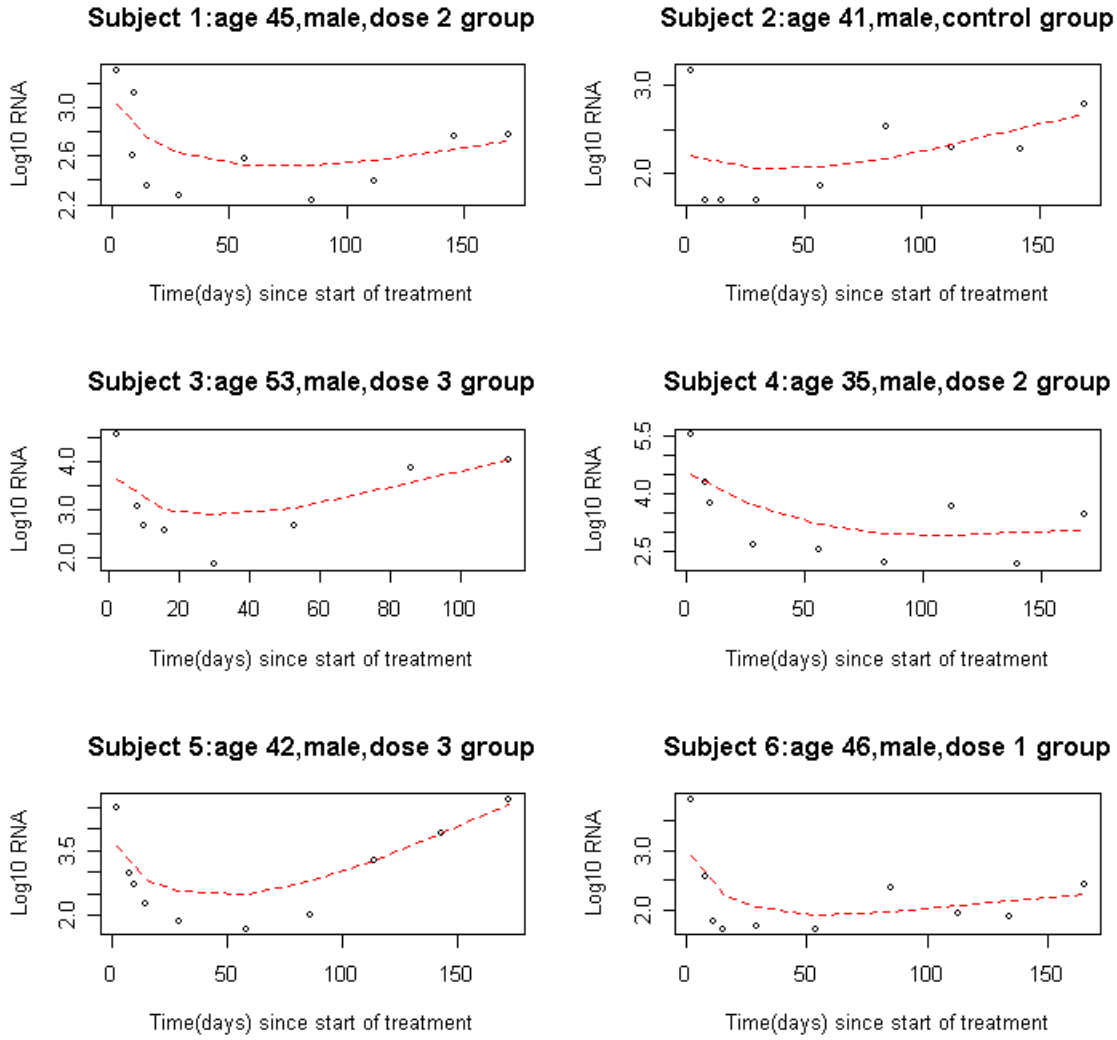

Figure 6.4: The fitted curves from 6 randomly selected patients whose plasma HIV-1 RNA rebounded. The dots are the observations, and the solid lines are Friedman's super smoother.

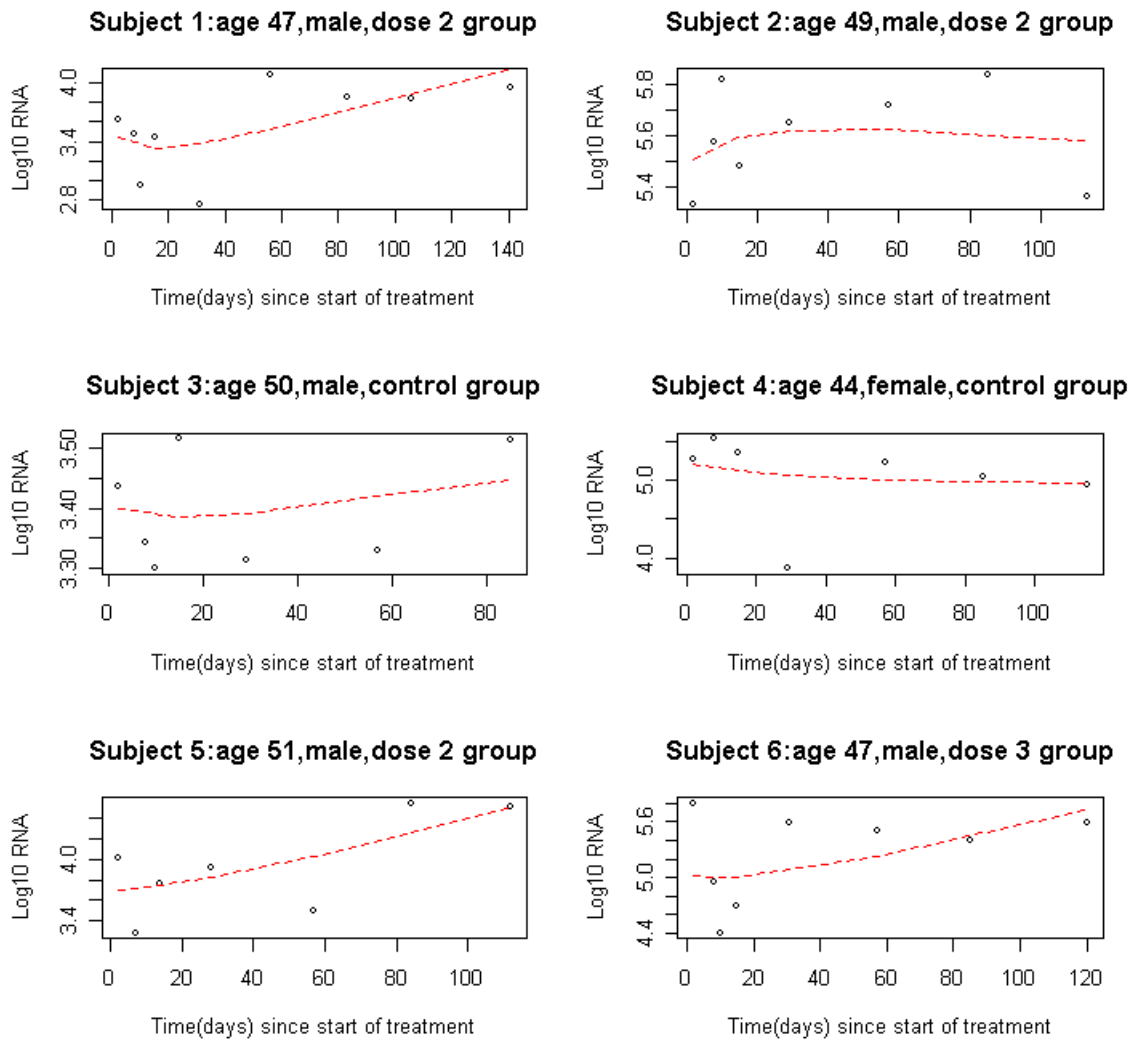

Figure 6.5: The fitted curves from 6 randomly selected patients demonstrating insufficient viral load response. The dots are the observations, and the solid lines are Friedman's super smoother.

The following plots (in figure 7) are for "windowed" $\log_{10}$ (original numeric result) through the day after starting treatment, by treatment with three potent antiviral drug groups and a control group.

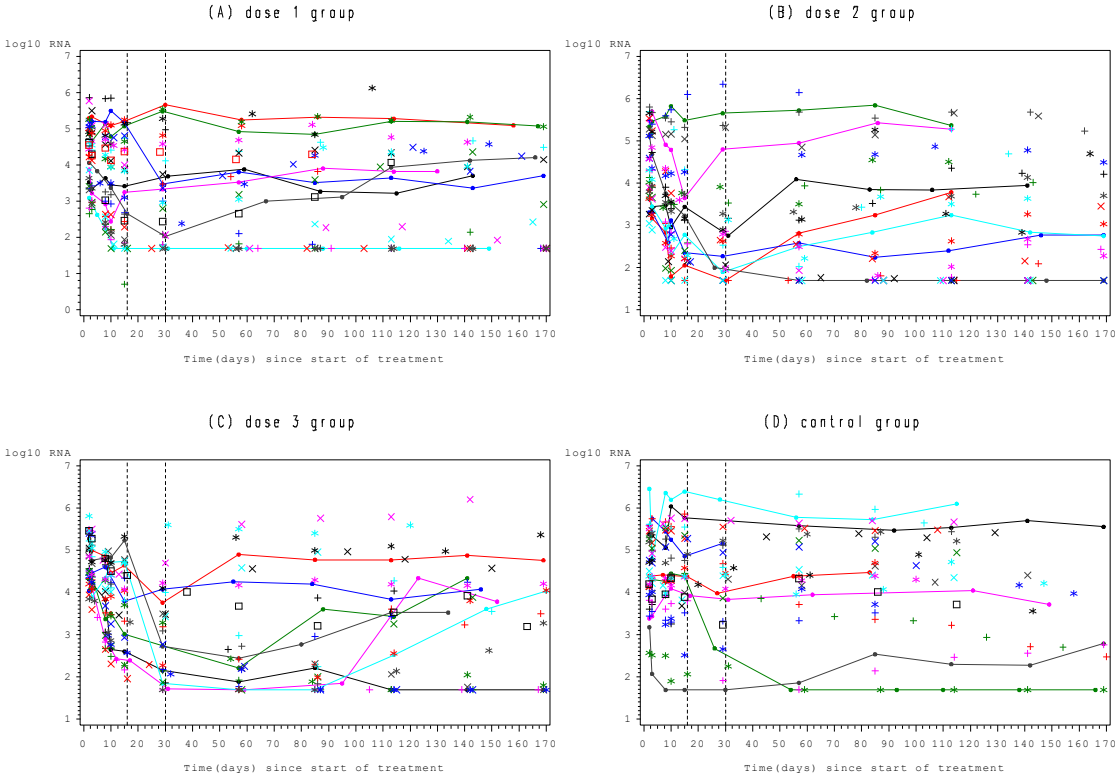

Figure 7: The fitted curves from 30, 28, 29 and 29 patients by actual treatment (A) dose 1, (B) dose 2, (C) dose 3 and (D) control group, clockwise from top left. The first reference vertical dotted line indicates 2 weeks (16 days) and the second reference vertical dotted line indicates 4 weeks (30 days) which is the end of period Wu *et al.* (2004) [13] derived viral decay rates.

4.2 Fitting Nonlinear Mixed Effect Models

4.2.1 Fitting Nonlinear Mixed Effect Single-phase Viral Dynamic Models

At first, since most of the data show rapid decay patterns from the day of starting treatment to approximately 10 days, a nonlinear mixed effect single-phase viral dynamic model based on (7) was considered to drive viral decay rates using HIV-1 RNA data from

day 2 ($t = 0$, the day of starting treatment) to day 16 ($t = 14$, 2 weeks) for the next model[14]:

$$\log_{10}[V(t)] = \log_{10}[\exp(P_{1i})\exp(-d_{1i}t)] + \varepsilon_i(t) \qquad (8)$$

where $\varepsilon_i(t)$ is a HIV-1 RNA measurement error (on the $\log_{10}$ scale) with a mean of zero, and $V_i(t)$ is HIV-1 RNA copies/mL plasma at treatment time t for the i th subject. The viral decay rates for the i th subject is $d_{1i} = d_1 + b_{1i}$, where the fixed-effect parameter, i.e. d_1 is the population decay rates for the first viral decay phases and b_{1i} is a random effect parameter assumed to be $iid\ N(0, \sigma_b^2)$, that quantify the between-subject variation of viral decay rates. Parameter $P_{1i} = P_1 + b_{2i}$ is a macroparameter, with $\exp(P_{1i})$ being baseline virus load at time $t = 0$ [13, 14, 20, 26, 27].

To get starting values for the model, a simple nonlinear regression model using the NLIN procedure was used: without fitting random factor, the NLIN procedure was used to generate appropriate values because the NLMIXED procedure is sensitive to starting values.

The model had converged by Gauss-Newton iterative Method, i.e. the residual sum of squares decreased until there was no improvement in model fit [Appendix 1, table 2.1].

[14] HIV-1 RNA data from day 2 ($t = 0$, the day of starting treatment) to day 10 ($t = 7$, 1 week) and day 2 to day 30 ($t = 28$, 4 weeks) were also considered and fitted to the models (result not shown, the main results are consistent with the result of this study). However, 2 weeks dataset were used to derive viral decay rates for the adequateness based on the viral decay patterns, because viral rebound occurred after 2 weeks and the viral dynamic model is valid only for the early stage of treatment [14, 21].

In the ANOVA table 2.1, the value of the residual sum of squares presents the value that the iterative fitting process converged to. The mean square error of the model fit is the estimate of variability in the data when adjusted for the non-linear logistic model trend we have assumed. The ANOVA table 2.2 is then followed by a table of parameter estimates. (In this case, there were P_1 10.488 and d_1 0.169 want to estimate, standard error and an asymptotic 95% confidence interval.)

Table 2.1: ANOVA table

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	2	8094.9	4047.5	3715.43	<.0001
Error	501	545.8	1.0894		
Uncorrected Total	503	8640.7			

Table 2.2: Parameter estimates table

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
P1	10.4882	0.2001	10.0951	10.8814
d1	0.1691	0.0222	0.1255	0.2127

A nonlinear mixed effect single-phase viral dynamic model with one random effect parameter, i.e. b_{1i}, was fitted after applying the estimated fixed-effect parameter estimates, i.e. d_1, P_1, and the mean squared error as starting values (using PROC NLMIXED statements) for the 107 subjects. For the Optimization Technique, Dual Quasi-Newton was used, and Adaptive Gaussian Quadrature was used as the Integration Method [Appendix 1, table 2.2].

The algorithm has converged successfully and the fitting information (table 2.3) lists the final maximized value of the log likelihood, i.e. -2 log likelihood 1083.5 as well as the

information criteria of Akaike and corrected Akaike (for small sample sizes) in two different forms, i.e. AIC and AICC: 1091.5 and 1091.6, respectively. These statistics can be used to compare different nonlinear mixed models. Also, the "Parameter Estimates" (table 2.4) lists the maximum likelihood estimates of the four parameters and their approximate standard errors computed using the final Hessian matrix [24].

Table 2.3: Fit Statistics table

Fit Statistics	
-2 Log Likelihood	1083.5
AIC (smaller is better)	1091.5
AICC (smaller is better)	1091.6
BIC (smaller is better)	1102.2

Table 2.4: Parameter Estimates table

Parameter Estimates									
Parameter	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**	**Gradient**
P1	10.4761	0.1041	106	100.65	<.0001	0.05	10.2698	10.6825	5.608E-6
d1	0.1662	0.02473	106	6.72	<.0001	0.05	0.1171	0.2152	2.06E-6
Error	0.5406	0.01918	106	28.19	<.0001	0.05	0.5025	0.5786	-0.00002
varcomp11	0.05043	0.007476	106	6.75	<.0001	0.05	0.03561	0.06526	0.000044

Using the fixed-effect parameter estimates, i.e. d_1, P_1, the mean squared error, and variance component of b_{1i} as starting values, a nonlinear mixed effect single-phase viral dynamic model with two random effect parameters, i.e. b_{1i}, P_{1i}, was fitted (using PROC NLMIXED statements) for the same subjects.

The algorithm has converged successfully and the final maximized value of the log likelihood, i.e. -2 log likelihood, was 768.3, AIC and AICC were 780.3 and 780.5, respectively (table 2.8). The maximum likelihood estimates of the six parameters, i.e. d_1,

P_1, the mean squared error, covariance of b_{1i} and b_{2i} (which is not significant in the result), and variance of b_{1i}, b_{2i} respectively, and also their approximate standard errors are listed in the "Parameter Estimates" (table 2.5).

Table 2.5: Parameter Estimates table

Parameter Estimates									
Parameter	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**	**Gradient**
P1	10.4925	0.1756	105	59.76	<.0001	0.05	10.1443	10.8406	0.001018
d1	0.1684	0.01524	105	11.05	<.0001	0.05	0.1382	0.1986	0.006712
error	0.3007	0.01252	105	24.02	<.0001	0.05	0.2759	0.3255	0.002293
varcomp11	0.02003	0.003425	105	5.85	<.0001	0.05	0.01323	0.02682	-0.00064
varcomp12	-0.03959	0.02775	105	-1.43	0.1568	0.05	-0.09462	0.01545	-0.00134
varcomp22	2.9176	0.4497	105	6.49	<.0001	0.05	2.0259	3.8092	0.000512

4.2.2 *Fitting Nonlinear Mixed Effect Biphasic Viral Dynamic Models*

Rapid decay pattern (from the day of starting treatment to approximately 10 days) and slow decay pattern are observed in the same dataset, i.e. HIV-1 RNA data from day 2 ($t = 0$, the day of starting treatment) to day 16 ($t = 14$, 2 weeks). With the data, a nonlinear mixed effect biphasic viral dynamic model (7) was fitted next.

Also, to get starting values for the model, a simple nonlinear regression model for the biphasic of model (using PROC NLIN statement) was used with the arbitrary initial values for fixed effect parameters without fitting random factors. Then, a biphasic model with two random effect parameters; d_1, P_1 was fitted using the estimated fixed-effect parameter estimates; d_1 , P_1 , d_2 , P_2 and the mean squared error. The algorithm has converged successfully and the estimated of the eight parameters, i.e. d_1, P_1, d_2, P_2, the mean squared

error, covariance of b_{1i} and b_{3i} (which is not significant in the result), variance of b_{1i} and b_{3i} , are given the "Parameter Estimates" table 2.6.

Table 2.6: Parameter Estimates table

Parameter Estimates									
Parameter	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**	**Gradient**
P1	10.1026	0.2630	83	38.41	<.0001	0.05	9.5795	10.6257	-1.76E-6
P2	9.6232	0.6032	83	15.95	<.0001	0.05	8.4235	10.8230	-5.09E-6
d1	0.1691	0.01812	83	9.33	<.0001	0.05	0.1330	0.2051	0.000557
d2	0.8850	0.2497	83	3.54	0.0006	0.05	0.3884	1.3816	0.000027
error	0.2868	0.01340	83	21.40	<.0001	0.05	0.2602	0.3135	-0.00024
varcomp11	0.01885	0.003801	83	4.96	<.0001	0.05	0.01129	0.02641	-0.00166
varcomp13	0.05282	0.04726	83	1.12	0.2669	0.05	-0.04117	0.1468	0.000028
varcomp33	4.7793	0.9722	83	4.92	<.0001	0.05	2.8457	6.7130	8.273E-6

Using approximate values of estimated fixed and random effect parameters from the previous model, a biphasic model with three random effects was fitted next with zero covariance components. (For the variance of d_2, parameter value searching function in NLMIXED was used.) The algorithm has converged successfully, and the fitting information table lists the final maximized value of the log likelihood, i.e. -2 log likelihood 735.0, AIC and AICC: 751.0 and 751.3, respectively [Appendix 1, table 2.3]. Also, the "Parameter Estimates" table 2.7 lists the maximum likelihood estimates of the eight parameters; d_1 , P_1 , d_2 , P_2 , the mean squared error, variance of b_{1i} , b_{2i} and b_{3i} respectively, and their approximate standard.

Table 2.7: Parameter Estimates table

Parameter Estimates									
Parameter	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**	**Gradient**
P1	10.1761	0.2129	104	47.80	<.0001	0.05	9.7539	10.5982	0.000668
P2	11.2454	0.5086	104	22.11	<.0001	0.05	10.2369	12.2539	-0.00147
d1	0.1423	0.01511	104	9.42	<.0001	0.05	0.1124	0.1723	-0.0029
d2	1.7727	0.2487	104	7.13	<.0001	0.05	1.2796	2.2658	0.003581
error	0.2776	0.01178	104	23.57	<.0001	0.05	0.2542	0.3009	0.006208
varcomp11	0.01928	0.003195	104	6.04	<.0001	0.05	0.01295	0.02562	0.071378
varcomp22	0.1934	0.06813	104	2.84	0.0055	0.05	0.05826	0.3285	-0.00593
varcomp33	4.3740	0.6899	104	6.34	<.0001	0.05	3.0058	5.7422	-0.00021

Since the biphasic model with three random effects and zero covariance components was successfully converged, using previous estimated initial parameter values, the same model with covariance components was fitted next. (For the variance of d_2, parameter value searching function in NLMIXED was used also.) The algorithm has converged successfully and the fitting information table lists the final maximized value of the log likelihood, i.e. -2 log likelihood 702.8, AIC and AICC: 724.8 and 725.3, respectively which are slightly better than the previous model (table 2.8). (Model diagnoses support this result [Appendix 1, table 2.4- 2.5]). Also, the "Parameter Estimates" (table 2.9) lists the maximum likelihood estimates of the eleven parameters; d_1, P_1, d_2, P_2, the mean squared error, variance of b_{1i}, b_{2i} and b_{3i} respectively, covariance of b_{1i} and b_{2i} (which is not significant in the result), b_{2i} and b_{3i}, b_{1i} and b_{3i} respectively and their approximate standard.

Table 2.8: Fit Statistics for both models

Fit Statistics		
	The single-phase model	**The biphasic model**
-2 Log Likelihood	768.3	702.8
AIC (smaller is better)	780.3	724.8
AICC (smaller is better)	780.5	725.3
BIC (smaller is better)	796.3	754.2

Table 2.9: Parameter Estimates table

Parameter Estimates									
Parameter	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**Alpha**	**Lower**	**Upper**	**Gradient**
P1	9.5657	0.2278	104	41.98	<.0001	0.05	9.1139	10.0176	0.000304
P2	13.3343	0.2344	104	56.89	<.0001	0.05	12.8695	13.7992	0.00027
d1	0.09568	0.01432	104	6.68	<.0001	0.05	0.06729	0.1241	0.001695
d2	1.7545	0.1488	104	11.79	<.0001	0.05	1.4594	2.0496	0.000999
error	0.2615	0.01443	104	18.11	<.0001	0.05	0.2328	0.2901	0.003725
varcomp11	0.01090	0.003285	104	3.32	0.0012	0.05	0.004389	0.01742	0.015094
varcomp22	0.7450	0.1586	104	4.70	<.0001	0.05	0.4306	1.0595	-0.00055
varcomp33	2.7611	0.4880	104	5.66	<.0001	0.05	1.7933	3.7288	0.000469
varcomp13	-0.09803	0.03071	104	-3.19	0.0019	0.05	-0.1589	-0.03713	0.006636
varcomp23	-1.0988	0.2269	104	-4.84	<.0001	0.05	-1.5487	-0.6488	0.000382
varcomp12	0.01134	0.01517	104	0.75	0.4566	0.05	-0.01875	0.04142	0.008288

The biphasic model with four random effects which contains b_{4i} had not successfully converged[15].

4.3 Analysis Results from the Models

4.3.1 *Initial Viral Decay Rates*

The estimated first-phase decay rates, i.e. d_1 from individual subjects (the empirical Bayesian estimates) of the nonlinear mixed effect single-phase and biphasic viral dynamic models are summarized for different ages, ethnicities and actual treatment groups in the table 4.1 and 4.2.

[15] It had not converged using alterative methods for the convergence. However, since we are interested in the first-phase viral decay rates rather than the other phase viral decay rates (for the multi-phase model), and we obtained the values with other main random effects of the biphasic model, it seems that it is not a principle problem in this study.

Table 4.1: Summary of estimated viral decay rates of the nonlinear mixed effect single-phase viral dynamic model.

Characteristic	Label	Decay rate, Mean d1 ± SD
		Phase1
Total	(n=107)	0.168 ± 0.129
Treatment regimen	dose1 (n=28)	0.182 ± 0.127
	dose2 (n=28)	0.212 ± 0.124
	dose3 (n=29)	0.193 ± 0.126
	control (n=22)	0.063 ± 0.085
Age range, years	Less than 20 (n=7)	0.180 ± 0.123
	30-34 (n=5)	0.184 ± 0.129
	35-39 (n=22)	0.181 ± 0.122
	40-44 (n=31)	0.162 ± 0.149
	45-49 (n=18)	0.185 ± 0.125
	50-59 (n=20)	0.119 ± 0.112
	Over 60 (n=4)	0.287 ± 0.067
Race/Ethnicity	White – White/Caucasian/European Heritage and Arabic/North African Heritage (n=87)	0.161 ± 0.126
	African American/African Heritage (n=20)	0.202 ± 0.143

Table 4.2: Summary of estimated viral decay rates of the nonlinear mixed effect biphasic viral dynamic model.

Characteristic	Label	Decay rate, Mean d1 ± SD	
		Phase1	Phase2
Total	(n=107)	0.097 ± 0.095	1.681 ± 0.810
Treatment regimen	dose1 (n=28)	0.114 ± 0.096	1.872 ± 0.915
	dose2 (n=28)	0.128 ± 0.091	1.690 ± 0.841
	dose3 (n=29)	0.109 ± 0.089	1.412 ± 0.520
	control (n=22)	0.022 ± 0.072	1.779 ± 0.90
Age range, years	Less than 20 (n=7)	0.112 ± 0.101	1.920 ± 1.155
	30-34 (n=5)	0.119 ± 0.109	2.252 ± 1.336
	35-39 (n=22)	0.103 ± 0.092	1.540 ± 0.805
	40-44 (n=31)	0.092 ± 0.109	1.660 ± 0.718
	45-49 (n=18)	0.109 ± 0.093	1.545 ± 0.827
	50-59 (n=20)	0.062 ± 0.073	1.729 ± 0.717
	Over 60 (n=4)	0.183 ± 0.035	1.856 ± 0.590
Race/Ethnicity	White – White/Caucasian/European Heritage and Arabic/North African Heritage (n=87)	0.093 ± 0.095	1.710 ± 0.856
	African American/African Heritage (n=20)	0.115 ± 0.098	1.552 ± 0.571

4.3.2 *Baseline Characteristics*

Baseline characteristics of study participants are summarized by actual treatment group in table 5.

Table 5: Baseline characteristics of study participants by actual treatment group for the completed dataset.

Characteristic	Label	Total (n=116)	Actual treatment group			
			control (n=29)	dose 1 (n=30)	dose 2 (n=28)	dose 3 (n=29)
Sex, no.(%)	Female	17 (14.7)	3	9	4	1
	Male	99 (85.3)	26	21	24	28
Age range, years, no.(%)	Less than 20	7 (6.0)	1	2	2	2
	30-34	5 (4.3)	2	2	1	0
	35-39	22 (19.0)	2	5	7	8
	40-44	36 (31.0)	13	7	5	11
	45-49	20 (17.2)	3	8	5	4
	50-59	22 (19.0)	8	6	5	3
	Over 60	4 (3.4)	0	0	3	1
Race/Ethnicity, no.(%)	African American/African Heritage	20 (17.2)	6	4	2	8
	White – White/Caucasian/European Heritage and Arabic/North African Heritage	96 (82.8)	23	26	26	21
HIV-1 RNA, $\log_{10}$ copies/mL	25th percentile	3.89	4.17	3.49	3.67	4.36
	Median	4.55	4.36	4.57	4.31	4.63
	75th percentile	5.17	5.45	5.09	5.18	5.00
CD4 cell count, cells/mm3	25th percentile	53.5	23	62	60.5	101
	Median	152	69	180.5	190.0	155
	75th percentile	281.5	175	337	325.5	280
CD4 cells, %	25th percentile	6.5	5	9	8	10
	Median	13	7	16.5	13	13
	75th percentile	19	13	25	22	17

ANOVA (1-way analysis of variance) analysis showed the effects of "age", "gender" and "ethnicity" on the first-phase viral decay rates of both models were not significant: p=0.733, p=0.257 and p=0.420 respectively for the single-phase model, and p=0.812, p=0.192 and p=0.593 respectively for the biphasic. (This result was marginally confirmed by Univariate Regression Analyses treating age, gender, and ethnicity as continuous covariates). However, there was a marginally significant difference among "actual treatment groups" for both models (p=0.0001 and p=0.0003) [Appendix 1, table 3.1.1- 3.4.2].

Results of the Wilcoxon rank sum and Kruskal-Wallis tests indicated that the first-phase viral decay rate in the control group (mean $d_1 \pm SD$, 0.063 $\pm$0.085 and 0.022 $\pm$0.072 for the single and biphasic model, respectively) is significantly lower than in other treatment groups, i.e. dose 2 (mean $d_1 \pm SD$, 0.212 $\pm$0.124 and 0.128 $\pm$0.091 for the single and biphasic model, respectively), dose 1 (mean $d_1 \pm SD$, 0.182 $\pm$0.127 and 0.114 $\pm$0.096, respectively), and dose 3 (mean $d_1 \pm SD$, 0.193 $\pm$0.126 and 0.109 $\pm$0.089, respectively), for both models (p=0.0002 and p=0.0005, respectively) [Appendix 1, table 3.5.1- 3.6.8]. The first-phase decay rates in dose 2 was higher than other groups, but there were no significant differences among three treatment groups according to Tukey's Studentized Range (HSD) in GLM for the both models [Appendix 1, table 3.7.1- 3.7.2]. In addition, the regression analysis confirmed that treatment assignment was a significant predictor for the first-phase viral decay rates, i.e. d_1 in both models (p=0.002 and p=0.001 respectively).

Multiple (Covariate) Regression Analyses (GLM), including the covariates baseline HIV-1 RNA levels, baseline CD4+ counts, age, ethnicity and gender of patients, number of mutations at baseline, introduction of T20 and treatment assignment, indicated that 'treatment assignment', and 'number of mutations at baseline' were significant predictors (p=0.014 and p=0.0023, respectively) of the first-phase viral decay rates for the single-phase model. In addition, 'baseline HIV-1 RNA levels', 'treatment assignment' and 'number of mutations at baseline' were significant predictors of d_1 (p=0.005, p=0.02 and p=0.002, respectively) for the biphasic model [Appendix 1, table 3.8.1- 3.8.2].

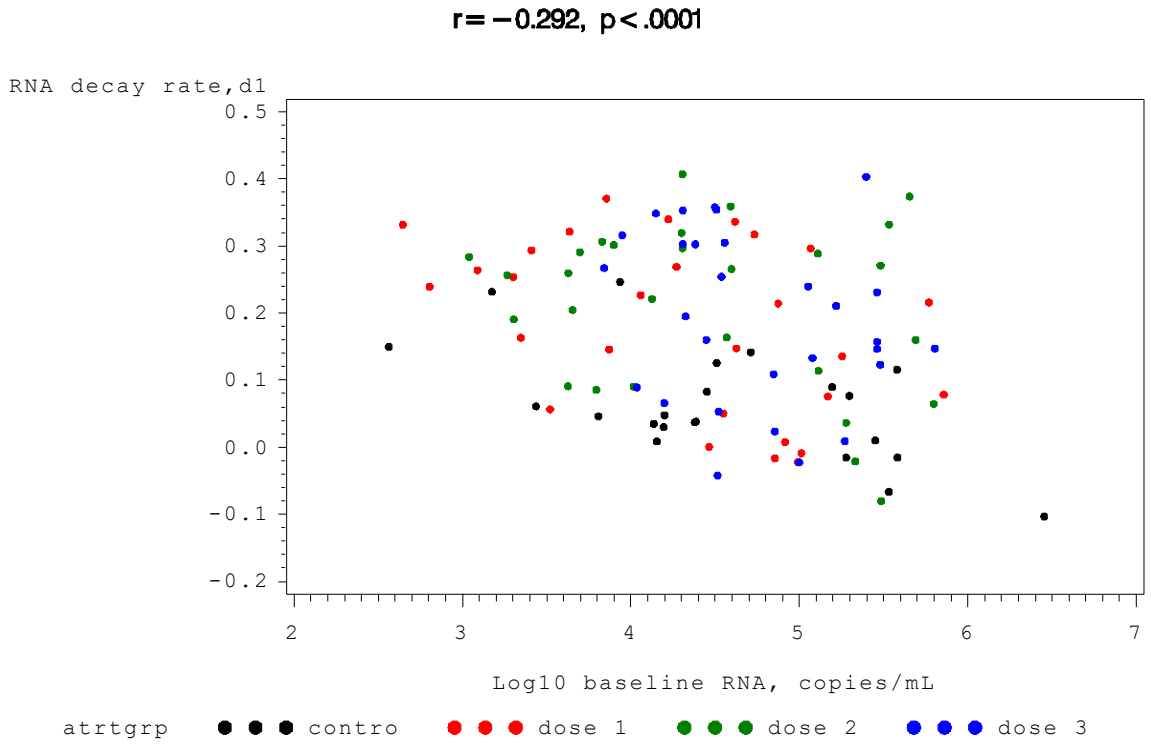

Figure 8.1: The correlation between the first-phase viral decay rates and baseline HIV-1 RNA levels at t=0 for the nonlinear mixed effect single-phase viral dynamic model. The correlation coefficient and P value from Spearman's rank tests are given.

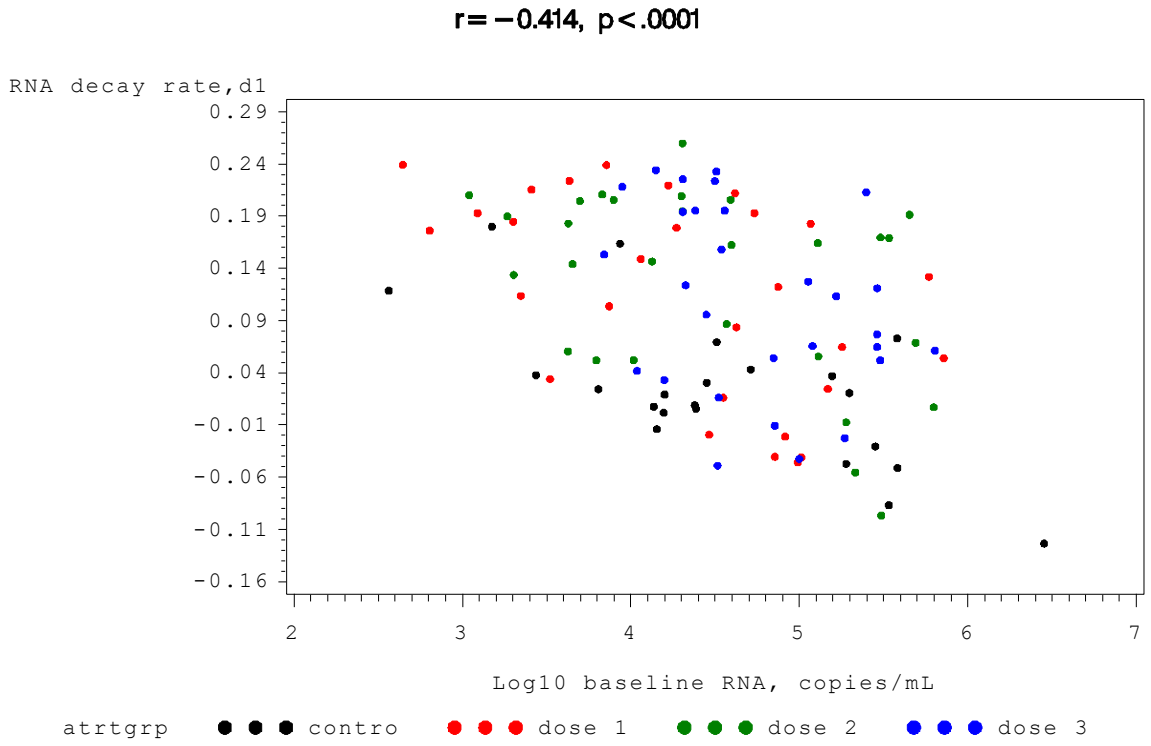

Figure 8.2: The correlation between the first-phase viral decay rates and baseline HIV-1 RNA levels at t=0 for the nonlinear mixed effect biphasic viral dynamic model. The correlation coefficient and P value from Spearman's rank tests are given.

Figure 8.1- 8.2 show the correlation of the first-phase viral decay rates, i.e. d_1 with baseline HIV-1 RNA levels for the two models. d_1 were somewhat negatively correlated with baseline virus load (r = -0.292, p=0.002) and (r = -0.414, p<0.0001) for the both models respectively [Appendix 1, table 3.9.1- 3.9.2] and positively correlated with baseline CD4+ cell counts (r = 0.428, p<0.001) and (r = 0.482, p<0.001) for the both models [Appendix 1, table 3.10.1- 3.10.2].

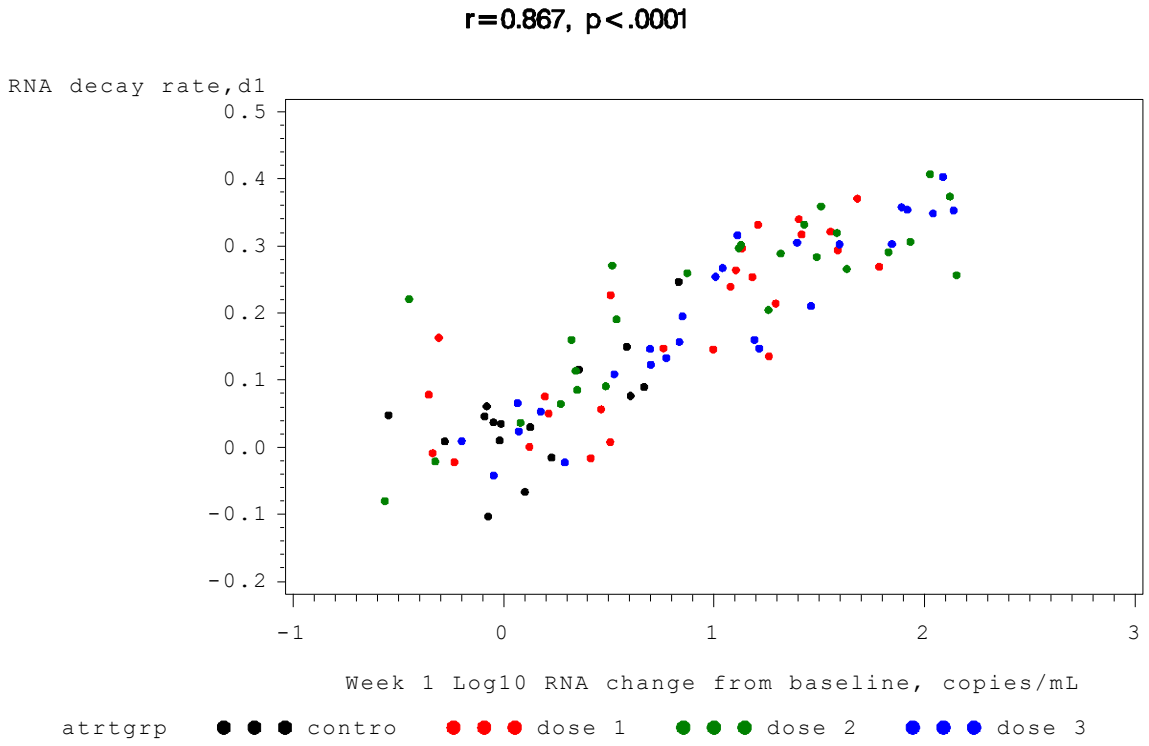

Figure 9.1: The correlation between the first-phase viral decay rates and week 1 virus load reduction from baseline for the single-phase viral dynamic model (for instance, 3 indicates $\log_{10}$ viral load reduction from baseline whereas, -1 indicates $\log_{10}$ viral load increase from baseline.). The correlation coefficient and P value from Spearman's rank tests are given.

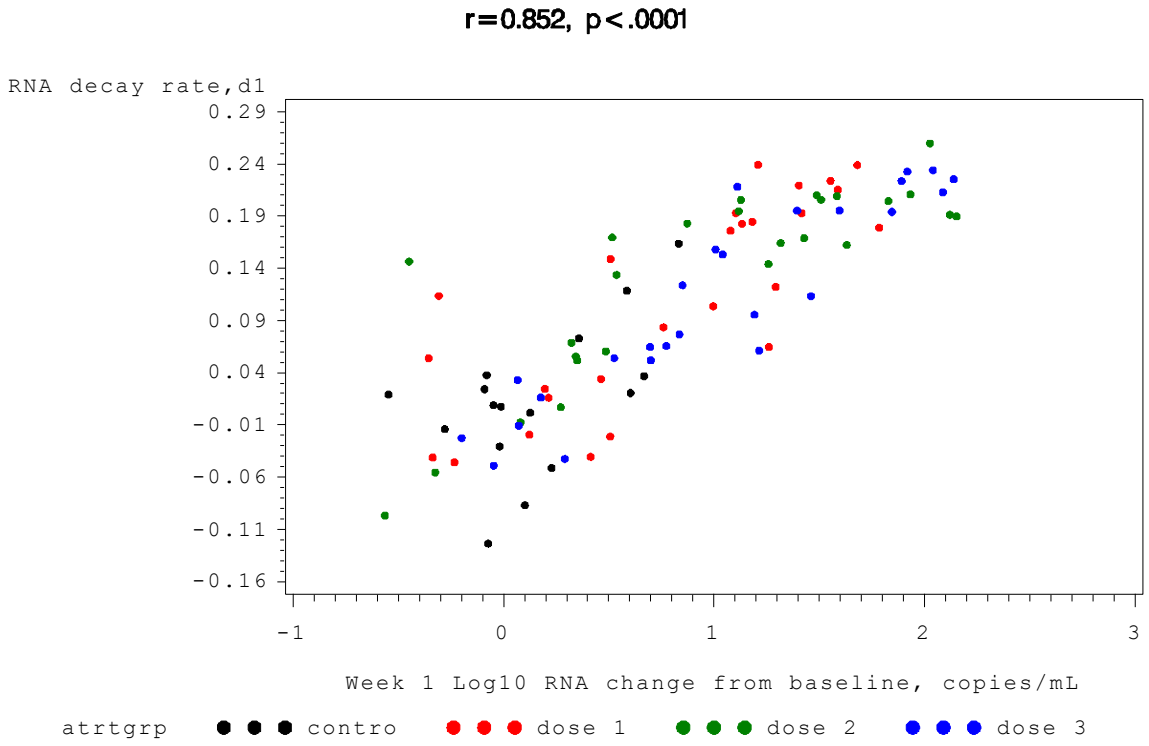

Figure 9.2: The correlation between the first-phase viral decay rates and week 1 virus load reduction from baseline for the single-phase viral dynamic model (for instance, 3 indicates $\log_{10}$ viral load reduction from baseline whereas, -1 indicates $\log_{10}$ viral load increase from baseline.). The correlation coefficient and P value from Spearman's rank tests are given.

Also, the strong positive correlations (r = 0.867, p<0.0001 and r = 0.852, p<0.0001) between week 1 virus load reduction, i.e. early virus dynamics, and the first-phase viral decay rates for both models were observed (figure 9.1- 9.2) [Appendix 1, table 3.11.1-3.11.2].

In addition, there were positive correlations (r = 0.415, p=0.0003 and r = 0.432, p=0.002, respectively) between week 20 and 24 virus load reduction and the first-phase viral decay rates for the single-phase model. Also, there were positive correlations (r = 0.346, p=0.003 and r = 0.394, p=0.006, respectively) between week 20 and 24 virus load reduction and the first-phase viral decay rates for the biphasic model. However, their correlations were less

than those of week 1 virus load reduction [Appendix 1, table 3.12.1- 3.12.4]. Also, week 1 virus load reduction, i.e. early viral dynamics, and week 20 and 24 virus load reduction have positive correlations (r = 0.445, p<0.0001 and r = 0.420, p=0.003, respectively) for both models [Appendix 1, table 3.13.1- 3.13.4]. Similarly, it is observed that the individuals with higher first-phase viral decay rates, i.e. subjects with d_1 > 0.168 (mean of d_1 for the single-phase model), were more likely to have suppressed virus load at week 20 (mean $d_1 \pm SD$, 2.743 ± 1.172) and at week 24 (2.56 ± 0.996) than the other subjects with $d_1 \leq 0.168$ (3.701 ± 1.164 at week 20 and 3.510 ± 1.313 at week 24, respectively) for the single-phase model. Consistently, the individuals with higher first-phase viral decay rates, i.e. subjects with d_1 > 0.097 (mean of d_1 for the biphasic model), were also more likely to have suppressed virus load at week 20 (2.722 ± 1.166) and at week 24 (2.503 ± 0.987) than the other subjects with $d_1 \leq 0.097$ (3.818 ± 1.102 at week 20 and 3.692 ± 1.234 at 24 week) for the biphasic model [Appendix 1, table 3.14.1- 3.14.8].

The subjects with higher first-phase viral decay rates were also more likely to show a smaller number of mutations at baseline than the other subjects in both models [Chart 1.1-1.2].

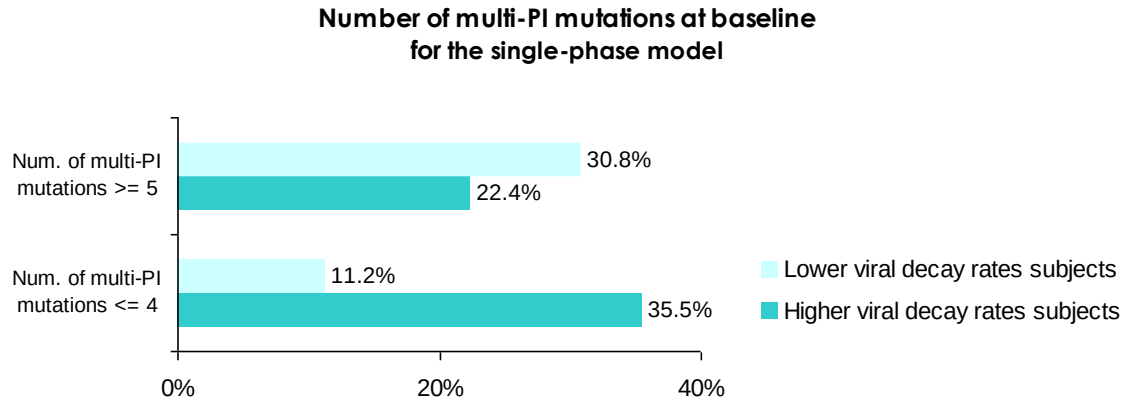

Chart 1.1: Number of multi-PI mutations at baseline by the first-phase viral decay rates for the single-phase model.

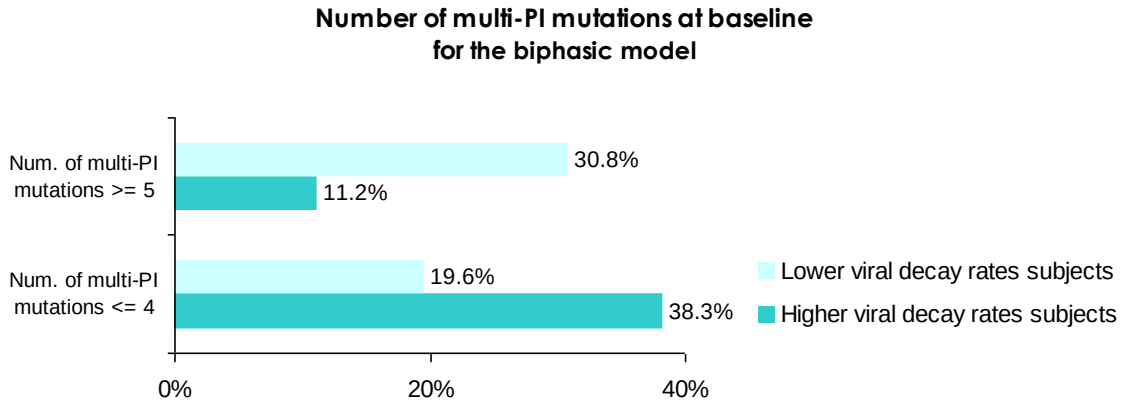

Chart 1.2: Number of multi-PI mutations at baseline by the first-phase viral decay rates for the biphasic model.

4.3.3 *Initial Viral Decay Rates and Long-Term Response*

Among the 107 subjects tested for in this study, 81 were classified as virological non-responders, and 26 were classified as responders (P400_TLO). The viral decay rates between the virological responders and non-responders using the Wilcoxon rank sum and Kruskal-Wallis tests (using the NPAR1WAY procedure in SAS) was performed [Appendix 1, table 3.15.1- 3.15.6]. It was found that the first-phase viral decay rates in responders (mean $d_1 \pm SD$, 0.267 ± 0.103) were significantly higher (both $p<0.001$ respectively) than those in non-responders (mean $d_1 \pm SD$, 0.137 ± 0.121) in the single-phase model. Also, the first-phase viral decay rates in responders (mean $d_1 \pm SD$, 0.171 ± 0.073) were significantly higher (both $p<0.001$ respectively) than those in non-responders (mean $d_1 \pm SD$, 0.074 ± 0.09) in the biphasic model.

The subjects with higher first-phase viral decay rates were more likely to be responders. For example, if $d_1 > 0.168$ (mean of d_1 for the single-phase model), 20 (40.0%) of 50 were responders and only $d_1 \leq 0.168$, 6 (10.53%) of 57 were responders for the single-phase

model. Also, if $d_1 > 0.097$ (mean of d_1 for the biphasic model), 21 (39.62%) of 53 were responders and $d_1 \leq 0.097$, only 5 (9.30%) of 54 were responders) for the biphasic model [Appendix 1, table 3.16.1- 3.16.2].

The week 1 virus load reduction in the responder group was also higher (mean $d_1 \pm SD$, 1.323 ± 0.689) than that in the non-responder group (mean $d_1 \pm SD$, 0.614 ± 0.675) for both models identically. The differences were statistically significant ($p<0.0001$) for the Wilcoxon rank sum and Kruskal-Wallis tests for both models [Appendix 1, table 3.17.1- 3.17.6].

In addition, Univariate Logistic Regression analyses (using the CATMOD procedure in SAS) showed that the control group was a significant predictor for 24 weeks virological response for both the models (both $p = 0.22$) [Appendix 1, table 3.18.1- 3.18.4].

SECTION 5: DISCUSSION

The purpose of this project was to model viral decay rates, and check the validity of the model for the set of data and investigate if the relationships found with baseline covariates and long-term response were consistent with Wu *et al.*'s (2004) findings [13].

The Nonlinear Mixed Effect Single and Biphasic Viral Dynamic Models for the HIV data were fitted using PROC NLMIXED statement in SAS. And the first-phase viral decay rates for each subject for both models were derived. Baseline characteristics which were correlated with the viral decay rates were identified, and whether the initial, i.e. the first-phase, viral decay rates can predict long-term response was examined along with other relevant analyses.

5.1 Actual Treatment Group for the Viral Decay Rates

The results indicate that the actual treatment groups were more potent, i.e. higher first-phase viral decay rates than control group. However, there were no significant differences among treatment groups, dose 1, 2 and 3. (This was consistent with the result without control group in the data.)

There were no significant differences in the first-phase viral decay rates for different ages, ethnicities, and gender groups. The actual treatment effect and the number of multi-PI mutations at baseline had impact on the first-phase viral decay rates for the single-phase model. Also, the actual treatment effect, baseline HIV-1 RNA levels and the number of multi-PI mutations at baseline had impact on the first-phase viral decay rates for the

biphasic model. Besides, the first-phase viral decay rates were somewhat positively correlated with baseline CD4+ cell counts in both models.

5.2 Initial Viral Decay Rates and Viral Load Reduction

The first-phase viral decay rates were somewhat negatively correlated with the baseline HIV-1 RNA levels for both models ($r = -0.292$, $p=0.002$ and $r = -0.414$, $p<0.0001$, respectively). These negative correlations confirm the results of Wu *et al.* [41, 42, 20, 13], but differs from the results of Wu *et al.* (2004) [13] ($r = 0.44$, $p<0.001$). Based on the equations (6), three possible explanations for this negative correlation between the first-phase viral decay rates and baseline viral load were derived by Wu *et al.* (2003) [20]. However, the biological mechanisms behind these correlations are considered still unclear [13]. Wu *et al.* (2004) [13] guessed that the direction of the correlation may depend on many factors such as the potency of treatment regimens, pretreatment virus production: clearance ratio, and turnover rate of infected cells [44, 13].

In addition, strong correlations between the first-phase viral decay rates and week 1 virus load reduction from baseline were observed from both models ($r = 0.867$, $p<0.0001$ and $r = 0.852$, $p<0.0001$, respectively). This result consists with the findings of Wu *et al.* (2004) [13] ($r = 0.89$, $p<0.001$). There was positive correlation between week 20 and 24 virus load reduction and the first-phase viral decay rates for both models. However, their correlations were less than those of week 1 virus load reduction. It seems that the week 1 virus load reduction could be used to replace more complex viral decay rates for the assessment of the potency of antiretroviral regimens [13]. Wu *et al.* (2004) [13] suggested this simplification can avoid complicated viral dynamic model fitting and frequent clinical visits for HIV-1 RNA measurements. However, to compensate for the power loss using

the week 1 virus load reduction, a larger number of subjects (sample size) may be required [13, 40].

Individuals with higher viral decay rates were likely to show long-term viral load suppression (week 20 and 24 virus load reduction). Also, individuals with higher week 1 virus load reduction, i.e. early viral dynamics, were more likely to have suppressed virus load at week 24, which suggests that the antiviral potency or the initial viral decay rates are predictive of long-term viral load response [13].

5.3 Approaching NLMIXED Models

The Nonlinear Mixed Effect Viral Dynamic Model can easily handle unbalanced repeated and continuous measurements data on each of individuals when interest focuses on individual-specific characteristics and allows for flexible variance-covariance and non-independent error structures of the response vector, whereas GEEs are only feasible for the linear in their parameters.

However, the model fitting with the fixed and random effects entering nonlinearly is not easy to implement frequently, since standard likelihood approaches are considered much more difficult to implement than the linear mixed models[16]. Also, it seems to be very sensitive for initial values. This is not a surprising thing when we consider the data repeatedly and continuously measured from various among individuals, and individuals' responses all follow a similar functional form in the model (although it has parameters that vary among individuals). Changing the initial values or sectional iteration searching method for the initial values can be used, the work is sometimes not trivial and even after

[16] Pseudo-Data calculation can be implemented by modifying a standard nonlinear least squares estimation routine [16].

finding the adequate values, and quite a lot of time is required for the complex models which contain many parameters or a large dataset.

In general, there are some suggestions for the difficulties or filatures in converging. They are as follows: Rescale the data and model so that all parameters are of the same order of magnitude for the stability of the algorithm. Otherwise, use boundary constraints to avoid the region where overflows may happen, or delete outlying observations which are reasonable. Also, if the convergence criterion appears to be descending favourably, it might be needed to increase the maximum number of iterations using MAXITER=option in SAS NLINMIX procedure. Changing starting values by using a grid search specification or changing the optimization technique using TECH=option in SAS NLINMIX, and skipping RANDOM before getting accurate starting values can also be useful[17] [11]. For the long run times, it is important to check whether the model is specified correctly. The scaled parameters with same order of magnitude, and the data reasonably match the model are required because ill-posed or miss-specified models can cause the algorithms to use more extensive calculations designed to achieve convergence [11].

5.4 Conclusions and Suggestions

The Nonlinear Mixed Effect Single and Biphasic Viral Dynamic Models for the HIV data were fitted using the NLMIXED procedure in SAS. The main findings with the initial viral decay rates, i.e. the first-phase viral decay rates, from the models were almost identical. For the model comparison aspect, the biphasic viral dynamic model seems

[17] Besides, for SAS NLINMIX macro procedure, using of OPTION=SKIPNLIN, TOL= options, trying RIDGE=option instead of PROC MIXED itself, and using EXPAND=ZERO option but when EXPAND=EBLUP option, also trying GAUSS=, MAXSUBIT=, FRACTION=, and SUBCONV=options which request to take extra Gauss-Newton steps within each iteration can be recommended [22].

slightly better in terms of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) in this study.

Further studies could focus on model fitting using Generalised additive models (GAMs) with splines which consist of individual segments that are joined smoothly. A GAM is defined as a Generalized Linear Model (GLM) with a linear predictor involving a sum of smooth functions of covariates. It consists of a random component, an additive component and a link function relating the two [45]. Compare than traditional parametric modeling tools such as linear or nonlinear regression, the methodology behind the GAM procedure has greater flexibility. It relaxes the usual parametric assumption and enables us to uncover structure in the relationship between the independent variables and the dependent variable [50]. However, this increased flexibility can reduce the interpretability of the modeling output [45].

REFERENCES

[1] Burton, P., Gurrin, L., Sly, P., “Tutorial in biostatistics, extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modelling”, 17, 1261-1291, 1998.

[2] Armitage, P., & Berry, G., “Statistical methods in medical research”, 2nd edn, ***Blackwell Scientific Publications***, Oxford, 70, 82, 86, 90, 104-106, 109, 143-150, 156, 273-295, 302, 307, 314-316, 318, 321, 347-357, 506, 1987.

[3] Feldman, H. A., “Families of lines: random effects in linear regression analysis”, ***Journal of Applied Physiology***, 64, 1721-1732, 1988.

[4] Goldstein, H., “Multilevel models in educational and social research”, ***Charles Griffin & Company Ltd,*** London, 1, 10-31, 32-33, 48, 51-60, 83-84, 1987.

[5] Laird, N., “Longitudinal studies with continuous responses, Statistical Methods in Medical Research”, 1, 225-247, 1992.

[6] Neuhaus, J., “Statistical methods for longitudinal and clustered designs with binary responses”, ***Statistical Methods in Medical Research***, 1, 249-273, 1992.

[7] Zeger, S. L., & Liang, K-Y., “An overview of methods for the analysis of longitudinal data”, ***Statistics in Medicine***, 11, 1825-1839, 1992.

[8] Liang, K-Y., & Zeger, S. L., “Longitudinal data analysis using generalized linear models”, ***Biometrika***, 73, 13-22, 1986.

[9] Zeger, S. L., & Liang, K-Y., “Longitudinal data analysis for discrete and continuous outcomes”, ***Biometrics***, 42, 121-130, 1986.

[10] Wu, H., & Wu, L., “Identification of significant host factors for HIV dynamics modelled by nonlinear mixed-effect models”, ***Stat Med***., 21, 753–771, 2002.

[11] NLMIXED procedure, assumptions, SAS Help and Documentation, SAS v9.1 (SAS Institute Inc., 2003).

[12] Bishop, J., “Nonlinear Regression”, MT5756 DATA analysis, St Andrews University, 2007.

[13] Wu, H., Lathey, J., Ruan, P., Douglas, S D., Spector, S A., Lindsey, J, Hughes, M D., Rudy, B J., & Flynn, P M., for the PACTG 381 Team, “Relationship of plasma HIV-1 RNA dynamics to baseline factors and virological responses to highly active antiretroviral therapy in adolescents (Aged 12–22 Years) infected through high-risk behaviour”, ***Journal of Infectious Diseases***, 189, 593-601, 2004.

[14] Wu, H., & Ding, A A., “Population HIV-1 Dynamics In Vivo, Applicable models and inferential tools for virological data from AIDS clinical trials”, ***Biometrics***, 55, 410-418, 1999.

[15] Ding, A A., & Wu, H., “Relationship between antiviral treatment effects and biphasic viral decay in modelling HIV dynamics”, ***Mathematical Biosciences***, 160, 63-85, 1999.

[16] Lindstrom, M J., & Bates D M., “Nonlinear mixed effects models for repeated measures data”, ***Biometrics***, 46, 673-687.

[17] Pinheiro, J C., & Bates D M., “Approximations to the log-likelihood function in nonlinear mixed-effects models”, ***Journal of Computational and Graphical Statistics***, 4, 12-35, 1995.

[18] Davidian, M., & Giltinan, D M., “Nonlinear models for repeated measurement data: An Overview and Update”, Chapman & Hall, London.

[19] Pinheiro, J C., & Bates, D M., “Model building for nonlinear mixed-effects models”, Department of Biostatistics, 1994

[20] Wu, H., Mellors, J., Ruan, P., McMahon, D., Kelleher, D., & Lederman, M M., “Viral dynamics and their relations to baseline factors and longer”, ***JAIDS (Journal of Acquired Immune Deficiency Syndromes)***, Lippincott Williams & Wilkins, Inc., Philadelphia, 33, 557–563, 2003.

[21] Perelson, A.S., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., & Ho, D.D., “Decay characteristics of HIV-1-infected compartments during combination therapy”, ***Nature***, 387, 188-191, 1997.

[22] Littell R C., Milliken G A., Stroup W W., & Wolfinger R D., “SAS system for mixed models”, SAS Institute Inc., 1996

[23] Anthony, P. F., Victor G. D. & Florin V., “Modelling HIV viral rebound using non-linear mixed effects models”, ***Statist. Med***, 21, 2093–2108, 2002.

[24] Wolfinger R D., “Fitting nonlinear mixed models with the NLMIXED Procedure”, SAS Institute Inc., Cary, NC.

[25] A. S. Perelson, P. Essunger, Y. Cao, M. Vesanen, A. Hurley, K. Saksela, M. Markowitz, & D.D. Ho, “Decay characteristics of HIV-1-infected compartments during combination therapy”, ***Nature***, 387, 188-191, 1997.

[26] Ding A A., & Wu, H. “A comparison study of models and fitting procedures for biphasic viral dynamics in HIV-1 infected patients treated with antiviral therapies”, ***Biometrics***; 56, 293–300, 2000.

[27] Ding, A A., & Wu H., “Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral Dynamic Model”, ***Biostatistics,*** 2, 13–29, 2001.

[28] Perelson, A S., Essunger, P., & Cao, Y. *et al.,* “Decay characteristics of HIV 1–infected compartments during combination therapy”, ***Nature***, 387, 188–91, 1997.

[29] Lalezari, J P., Ward, D J., Tomkins, S A., & Garges, H P., “Preliminary safety and efficacy data of brecanavir, a novel HIV-1 protease inhibitor: 24 week data from study HPR10006”, ***Journal of Antimicrobial Chemotherapy***, 60, 170-174, 2007

[30] Glaxosmithkline Research and Development proposed project description of project title: analysis of HIV data.

[31] Glaxosmithkline Research and Development proposed project dataset description of project title: analysis of HIV data.

[32] Sheiner, L. B., & Beal, S. L., "Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten Model: Routine clinical pharmacokinetic data", ***Journal of Pharmacokinetics and Biopharmaceutics***, 8, 553-571, 1980.

[33] Vonesh, E. F., & Carter, R. L., "Mixed-effects nonlinear regression for unbalanced repeated measures", ***Biometrics***, 48, 1-18, 1992.

[34] Davidian, M., & Gallant, A. R., "Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with Application to Quinidine", ***Journal of pharmacokinetics and biopharma ceutics***, 20, 529-556, 1992.

[35] Lindstrom, M. J., & Bates, D. M., "Newton-raphson and EM algorithms for linear mixed-effects models for repeated-measures data", ***Journal of the American Statistical Association***, 83, 1014-1022. 1988.

[36] Laird, N. M., & Ware, J. H., "Random-effects models for longitudinal data", ***Biometrics***, 38, 963- 974, 1982.

[37] Bates, D. M., & Watts, D. G., “Relative curvature measures of nonlinearity”, ***Journal of the Royal Statistical Society***, Ser. B, 42, 1-25, 1980.

[38] Sheiner, L. B., & Beal, S. L., “Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-menten model: Routine clinical pharmacokinetic data", ***Journal of Pharmacokinetics and Biopharmaceutics***, 8, 553-571, 1980.

[39] Mallet, A., Mentre, F., Steimer, J.-L., & Lokiek, F., “Nonparametric maximum likelihood estimation for population pharmacokinetics, with applications to cyclosporine”, ***Journal Pharmacokinetics and Biopharmeutics***, 16, 311-327, 1988.

[40] Wu, H. & Ding A A., “Design of viral dynamic studies for efficiently assessing anti-HIV therapies in AIDS clinical trials”, ***Biom J***, 2, 175–96, 2002.

[41] Wu, H., Kuritzkes, DR., & McClernon, DR. *et al.*, “Characterization of viral dynamics in human immunodeficiency virus type 1–infected patients treated with combination antiretroviral therapy: relationships to host factors, cellular restoration and virological endpoints”, ***J Infect Dis***, 179, 799–807, 1999.

[42] Wu, H. & Wu, L., “Identification of significant host factors for HIV dynamics modelled by nonlinear mixed-effect models”, ***Stat Med,*** 21, 753–71, 2002.

[43] Perelson, A S., & Nelson, P W., “Mathematical analysis of HIV-1 dynamics in vivo”, ***SIAM Rev***., 41, 3–44, 1999.

[44] Wu, H., Ding, A., & DeGruttola, V., "Why are the decay rates in plasma HIV-1 different for different treatments and in different patient populations?", ***AIDS,*** 13, 429–30, 1999.

[45] Bishop, J., "Smooths, Splines and Generalized Additive Models", MT5757 Advanced Data Analysis, St Andrews university, 2007.

[46] Vonesh, E. F., & Carter, R. L., "Mixed-Effects Nonlinear Regression for Unbalanced Repeated Measures," ***Biometrics***, 48, 1–18, 1992.

[47] Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., & Gelfand, A. E., "Bayesian Analysis of Linear and Nonlinear Population Models Using the Gibbs Sampler," ***Applied Statistics***, 1994.

[48] AIDSinfo, a U.S. Department of Health and Human Services (DHHS) project that offers the latest federally approved information on HIV/AIDS clinical research, treatment and prevention, and medical practice guidelines for people living with HIV/AIDS, their families and friends, health care providers, scientists, and researchers.

[49] Hollis, S., & Campbell, F., "What is meant by intention to treat analysis? Survey of published randomised controlled trials", ***British Medical Journal***, 319(7211), 670–674, 1999.

[50] Xiang, D., "Fitting Generalized Additive Models with the GAM Procedure", SAS Institute Inc., Cary, NC.

APPENDIX 1

Chart 2.1: Column chart for classification of type of failure/success at week 24 by sex.

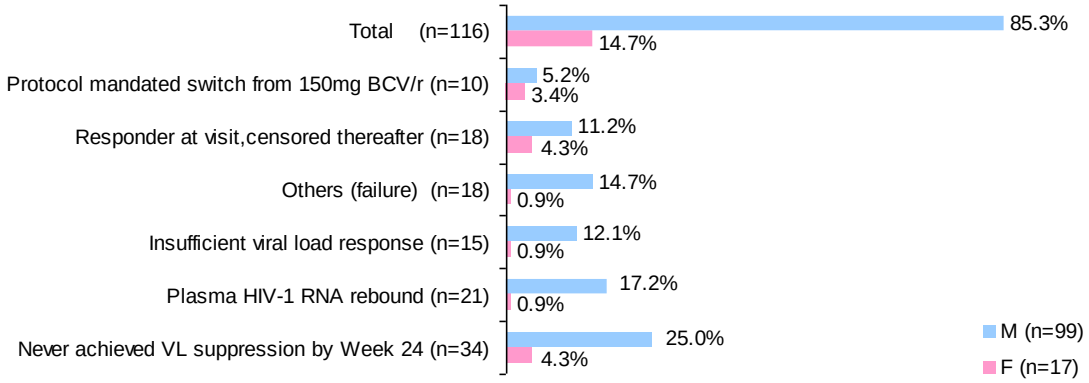

Chart 2.2: Column chart for classification of type of failure/success at week 24 by ethnic origin.

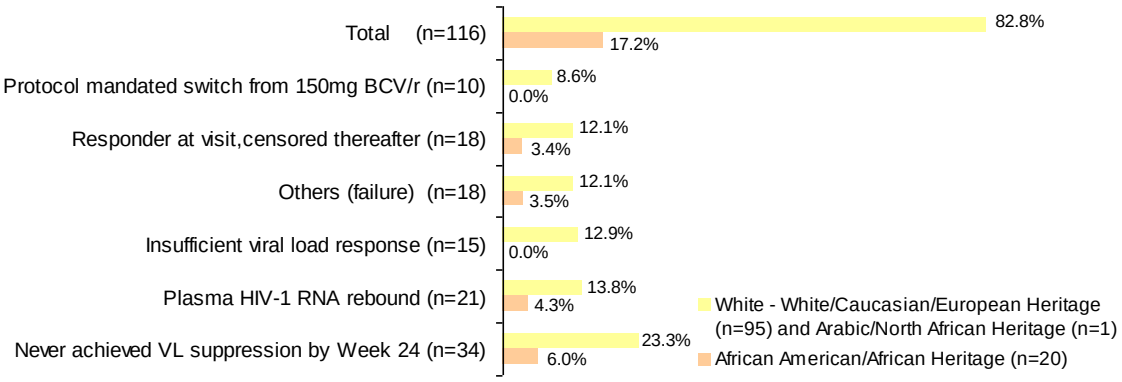

Chart 2.3: Column chart for classification of type of failure/success at week 24 by age group.

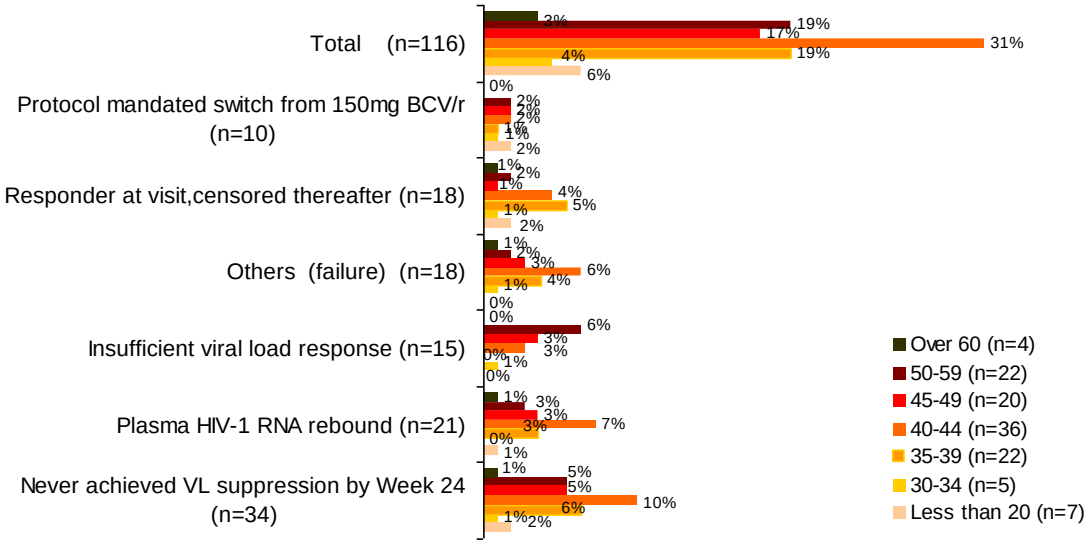

Table 2.1: Iterative Phase table

Iterative Phase			
Iter	**P1**	**d1**	**Sum of Squares**
0	1.0000	1.0000	25802.7
1	10.4882	0.1691	545.8

NOTE: Convergence criterion met.

Table 2.2: Specifications table

Specifications	
Data Set	WORK.DATA16DAYS
Dependent Variable	LBORRLG
Distribution for Dependent Variable	Normal
Random Effects	b1
Distribution for Random Effects	Normal
Subject Variable	SUBJID
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Table 2.3: Fit Statistics table

Fit Statistics	
-2 Log Likelihood	735.0
AIC (smaller is better)	751.0
AICC (smaller is better)	751.3
BIC (smaller is better)	772.4

Table 2.4: Fit Diagnostics for the single-phase model

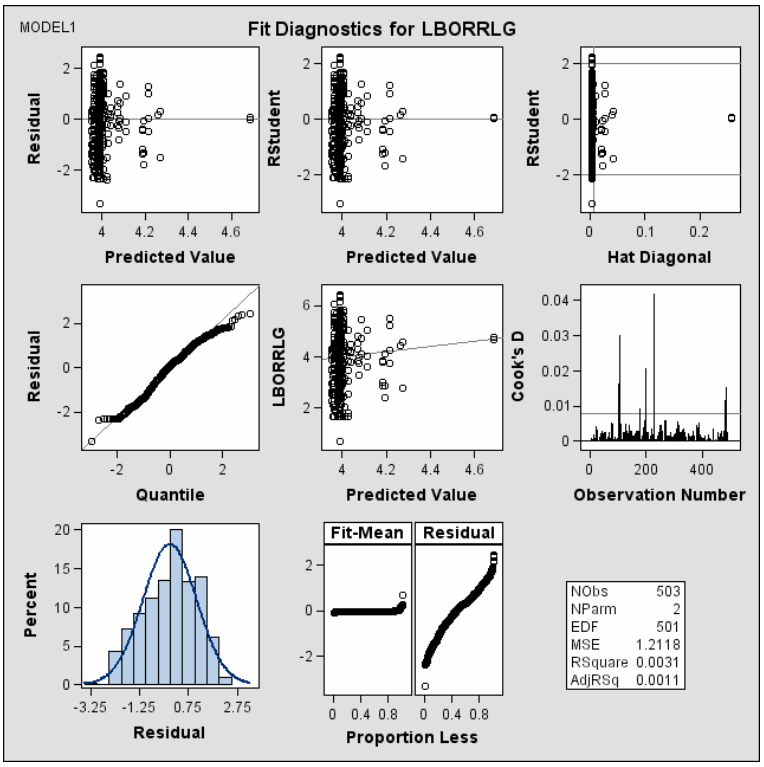

Table 2.5: Fit Diagnostics for the biphasic model

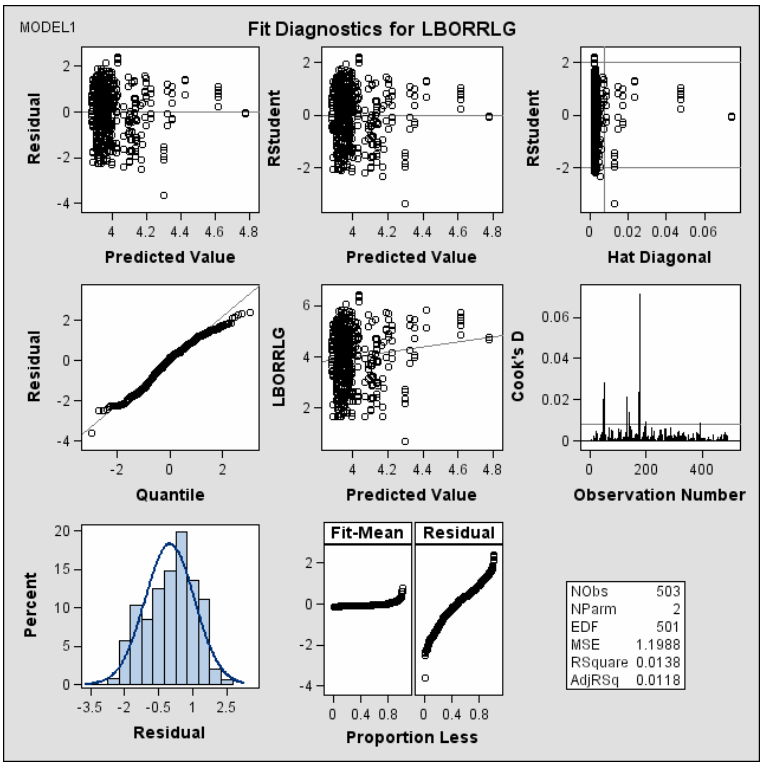

Table 3.1.1: ANOVA table of age variable for the single-phase model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	31	0.44641255	0.01440040	0.82	0.7326
Error	75	1.32453849	0.01766051		
Corrected Total	106	1.77095104			

Table 3.1.2: ANOVA table of age variable for the biphasic model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	31	0.22688636	0.00731891	0.75	0.8115
Error	75	0.73122736	0.00974970		
Corrected Total	106	0.95811372			

Table 3.2.1: ANOVA table of sex variable for the single-phase model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.02164941	0.02164941	1.30	0.2569
Error	105	1.74930163	0.01666002		
Corrected Total	106	1.77095104			

Table 3.2.2: ANOVA table of sex variable for the biphasic model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.01547787	0.01547787	1.72	0.1920
Error	105	0.94263586	0.00897748		
Corrected Total	106	0.95811372			

Table 3.3.1: ANOVA table of ethnicity variable for the single-phase model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.02932355	0.01466178	0.88	0.4197
Error	104	1.74162748	0.01674642		
Corrected Total	106	1.77095104			

Table 3.3.2: ANOVA table of ethnicity variable for the biphasic model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.00958037	0.00479019	0.53	0.5930
Error	104	0.94853335	0.00912051		
Corrected Total	106	0.95811372			

Table 3.4.1: ANOVA table of actual treatment group variable for the single-phase model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.32256405	0.10752135	7.65	0.0001
Error	103	1.44838699	0.01406201		
Corrected Total	106	1.77095104			

Table 3.4.2: ANOVA table of actual treatment group variable for the biphasic model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.16197366	0.05399122	6.99	0.0003
Error	103	0.79614007	0.00772952		
Corrected Total	106	0.95811372			

Table 3.5.1: Wilcoxon Scores (Rank Sums) table for the single-phase model

Wilcoxon Scores (Rank Sums) for Variable Pred Classified by Variable atrtgrp1					
atrtgrp1	**N**	**Sum of Scores**	**Expected Under H0**	**Std Dev Under H0**	**Mean Score**
2	28	1815.0	1512.0	141.095712	64.821429
1	28	1596.0	1512.0	141.095712	57.000000
3	29	1734.0	1566.0	142.681463	59.793103
4	22	633.0	1188.0	129.730490	28.772727

Table 3.5.2: Wilcoxon Scores (Rank Sums) table for the biphasic model

Wilcoxon Scores (Rank Sums) for Variable Pred Classified by Variable atrtgrp1					
atrtgrp1	**N**	**Sum of Scores**	**Expected Under H0**	**Std Dev Under H0**	**Mean Score**
2	28	1775.0	1512.0	141.095712	63.392857
1	28	1661.0	1512.0	141.095712	59.321429
3	29	1695.0	1566.0	142.681463	58.448276
4	22	647.0	1188.0	129.730490	29.409091

Table 3.5.3: Kruskal-Wallis Test for the single-phase model

Kruskal-Wallis Test	
Chi-Square	19.2163
DF	3
Asymptotic Pr > Chi-Square	0.0002
Exact Pr >= Chi-Square	.

Table 3.5.4: Kruskal-Wallis Test for the biphasic model

Kruskal-Wallis Test	
Chi-Square	17.7993
DF	3
Asymptotic Pr > Chi-Square	0.0005
Exact Pr >= Chi-Square	.

Table 3.6.1: Mean of first-phase viral decay rates of dose 1 group for the single-phase model

Basic Statistical Measures			
Location		**Variability**	
Mean	0.182450	**Std Deviation**	0.12727
Median	0.215252	**Variance**	0.01620
Mode	.	**Range**	0.39265
		Interquartile Range	0.22878

Table 3.6.2: Mean of first-phase viral decay rates of dose 2 group for the single-phase model

Basic Statistical Measures			
Location		**Variability**	
Mean	0.212025	**Std Deviation**	0.12417
Median	0.258296	**Variance**	0.01542
Mode	.	**Range**	0.48697
		Interquartile Range	0.19672

Table 3.6.3: Mean of first-phase viral decay rates of dose 3 group for the single-phase model

Basic Statistical Measures			
Location		**Variability**	
Mean	0.193033	**Std Deviation**	0.12603
Median	0.195224	**Variance**	0.01588
Mode	.	**Range**	0.44519
		Interquartile Range	0.19377

Table 3.6.4: Mean of first-phase viral decay rates of dose 4 group for the single-phase model

Basic Statistical Measures			
Location		**Variability**	
Mean	0.062623	**Std Deviation**	0.08451
Median	0.047161	**Variance**	0.00714
Mode	.	**Range**	0.34973
		Interquartile Range	0.10569

Table 3.6.5: Mean of first-phase viral decay rates of dose 1 group for the biphasic model

Basic Statistical Measures			
Location		**Variability**	
Mean	0.113850	**Std Deviation**	0.09562
Median	0.127144	**Variance**	0.00914
Mode	.	**Range**	0.28513
		Interquartile Range	0.16380

Table 3.6.6: Mean of first-phase viral decay rates of dose 2 group for the biphasic model

Basic Statistical Measures			
Location		**Variability**	
Mean	0.127813	**Std Deviation**	0.09049
Median	0.163327	**Variance**	0.00819
Mode	.	**Range**	0.35659
		Interquartile Range	0.14152

Table 3.6.7: Mean of first-phase viral decay rates of dose 3 group for the biphasic model

Basic Statistical Measures			
Location		**Variability**	
Mean	0.109193	**Std Deviation**	0.08857
Median	0.113465	**Variance**	0.00784
Mode	.	**Range**	0.28308
		Interquartile Range	0.14345

Table 3.6.8: Mean of first-phase viral decay rates of dose 4 group for the biphasic model

Basic Statistical Measures			
Location		**Variability**	
Mean	0.022183	**Std Deviation**	0.07190
Median	0.019937	**Variance**	0.00517
Mode	.	**Range**	0.30368
		Interquartile Range	0.05714

Table 3.7.1: Tukey's Studentized Range (HSD) Test for the actual treatment group in the single-phase model

Comparisons significant at the 0.05 level are indicated by ***.				
atrtgrp Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
dose 2 - dose 3	0.01995	-0.01690	0.05681	
dose 2 - dose 1	0.03075	-0.00637	0.06788	
dose 2 - contro	0.15171	0.11133	0.19208	***
dose 3 - dose 2	-0.01995	-0.05681	0.01690	
dose 3 - dose 1	0.01080	-0.02598	0.04759	
dose 3 - contro	0.13175	0.09168	0.17182	***
dose 1 - dose 2	-0.03075	-0.06788	0.00637	
dose 1 - dose 3	-0.01080	-0.04759	0.02598	
dose 1 - contro	0.12095	0.08064	0.16127	***
contro - dose 2	-0.15171	-0.19208	-0.11133	***
contro - dose 3	-0.13175	-0.17182	-0.09168	***
contro - dose 1	-0.12095	-0.16127	-0.08064	***

Table 3.7.2: Tukey's Studentized Range (HSD) Test for the actual treatment group in the biphasic model

Comparisons significant at the 0.05 level are indicated by ***.				
atrtgrp Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
dose 2 - dose 1	0.01547	-0.01296	0.04391	
dose 2 - dose 3	0.01999	-0.00824	0.04822	
dose 2 - contro	0.10704	0.07611	0.13797	***
dose 1 - dose 2	-0.01547	-0.04391	0.01296	
dose 1 - dose 3	0.00452	-0.02365	0.03269	
dose 1 - contro	0.09157	0.06069	0.12245	***
dose 3 - dose 2	-0.01999	-0.04822	0.00824	
dose 3 - dose 1	-0.00452	-0.03269	0.02365	
dose 3 - contro	0.08705	0.05636	0.11774	***
contro - dose 2	-0.10704	-0.13797	-0.07611	***
contro - dose 1	-0.09157	-0.12245	-0.06069	***
contro - dose 3	-0.08705	-0.11774	-0.05636	***

Table 3.8.1: Type III sums of squares table from GLM analysis for the single-phase model

Source	DF	Type III SS	Mean Square	F Value	Pr > F
AGE	31	0.21076183	0.00679877	0.59	0.9463
SEX	1	0.00224574	0.00224574	0.19	0.6606
RACE	2	0.02524211	0.01262105	1.09	0.3412
CD4__BLC	1	0.02114137	0.02114137	1.83	0.1806
LBORRLG	1	0.02894062	0.02894062	2.51	0.1182
atrtgrp	3	0.13199133	0.04399711	3.81	0.0140
MULPICD2	1	0.11675334	0.11675334	10.11	0.0023
ACT20GCD	1	0.01048424	0.01048424	0.91	0.3441

Table 3.8.2: Type III sums of squares table from GLM analysis for the biphasic model

Source	DF	Type III SS	Mean Square	F Value	Pr > F
AGE	31	0.09743283	0.00314299	0.56	0.9624
SEX	1	0.00065033	0.00065033	0.12	0.7355
RACE	2	0.00933769	0.00466885	0.83	0.4421
CD4__BLC	1	0.01232009	0.01232009	2.18	0.1445
LBORRLG	1	0.04677750	0.04677750	8.28	0.0054
atrtgrp	3	0.06022835	0.02007612	3.55	0.0190
MULPICD2	1	0.05811690	0.05811690	10.29	0.0021
ACT20GCD	1	0.00364671	0.00364671	0.65	0.4246

Table 3.9.1: Spearman Correlation Coefficients table of first-phase viral decay rates and log10 baseline RNA for the single-phase model

Spearman Correlation Coefficients, N = 107 Prob > |r| under H0: Rho=0		
	Pred	**LBORRLG**
Pred	1.00000	-0.29217 0.0023
LBORRLG	-0.29217 0.0023	1.00000

Table 3.9.2: Spearman Correlation Coefficients table of first-phase viral decay rates and log10 baseline RNA for the biphasic model

Spearman Correlation Coefficients, N = 107 Prob > |r| under H0: Rho=0		
	Pred	**LBORRLG**
Pred	1.00000	-0.41440 <.0001
LBORRLG	-0.41440 <.0001	1.00000

Table 3.10.1: Spearman Correlation Coefficients table of first-phase viral decay rates and CD4__BLC for the single-phase model

Spearman Correlation Coefficients, N = 107 Prob > |r| under H0: Rho=0		
	Pred	**CD4__BLC**
Pred	1.00000	0.42794 <.0001
CD4__BLC	0.42794 <.0001	1.00000

Table 3.10.2: Spearman Correlation Coefficients table of first-phase viral decay rates and CD4__BLC for the biphasic model

Spearman Correlation Coefficients, N = 107 Prob > |r| under H0: Rho=0		
	Pred	**CD4__BLC**
Pred	1.00000	0.48238 <.0001
CD4__BLC	0.48238 <.0001	1.00000

Table 3.11.1: Spearman Correlation Coefficients table of first-phase viral decay rates and week 1 log10 RNA change from baseline for the single-phase model

Spearman Correlation Coefficients, N = 94 Prob > |r| under H0: Rho=0		
	Pred	**LBORLGCW**
Pred	1.00000	0.86714 <.0001
LBORLGCW	0.86714 <.0001	1.00000

Table 3.11.2: Spearman Correlation Coefficients table of first-phase viral decay rates and week 1 log10 RNA change from baseline for the biphasic model

Spearman Correlation Coefficients, N = 94 Prob > |r| under H0: Rho=0		
	Pred	**LBORLGCW**
Pred	1.00000	0.85162 <.0001
LBORLGCW	0.85162 <.0001	1.00000

Table 3.12.1: Spearman Correlation Coefficients table for the single-phase model

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Pred	**LBORLGCW1_S**
Pred	1.00000 94	0.41533 0.0003 72
LBORLGCW1_S	0.41533 0.0003 72	1.00000 80

Table 3.12.2: Spearman Correlation Coefficients table for the biphasic model

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Pred	**LBORLGCW1_S**
Pred	1.00000 94	0.34594 0.0029 72
LBORLGCW1_S	0.34594 0.0029 72	1.00000 80

Table 3.12.3: Spearman Correlation Coefficients table for the single-phase model

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Pred	**LBORLGCW1_S**
Pred	1.00000 94	0.43161 0.0022 48
LBORLGCW1_S	0.43161 0.0022 48	1.00000 51

Table 3.12.4: Spearman Correlation Coefficients table for the biphasic model

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	Pred	**LBORLGCW1_S**
Pred	1.00000 94	0.39416 0.0056 48
LBORLGCW1_S	0.39416 0.0056 48	1.00000 51

Table 3.13.1: Spearman Correlation Coefficients table for the single-phase model

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	LBORLGCW_S	**LBORLGCW1_S**
LBORLGCW_S	1.00000 94	0.44488 <.0001 72
LBORLGCW1_S	0.44488 <.0001 72	1.00000 80

Table 3.13.2: Spearman Correlation Coefficients table for the biphasic model

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	LBORLGCW_S	**LBORLGCW1_S**
LBORLGCW_S	1.00000 94	0.44488 <.0001 72
LBORLGCW1_S	0.44488 <.0001 72	1.00000 80

Table 3.13.3: Spearman Correlation Coefficients table for the single-phase model

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	LBORLGCW_S	**LBORLGCW1_S**
LBORLGCW_S	1.00000 94	0.41967 0.0030 48
LBORLGCW1_S	0.41967 0.0030 48	1.00000 51

Table 3.13.4: Spearman Correlation Coefficients table for the biphasic model

Spearman Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	LBORLGCW_S	**LBORLGCW1_S**
LBORLGCW_S	1.00000 94	0.41967 0.0030 48
LBORLGCW1_S	0.41967 0.0030 48	1.00000 51

Table 3.14.1: Mean of viral load at 20 week – higher first-phase viral decay group for the single-model

Moments			
N	44	**Sum Weights**	44
Mean	2.74250455	**Sum Observations**	120.6702
Std Deviation	1.17205848	**Variance**	1.37372108
Skewness	0.86309756	**Kurtosis**	-0.0578427
Uncorrected SS	390.008578	**Corrected SS**	59.0700063
Coeff Variation	42.7367926	**Std Error Mean**	0.17669446

Table 3.14.2: Mean of viral load at 20 week – lower first-phase viral decay group for the single-model

Moments			
N	36	**Sum Weights**	36
Mean	3.70135	**Sum Observations**	133.2486
Std Deviation	1.16427539	**Variance**	1.35553719
Skewness	-0.3718013	**Kurtosis**	-0.6894897
Uncorrected SS	540.643507	**Corrected SS**	47.4438016
Coeff Variation	31.4554255	**Std Error Mean**	0.1940459

Table 3.14.3: Mean of viral load at 24 week – higher first-phase viral decay group for the single-model

Moments			
N	29	**Sum Weights**	29
Mean	2.55901034	**Sum Observations**	74.2113
Std Deviation	0.99615883	**Variance**	0.99233242
Skewness	0.71551408	**Kurtosis**	-0.8477826
Uncorrected SS	217.692792	**Corrected SS**	27.7853077
Coeff Variation	38.9275031	**Std Error Mean**	0.18498205

Table 3.14.4: Mean of viral load at 24 week – lower first-phase viral decay group for the single-model

Moments			
N	22	**Sum Weights**	22
Mean	3.51029091	**Sum Observations**	77.2264
Std Deviation	1.31322081	**Variance**	1.72454888
Skewness	-0.2112281	**Kurtosis**	-1.4838382
Uncorrected SS	307.302656	**Corrected SS**	36.2155265
Coeff Variation	37.4105976	**Std Error Mean**	0.27997962

Table 3.14.5: Mean of viral load at 20 week – higher first-phase viral decay group for the biphasic model

Basic Statistical Measures			
Location		**Variability**	
Mean	2.721598	**Std Deviation**	1.16585
Median	2.152300	**Variance**	1.35920
Mode	1.690200	**Range**	4.51390
		Interquartile Range	2.17430

Table 3.14.6: Mean of viral load at 20 week – lower first-phase viral decay group for the biphasic model

Basic Statistical Measures			
Location		**Variability**	
Mean	3.818294	**Std Deviation**	1.10222
Median	3.944500	**Variance**	1.21489
Mode	1.690200	**Range**	3.98470
		Interquartile Range	1.30390

Table 3.14.7: Mean of viral load at 24 week – higher first-phase viral decay group for the biphasic model

Basic Statistical Measures			
Location		**Variability**	
Mean	2.502958	**Std Deviation**	0.98654
Median	1.690200	**Variance**	0.97325
Mode	1.690200	**Range**	3.00260
		Interquartile Range	1.57930

Table 3.14.8: Mean of viral load at 24 week – lower first-phase viral decay group for the biphasic model

Basic Statistical Measures			
Location		**Variability**	
Mean	3.692300	**Std Deviation**	1.23390
Median	4.039300	**Variance**	1.52250
Mode	1.690200	**Range**	3.67340
		Interquartile Range	2.00860

Table 3.15.1: Wilcoxon Scores (Rank Sums) table for viral decay rates and long-term response for the single-model

Wilcoxon Scores (Rank Sums) for Variable Pred Classified by Variable P400_TLO					
P400_TLO	**N**	**Sum of Scores**	**Expected Under H0**	**Std Dev Under H0**	**Mean Score**
0	81	3762.0	4374.0	137.673527	46.444444
1	26	2016.0	1404.0	137.673527	77.538462

Table 3.15.2: Wilcoxon Scores (Rank Sums) test for viral decay rates and long-term response table for the single-phase model

Wilcoxon Two-Sample Test	
Statistic (S)	2016.0000
Normal Approximation	
Z	4.4417
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001
t Approximation	
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001
Exact Test	
One-Sided Pr >= S	1.863E-06
Two-Sided Pr >= |S - Mean|	3.726E-06
Z includes a continuity correction of 0.5.	

Table 3.15.3: Kruskal-Wallis test for viral decay rates and long-term response for the single-phase model

Kruskal-Wallis Test	
Chi-Square	19.7607
DF	1
Pr > Chi-Square	<.0001

Table 3.15.4: Wilcoxon Scores (Rank Sums) test for viral decay rates and long-term response for the biphasic model

Wilcoxon Scores (Rank Sums) for Variable Pred Classified by Variable P400_TLO					
P400_TLO	**N**	**Sum of Scores**	**Expected Under H0**	**Std Dev Under H0**	**Mean Score**
0	81	3729.0	4374.0	137.673527	46.037037
1	26	2049.0	1404.0	137.673527	78.807692

Table 3.15.5: Wilcoxon Scores (Rank Sums) test for viral decay rates and long-term response for the biphasic model

Wilcoxon Two-Sample Test	
Statistic (S)	2049.0000
Normal Approximation	
Z	4.6814
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001
t Approximation	
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001
Exact Test	
One-Sided Pr >= S	4.639E-07
Two-Sided Pr >= |S - Mean|	9.278E-07
Z includes a continuity correction of 0.5.	

Table 3.15.6: Kruskal-Wallis test for viral decay rates and long-term response for the biphasic model

Kruskal-Wallis Test	
Chi-Square	21.9492
DF	1
Pr > Chi-Square	<.0001

Table 3.16.1: Confusion Matrix for the single-phase model

	Responder	Non-responder	Total
Correct	**20**	51	71
False	30	**6**	36
Total	50	57	107

Table 3.16.2: Confusion Matrix for the biphasic model

	Responder	Non-responder	Total
Correct	**21**	49	70
False	32	**5**	37
Total	53	54	107

Table 3.17.1: Wilcoxon Scores (Rank Sums) table for week 1 virus load reduction and long-term response for the single-phase model

Wilcoxon Scores (Rank Sums) for Variable LBORLGCW_S Classified by Variable P400_TLO					
P400_TLO	**N**	**Sum of Scores**	**Expected Under H0**	**Std Dev Under H0**	**Mean Score**
0	70	2864.0	3325.0	115.325626	40.914286
1	24	1601.0	1140.0	115.325626	66.708333

Table 3.17.2: Wilcoxon Scores (Rank Sums) test for week 1 virus load reduction and long-term response for the single-phase model

Wilcoxon Two-Sample Test	
Statistic (S)	1601.0000
Normal Approximation	
Z	3.9930
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001
t Approximation	
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	0.0001
Exact Test	
One-Sided Pr >= S	1.813E-05
Two-Sided Pr >= |S - Mean|	3.626E-05
Z includes a continuity correction of 0.5.	

Table 3.17.3: Kruskal-Wallis test for week 1 virus load reduction and long-term response for the single-phase model

Kruskal-Wallis Test	
Chi-Square	15.9790
DF	1
Pr > Chi-Square	<.0001

Table 3.17.4: Wilcoxon Scores (Rank Sums) table for week 1 virus load reduction and long-term response for the biphasic model

Wilcoxon Scores (Rank Sums) for Variable LBORLGCW_S Classified by Variable P400_TLO					
P400_TLO	**N**	**Sum of Scores**	**Expected Under H0**	**Std Dev Under H0**	**Mean Score**
0	70	2864.0	3325.0	115.325626	40.914286
1	24	1601.0	1140.0	115.325626	66.708333

Table 3.17.5: Wilcoxon Scores (Rank Sums) test for week 1 virus load reduction and long-term response for the biphasic model

Wilcoxon Two-Sample Test	
Statistic (S)	1601.0000
Normal Approximation	
Z	3.9930
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001
t Approximation	
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	0.0001
Exact Test	
One-Sided Pr >= S	1.813E-05
Two-Sided Pr >= |S - Mean|	3.626E-05
Z includes a continuity correction of 0.5.	

Table 3.17.6: Kruskal-Wallis test for week 1 virus load reduction and long-term response for the biphasic model

Kruskal-Wallis Test	
Chi-Square	15.9790
DF	1
Pr > Chi-Square	<.0001

Table 3.18.1: Maximum Likelihood Analysis of Variance table for the single-phase model

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
Intercept	1	20.64	<.0001
atrtgrp	3	4.42	0.2192
Likelihood Ratio	0	.	.

Table 3.18.2: Analysis of Maximum Likelihood Estimates for the single-phase model

Analysis of Maximum Likelihood Estimates					
Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
Intercept		1.4183	0.3121	20.64	<.0001
atrtgrp	**contro**	1.6262	0.7882	4.26	0.0391
	dose 1	-0.5020	0.4300	1.36	0.2431
	dose 2	-0.6711	0.4234	2.51	0.1130

Table 3.18.3: Maximum Likelihood Analysis of Variance table for the biphasic model

Maximum Likelihood Analysis of Variance			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
Intercept	1	20.64	<.0001
atrtgrp	3	4.42	0.2192
Likelihood Ratio	0	.	.

Table 3.18.4: Analysis of Maximum Likelihood Estimates for the biphasic model

Analysis of Maximum Likelihood Estimates					
Parameter		**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
Intercept		1.4183	0.3121	20.64	<.0001
atrtgrp	**contro**	1.6262	0.7882	4.26	0.0391
	dose 1	-0.5020	0.4300	1.36	0.2431
	dose 2	-0.6711	0.4234	2.51	0.1130

APPENDIX 2

< SAS code for the analyses >

```
/* Dataset1 Importing */
Proc IMPORT OUT= viral
            DataFILE= "C:\thesis\viral.xls"
            DBMS=EXCEL REPLACE;
     SHEET="viral$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;

/* Dataset2 Importing */
Proc IMPORT OUT= base_cov
            DataFILE= "C:\thesis\base_cov.xls"
            DBMS=EXCEL REPLACE;
     SHEET="base_cov$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;

/* Dataset3 Importing */
Proc IMPORT OUT= prop
            DataFILE= "C:\thesis\prop.xls"
            DBMS=EXCEL REPLACE;
     SHEET="prop$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;

/* Making merged dataset replicating subject observations*/
Data test;
Merge  base_cov(keep= MULPICD2 CD4__BLC CD4__BLQ ACT20GCD subjid)
       prop(keep= P400_OBS P400_TLO r400_tlc subjid)
       viral;
by subjid;
Run;

/* Age grouping */
Data test;
Set test;
If (AGE < 20) then AGE1 = 1;
If (AGE >= 30 AND AGE =< 34) then AGE1 = 2;
If (AGE >= 35 AND AGE =< 39) then AGE1 = 3;
If (AGE >= 40 AND AGE =< 44) then AGE1 = 4;
If (AGE >= 45 AND AGE =< 49) then AGE1 = 5;
If (AGE >= 50 AND AGE =< 59) then AGE1 = 6;
If (AGE >= 60) then AGE1 = 7;
Run;


/* Data manipulation for excluding missing cases of response variable
   and dropping pre-treatment measurements */
Data test1;
Set test;
If (LBACTDY < 0) then delete;
If (KEEPLOD < 1) then delete;
If (LBEVFLG < 1) then delete;
Where ATTYPE= 'Treatment';
Run;
```

```
/* Data manipulation for excluding observations over 24 weeks */
Data Data24weeks;
Set test1;
If LBACTDY > 170 then delete;
Run;


/* Data excluding who had no initial viral decline (seriously) */
Data test2;
Set Data24weeks;
If (SUBJID=61) then delete;
If (SUBJID=72) then delete;
If (SUBJID=223) then delete;
If (SUBJID=512) then delete;
If (SUBJID=1249) then delete;
If (SUBJID=1300) then delete;
If (SUBJID=1530) then delete;
If (SUBJID=1665) then delete;
If (SUBJID=1817) then delete;
Run;


/* Data manipulation including Data by 16 days(2 weeks) */
Data Data16days;
Set test2;
If LBACTDY > 16 then delete;
Run;


ods rtf;
/* A single-phase nonlinear decay Model using nlmixed
   to drive viral decay rates for 16 days (2 weeks)*/
/* Getting the initial values without fitting random effects using nlin */
Proc nlin Data=data16days;
Parms P1=1 d1=1;
Model LBORRLG = log10(exp(p1)*exp(-d1*LBACTDY));
Run;

/* Fitting nlmixed with one random effect to get the initial parameter values */
Proc nlmixed Data=data16days;
Parms P1=10.488 d1=0.169 error=1.089 varcomp11=1;
u1=d1+b1;
expected = log10(exp(P1)*exp(-u1*LBACTDY));
Model LBORRLG ~ normal(expected, error**2);
      Random b1 ~normal([0], [varcomp11]) subject=SUBJID;
      Predict u1 out= ratepredic1_1_0;
Run;

/* Fitting nlmixed with full random effect of a single-phase model using the initial
values */
 Proc nlmixed Data=Data16days;
 Parms P1=10.476
       d1=0.167
       error=0.541
       varcomp11=0.05
       varcomp12=0.01 to 1 by 0.01
       varcomp22=0.01 to 1 by 0.01;
u1=d1+b1;
u2=P1+b2;
expected = log10(exp(u2)*exp(-u1*LBACTDY));
Model LBORRLG ~ normal(expected, error**2);
      Random b1 b2 ~normal([0,0], [varcomp11,varcomp12,varcomp22]) subject=SUBJID;
      Predict u1 out= ratepredic1_1;
      Predict u2 out= ratepredic1_2;
Run;
ods rtf close;


/* A bi-phasic nonlinear decay Model using nlmixed
   to drive viral decay rates by 16 days (2 weeks) */
/* Getting the initial values without fitting random effects using nlin */
Proc nlin Data=data16days;
Parms P1=10.476 P2=1 d1=0.169 d2=1;
Model LBORRLG = log10(exp(P1)*exp(-d1*LBACTDY)+exp(P2)*exp(-d2*LBACTDY));
Run;
```

```
/* Data manipulation to get initial values for nlmixed model */
data Data16days1;
set Data16days;
where atrtgrp='dose 1' or atrtgrp='dose 2' or atrtgrp='dose 3';
run;
/* Fitting nlmixed with two random effects of biphasic model using the initial values from
nlin */
ods rtf;
Proc nlmixed Data=Data16days1;
Parms P1=9.024 P2=11.153 d1=0.054 d2=1 error=0.509 varcomp11=1.08 varcomp13=1
varcomp33=1;
u1=d1+b1;
u3=P1+b3;
expected = log10(exp(u3)*exp(-u1*LBACTDY)+exp(P2)*exp(-d2*LBACTDY));
Model LBORRLG ~ normal(expected, error**2);
      Random b1 b3 ~normal([0,0],
      [varcomp11,varcomp13,varcomp33]) subject=SUBJID;
      Predict u1 out= ratepredic2_1_0;
          Predict u3 out= ratepredic2_3_0;
Run;
ods rtf close;
/* Fitting nlmixed with three random effects but without covariance of biphasic model
   using the estimated values from the previous nlmixed model */
ods rtf;
Proc nlmixed Data=data16days;
Parms P1=10.1 P2=9.62 d1=0.169 d2=0.885 error=0.2868
      varcomp11=0.019
      varcomp22=0.01 to 1 by 0.01
      varcomp33=4.77;
u1=d1+b1;
u2=d2+b2;
u3=P1+b3;
expected = log10(exp(u3)*exp(-u1*LBACTDY)+exp(P2)*exp(-u2*LBACTDY));
Model LBORRLG ~ normal(expected, error**2);
      Random b1 b2 b3 ~normal([0,0,0],
      [varcomp11,0 ,varcomp22 ,0 ,0 ,varcomp33]) subject=SUBJID;
      Predict u1 out= ratepredic2_1_1_0;
          Predict u2 out= ratepredic2_2_1_0;
      Predict u3 out= ratepredic2_3_3_0;
Run;
ods rtf close;
/* Fitting nlmixed with three random effect and covariances of biphasic model
   using the estimated values from the previous nlmixed model */
ods rtf;
Proc nlmixed Data=data16days;
Parms P1=10.1 P2=9.62 d1=0.169 d2=0.885 error=0.2868
      varcomp11=0.019
      varcomp22=0.01 to 1 by 0.01
      varcomp33=4.77
      varcomp13=0.000001
      varcomp23=0.000001
      varcomp12=0.000001;
u1=d1+b1;
u2=d2+b2;
u3=P1+b3;
expected = log10(exp(u3)*exp(-u1*LBACTDY)+exp(P2)*exp(-u2*LBACTDY));
Model LBORRLG ~ normal(expected, error**2);
      Random b1 b2 b3 ~normal([0,0,0],
      [varcomp11,varcomp12,varcomp22,varcomp13,varcomp23,varcomp33]) subject=SUBJID;
      Predict u1 out= ratepredic2_1_1;
          Predict u2 out= ratepredic2_2_2;
      Predict u3 out= ratepredic2_3_3;
Run;
ods rtf close;


/* Changing into a numeric value */
Data RatePredic1_11;
Set RatePredic1_1;
If SEX='M' then SEX1=1;
If SEX='F' then SEX1=2;
If RACE='White - White/Caucasian/European Heritage' then RACE1=1;
If RACE='White - Arabic/North African Heritage' then RACE1=1;
If RACE='African American/African Heritage' then RACE1=2;
If atrtgrp='dose 1' then atrtgrp1=1;
If atrtgrp='dose 2' then atrtgrp1=2;
If atrtgrp='dose 3' then atrtgrp1=3;
```

```
If atrtgrp='contro' then atrtgrp1=4;
Run;

/* Changing into a numeric value */
Data RatePredic2_11;
Set RatePredic2_1_1;
If SEX='M' then SEX1=1;
If SEX='F' then SEX1=2;
If RACE='White - White/Caucasian/European Heritage' then RACE1=1;
If RACE='White - Arabic/North African Heritage' then RACE1=1;
If RACE='African American/African Heritage' then RACE1=2;
If atrtgrp='dose 1' then atrtgrp1=1;
If atrtgrp='dose 2' then atrtgrp1=2;
If atrtgrp='dose 3' then atrtgrp1=3;
If atrtgrp='contro' then atrtgrp1=4;
Run;


/* Changing into a numeric value */
Data RatePredic2_22;
Set RatePredic2_2_2;
If SEX='M' then SEX1=1;
If SEX='F' then SEX1=2;
If RACE='White - White/Caucasian/European Heritage' then RACE1=1;
If RACE='White - Arabic/North African Heritage' then RACE1=1;
If RACE='African American/African Heritage' then RACE1=2;
If atrtgrp='dose 1' then atrtgrp1=1;
If atrtgrp='dose 2' then atrtgrp1=2;
If atrtgrp='dose 3' then atrtgrp1=3;
If atrtgrp='contro' then atrtgrp1=4;
Run;


/* Dataset4 Importing contains results of predicted d1,baseline RNA
   for each subject from the single-phase model */
Proc IMPORT OUT= pred11
            DataFILE= "C:\thesis2\pred11.xls"
            DBMS=EXCEL REPLACE;
     SHEET="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;


/* Dataset5 Importing contains results of predicted d1,baseline RNA
   for each subject from the biphasic model */
Proc IMPORT OUT= pred2_11
            DataFILE= "C:\thesis2\pred2_11.xls"
            DBMS=EXCEL REPLACE;
     SHEET="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;

/* Dataset6 Importing contains results of predicted d2,baseline RNA
   for each subject from the biphasic model */
Proc IMPORT OUT= pred2_22
            DataFILE= "C:\thesis2\pred2_22.xls"
            DBMS=EXCEL REPLACE;
     SHEET="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;
```

```
/* Spearman's rank tests for initial viral decay rates(d1) and log10 baseline RNA  */
ods rtf;
Proc corr data=Pred11 spearman plots;
      var Pred LBORRLG;
   run;
Proc corr data=Pred2_11 spearman plots;
      var Pred LBORRLG;
   run;
ods rtf close;


/* Plot of initial viral decay rates(d1) and log10 baseline RNA
   by actual treatment groups for the single-phase model */
ods rtf;
TITLE height=10pt "r=-0.292, p<.0001";
Proc gplot Data=Pred11;
      symbol v=dot h=0.5;
      LABEL Pred='RNA decay rate,d1';
      LABEL LBORRLG='Log10 baseline RNA, copies/mL';
      plot Pred * LBORRLG=ATRTGRP;
   Run;
/* Plot of initial viral decay rates(d1) and log10 baseline RNA
   by actual treatment groups for the biphasic model */
TITLE height=10pt "r=-0.414, p<.0001";
Proc gplot Data=Pred2_11;
      symbol v=dot h=0.5;
      LABEL Pred='RNA decay rate,d1';
      LABEL LBORRLG='Log10 baseline RNA, copies/mL';
      plot Pred * LBORRLG=ATRTGRP;
   Run;
ods rtf close;


/* Dataset7 Importing contains results of predicted d1, week 1 log10 RNA change from
baseline for each subject from the single-phase model */
Proc IMPORT OUT= pred11_change
            DataFILE= "C:\thesis2\pred11_change.xls"
            DBMS=EXCEL REPLACE;
     SHEET="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;

/* Dataset8 Importing contains results of predicted d1, week 1 log10 RNA change from
baseline for each subject from the biphasic model */
Proc IMPORT OUT= pred2_11_change
            DataFILE= "C:\thesis2\pred2_11_change.xls"
            DBMS=EXCEL REPLACE;
     SHEET="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;


/* Spearman's rank tests for initial viral decay rates(d1) and week 1 log10 RNA change
from baseline  */
ods rtf;
Proc corr data=Pred11_change spearman plots;
      var Pred LBORLGCW_S;
   run;
Proc corr data=Pred2_11_change spearman plots;
      var Pred LBORLGCW_S;
   run;
ods rtf close;


/* Dataset9 Importing contains results of week 20 log10 RNA change from baseline */
Proc IMPORT OUT= d20week
            DataFILE= "C:\thesis2\d20week.xls"
            DBMS=EXCEL REPLACE;
     SHEET="sheet1$";
```

```
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;

/* Making merged dataset replicating subject observations */
Data Pred11_change_1;
Merge  d20week(keep= LBORLGCW1_S subjid)
       Pred11_change;
by subjid;
Run;
Data Pred2_11_change_1;
Merge  d20week(keep= LBORLGCW1_S subjid)
       Pred2_11_change;
by subjid;
Run;

/* Spearman's rank tests for week 1 log10 RNA change from baseline and week 20 log10 RNA
change from baseline */
ods rtf;
Proc corr data=Pred11_change_1 spearman plots;
      var LBORLGCW_S LBORLGCW1_S;
   run;
Proc corr data=Pred2_11_change_1 spearman plots;
      var LBORLGCW_S LBORLGCW1_S;
   run;
ods rtf close;


/* Spearman's rank tests for initial viral decay rates and week 20 log10 RNA change from
baseline */
ods rtf;
Proc corr data=Pred11_change_1 spearman plots;
      var pred LBORLGCW1_S;
   run;
Proc corr data=Pred2_11_change_1 spearman plots;
      var Pred LBORLGCW1_S;
   run;
ods rtf close;

/* Dataset10 Importing contains results of week 24 log10 RNA change from baseline */
Proc IMPORT OUT= d24week
            DataFILE= "C:\thesis2\d24week.xls"
            DBMS=EXCEL REPLACE;
     SHEET="sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;

/* Making merged dataset replicating subject observations */
Data Pred11_change_11;
Merge  d24week(keep= LBORLGCW1_S subjid)
       Pred11_change;
by subjid;
Run;
Data Pred2_11_change_11;
Merge  d24week(keep= LBORLGCW1_S subjid)
       Pred2_11_change;
by subjid;
Run;

/* Spearman's rank tests for week 1 log10 RNA change from baseline and week 24 log10 RNA
change from baseline */
ods rtf;
Proc corr data=Pred11_change_11 spearman plots;
      var LBORLGCW_S LBORLGCW1_S;
   run;
Proc corr data=Pred2_11_change_11 spearman plots;
      var LBORLGCW_S LBORLGCW1_S;
   run;
ods rtf close;
```

```
/* Spearman's rank tests for initial viral decay rates and week 24 log10 RNA change from
baseline */
ods rtf;
Proc corr data=Pred11_change_11 spearman plots;
      var pred LBORLGCW1_S;
   run;
Proc corr data=Pred2_11_change_11 spearman plots;
      var Pred LBORLGCW1_S;
   run;
ods rtf close;



/* Plot of initial viral decay rates(d1) and week 1 log10 RNA change from baseline
   by actual treatment groups for the single-phase model */
ods rtf;
TITLE height=10pt "r=0.867, p<.0001";
Proc gplot Data=Pred11_change;
      symbol v=dot h=0.5;
      LABEL Pred='RNA decay rate,d1';
      LABEL LBORLGCW_S='Week 1 Log10 RNA change from baseline, copies/mL';
      plot Pred * LBORLGCW_S=ATRTGRP;
   Run;
/* Plot of initial viral decay rates(d1) and week 1 log10 RNA change from baseline
   by actual treatment groups for the biphasic model */
TITLE height=10pt "r=0.852, p<.0001";
Proc gplot Data=Pred2_11_change;
      symbol v=dot h=0.5;
      LABEL Pred='RNA decay rate,d1';
      LABEL LBORLGCW_S='Week 1 Log10 RNA change from baseline, copies/mL';
      plot Pred * LBORLGCW_S=ATRTGRP;
   Run;
ods rtf close;



/* Spearman's rank tests for the initial viral decay rates(d1) and CD4__BLC */
ods rtf;
Proc corr data=Pred11 spearman plots;
      var Pred CD4__BLC;
   run;
Proc corr data=Pred2_11 spearman plots;
      var Pred CD4__BLC;
   run;
ods rtf close;


ods rtf;
/* ANOVA for age effect on d1 decay rates */
Proc anova data=pred11;
class age;
model pred=age;
run;
Proc anova data=pred2_11;
class age;
model pred=age;
run;
/* ANOVA for gender effect on d1 decay rates */
Proc anova data=pred11;
class sex;
model pred=sex;
run;
Proc anova data=pred2_11;
class sex;
model pred=sex;
/* ANOVA for ethnicity effect on d1 decay rates */
Proc anova data=pred11;
class race;
model pred=race;
run;
Proc anova data=pred2_11;
class race;
model pred=race;
run;
/* ANOVA for actual treatment group effect on d1 decay rates */
Proc anova data=pred11;
class atrtgrp;
```

```
model pred=atrtgrp;
run;
Proc anova data=pred2_11;
class atrtgrp;
model pred=atrtgrp;
run;
ods rtf close;


/* A univariate regression analysis for actual treatment, age, gender, and ethnicity
groups */
Proc reg data=pred11;
model pred=atrtgrp1; run;
Proc reg data=pred2_11;
model pred=atrtgrp1; run;
Proc reg data=pred11;
model pred=AGE; run;
Proc reg data=pred2_11;
model pred=AGE; run;
Proc reg data=pred11;
model pred=SEX1; run;
Proc reg data=pred2_11;
model pred=SEX1; run;
Proc reg data=pred11;
model pred=RACE1; run;
Proc reg data=pred2_11;
model pred=RACE1; run;


/* Npar1way for examining relationship between the initial decay rates and actual
treatment group */
ods rtf;
Proc npar1way wilcoxon data=Pred11;
      class atrtgrp1 ;
      var Pred; /* response variable */
   run;
Proc npar1way wilcoxon data=Pred2_11;
      class  atrtgrp1;
      var Pred; /* response variable */
   run;
ods rtf close;

/* Mean of each actual treatment group */
Proc sort data=pred11;
by atrtgrp1; run;
ods rtf;
Proc univariate data=pred11;
var pred;
by atrtgrp1; run;
ods rtf close;

Proc sort data=pred2_11;
by atrtgrp1; run;
ods rtf;
Proc univariate data=pred2_11;
var pred;
by atrtgrp1; run;
ods rtf close;


/* GLM for examining relationship between the initial decay rates and actual treatment
group */
ods rtf;
Proc glm Data=Ratepredic1_11;
      class SUBJID ATRTGRP;
      Model Pred= ATRTGRP;
      manova h=_all_ / /* printe printh */;
      means ATRTGRP/CLDIfF bon tukey;
   run;
Proc glm Data=Ratepredic2_11;
      class SUBJID ATRTGRP;
      Model Pred= ATRTGRP;
      manova h= all  / /* printe printh */;
      means ATRTGRP/CLDIfF bon tukey;
   Run;
ods rtf close;
```

```
/* GLM for identifying baseline characteristics which are correlated with decay rates */
ods rtf;
Proc glm Data=Pred11;
      class SUBJID AGE SEX RACE ATRTGRP ACT20GCD MULPICD2 ;
      Model Pred= AGE SEX RACE CD4__BLC LBORRLG ATRTGRP MULPICD2 ACT20GCD ;
      manova h=_all_ / /* printe printh */;
      means AGE SEX RACE ATRTGRP MULPICD2 ACT20GCD/CLDIfF tukey;
   Run;
Proc glm Data=Pred2_11;
      class SUBJID AGE SEX RACE ATRTGRP ACT20GCD MULPICD2 ;
      Model Pred= AGE SEX RACE CD4__BLC LBORRLG ATRTGRP MULPICD2 ACT20GCD ;
      manova h=_all_ / /* printe printh */;
      means AGE SEX RACE ATRTGRP MULPICD2 ACT20GCD/CLDIfF tukey;
   Run;
ods rtf close;


/* Nparlway for examining whether the initial decay rates Predict long term response */
ods rtf;
Proc nparlway wilcoxon data=Pred11;
      class  P400_TLO;
      var Pred; /* response variable */
      exact;
   run;
Proc nparlway wilcoxon data=Pred2_11;
      class  P400_TLO;
      var Pred; /* response variable */
      exact;
   run;
ods rtf close;

Proc insight data=Pred11; run;
Proc insight data=Pred2_11; run;
Proc insight data=Pred2_22; run;


/* Making merged dataset with the variable of week 20 viral load */
data pred11_a;
Merge  d20week(keep= LBORRLG1 subjid)
       pred11;
by subjid;
Run;
/* Subjects who have higher decay rates for the single-phase model */
Data Pred11_aa;
set Pred11_a;
if pred < 0.168 then delete; run;
/* Subjects who have lower decay rates for the single-phase model */
Data Pred11_aaa;
set Pred11_a;
if pred >= 0.168 then delete; run;

Proc insight data=Pred11_aa; run;
Proc insight data=Pred11_aaa; run;

/* Making merged dataset with the variable of week 24 viral load */
data pred11_b;
Merge  d24week(keep= LBORRLG1 subjid)
       pred11;
by subjid;
Run;
/* Subjects who have higher decay rates for the single-phase model */
Data Pred11_bb;
set Pred11_b;
if pred < 0.168 then delete; run;
/* Subjects who have lower decay rates for the single-phase model */
Data Pred11_bbb;
set Pred11_b;
if pred >= 0.168 then delete; run;

Proc insight data=Pred11_bb; run;
Proc insight data=Pred11_bbb; run;
```

```
/* Making merged dataset with the variable of week 20 viral load */
data pred2_11_a;
Merge  d20week(keep= LBORRLG1 subjid)
       pred2_11;
by subjid;
Run;
/* Subjects who have higher decay rates for the biphasic model */
Data Pred2_11_aa;
set Pred2_11_a;
if pred < 0.097 then delete; run;
/* Subjects who have lower decay rates for the biphasic model */
Data Pred2_11_aaa;
set Pred2_11_a;
if pred >= 0.097 then delete; run;

Proc insight data=Pred2_11_aa; run;
Proc insight data=Pred2_11_aaa; run;

/* Making merged dataset with the variable of week 24 viral load */
data pred2_11_b;
Merge  d24week(keep= LBORRLG1 subjid)
       pred2_11;
by subjid;
Run;
/* Subjects who have higher decay rates for the biphasic model */
Data Pred2_11_bb;
set Pred2_11_b;
if pred < 0.097 then delete; run;
/* Subjects who have lower decay rates for the biphasic model */
Data Pred2_11_bbb;
set Pred2_11_b;
if pred >= 0.097 then delete; run;

Proc insight data=Pred2_11_bb; run;
Proc insight data=Pred2_11_bbb; run;


/* Mean of week 20/24 viral load */
ods rtf;
Proc univariate data=Pred11_aa;
var LBORRLG1; run;
Proc univariate data=Pred11_aaa;
var LBORRLG1; run;
Proc univariate data=Pred11_bb;
var LBORRLG1; run;
Proc univariate data=Pred11_bbb;
var LBORRLG1; run;
Proc univariate data=Pred2_11_aa;
var LBORRLG1; run;
Proc univariate data=Pred2_11_aaa;
var LBORRLG1; run;
Proc univariate data=Pred2_11_bb;
var LBORRLG1; run;
Proc univariate data=Pred2_11_bbb;
var LBORRLG1; run;
ods rtf close;

/* Dataset11 Importing contains results of week 24 log10 RNA change from baseline */
Proc IMPORT OUT= Pred11_test222
            DataFILE= "C:\thesis2\Pred11_test222.xls"
            DBMS=EXCEL REPLACE;
     SHEET="Pred11_test222$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;
/* Making merged dataset replicating subject observations */
data pred11_test2222;
Merge  pred11_test222(keep= LBORRLG2 subjid)
       pred11_test333;
by subjid;
Run;
```

```
/* Confusion matrix for the initial decay rates and long-term response variable
   for the single-phase model */
data Pred11_test;
set pred11;
predicted=0;
if pred>=0.168 then predicted=1;
run;
data compare1;
set pred11_test;
correctnonresp=0;
correctresp=0;
falsenonresp=0;
falseresp=0;
if predicted=1 AND P400_TLO=1 then correctresp=1;
if predicted=0 AND P400_TLO=0 then correctnonresp=1;
if predicted=1 AND P400_TLO=0 then falseresp=1;
if predicted=0 AND P400_TLO=1 then falsenonresp=1;
run;
Proc insight data=compare1;
run;
Proc freq data=compare1;
tables correctresp correctnonresp falseresp falsenonresp/out=FreqCnt;
run;


/* Confusion matrix for the initial decay rates and long-term response variable
   for the biphasic model */
data Pred2 11_test;
set pred2_11;
predicted=0;
if pred>=0.097 then predicted=1;
run;
data compare2;
set pred2_11_test;
correctnonresp=0;
correctresp=0;
falsenonresp=0;
falseresp=0;
if predicted=1 AND P400_TLO=1 then correctresp=1;
if predicted=0 AND P400_TLO=0 then correctnonresp=1;
if predicted=1 AND P400_TLO=0 then falseresp=1;
if predicted=0 AND P400_TLO=1 then falsenonresp=1;
run;
Proc insight data=compare2;
run;
Proc freq data=compare2;
tables correctresp correctnonresp falseresp falsenonresp/out=FreqCnt;
run;

ods rtf;


/* Npar1way for examining relationship between week 1 virus load reduction and long-term
response  */
Proc npar1way wilcoxon data=Pred11_change;
      class  P400_TLO;
      var LBORLGCW_S; /* response variable */
      exact;
   run;
Proc npar1way wilcoxon data=Pred2_11_change;
      class  P400_TLO;
      var LBORLGCW_S; /* response variable */
      exact;
   run;
ods rtf close;


/* The univariate logistic regression analyses for the long-term response
   and actual treatment group */
ods rtf;
Proc catmod data=pred11;
model P400_TLO=atrtgrp;
run;
Proc catmod data=pred2 11;
model P400_TLO=atrtgrp;
run;
ods rtf close;
```

```
Proc insight data=Ratepredic1_11; run;
Proc insight data=pred11; run;
Proc insight data=pred2_11; run;
Proc insight data=pred2_22; run;
Proc insight data=Pred11_change; run;
Proc insight data=Pred2_11_change; run;



/* Graph Section */
ods rtf;
/* Gplots 1-1 of viral load result for 7 randomly selected individual */
DATA data24weeks_1_1; /* New SAS data file name */
SET data24weeks;
IF (SUBJID < 251) THEN DELETE;
IF (SUBJID > 355) THEN DELETE;
run;
/* TITLE "Plot of Windowed original numeric result" */
/* Proc gplot data=data24weeks 1;
      symbol i=line v=dot h=0.5;
      LABEL LBORRSNW='Windowed original N/R';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRSNW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run; */
/* TITLE "Plot of Windowed log10 original numeric result" */
Proc gplot data=data24weeks_1_1;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;
/* TITLE "Plot of Windowed change from baseline log 10" */
/* Proc gplot data= data24weeks_1;
      symbol i=line v=dot h=0.5;
      LABEL LBORLGCW='W/C from baseline log 10';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORLGCW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run; */

/* Gplots 1-2 of viral load result for 7 randomly selected individual */
DATA data24weeks_1_2; /* New SAS data file name */
SET data24weeks;
IF (SUBJID < 565) THEN DELETE;
IF (SUBJID > 574) THEN DELETE;
run;
/* TITLE "Plot of Windowed log10 original numeric result" */
Proc gplot data=data24weeks_1_2;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;

/* Gplots 1-3 of viral load result for 7 randomly selected individual */
DATA data24weeks_1_3; /* New SAS data file name */
SET data24weeks;
IF (SUBJID < 831) THEN DELETE;
IF (SUBJID > 1250) THEN DELETE;
run;
/* TITLE "Plot of Windowed log10 original numeric result" */
Proc gplot data=data24weeks_1_3;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;


/* Gplots 1-4 of viral load result for 7 randomly selected individual */
DATA data24weeks_1_4; /* New SAS data file name */
SET data24weeks;
IF (SUBJID < 1570) THEN DELETE;
IF (SUBJID > 1661) THEN DELETE;
run;
```

```
/* TITLE "Plot of Windowed log10 original numeric result" */
Proc gplot data=data24weeks_1_4;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;

/* Gplots 1-5 of viral load result for 7 randomly selected individual */
DATA data24weeks_1_5; /* New SAS data file name */
SET data24weeks;
IF (SUBJID < 1701) THEN DELETE;
IF (SUBJID > 1832) THEN DELETE;
run;
/* TITLE "Plot of Windowed log10 original numeric result" */
Proc gplot data=data24weeks_1_5;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;

/* Gplots 1-6 of viral load result for 7 randomly selected individual */
DATA data24weeks_1_6; /* New SAS data file name */
SET data24weeks;
IF (SUBJID < 1890) THEN DELETE;
IF (SUBJID > 2020) THEN DELETE;
run;
/* TITLE "Plot of Windowed log10 original numeric result" */
Proc gplot data=data24weeks_1_6;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;
ods rtf close;


/* Gplots 2 by responder/non responders */
ods rtf;
DATA data24weeks_1_7;
SET data24weeks;
WHERE r400_tlc = '6';
run;
TITLE height=12pt font=regular "(A) Responder at visit and censored thereafter";
Proc gplot data=data24weeks_1_7;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;
ods rtf close;
DATA data24weeks_1_8;
SET data24weeks;
WHERE r400_tlc = '2.27';
run;
TITLE height=12pt font=regular "(B) Protocol mandated switch from 150mg BCV/r";
Proc gplot data=data24weeks_1_8;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;

DATA data24weeks_1_9;
SET data24weeks;
WHERE r400_tlc = '3';
run;
TITLE height=12pt font=regular "(C) Never achieved VL suppression by Week 24";
Proc gplot data=data24weeks_1_9;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;
```

```
DATA data24weeks_1_10;
SET data24weeks;
WHERE r400_tlc = '5';
run;
TITLE height=12pt font=regular "(D) Plasma HIV-1 RNA rebound";
Proc gplot data=data24weeks_1_10;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;

DATA data24weeks_1_11;
SET data24weeks;
WHERE r400_tlc = '2.10';
run;
TITLE height=12pt font=regular "(E) Insufficient viral load response";
Proc gplot data=data24weeks_1_11;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;
ods rtf close;


/* Gplots 3 for treatment group*/
ods rtf;
DATA data24weeks_1_12;
SET data24weeks;
WHERE ATRTGRP = 'dose 1';
run;
TITLE height=12pt font=regular "(A) dose 1 group";
Proc gplot data=data24weeks_1_12;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;

DATA data24weeks_1_13;
SET data24weeks;
WHERE ATRTGRP = 'dose 2';
run;
TITLE height=12pt font=regular "(B) dose 2 group";
Proc gplot data=data24weeks_1_13;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;

DATA data24weeks_1_14;
SET data24weeks;
WHERE ATRTGRP = 'dose 3';
run;
TITLE height=12pt font=regular "(C) dose 3 group";
Proc gplot data=data24weeks_1_14;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;

DATA data24weeks_1_15;
SET data24weeks;
WHERE ATRTGRP = 'contro';
run;
TITLE height=12pt font=regular "(D) control group";
Proc gplot data=data24weeks_1_15;
      symbol i=line v=dot h=0.5;
      LABEL LBORRLGW='log10 RNA';
      LABEL LBACTDY='Time(days) since start of treatment';
      plot LBORRLGW * LBACTDY=SUBJID/NOLEGEND href=16 href=30 lhref=2;
   run;
ods rtf close;
```