7-10-2014

Statistical Modeling and Prediction of HIV/AIDS Prognosis: Bayesian Analyses of Nonlinear Dynamic Mixtures

Xiaosun Lu

University of South Florida, xiaosunlu@gmail.com

Follow this and additional works at: https://scholarcommons.usf.edu/etd

Part of the Biostatistics Commons

Scholar Commons Citation


This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Statistical Modeling and Prediction of HIV/AIDS Prognosis: Bayesian Analyses of Nonlinear Dynamic Mixtures

by

Xiaosun Lu

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
Department of Epidemiology & Biostatistics
College of College of Public Health
University of South Florida

Major Professor: Yiliang Zhu, Ph.D.
Yangxin Huang, Ph.D.
Todd Wills, M.D.
Henian Chen, Ph.D.

Date of Approval:
July 10, 2014

Keywords: HIV/AIDS, mixtures, longitudinal data, Bayesian inference, Markov chain Monte Carlo

Copyright © 2014, Xiaosun Lu
## Table of Contents

List of Tables ........................................................................................................ iii
List of Figures ........................................................................................................ v
Abstract .................................................................................................................. vi

Chapter 1 Introduction ............................................................................................ 1
  1.1 HIV/AIDS background .................................................................................. 1
  1.2 Research motivation ..................................................................................... 5
  1.3 Specific aims ................................................................................................ 7

Chapter 2 Finite mixture of nonlinear trajectories and Bayesian inference .......... 9
  2.1 Mixture of nonlinear mixed-effects trajectories .......................................... 9
  2.2 Bayesian inference approaches ..................................................................... 11
    2.2.1 Prior distributions .................................................................................. 12
    2.2.2 Posterior distributions .......................................................................... 12
    2.2.3 MCMC algorithm based on conditional posterior distributions .......... 14
    2.2.4 MCMC implementation and convergence diagnosis .............................. 15
    2.2.5 Model selection criteria ........................................................................ 15
  2.3 Class membership probability depending on covariates ............................. 17

Chapter 3 Modeling of CD4 Trajectories .............................................................. 18
  3.1 Motivating data set ....................................................................................... 18
  3.2 Mixture components specification based on Michaelis Menten function ...... 22
  3.3 Regression Models ....................................................................................... 24
  3.4 Candidate models ........................................................................................ 26
  3.5 MCMC implementation .............................................................................. 27
  3.6 Model comparison and selection .................................................................. 31
  3.7 Further results from model VI ...................................................................... 34
  3.8 Class membership prediction ....................................................................... 37
    3.8.1 Current class membership probability ............................................... 38
    3.8.2 Class membership probability based on truncated year 2 data .......... 38
    3.8.3 Predicting class membership at the end of year 2 ............................... 40
    3.8.4 Evaluation of prediction ...................................................................... 41
    3.8.5 Class membership transition matrices .............................................. 42
Chapter 4   Viral load trajectories analysis   .................................................. 45
  4.1   HIV viral load dynamic models   ......................................................... 45
  4.2   Motivating data set   ......................................................................... 49
  4.3   Mixture components specification and covariate model for missing values 49
        4.3.1   Mixture components specification   .............................................. 49
        4.3.2   Covariate models with missing and mismeasured data .......... 51
  4.4   MCMC implementation   ................................................................. 52
  4.5   Results   ...................................................................................... 55
        4.5.1   The dependence of posterior probability on covariates ........ 57

Chapter 5   Discussion   ........................................................................... 59

References   .............................................................................................. 62
List of Tables

Table 1  Demographics characteristics by the number of years patients staying in the study. 19
Table 2  Baseline HAART regimens and patients’ characteristics by the number of years patients staying in the study. 21
Table 3  The deviance information criterion (DIC), expected predictive deviance (EPD), residual sum of squares (RSS) based on all candidate models. 31
Table 4  Posterior mean ($PM$), standard deviation ($SD$), 95% credible interval ($CI$) for intercepts $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ and $\sigma^2$. 32
Table 5  Posterior mean ($PM$), standard deviation ($SD$), 95% credible interval ($CI$) for coefficients of covariates in $\beta_2$. 33
Table 6  Posterior mean ($PM$), standard deviation ($SD$) and credible interval ($CI$) for regression coefficients of covariates in the model for $\beta_3$, and corresponding standard deviation ($SD$), lower limit ($L_{CI}$) and upper limit ($U_{CI}$) of 95% equal-tail credible interval ($CI$). 34
Table 7  Population proportion and 95% equal-tail credible interval ($CI$). 35
Table 8  Individual posterior probabilities of belonging to 4 trajectory classes for four representative patients. 36
Table 9  Posterior mean ($PM$) for $\gamma_1 - \gamma_3$ (95% CI) in multinomial logistic predictive model for year two membership. 40
Table 10 Test results based on non-central Chisq distribution. 42
Table 11 Closeness evaluation between mean transition matrices $T_{(12)}$ and $T_{(12')}$ 43
Table 12 Summary of estimated posterior mean ($PM$) for $\gamma_1 - \gamma_3$, and corresponding 95% equal-tail credible interval (in the parenthesis) in membership prediction multinomial logistic regression. 44
Table 13 Summary of the estimated posterior mean (PM) of population (fixed-effects), precision parameters, and corresponding standard deviation (SD), lower limit ($L_{CI}$) and upper limit ($U_{CI}$) of 95% equal-tail credible interval (CI). 56
Table 14 Summary of population proportion and 95% equal-tail credible interval (CI) for 3 trajectory classes. 56
Table 15 Individual posterior probabilities of belonging to 3 trajectory classes for six representative patients. 56
Table 16  Estimated parameters in class membership indicator logistic regression (Treatment 4 group is reference).
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Diagram of HIV (from the website of US National Institute of Health).</td>
<td>2</td>
</tr>
<tr>
<td>Figure 2</td>
<td>The process of HIV replication (from the website of Baylor College of Medicine)</td>
<td>3</td>
</tr>
<tr>
<td>Figure 3</td>
<td>(a) Profile of CD4 cell count from a clinical management database. (b) Profile of viral load in $\log_{10}$ scale for six representative patients from a clinical trial ACTG398.</td>
<td>6</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Profile of CD4 cell count from a clinical management database.</td>
<td>20</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Reaction velocity under the Michaelis-Menten equation with $V_{max} = 200$ and $K_M = 200$.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Behaviour of mean functions 4 classes of trajectories.</td>
<td>25</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Gelman-Rubin (GR) diagnostic plot based on the NLME mixture model with three Markov chains as obtained from the WinBUGS software for representative parameters.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Traceplots of MCMC parameter samples ($\beta_1, \beta_{20}, \beta_{30}, \beta_4, \beta_5, \beta_6$ and $\sigma^2$).</td>
<td>29</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Histograms of MCMC parameter samples ($\beta_1, \beta_{20}, \beta_{30}, \beta_4, \beta_5, \beta_6$ and $\sigma^2$).</td>
<td>30</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Posterior probabilities of belonging to 4 trajectory classes for 30 patients (from the 51st to the 80th patient).</td>
<td>36</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Class membership probabilities for the first 30 patients at the end of year one.</td>
<td>39</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Class membership probability based truncated year two data.</td>
<td>39</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Predicted class membership probabilities at year 2 using truncated year one data and model.</td>
<td>40</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Gelman-Rubin (GR) diagnostic plot based on the NLME mixture model with three Markov chains as obtained from the WinBUGS software for representative parameters.</td>
<td>53</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Traceplots of MCMC parameter samples ($\beta_1 - \beta_6$).</td>
<td>54</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Histograms of MCMC parameter samples ($\beta_1 - \beta_6$).</td>
<td>54</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Posterior probabilities of belonging to 3 trajectory classes for 20 patients (from the 21st to the 40th patient).</td>
<td>56</td>
</tr>
</tbody>
</table>
Abstract

Statistical analyses and modeling have contributed greatly to our understanding of the pathogenesis of HIV-1 infection; they also provide guidance for the treatment of AIDS patients and evaluation of antiretroviral (ARV) therapies. Various statistical methods, nonlinear mixed-effects models in particular, have been applied to model the CD4 and viral load trajectories. A common assumption in these methods is all patients come from a homogeneous population following one mean trajectories. This assumption unfortunately obscures important characteristic difference between subgroups of patients whose response to treatment and whose disease trajectories are biologically different. It also may lack the robustness against population heterogeneity resulting misleading or biased inference.

Finite mixture models, also known as latent class models, are commonly used to model non-predetermined heterogeneity in a population; they provide an empirical representation of heterogeneity by grouping the population into a finite number of latent classes and modeling the population through a mixture distribution. For each latent class, a finite mixture model allows individuals in each class to vary around their own mean trajectory, instead of a common one shared by all classes. Furthermore, a mixture model has ability to cluster and estimate class membership probabilities at both population and individual levels. This important feature may help physicians to better understand a particular patient disease progression and refine the therapeutical strategy in advance.

In this research, we developed mixture dynamic model and related Bayesian inferences via Markov chain Monte Carlo (MCMC). One real data set from HIV/AIDS clinical management and another from clinical trial were used to illustrate the proposed models and methods. This dissertation explored three topics. First, we modeled the CD4 trajectories using a finite mixture model with four distinct components of which the mean functions are designed based on Michaelis-Menten function. Relevant covariates both baseline and time-varying were considered and model comparison and selection were based on such-criteria as Deviance Information Criteria (DIC). Class
membership model was allowed to depend on covariates for prediction. Second, we explored disease status prediction HIV/AIDS using the latent class membership model. Third, we modeled viral load trajectories using a finite mixture model with three components of which the mean functions are designed based on published HIV dynamic systems. Although this research is motivated by HIV/AIDS studies, the basic concepts and methods developed here have much broader applications in management of other chronic diseases; they can also be applied to dynamic systems in other fields. Implementation of our methods using the publicly-available WinBUGS package suggest that our approach can be made quite accessible to practicing statisticians and data analysts.
Chapter 1
Introduction

1.1 HIV/AIDS background

In 1981, doctors in Los Angeles, San Francisco, Atlanta and New York reported that a small number of homosexual men had been diagnosed with rare forms of Kaposi’s sarcoma and Pneumocystis carinii pneumonia, which are generally found in people with seriously compromised immune systems. It became clear that an unknown disease, for which we did not know the mechanism, had appeared. In September 1982, Centers for Disease Control and Prevention (CDC) used the term acquired immune deficiency syndrome (AIDS) as an official diagnosis for this disease, characterized by a severe impairment of the immune system.

Acquired immune deficiency syndrome (AIDS) is the final stage of human immunodeficiency virus (HIV) infection. HIV infection is acquired primarily by unprotected sexual intercourse, exposure to contaminated blood or plasma, or maternal-fetal transmission. The risk of transmission after single encounter with an HIV source has been estimated to be 1 in 150 with needle sharing, 1 in 300 with occupational percutaneous exposure, 1 in 300-1000 with insertive vaginal intercourse, and 1 in 3000 with insertive anal intercourse [1].

Since the beginning of the epidemic, almost 75 million people have been infected with the HIV and about 36 million people have died of AIDS. By the end of 2012 (UNAIDS 2013), 35.3 million people were living with HIV approximately. An estimated 0.8% of adults aged 15-49 years worldwide are living with HIV, although the burden of the epidemic continues to vary considerably between countries and regions. Sub-Saharan Africa remains most severely affected, with nearly 1 in every 20 adults living with HIV and accounting for 71% of the people living with HIV worldwide.

HIV belongs to a class of viruses known as retroviruses which use ribonucleic acid (RNA) to encode their genetic information. The RNA is translated into deoxyribonucleic acid (DNA) during its life-cycle by a specific viral enzyme called reverse transcriptase. Viruses cannot grow or reproduce
on their own so they must infect cells of a living organism in order to survive and make new copies.

**Figure 1.** Diagram of HIV (from the website of US National Institute of Health).

Figure 1 shows the basic structure of HIV, which is roughly spherical and has a diameter of about 1/10,000 mm. It has a lipid membrane, which is the outer envelope of the virus and consists of two layers of lipids. Different proteins are embedded in this viral envelope consisting of glycoprotein (gp) 120, needed to attach the virion to the host cell, and transmembrane gp41, needed for the cell fusion process. Between the envelope and core, there lie matrix proteins. The viral core contains the viral capsule protein p24 which surrounds two single strands of RNA and the enzymes needed for HIV replication, such as reverse transcriptase, protease, ribonuclease, and integrase. Nine virus genes, including gag, pol and env, coded on one long stand of RNA are needed to make structural proteins for new virus copies.

Figure 2 shows the six steps of the HIV infection and replication process. (i) By binding specific receptors on the surface of a target cell, such as CD4 positive T cells (i.e., CD4 cells), macrophages and microglial cells, HIV enters the host cells. The CD4 receptor is necessary but not sufficient to permit virus entry. The secondary receptors are “chemokine receptors” that bind to chemokines and are needed to facilitate the entering [2]. (ii) HIV uses an enzyme known as reverse transcriptase to convert its RNA into DNA. (iii) HIV DNA enters the nucleus of the target cell and inserts itself into the cells DNA, where it may stay inactive for years. (iv) The infected cell makes many copies of the original virus, along with some more specialized genetic materials for making longer proteins. (v)
The longer HIV proteins is cut by an enzyme called protease into individual proteins. A new virus is assembled as long as all components come together. The virus pushes itself out of the host cell and takes with it part of the cell membrane. This outer part covers the virus and contains all of the structures necessary for the virus to bind to a new CD4 cell and begin the virus life cycle process again. Current treatment strategy involves a combination of drugs that target different steps of HIV life cycle such as entry inhibitors that prevent binding of HIV to the CD4 receptor, reverse transcriptase inhibitors that prevent the HIV RNA from being transcribed into DNA and protease inhibitors that prevent the assembly.

A T lymphocyte, called CD4 cell, is the major target cell for HIV. The CD4 cell is a subset of T cells, also known as T helper cell, which express the cluster of differentiation 4 (CD4). These cells assist other white blood cells in immunologic processes. The normal CD4 cells account for 32% to 68% of total number of lymphocytes and range between 500-1600/mL. Without any effective treatment, the dramatic decrease in CD4 cells results in such a weakened immune system that the body can no longer fight infections or certain cancers. The mechanisms of CD4 cell death in HIV infection are still not fully understood. The mechanisms by which HIV can directly induce infected
cell death include plasma membrane disruption or increased permeability due to continuous budding of the virion [3], increasing cellular toxicity due to build up of un-integrated liner viral DNA [4] and inactivation of anti-apoptotic genes [5]. However, a longstanding question in HIV biology is how HIV viruses kill so many CD4 cells, despite the fact that most of them appear to be “bystander” cells that are not infected [6]. Recent researches demonstrate that the majority uninfected CD4 cells in peripheral blood and lymph nodes undergo three types of apoptosis [7], which is a tightly regulated programmed cell death [8]. Several HIV proteins, such as Env and Vpr, have been found to be able to up-regulate Fas/FasL gene expression either on the infected cells or neighboring uninfected cells [9], and these two genes will send signal of apoptosis to these cells.

Without treatment, the average time from acquisition of HIV to an AIDS-defining opportunistic infection is about 10 years, which is the reason why many people originally thought the rate of HIV replication and disease process would be slow. But it is not true. Several researches [12, 13, 14] suggested that HIV replication and the disease process are very vibrant. On average, plasma virions have a mean lifespan of 0.3 days (half-life = 0.24 days), and the average total HIV-1 production is $10^{3} \times 10^{9}$ per day, the minimum duration of the HIV-1 life cycle in vivo is 1.2 days, and the average HIV-1 generation time is 2.6 days (generation time is defined as the time from release of a virion until it infects another cell and causes the release of a new generation of viral particles). Because the high viral replication rate may result in a high mutation rate, Ho [12] proposed the treatment strategy of “Hit Hard, Hit Early”. “Hit Hard” requires simultaneously combining different medications in the treatment, in which “Hit Early” means the treatment should start as early as HIV infection has been confirmed. Based on 2012 U.S. Department of Health and Human Services Panel on Antiretroviral Guidelines for Adults and Adolescents [15], the initiation of antiretroviral therapy (ART) is optional if the CD4 cell count is $> 500/mL$, moderately recommended if the CD4 cell count is 350 to 500/mL and strongly recommended if the value is $< 350/mL$. Regardless of the CD4 cell count, ART is strongly recommended if patients have certain conditions such as pregnancy, history of an AIDS defining illness or hepatitis B (HBV) co-infection. The usual highly active antiretroviral therapy (HAART) combines three or more different medications such as two nucleoside reverse transcriptase inhibitors (NRTIs) and a protease inhibitor (PI), a non-nucleoside reverse transcriptase inhibitor (NNRTI) or other such combinations. These HAART regimens have been proven to be able to reduce the amount of active viruses and in some cases can lower the...
number of active viruses until it is undetectable by current blood testing techniques.

1.2 Research motivation

A longitudinal study refers to an investigation where participant outcomes are collected at multiple follow-up times yielding multiple measurements on each subject. In AIDS longitudinal studies, HIV infected patients may be followed over time and monthly measures such as CD4 counts, or viral load are collected to characterize immune status and disease burden respectively. Such longitudinal data are correlated within subjects and thus require special statistical techniques for valid analysis and inference. Further, it is not uncommon for the relationship between an explanatory variable (e.g., time) and a response variable (e.g., CD4 or viral load) to be nonlinear in the parameters. Nonlinear mixed-effects (NLME) models provide a tool for analyzing repeated measurements data by taking into consideration intra- and inter-variability as well as the nonlinear relationship between the explanatory variable and the response variable.

Various models and inference methods have been used to analyze longitudinal CD4 trajectories including, but not limited to, parametric linear growth curve model with random effect [17, 16], piecewise linear with random change-point[18]; and longitudinal viral load trajectories including, but not limited to, linear and nonlinear regression [19], NLME modeling approach [20, 21], nonparametric NLME modeling approach [22, 23], joint modeling approach via Monte Carlo EM algorithm [24], and Bayesian NLME modeling approach via Markov chain Monte Carlo (MCMC) procedure [25]. Those models and methods have provided feasible modeling choice to better understand the treatment effects of ART.

However, the majority of existing statistical models is based on the assumption that patients are from a common homogenous population governed by the same mean trajectory; random effects are used to characterize the large inter-individual variation in addition to time-varying covariates. Figure 3(a) shows the inter-individual variation in CD4 trajectory through four patients chosen from a clinical management data set (see Section 3.1 for details of this data set), and Figure 3 (b) shows the same with viral load trajectory using six patients in an AIDS Clinical Trial (ACTG398) study [26](see Section 4.2 for details of this study and data). Figure 3(a) suggests four main classes of CD4 trajectories can be roughly classified into four classes: (i) stable, (solid line, ID: 102), (ii) increasing steadily, (dashed line, ID: 143), (iii) increasing then stable, (dash-dotted line, ID: 328),
and (iv) increasing then decreasing, (dotted line, ID: 970). In comparison, Figure 3(b) reveals 3 classes for viral load trajectories: (i) rapid decreasing in a short-term period (solid lines, ID: 31, 105), (ii) rapid decreasing then stable at a low level (dashed lines, ID: 29, 132), and (iii) decreasing at the beginning, followed by rebound (dotted line, ID: 33, 99).

(a) Profiles of CD4

(b) Profiles of viral load in log10 scale

Figure 3: (a) Profile of CD4 cell count from a clinical management database. (b) Profile of viral load in log10 scale for six representative patients from a clinical trial ACTG398.

Along with these observations, we can assume that, CD4 trajectories can be described with four distinct classes with each the patients are relatively homogeneous, and viral load trajectories can be characterized in three classes. A patient’s outcome trajectory thus can be from one of these plausible empirical classes with some uncertainty. This suggests that we consider a finite mixture of NLME models to describe the population trajectories of the patients. Instead of assuming individual variation around a single common mean trajectory, a finite mixture model recognizes that the population consists of different classes of individuals whose trajectories vary around their own group-specific mean trajectory. In this sense, finite mixture modeling better captures inter-individual variation and heterogeneity. Specific applications of finite mixture models such as growth mixture model (GMM) and latent curve model (LCM) are often employed in social sciences studies [27, 28, 29] to explicitly model clustered or grouped individual behaviors. However, most finite mixture models are based on linear (polynomial) [27, 28] or piecewise linear [29] mean functions. Linear models have the advantages that the associated likelihood function has a closed form [27]. When a mixture model is extended to incorporate nonlinear mean functions, inferential procedures becomes complex because a closed form of likelihood function is no longer available. We adopt a Bayesian inferential
in our effort to model the nonlinear trajectories of CD4 and viral load. It was noticed that, in analysis of heterogeneous data, finite mixture models not only fit data “better” by recognizing different classes beyond the influence of covariates, but also provide an efficient modeling-based clustering and classification. As a co-product of the modeling process, we also obtain probability of an subject belonging to each class. The estimated probabilities may inform clinicians of the patient’s disease prognosis. For example, the information could help clinicians would be able to plan for a follow-up treatment strategy if they know the proportion of patients who could fall into a particular class under certain treatment regimen within a given time window.

1.3 Specific aims

Modeling CD4 and viral load trajectories are important in understanding HIV/AIDS prognosis and treatment effects in HIV/AIDS studies, the CD4 and viral load trajectories exhibit obvious and sizeable heterogeneity. CD4 and viral load trajectories can identify patients whose trajectories are of distinct clinical presentation and their explicit characterization in different classes can inform clinical decision. This research is focused on mixture modeling of HIV/AIDS outcomes with three specific questions to address:

- First, the finite mixture models for longitudinal data are commonly for linear mean functions. To extend the mixture models to incorporate nonlinear mean functions the use of traditional inference methods such as the EM algorithm needs to be re-evaluated. Computational complexity may render certain methods less useful and ineffective. A finite mixture model with nonlinear mean functions for longitudinal data and associated inference method need better understanding both theoretically and application wise.

- Second, in order to address the heterogeneity in the CD4 trajectories, specific nonlinear mean functions are needed for different classes. The nonlinear behavior of CD4 trajectories has not been well quantified in the literature, and nonlinear functional forms with meaningful interpretation are useful. Besides, it is of clinical interest to able to predict a patient’s membership, say, at the end of second year based the data from the first year. Developing a model based prediction of the class membership is useful.

- Third, the approach and methodology developed for CD4 trajectory analysis, can be applied
for the analysis of viral load with classes of models specific to viral load trajectories. In this spirit, the methodology developed here can be utilized for modeling progression of other chronic diseases.

This dissertation research thus is focused on three specific aims:

Aim 1. In Chapter 2, we develop a finite mixture model with distinct nonlinear mean functions and associated Bayesian inference method.

Aim 2. In Chapter 3, we designed four mean functions based on the Michaelis Menten function for four trajectory classes for CD4 counts in conjunction with a model-based membership prediction approach.

Aim 3. In Chapter 4, we apply the methodology to analyze viral load using three latent classes differentiating viral load with or without a rebound.
Chapter 2
Finite mixture of nonlinear trajectories and Bayesian inference

Finite mixture of nonlinear trajectory models and associated Bayesian inference method are presented in this chapter in a general form. Although this methodology is motivated by AIDS studies, the basic concepts of the newly-developed mixture modeling approach can be readily applied under similar circumstances, especially those of chronic diseases management where we observe disease management longitudinally with continuous measurement.

2.1 Mixture of nonlinear mixed-effects trajectories

Let \( y_i = (y_{i1}, \ldots, y_{in_i})^T \) be a vector of observed outcomes (e.g. CD4 counts or viral load) on the \( i \)th individual \((i = 1, 2, \ldots, n)\) at times \( t_{ij} \) \((j = 1, 2, \ldots, n_i)\). Assume there is a vector of \( p \) covariates, \( x \), which are measured repeatedly on the \( i \)th individual at each time point \( t_{ij} \). Some of the covariates may be time-varying in value, including time \( t_{ij} \) for example. Therefore individual \( i \) has a \( n_i \times p \) covariate matrix, \( X_i = (x_{i1}, \ldots, x_{in_i})^T \), \( x_{ij} \) is the \( j \)th column of \( X_i \) for individual \( i \) at time \( t_{ij} \). Note for a non-time-varying covariate, its value is a constant repeated \( n_i \) times. Further note that the set of covariates \( x \) includes all available variables. In practice, however, only a subset will be chosen for a particular model.

Assume that \( y_i \) follows one of \( K \) plausible latent trajectories which are described by mean functions \( g_k(\cdot), k = 1, \ldots, K \). However, it is not predetermined which group the \( i \)th individual’s trajectory belongs to. There is an unknown probability that the person’s trajectory belongs to group \( k \). Let \( c_i \) be a latent class indicator of individual \( i \), of which the value is unobservable but can assume only one value among \( 1 : K \), with the probability \( \pi_{ik} = P(c_i = k) \) \((k = 1, \ldots, K)\) and \( \sum_{k=1}^{K} \pi_{ik} = 1 \). Given \( c_i = k \) a statistical model of growth trajectory can be formulated as follows:

\[
E(y_i|X_i, \beta, b_i; c_i = k) = g_k(A_k, B_k, \beta, b_i, X_i) = (g_k(A_k, B_k, \beta, b_i, x_{i1}), \ldots, g_k(A_k, B_k, \beta, b_i, x_{in_i}))^T
\]
where $A_k$ is a $(p \times p)$ indicator matrix that determines what covariates warrant inclusion in the given model. $A_k$ is essentially an identity matrix $I_p$ with chosen diagonal elements set to 0, corresponding to the covariates excluded from the model. $B_k$ is also a $(p \times p)$ indicator matrix that selects among the covariates already chosen by $A_k$ those that warrant random effects. Because the choices can be distinct for each trajectory class $k$, $A_k$ and $B_k$ have subscript $k$.

For model $g_k$, the vector of regression coefficients for individual $i$ can now be expressed as

$$\beta_{ki} = A_k \beta + B_k b_i,$$

(2.1)

where $\beta = (\beta_1, ..., \beta_p)^T$ is the vector of population coefficients for all plausible covariates and $b_i = (b_{i1}, ..., b_{ip})^T$ is the individual-specific regression coefficients (random effects) assigned to the entire set of the plausible covariates. Note that the matrix operations $A_k \beta$ and $B_k b_i$ effectively set the unwanted coefficients to zero to arrive at the desirable subsets. This is exactly the same as in the classic regression setup,

$$X_i \beta_{ki} = X_i A_k \beta + X_i B_k b_i = X_{ki}^* \beta_k^* + Z_{ki}^* b_{ki},$$

(2.2)

where $X_{ki}^* = X_i A_k$ is a design matrix for class $k$ after removing the corresponding zero-columns; and $\beta_k^* = A_k \beta$ is a vector after taking out the zero elements from the product. (Note that $A_k A_k = A_k$.) Similarly $Z_{ki}^* = X_i B_k$ is a design matrix for the random effects, and $b_{ki}^* = B_k b_i$ is a vector of random coefficients.

The random effects are typically used to describe between-patients variation that are not explained by difference in patient physiological or clinical characters. In practice it is uncommon that random effects are attached to a large number of covariates because it is neither biologically necessary nor statistically supported by the data. Instead only a small number of covariates in the model will be assessed for potential between-patient variation. In other words, $r$ and $s$ are the number of fixed effects and random effects in a particular class. In our case we consider the case $p > r > s$. We will give examples of $A_k$ and $B_k$ in Section 3.2 and 4.3.

Random effects are typically assumed to follow a multivariate normal distributions

$$b_i \overset{iid}{\sim} N_p(0, \Sigma),$$

(2.3)

where $\Sigma (p \times p)$ is the variance-covariance matrix to be estimated.
By adding error terms to the (mean) model above we have

\[
(y_i | A_k, B_k, \beta, b_i, X_i; c_i = k) = g_k(A_k, B_k, \beta, b_i, X_i) + e_i,
\]

\[e_i \overset{iid}{\sim} N_{n_i}(0, \sigma^2 I_{n_i});\] \hspace{1cm} (2.4)

or equivalently

\[
(y_i | A_k, B_k, \beta, b_i, X_i; c_i = k) \sim N_{n_i}(g_k(A_k, B_k, \beta, b_i, X_i), \sigma^2 I_{n_i}). \]

The marginal distribution of the model above over all latent classes is given by

\[
y_i \sim \sum_{k=1}^{K} \pi_{ik} N_{n_i}(g_k(A_k, B_k, \beta, b_i, X_i), \sigma^2 I_{n_i}). \] \hspace{1cm} (2.6)

This distribution is a mixture of \( K \) non-linear mixed-effects (NLME) regression or trajectory models. The mixture probabilities \( \pi_i = (\pi_{i1}, \ldots, \pi_{iK})^T \), \( (i = 1, \ldots, n) \), can be also viewed as the mixture weights. Model (2.6) is identifiable as long as each component model is identifiable. Here identifiability also implies the models \( g_k(\cdot) \) are distinguishable from one to another [36].

This model can be interpreted as a missing data model if the indicator vector \( c = (c_1, \ldots, c_n)^T \) is treated as missing. The corresponding complete likelihood is then

\[
(y_i, c_i) \sim \prod_{k=1}^{K} [N_{n_i}(g_k(A_k, B_k, \beta, b_i, X_i), \sigma^2 I_{n_i}) Pr(c_i = k)]^{I(c_i = k)}. \] \hspace{1cm} (2.7)

It can be shown that

\[
f(y_i, c_i) = f(y_i | c_i = k) Pr(c_i = k) = f(y_i) Pr(c_i = k | y_i). \]

2.2 Bayesian inference approaches

Let \( \theta = \{ \beta, \Sigma, \sigma^2 \} \) be the collection of all unknown population parameters in models (2.2), (2.3) and (2.6). The mixture weights \( \pi_i \) in (2.6) will be dealt with separately. To make inference for the models (2.2), (2.3) and (2.6) and class membership probabilities \( \pi_i \), we take a Bayesian approach using the Markov chain Monte Carlo (MCMC) procedure [37].

Under the Bayes framework, it is assumed that there are prior information about the parameters that we are interested in. The prior information is often given in the form of a prior distribution. The prior distribution may reflect a strong belief so that a parameter is well-centered with a narrow
range of variation (informative prior, such as a normal distribution with small variance) or lack of specific knowledge so a non-informative prior (e.g. an uniform distribution or a normal distribution with large variance) is used. Upon observing the data, a posterior distribution is obtained combing the data with the prior to update the parameter with newly obtained information.

2.2.1 Prior distributions

Under the Bayes framework, we specify prior distributions for $\theta$ as follows.

$$
\beta \sim N_p(\beta_0, \Lambda), \Sigma \sim IW(\Omega, \nu), \sigma^2 \sim IG(\omega_1, \omega_2),
$$

(2.8)

where the prior distributions are normal ($N$), inverse Gamma ($IG$), and inverse Wishart ($IW$) respectively. They are chosen to be mutually independent to facilitate computations [38]. We assume the hyper-parameter matrices $\Lambda$ and $\Omega$ to be diagonal for convenience. Note that we can choose non-informative prior such as uniform distributions for selected elements of $\beta$ when needed.

By definition, the latent indicating variable $c_i$ ($i = 1, \ldots, n$) follows a Multinomial distribution ($Mul$) given the class probabilities $\pi_i$:

$$
c_i \sim Mul((1, \ldots, K), (\pi_{i1}, \ldots, \pi_{iK})).
$$

(2.9)

The class probabilities $\pi_i = (\pi_{i1}, \ldots, \pi_{iK})^T$ are assumed to initially follow a prior Dirichlet distribution ($Dir$) [39, 40, 41],

$$
\pi_i \sim Dir(\phi_1, \ldots, \phi_K).
$$

(2.10)

where $\phi_k > 0$ ($k = 1, \ldots, K$) are the hyper parameters to be assigned. In the prior, we do not consider between-individual variation, i.e. $\phi_k, (k = 1, \ldots, K)$, are free of any covariates. Since the Dirichlet is a conjugate prior to a Multinomial distribution, the posterior distribution of $\pi_i$ given $c_i$ ($i = 1, \ldots, n$) is again a Dirichlet distribution.

2.2.2 Posterior distributions

Conditional on $c_i$ and the random effects $b_i$, the conditional distribution of $y_i$ is,

$$
(y_i | b_i, A_{c_i}, B_{c_i}, X_i, \beta, \sigma^2, c_i) \sim N_{n_i}(g_{c_i}(A_{c_i}, B_{c_i}, \beta, b_i, X_i), \sigma^2 I_{n_i}).
$$

(2.11)
For notational convenience, let the observed data of the $i$th individual $\text{DATA}_i = \{y_i, X_i\}$, observed data $\text{DATA} = \{\text{DATA}_i; (i = 1, \ldots, n)\}$, $b = \{b_i, i = 1, \ldots, n\}$, $c = (c_1, \ldots, c_n)^T$.

To draw samples from posterior distributions, we derive conditional posterior distributions for all unknown parameters. Based on Bayes’ theorem, the posterior probability $c_i = k$ is given by,

$$P(c_i = k|b_i; \theta, y_i, X_i) = \frac{P(c_i = k\mid y_i \mid b_i, c_i = k, \theta, X_i) f(y_i|b_i, \theta, X_i)}{\sum_{m=1}^{\kappa} P(c_i = m\mid y_i \mid b_i, c_i = m, \theta, X_i)}$$

(2.12)

where $f(y_i|b_i, c_i = k, \theta, X_i) (k = 1, \ldots, K)$ are conditional density functions of $y_i$ based on (2.5) and $f(y_i|b_i, \theta, X_i)$ is a marginal density function of $y_i$ based on (2.6). Assuming the prior class membership probability are the same for all individuals, $P(c_1 = k) = \ldots = P(c_n = k) = \pi_k$, model 2.12 becomes

$$P(c_i = k|b_i; \theta, y_i, X_i) = \frac{\pi_k f(y_i|b_i, c_i = k, \theta, X_i)}{\sum_{m=1}^{\kappa} \pi_m f(y_i|b_i, c_i = m, \theta, X_i)},$$

(2.13)

which means in each iteration of MCMC the prior membership probability is generated from 2.10. Note that although, to facilitate the implementation, the prior class membership probability are assumed to be the same for all individuals, the posterior distribution 2.13 can be readily extended to be dependent on individual covariates, for example,

$$P(c_i = k|b_i; \theta, y_i, X_i, x_i) = \frac{P(c_i = k\mid x_i) f(y_i|b_i, c_i = k, \theta, X_i)}{\sum_{m=1}^{K} P(c_i = m\mid x_i) f(y_i|b_i, c_i = m, \theta, X_i)},$$

in which $x_i$ is an individual covariate. In this way the posterior membership probability in the previous iteration is used to be a new prior to update the membership probability.

In this case, 2.13 is used in the MCMC process. So the dependence of $\pi_{ik}$ on $y_i$ and $x$ becomes simplified in that $\pi_{ik}$ it is no longer individual specific, so the notation $\pi_k$ will be used instead. In this way, each iteration generates $\pi$ given $c_i, (i = 1, \ldots, n)$, using the following distribution

$$(\pi|\nu_1, \ldots, \nu_K) \sim \text{Dir}(\phi_1 + \nu_1, \ldots, \phi_K + \nu_K),$$

(2.14)

where $\nu_k = \sum_{i=1}^{n} I(c_i = k), (k = 1, \ldots, K)$, in which $I(\cdot)$ is an indicator function[39, 40, 41]. Since Dirichlet distribution is a conjugate prior to multinomial distribution, as mentioned above, the posterior distribution of $\pi$ given $c_i, (1, \ldots, n)$, which are generated based on 2.13, is also a Dirichlet distribution.

$$(\beta|b, \Sigma, c, \text{DATA}) \sim N_p(\beta'_0, \Lambda');$$

(2.15)
\[(\sigma^2|b, c, \text{DATA}) \sim IG(\omega_1', \omega_2'); \quad (2.16)\]
\[(\Sigma|b) \sim IW(\Omega', v'). \quad (2.17)\]

In (2.15), (2.16) and (2.17),
\[
\Lambda' = \left\{ \sum_i [A_{ci}^T (B_{ci} \Sigma B_{ci}^T)^{-1} A_{ci}] + \Lambda^{-1} \right\}^{-1},
\]
\[
\beta_0' = \Lambda' \left\{ \sum_i [A_{ci}^T (B_{ci} \Sigma B_{ci}^T)^{-1} (A_{ci} \beta + B_{ci} b_i)] + \Lambda^{-1} \beta_0 \right\};
\]
\[
\omega_1' = \omega_1 + 2^{-1} \sum_i n_i,
\]
\[
\omega_2' = \omega_2 + 2^{-1} \sum_i [y_i - g_{ci} (A_{ci}, B_{ci}, \beta, b_i, X_i)]^T [y_i - g_{ci} (A_{ci}, B_{ci}, \beta, b_i, X_i)];
\]
\[
\Omega' = \Omega + \sum_i b_i b_i^T,
\]
\[
v' = v + n.
\]

The full conditional distribution of each \(b_i\) given the remaining parameters and data, however, cannot explicitly be expressed. The distribution of \((b_i|\beta, \Sigma, \sigma^2, c_i, \text{DATA}_i)\) has a density function that is proportional to
\[
\exp\left\{ -(2\sigma^2)^{-1}[y_i - g_{ci} (A_{ci}, B_{ci}, \beta, b_i, X_i)]^T [y_i - g_{ci} (A_{ci}, B_{ci}, \beta, b_i, X_i)] - 2^{-1} b_i^T \Sigma^{-1} b_i \right\}. \quad (2.18)
\]

2.2.3 MCMC algorithm based on conditional posterior distributions

The idea of MCMC lies in that we draw samples for the parameters based on their posterior distributions given the data and the latent membership indicators. Sufficiently large MCMC samples reveal the posterior distribution numerically. In the present context, the Gibbs sampler along with the Metropolis-Hastings (M-H) algorithm can be used to draw samples.

Our mixture model requires two MCMC sequences be drawn simultaneously. One is for the latent class indicators and the other is for the parameters and individual random effects given the latent class indicators. We alternate between these two sequence in each MCMC run because it is necessary to update the latent class indicators as the data models shape up, and derive the posterior distribution based on the sampled latent class. An important advantage of the above representations
based on the three-level hierarchical models is that they can be very easily implemented using
the freely available WinBUGS software [42] interacted with a function called bugs in package
R2WinBUGS of the software R.

2.2.4 MCMC implementation and convergence diagnosis

The MCMC sampler was implemented using WinBUGS software [42] interacted with a function
called bugs in a package R2WinBUGS of R. When the MCMC procedure was applied to the
actual clinical data, convergence of the generated samples was assessed using standard tools within
WinBUGS software such as trace plots and Gelman-Rubin (GR) diagnostics [43]. We will illustrate
the use of GR diagnostic in the Section 3.5 and 4.4.

2.2.5 Model selection criteria

The MCMC algorithm under Bayesian framework has made it possible to fit increasingly complex
statistical models with larger number of parameters and determine the best-fitting model candidates.
To determine relevant covariates in the model, a Bayesian selection criterion, known as deviance in-
formation criterion (DIC) suggested by Spiegelhalter et al. [44], can be used. DIC is applicable to
a wide range of statistical models. There are other Bayesian approaches to model selection includ-
ing, for example, posterior model probabilities, Bayes factor, posterior predictive checks (expected
predictive deviance). However, some of these methods are not automatic nor easily reduced to a
unique and single value summary [45]. To compare the candidate models and select covariates, we
examine their DICs.

Assume that the distribution of the data, \( Y \), depends on the parameter vector \( \theta \). Spiegelhal-
ter et al. [44] suggested examining the posterior distribution of the deviance statistic defined by
\( D(\theta) = -2\log f(Y|\theta) + 2\log h(Y) \) for Bayesian model comparison, where \( f(Y|\theta) \) is the likeli-
hood function and \( h(Y) \) denotes a fully specified standardizing term that is a function of the data
alone, which has no impact on model selection. For model comparison, we can set \( h(Y) = 1 \), so
we take,

\[
D(\theta) = -2\log f(Y|\theta).
\]

Based on posterior distribution of \( D(\theta) \), DIC consists of two components as follows:

\[
DIC = \overline{D} + p_D = 2\overline{D} - D(\overline{\theta}),
\]  

\[(2.19)\]
where $\overline{D} = E_{\theta|Y}[D(\theta)] = E_{\theta|Y}[-2\log f(Y|\theta)]$ is the posterior mean (PM) of deviance, and $p_D$ is the effective number of parameters, defined as the difference between the PM of deviance and deviance evaluated at the PM of $\overline{\theta}$ of the parameters. $p_D$ can also be considered as a “mean deviance minus the deviance of the means”. Spiegelhalter et al. [44] showed that such a difference, between the average of log-likelihood ratios and the likelihood ratio evaluated at the average of the parameters, is the key quantity in estimating the degrees of freedom of a test. The effective number of parameters is the sum of the intraclass correlation coefficients, which essentially measures the sum of the ratios of the precision in the likelihood to the precision in the posterior. This fact motivates using $p_D$ to be a complexity measure and the effective number of parameters of a model.

Software WinBUGS has a built-in function to compute DIC for general Bayes models, but the built-in functions cannot compute DIC for mixture models, due to their complex nature. Note that the likelihood function of a mixture model is a mean of $K$ components weighted by mixture weights $\pi = (\pi_1, \ldots, \pi_K)^T$. Extra efforts must be taken to write R code interacted with WinBUGS to compute DIC for mixture models. Celeux et al.[46] presented the calculation of DIC for mixture models. Following Celeux’s notation, an archetypical example of a $K$-component mixture model can be expressed as

$$f(Y|\theta) = \sum_{k=1}^K \pi_k f_k(Y|\theta_k), \quad \sum_{k=1}^K \pi_k = 1,$$

in which $\theta = \{\pi_k, \theta_k, (k = 1, \ldots, K)\}$, $f_k, (k = 1, \ldots, K)$ are probability density functions. The observed likelihood of model 2.20 is

$$f(Y|\theta) = \prod_{i=1}^n \sum_{k=1}^K \pi_k f_k(Y_i|\theta_k).$$

The term $\overline{D}$ in 2.19 is therefore approximated by MCMC algorithm as

$$\overline{D} \approx -2M^{-1} \sum_{m=1}^M \log f(Y|\theta^{(m)})$$

$$= -2M^{-1} \sum_{m=1}^M \sum_{i=1}^n \log \left[ \sum_{k=1}^K \pi_k^{(m)} f_k(Y_i|\theta_k^{(m)}) \right],$$

where $m$ and $M$ are iteration number and total iteration numbers, $\{\pi_k^{(m)}, \theta_k^{(m)}, (k = 1, \ldots, K)\}$ are the simulated values of parameters in the $m$th iteration. The term $D(\overline{\theta})$ in 2.19 is, straightforwardly,

$$D(\overline{\theta}) = -2\log f(Y|\overline{\theta})$$

$$= -2 \sum_{i=1}^n \log \left[ \sum_{k=1}^K \pi_k^{(m)} f_k(Y_i|\overline{\theta}_k^{(m)}) \right],$$

where $\overline{\pi}_k = M^{-1} \sum_{m=1}^M \pi_k^{(m)}$ and $\overline{\theta}_k = M^{-1} \sum_{m=1}^M \theta_k^{(m)}$ are the MCMC sample means of simulated values. Based on $\overline{D}$ and $D(\overline{\theta})$ from 2.22 and 2.23, DIC can be obtained according to 2.19.
Models with smaller DIC should be preferred to those with larger DIC. Models are penalized both by the value of $D$, which favors a good fit, but also (like $AIC$ and $BIC$) by the effective number of parameters $p_D$. Since $D$ will decrease as the number of parameters in a model increases, the $p_D$ term compensates for this effect by favoring models with a smaller number of parameters.

Besides DIC, we also evaluate model fitting by comparing the values of expected predictive deviance (EPD) and residual sum of squares (RSS) obtained from each model. EPD is formulated by $EPD = E[\sum_{i,j} (y_{rep,ij} - y_{obs,ij})^2]$, where the predictive value $y_{rep,ij}$ is a replicate of the observed $y_{obs,ij}$. In other words, $y_{rep,ij}$ is a random sample from the distribution of $y$ given simulated parameters. The expectation is taken over the posterior distribution of the model parameters $\theta$ [47]. 

RSS is given by $\sum_{i,j} (y_{obs,ij} - y_{fitted,ij})^2$ and it is a measure of the discrepancy between the data and an estimation model. The smaller the value of EPD and RSS, the better fit of the model to the data.

### 2.3 Class membership probability depending on covariates

In 2.12, the individual classification depends on individual covariates $X_i$ in an inexplicit way. In order to predict class membership in the future based currently available data (see Section 3.8 for details) the proposed finite mixture model can be extended to allow the probabilities of class membership to depend on covariates explicitly. This could be accomplished by extending $\pi_k$ to be a function $\pi_k(x_i)$, in which $x_i$ is a predictor vector [48]. The odds of belonging to class $k$ to belonging to the last class $K$ is modeled in a multinomial logistic regression [27, 29],

$$\ln(p_{ik}/p_{iK}) = x_i^T \gamma_k, (k = 1, ..., K - 1), \quad (2.24)$$

where $p_{ik}$ is posterior class membership probability defined in (2.12), $x_i$ ($l \times 1$) is a vector of intercept and other covariates for subject $i$, and $\gamma_k = (\gamma_{k1}, ..., \gamma_{kl})^T$, ($k = 1, ..., K - 1$), are unknown parameter vectors. In practice, model (4.21) can be fitted for the $m$th iteration of MCMC procedure based on class membership indicators $c_i^{(m)} = (c_1^{(m)}, ..., c_m^{(m)})^T$ and inference for $\gamma_k$ is then made based on all estimated $\gamma_k^{(m)}$, ($m = 1, ..., M$), in which $c_i^{(m)}$ is class membership indicator vector in the $m$th MCMC iteration and $M$ is the total iteration number of posterior samples.
Chapter 3
Modeling of CD4 Trajectories

In this chapter, we apply the proposed finite mixture model CD4 trajectories.

3.1 Motivating data set

The dataset used for this research is from the clinical database of the Hillsborough County Health Department (HCHD) HIV/AIDS Special Clinic which is a site of the HIV Research Network. Patients received care at the HCHD clinic lacked private insurance and depended largely on the Ryan-White program, a federal program designed specifically for people with HIV/AIDS. The sample of patients were enrolled into the program between January 2000 and December 2006. Included in the data is information on antiretroviral drugs prescribed to each patient and the prescription dates. The stop date of a prescription is also recorded. The use of HAART regimen followed the department of health and human services (DHHS) guideline. Because many of the tests such as phenotype and genotype were optional, only a fraction of the patients had such information. The incompleteness of such data rendered it less useful. This is a typical limitation of clinical management databases. However, the data set does contain demographics information such as gender, age, risk factors such as drug use in the past. Table 1 summarized the baseline demographics characteristics and risk factors by the number of years patients staying in the study. Given CD4 count as outcome and lab tests for CD4 were ordered quarterly, our inclusion criteria for this particular analysis include: patients were likely HAART-naive upon enrolling into the program; they had at least five CD4 measures while staying on in the program. As a result, the final dataset included 1011 patients. Overall, 998 previously naive patients have stayed in study for more than 1 year. Male patients were basically twice as many as female patients. Nearly half of total patients were black, and mean age was 40. Heterosexual risk factor was reported in more than 50% of patients. Men having sex with men accounted for 30% of infections; while less than 10% of patients were declared to have acquired HIV
infection through other risk behaviors.

**Table 1**: Demographics characteristics by the number of years patients staying in the study.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>1011</td>
<td>998</td>
<td>810</td>
<td>652</td>
<td>459</td>
<td>331</td>
<td>254</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>321(31.7%)</td>
<td>314(31.4%)</td>
<td>263(32.5%)</td>
<td>221(33.4%)</td>
<td>143(31.2%)</td>
<td>116(35.0%)</td>
<td>76(29.9%)</td>
</tr>
<tr>
<td>Male</td>
<td>689(68.3%)</td>
<td>684(68.6%)</td>
<td>547(67.5%)</td>
<td>431(66.6%)</td>
<td>316(68.8%)</td>
<td>215(65.0%)</td>
<td>178(70.1%)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>463(45.8%)</td>
<td>460(46.1%)</td>
<td>416(51.4%)</td>
<td>290(44.5%)</td>
<td>220(47.9%)</td>
<td>142(42.9%)</td>
<td>101(39.8%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>193(19.1%)</td>
<td>189(18.9%)</td>
<td>142(17.5%)</td>
<td>132(20.2%)</td>
<td>90(19.6%)</td>
<td>72(21.8%)</td>
<td>46(18.1%)</td>
</tr>
<tr>
<td>White</td>
<td>334(33.0%)</td>
<td>329(33.0%)</td>
<td>234(28.9%)</td>
<td>217(33.3%)</td>
<td>144(31.4%)</td>
<td>117(35.3%)</td>
<td>107(42.1%)</td>
</tr>
<tr>
<td>Other</td>
<td>21(2.1%)</td>
<td>20(2.0%)</td>
<td>18(2.2%)</td>
<td>13(2.0%)</td>
<td>5(1.1%)</td>
<td>0(0%)</td>
<td>0(0%)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean(SD)</td>
<td>40.1(9.2)</td>
<td>40.0(9.3)</td>
<td>40.7(9.3)</td>
<td>41.7(8.9)</td>
<td>42.0(9.7)</td>
<td>43.0(8.5)</td>
<td>43.8(9.8)</td>
</tr>
<tr>
<td>HIV Risk Behavior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavior 1</td>
<td>543(53.7%)</td>
<td>538(53.9%)</td>
<td>450(55.5%)</td>
<td>345(52.9%)</td>
<td>229(49.9%)</td>
<td>166(50.2%)</td>
<td>128(50.4%)</td>
</tr>
<tr>
<td>Behavior 2</td>
<td>81(8.0%)</td>
<td>80(8.0%)</td>
<td>63(7.8%)</td>
<td>42(6.4%)</td>
<td>36(7.8%)</td>
<td>30(9.1%)</td>
<td>20(7.8%)</td>
</tr>
<tr>
<td>Behavior 3</td>
<td>322(31.8%)</td>
<td>320(32.1%)</td>
<td>251(31.0%)</td>
<td>235(36.0%)</td>
<td>173(37.7%)</td>
<td>122(36.7%)</td>
<td>97(38.2%)</td>
</tr>
<tr>
<td>Behavior 4</td>
<td>10(1.0%)</td>
<td>8(0.8%)</td>
<td>6(0.7%)</td>
<td>2(0.3%)</td>
<td>1(0.2%)</td>
<td>1(0.3%)</td>
<td>1(0.4%)</td>
</tr>
<tr>
<td>Behavior 5</td>
<td>55(5.4%)</td>
<td>52(5.2%)</td>
<td>40(4.9%)</td>
<td>28(4.3%)</td>
<td>20(4.4%)</td>
<td>12(3.6%)</td>
<td>8(3.1%)</td>
</tr>
</tbody>
</table>

HIV Risk Behavior
1. Heterosexual
2. Intravenous Drug Usage
3. Men having sex with Men
4. Men having sex with Men + intravenous drug use.
5. Other

Give that a large number of HAART regimens were prescribed to this sample of patients, we grouped HAART regimens according to combination of nucleoside/nucleotide reverse transcriptase inhibitor (NRTI), non-nucleoside reverse transcriptase inhibitors (NNRTI), protease inhibitor (PI) and entry inhibitors (EI), also known as fusion inhibitors. This grouping revealed the following most common HAART regimens in order of decreasing prescription frequency: 2 NRTIs + 1 NNRTI, 2 NRTIs + 2 PIs, 2 NRTIs + 1 PI, 3 NRTIs. The remaining HAART regimens were prescribed less frequently, and were grouped as “Others”, including 3 NRTIs + 1 NNRTI, 3 NRTIs + 1 NNRTI + 3 PIs, 4 NRTIs, 4 NRTIs + 2 PIs, 2 NRTIs + 2 PIs + 1 EI and others. Those periods in which HAART regimen was not prescribed were denoted “no drug”. Note, however, DHHS guideline recommends an HIV/AIDS patient be on HAART regimen all the time once HAART started. Table 2 shows 332
of 1011 patients (32.8%) initial treatment combinations were based on 2 NRTIs + 1 NNRTI; 25.7% (260) and 15.1% (153) were based on 2 NRTIs +2 PIs and 2 NRTIs + 1 PI, respectively. 85 patients (8.4%) started with 3 NRTIs. Hepatitis C was reported in 16.9% of patients. 73.0% of patients did not have AIDS defining illness when the initial regimen started. 14.2% of patients had one or more than one AIDS Co-Morbidity. Figure 4 shows the 4 typical trajectories with HAART starting date and type information.

Figure 4.: Profile of CD4 cell count from a clinical management database. Viral loads were also measured in clinical management, but on a time scheme more frequent than
Table 2: Baseline HAART regimens and patients’ characteristics by the number of years patients staying in the study.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>1011</td>
<td>998</td>
<td>810</td>
<td>652</td>
<td>439</td>
<td>331</td>
<td>254</td>
</tr>
<tr>
<td>The first HAART</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 NRTIs + 1 PI</td>
<td>153(15.1%)</td>
<td>150(15.0%)</td>
<td>115(14.2%)</td>
<td>101(15.5%)</td>
<td>77(16.7%)</td>
<td>60(18.1%)</td>
<td>55(21.7%)</td>
</tr>
<tr>
<td>2 NRTIs + 2 PIs</td>
<td>260(25.7%)</td>
<td>259(25.0%)</td>
<td>205(25.3%)</td>
<td>161(24.7%)</td>
<td>101(22.0%)</td>
<td>81(24.5%)</td>
<td>61(24.0%)</td>
</tr>
<tr>
<td>2 NRTIs + 1 NNRTI</td>
<td>332(32.8%)</td>
<td>329(32.9%)</td>
<td>259(32.0%)</td>
<td>195(29.9%)</td>
<td>120(26.1%)</td>
<td>94(28.4%)</td>
<td>87(34.2%)</td>
</tr>
<tr>
<td>3 NRTIs</td>
<td>85(8.4%)</td>
<td>80(8.0%)</td>
<td>72(8.9%)</td>
<td>65(10.0%)</td>
<td>57(12.4%)</td>
<td>40(12.1%)</td>
<td>31(12.2%)</td>
</tr>
<tr>
<td>Other</td>
<td>181(17.9%)</td>
<td>180(18.0%)</td>
<td>159(19.6%)</td>
<td>130(19.9%)</td>
<td>104(22.6%)</td>
<td>56(16.9%)</td>
<td>20(7.9%)</td>
</tr>
<tr>
<td>Hepatitis C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No co-infection</td>
<td>465(46.0%)</td>
<td>462(46.3%)</td>
<td>364(44.9%)</td>
<td>312(47.8%)</td>
<td>220(47.9%)</td>
<td>158(47.7%)</td>
<td>112(44.1%)</td>
</tr>
<tr>
<td>Co-infection</td>
<td>171(16.9%)</td>
<td>165(16.5%)</td>
<td>145(19.0%)</td>
<td>136(20.9%)</td>
<td>88(19.2%)</td>
<td>59(17.8%)</td>
<td>51(20.1%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>375(37.1%)</td>
<td>371(37.2%)</td>
<td>301(37.2%)</td>
<td>204(31.3%)</td>
<td>151(32.9%)</td>
<td>114(34.4%)</td>
<td>91(35.8%)</td>
</tr>
<tr>
<td>AIDS Defining Illness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No ADI</td>
<td>738(73.0%)</td>
<td>732(73.3%)</td>
<td>583(72.0%)</td>
<td>449(68.9%)</td>
<td>321(69.9%)</td>
<td>248(75.0%)</td>
<td>198(77.9%)</td>
</tr>
<tr>
<td>One ADI</td>
<td>202(20.0%)</td>
<td>200(20.0%)</td>
<td>178(22.0%)</td>
<td>159(24.4%)</td>
<td>114(24.8%)</td>
<td>83(25.0%)</td>
<td>56(22.1%)</td>
</tr>
<tr>
<td>Two ADIs</td>
<td>40(3.9%)</td>
<td>39(3.9%)</td>
<td>30(3.7%)</td>
<td>26(4.0%)</td>
<td>15(3.3%)</td>
<td>0(0%)</td>
<td>0(0%)</td>
</tr>
<tr>
<td>Three or More ADIs</td>
<td>31(3.1%)</td>
<td>29(2.9%)</td>
<td>19(2.3%)</td>
<td>18(2.7%)</td>
<td>9(2.0%)</td>
<td>0(0%)</td>
<td>0(0%)</td>
</tr>
<tr>
<td>Co-Morbidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Co-Morbidity</td>
<td>868(85.9%)</td>
<td>863(85.4%)</td>
<td>697(86.0%)</td>
<td>571(87.6%)</td>
<td>396(86.3%)</td>
<td>280(84.6%)</td>
<td>233(91.2%)</td>
</tr>
<tr>
<td>One</td>
<td>110(10.9%)</td>
<td>105(10.5%)</td>
<td>98(12.1%)</td>
<td>74(11.3%)</td>
<td>60(13.1%)</td>
<td>51(15.4%)</td>
<td>21(8.8%)</td>
</tr>
<tr>
<td>Two or more</td>
<td>33(3.3%)</td>
<td>30(3.0%)</td>
<td>15(1.9%)</td>
<td>7(1.1%)</td>
<td>3(0.1%)</td>
<td>0(0%)</td>
<td>0(0%)</td>
</tr>
<tr>
<td>CD4 Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4 ≤ 50</td>
<td>327(32.3%)</td>
<td>325(32.1%)</td>
<td>270(33.2%)</td>
<td>223(34.2%)</td>
<td>179(39.0%)</td>
<td>119(35.9%)</td>
<td>98(38.6%)</td>
</tr>
<tr>
<td>50 &lt; CD4 ≤200</td>
<td>414(40.9%)</td>
<td>410(40.6%)</td>
<td>332(41.0%)</td>
<td>246(37.7%)</td>
<td>129(28.1%)</td>
<td>98(29.6%)</td>
<td>60(23.6%)</td>
</tr>
<tr>
<td>200 &lt; CD4 ≤350</td>
<td>170(16.8%)</td>
<td>168(16.6%)</td>
<td>145(17.9%)</td>
<td>125(19.2%)</td>
<td>96(20.9%)</td>
<td>72(21.7%)</td>
<td>61(24.0%)</td>
</tr>
<tr>
<td>CD4 &gt;350</td>
<td>100(9.9%)</td>
<td>95(9.4%)</td>
<td>63(7.7%)</td>
<td>58(8.9%)</td>
<td>55(12.0%)</td>
<td>42(12.7%)</td>
<td>35(13.8%)</td>
</tr>
</tbody>
</table>
CD4 count. Note however, that tests for viral load and CD4 might not be performed and recorded on the same date even if the tests were ordered on the same date. To reconcile this time misalignment, we used the viral load that was measured at a time closest to the current CD4 date within the prior 6 months or 2 days after as an “imputed” value corresponding to current CD4. The log-transformation of viral load was used in the analysis to stabilize the variation of the measurement errors, which tended to increase with the value of the viral load.

Because many of the tests such as phenotype and genotype were optional, only a fraction of the patients had such information, rendering it less useful, a typical limitation of many clinical management databases. Demographics information such as gender, age, risk factors such as drug use, were also tested in the model but they showed no significant effects on the CD4 trajectories. So that, they were not included in this analysis.

3.2 Mixture components specification based on Michaelis Menten function

The Michaelis-Menten model [49] is widely used to quantify enzyme kinetics. In enzyme kinetics a substrate $S$ binds reversibly to an enzyme $E$ to form an enzyme-substrate complex $ES$, which then reacts irreversibly to generate a product $P$ and to regenerate the free enzyme $E$. This system can be represented schematically as follows:

$$E + S \rightleftharpoons ES \rightarrow E + P$$

The Michaelis-Menten quantifies the reaction velocity as:

$$v = \frac{V_{\text{max}}[S]}{K_M + [S]}$$

where $V_{\text{max}}$ represents the maximum velocity achievable by the system, when substrate concentration $[S]$ increases to reach a saturation level; $K_M$ (Michaelis constant) is the substrate concentration at which the reaction velocity is 50% of the $V_{\text{max}}$.

Figure 5 presents the reaction velocity as a function of substrate concentration $[S]$ under the Michaelis-Menten equation, with $V_{\text{max}} = 200$ and $K_M = 200$.

We adopted this classical Michaelis-Menten model to describe different classes of CD4 trajectories as follows.
Figure 5: Reaction velocity under the Michaelis-Menten equation with $V_{max} = 200$ and $K_M = 200$.

1. Class 1: stable around a constant level $\beta_1$

$$CD4(t) = \beta_1,$$  
(3.1)

2. Class 2: steady increase from baseline level, $\beta_1$

$$CD4(t) = \beta_1 + \frac{\beta_2 t^{\beta_4} + \beta_5}{\beta_3 + t^{\beta_4}},$$  
(3.2)

3. Class 3: steady increase and remaining stable

$$CD4(t) = \beta_1 + \frac{\beta_2 t^{\beta_4}}{\beta_3 + t^{\beta_4}},$$  
(3.3)

4. Class 4: increase followed by a decrease

$$CD4(t) = \beta_1 + \frac{\beta_2 t^{\beta_4}}{\beta_3 + t^{\beta_4} + \beta_6}.$$  
(3.4)

In (3.1), (3.2), (3.3) and (3.4), all parameters are positive ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6 > 0$). Figure 6 presents the behaviour of functions (3.1), (3.2), (3.3) and (3.4), with $\beta_1 = 300$, $\beta_2 = 300$, $\beta_3 = 333$, $\beta_4 = 1$, $\beta_5 = 1$, and $\beta_6 = 1.3$. To convert the four trajectory classes (3.1-3.4) into regression type of models, it is necessary to incorporate covariates as well as random effects into the basic parameters $\beta_1 - \beta_6$. Although in principle covariates and random effects can be incorporated into each and every of the basic parameter, we will focus on $\beta_2$ and $\beta_3$ both for illustration purpose and for
practicality because beta2 and beta3 are the key kinetic parameters that dominate the shape of the trajectory. From clinical standpoint, the four classes may represent stages of disease progression in increasing severity. And these four classes appear to capture the main characteristics of the CD4 trajectories observed in this dataset.

### 3.3 Regression Models

To convert the kinetic models to regression models, let us consider including, in addition to time, three covariates in the models: \((z_1, z_2, z_3) = \text{(viral load, HAART regimens, baseline CD4)}\). Among them viral load and HAART are time varying. As we have pointed out at the end of Section 3.2, we focus on two kinetic parameters for setting up the regression models: \(\beta_2\) and \(\beta_3\). Specifically,

\[
\beta_2 = \beta_{20} + \beta_{21} z_1 + \beta_{22} z_2 + \beta_{23} z_3,
\]

\[
\beta_3 = \beta_{30} + \beta_{31} z_1 + \beta_{32} z_2 + \beta_{33} z_3.
\]

As a result the vector of mean parameters beta is given by

\[
\beta = (\beta_1, \beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{30}, \beta_{31}, \beta_{32}, \beta_{33}, \beta_4, \beta_5, \beta_6).
\]

Correspondingly, the design matrix for the \(i\)th patient is

\[
X_i = (x_{i1}, ..., x_{ij}, ..., x_{in_i})^T
\]

with the observation vector at time \(j\) to be \(x_{ij} = (t_{ij}, z_{1ij}, z_{2ij}, z_{3i}, z_{1ij}, z_{2ij}, z_{3i})^T\). The repetition columns of covariates in the design matrix tailor to the needs of nonlinear models. It is well recognized that individual CD4 at baselines can be substantially heterogeneous. In our sample, it varies from less than 50 to above 2300. The baseline CD4 not only reflected the disease prognosis at the point in time, but also influenced treatment regimens to be applied to the patient (reference to DHHS guideline[59]). We found it difficult to describe the variation in baseline CD4 using a single continuous function. As a result, we grouped patients according to their baseline CD4 level, starting from group 1 of CD4 0 to 100, with increment of 100 thereafter. We then replace \(\beta_1\) by group specific \(\beta_{1l}\). Although \(\beta_{1l}\) can be estimated for each group, a decision was made to fix it to the middle point of the group interval in conjunction with individual specific random effect. With appropriately defined \(A_k\) and \(B_k\), we can define regression models for CD4 trajectories as follows.
Figure 6: Behaviour of mean functions 4 classes of trajectories.

1. Class 1: stable around a constant level

\[ g_1(A_1, B_1, \beta_i, X_i) = \beta_{1l} + b_{1i}, \]  

(3.5)

where

\[ A_1 = \text{diag}(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0), \]

\[ B_1 = \text{diag}(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0). \]

Here diag represents a diagonal matrix. \( \beta_l \) the middle level of CD4 assuming patient’s baseline CD4 is in the \( l \)th-interval.

2. Class 2: steady increase from baseline level

\[ g_2(A_2, B_2, \beta, b_i, X_i) = \beta_{1l} + b_{1i} + \frac{\beta_{2i} + \beta_{3i} + \beta_{4i} + \beta_{5i}}{\beta_{3i} + \beta_{4i}}, \]  

(3.6)

where

\[ \beta_{2i} = \beta_{20} + b_{20i} + z_{1ij} \beta_{21} + z_{2ij} \beta_{22} + z_{3ij} \beta_{23}, \]

\[ \beta_{3i} = \beta_{30} + b_{30i} + z_{1ij} \beta_{31} + z_{2ij} \beta_{32} + z_{3ij} \beta_{33}, \]

\[ \beta_{4i} = \beta_{3i} + b_{4i}, \beta_{5i} = \beta_{3i} + b_{5i}, \]

\[ A_2 = \text{diag}(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1), \]

\[ B_2 = \text{diag}(1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0), \]
3. Class 3: steady increase and remaining stable

\[ g_3(A_3, B_3, \beta, b_i, X_i) = \beta_{1i} + b_{1i} + \frac{\beta_{2i} t^\beta_{4i}}{\beta_{3i} + t^\beta_{4i}}, \]  

(3.7)

where \( \beta_{2i}, \beta_{3i}, \) and \( \beta_{4i} \) are the same to those in (3.6),

\[ A_3 = \text{diag}(1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0). \]

4. Class 4: increase followed by a decrease

\[ g_4(A_4, B_4, \beta, b_i, X_i) = \beta_{1i} + b_{1i} + \frac{\beta_{2i} t^\beta_{4i}}{\beta_{3i} + t^\beta_{4i} + \beta_{6i}}, \]  

(3.8)

where \( \text{beta}_{1i}, \beta_{2i}, \beta_{3i}, \) and \( \beta_{4i} \) are the same as those in (3.6), and \( \beta_{6i} = \beta_{6} + b_{6i} \),

\[ A_4 = \text{diag}(1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1). \]

### 3.4 Candidate models

We considered the following six models with different combinations of covariates in regression models for \( \text{beta}_{2i} \) or \( \text{beta}_{3i} \).

- **Model I:** including viral load and HAART in parameter \( \beta_{2i} \).
- **Model II:** including viral load, HAART, and ordinal baseline CD4 in parameter \( \beta_{2i} \).
- **Model III:** including viral load and HAART in both parameters \( \beta_{2i} \) and \( \beta_{3i} \).
- **Model IV:** including viral load, HAART, and categorical baseline CD4 in parameter \( \beta_{2i} \).
- **Model V:** including viral load and HAART in parameter \( \beta_{3i} \).
- **Model VI:** including viral load, HAART, and logarithm transformed baseline CD4 in parameter \( \beta_{2i} \).

We investigated the following two scenarios. First, we investigated CD4 baseline as a covariate in Models I, II, IV, and VI. In Models II and IV, baseline CD4 was categorized into 4 groups, group 1: \( \leq 50 \), group 2: \( > 50 \) and \( \leq 200 \), group 3: \( > 200 \) and \( \leq 350 \), group 4: \( > 350 \), the ordinal score 1 to 4 was assigned to each individual in Model II, but a 4–level nominal scale was used in Model IV. Model VI used log-CD4 as the covariates, whereas Model II used the original CD4 in contrast with Model I which does not include CD4. Second, we investigated whether to put those covariates, in \( \beta_{2i} \) or \( \beta_{3i} \), or both. The comparison results will be shown in Section 3.6.
3.5 MCMC implementation

To carry out the Bayesian inference, we took weakly-informative prior distributions for the parameters in Models I, II, III, IV, V and VI. In particular, (i) the prior of $\beta_3$ and power parameters in the population parameter vector $\beta$, such as $\beta_4$, $\beta_5$ and $\beta_6$, were taken to be uniform distribution $U(0, 3)$; (ii) the prior for the remaining mean parameters of $\beta$ were taken to be independent normal distribution $N(0, 100)$ for each element; (iii) we assume a noninformative inverse Gamma prior distribution $IG(0.01, 0.01)$, which has mean 1 and variance 100, for variance parameter $\sigma^2$; (iv) the priors for the variance-covariance matrices of the random-effects $\Sigma$ was taken to be inverse Wishart distributions $IW(\Omega, \nu)$, where the diagonal elements for diagonal variance matrix $\Omega$ were 0.01, and $\nu = 4$; and (v) finally, we set hyper-parameters of Dirichlet distribution in 2.10, $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 1$, assuming individuals have equal probabilities of coming from any one of four classes initially.

The MCMC sampler was implemented using WinBUGS software [42] interacting with R through a function called bugs in a package R2WinBUGS. When the MCMC procedure was applied to the actual clinical data, convergence of the generated samples was assessed using standard tools within WinBUGS software such as trace plots and Gelman-Rubin (GR) diagnostics [43]. Figure 7 shows the dynamic version of GR diagnostics using results of Model VI as obtained from the WinBUGS software for the representative parameters where the three curves are given: the middle and bottom curves below the dashed horizontal line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$) and average within-sample variance ($\bar{W}$), respectively, and the top curve represents their ratio ($\hat{R}$). It is seen that $\hat{R}$ tends to 1, and $\hat{V}$ and $\bar{W}$ will stabilize as the number of iterations increase indicating that the algorithm has approached convergence. Figures 8 and 9 show the trace plots and the histograms of $\beta_{10}-\beta_{30}$, and $\beta_4-\beta_6$. While sampled values fluctuates, the fluctuation tends to stabilizes when the posterior distribution stabilizes. Thus when the horizontal line, which is the cumulative average of the parameter estimate, on each plot of the trace plots is stable we consider convergence reached. The histograms show the distribution of each parameter and all look normally distributed.

Upon convergence, observed, we proposed that, after an initial 50,000 burn-in iterations of three chains of length 100,000, we retained every 50th MCMC sample from the next 50,000 for each chain. Thus, we obtained a total of 3,000 samples of targeted posterior distributions of the unknown parameters for statistical inference.
Figure 7: Gelman-Rubin (GR) diagnostic plot based on the NLME mixture model with three Markov chains as obtained from the WinBUGS software for representative parameters.
Figure 8: Traceplots of MCMC parameter samples ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ and $\sigma^2$).
Figure 9: Histograms of MCMC parameter samples ($\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$ and $\sigma^2$).
3.6 Model comparison and selection

To select the “best” model for the data under consideration, we used DIC, a Bayesian selection criterion. As with other model selection criteria, DIC is not intended for identification of the “correct” model, but rather merely as a method of comparing a collection of alternative models. EPD and RSS as supplemental criteria were also used. The detailed information of DIC, EPD and RSS can be found in Section 2.2.5. Table 3 presents the DIC, EPD and RSS values among the six competing models. It can be seen that the Model VI with all covariates in parameter $\beta_2$ produce better fit than all other Models in terms of $DIC$, as well as $EPD$ and $RSS$. Thus, Model VI is selected as the model for subsequent inference such as class membership prediction.

Table 3: The deviance information criterion (DIC), expected predictive deviance (EPD), residual sum of squares (RSS) based on all candidate models.

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
<th>EPD</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>32851</td>
<td>7.91</td>
<td>62589</td>
</tr>
<tr>
<td>II</td>
<td>33149</td>
<td>8.24</td>
<td>63027</td>
</tr>
<tr>
<td>III</td>
<td>34480</td>
<td>8.81</td>
<td>63577</td>
</tr>
<tr>
<td>IV</td>
<td>34906</td>
<td>9.03</td>
<td>63621</td>
</tr>
<tr>
<td>V</td>
<td>33592</td>
<td>8.56</td>
<td>63189</td>
</tr>
<tr>
<td>VI</td>
<td>32367</td>
<td>7.32</td>
<td>62173</td>
</tr>
</tbody>
</table>

The population posterior mean (PM), the corresponding standard deviation (SD) and 95% credible interval (CI) for fixed-effect parameters based on the Models I-VI are given in Tables 4–6. The following findings are observed based on the estimated results. First, for the key kinetic parameters, $\beta_2$ and $\beta_3$, which are the “altitude” of CD4 trajectories and the time CD4 increase to the half of the “altitude”, not taking the effects of covariates into consideration, the results based on all six candidate models are comparable, even though the estimate of $\beta_2$ from Model I is higher than that from Model VI. Second, for $\beta_3$ which determine the time when the trajectories increase to the half of the altitude, $\beta_3$, estimates from all six models are around 0.5. Third, the viral load is associated with peak negatively, which mean the higher the viral load is the lower CD4 can go. The coefficients from all candidate models are negative and statistically significant, since the 95% CIs do not contain zero. Fourth, because parameters $\beta_2$ and $\beta_3$ are not independent in the Michaelis-Menten system it is not necessary to have covariates in regression models for both parameters. That Model III gives high $DIC$, $EPD$ and $RSS$ than other models is empirical evidence towards that direction. Fifth,
logarithm of baseline CD4 is marginally significant in its association with the trajectory, since the 95% credible limit excludes zero. The term was therefore retained in Model VI, since this model is the best in terms of all model selection criteria. Lastly but importantly, based on the coefficients of HAART treatments in $\beta_2$, the greatest peak CD4 attainable appears to be associated with 2 NRTIs + 2PIs, followed by 3 NRTIs, 2 NRTIs + 1PI and 2 NRTIs + 1 NNRTI. More specifically, the four regimens would improve the peak CD4 counts by 44.6, 33.5, 31.1 and 27.7, respectively. In contrast, the remaining HAART regimens “others” led to the least improvement in peak CD4 at 16.2 only.

**Table 4:** Posterior mean ($PM$), standard deviation ($SD$), 95% credible interval ($CI$) for intercepts $\beta_1, \beta_{20}, \beta_{30}, \beta_4, \beta_5, \beta_6$ and $\sigma^2$.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\beta_1$</th>
<th>$\beta_{20}$</th>
<th>$\beta_{30}$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>19.6</td>
<td>85.7</td>
<td>0.51</td>
<td>0.96</td>
<td>0.88</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td>15.8</td>
<td>78.8</td>
<td>0.36</td>
<td>0.79</td>
<td>0.76</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>22.4</td>
<td>90.4</td>
<td>0.69</td>
<td>1.21</td>
<td>1.01</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>2.32</td>
<td>3.55</td>
<td>0.09</td>
<td>0.14</td>
<td>0.11</td>
<td>0.24</td>
</tr>
<tr>
<td>II</td>
<td>25.8</td>
<td>72.6</td>
<td>0.44</td>
<td>0.89</td>
<td>0.85</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>22.0</td>
<td>65.0</td>
<td>0.26</td>
<td>0.73</td>
<td>0.70</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>28.1</td>
<td>76.9</td>
<td>0.55</td>
<td>1.08</td>
<td>0.99</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td>2.26</td>
<td>3.37</td>
<td>0.10</td>
<td>0.08</td>
<td>0.08</td>
<td>0.27</td>
</tr>
<tr>
<td>III</td>
<td>31.1</td>
<td>83.4</td>
<td>0.55</td>
<td>0.88</td>
<td>1.12</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>28.0</td>
<td>77.7</td>
<td>0.40</td>
<td>0.71</td>
<td>0.98</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>33.9</td>
<td>90.5</td>
<td>0.73</td>
<td>0.99</td>
<td>1.23</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>2.17</td>
<td>3.57</td>
<td>0.12</td>
<td>0.07</td>
<td>0.10</td>
<td>0.34</td>
</tr>
<tr>
<td>IV</td>
<td>25.6</td>
<td>78.7</td>
<td>0.49</td>
<td>0.92</td>
<td>0.98</td>
<td>1.74</td>
</tr>
<tr>
<td></td>
<td>24.0</td>
<td>73.5</td>
<td>0.30</td>
<td>0.77</td>
<td>0.86</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>29.1</td>
<td>84.1</td>
<td>0.65</td>
<td>1.14</td>
<td>1.17</td>
<td>2.31</td>
</tr>
<tr>
<td></td>
<td>2.12</td>
<td>3.46</td>
<td>0.11</td>
<td>0.12</td>
<td>0.09</td>
<td>0.23</td>
</tr>
<tr>
<td>V</td>
<td>18.4</td>
<td>80.2</td>
<td>0.42</td>
<td>0.91</td>
<td>1.04</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>14.8</td>
<td>75.3</td>
<td>0.27</td>
<td>0.74</td>
<td>0.91</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>22.1</td>
<td>86.1</td>
<td>0.59</td>
<td>1.42</td>
<td>1.24</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>1.79</td>
<td>3.07</td>
<td>0.07</td>
<td>0.11</td>
<td>0.09</td>
<td>0.22</td>
</tr>
<tr>
<td>VI</td>
<td>22.5</td>
<td>76.4</td>
<td>0.47</td>
<td>0.84</td>
<td>0.93</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>18.1</td>
<td>70.1</td>
<td>0.31</td>
<td>0.66</td>
<td>0.79</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>25.9</td>
<td>81.8</td>
<td>0.62</td>
<td>1.02</td>
<td>1.06</td>
<td>2.26</td>
</tr>
<tr>
<td></td>
<td>1.92</td>
<td>3.23</td>
<td>0.08</td>
<td>0.09</td>
<td>0.07</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Table 5: Posterior mean (PM), standard deviation (SD), 95% credible interval (CI) for coefficients of covariates in \( \beta_2 \).

<table>
<thead>
<tr>
<th>Model</th>
<th>HAAR</th>
<th>log(VL)</th>
<th>N2NN1</th>
<th>N2P2</th>
<th>N2P1</th>
<th>N3</th>
<th>Others</th>
<th>CD4(ordinal)</th>
<th>log(CD4o)</th>
<th>CD4 o (categories)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
<td>PM</td>
<td>-15.3</td>
<td>25.4</td>
<td>44.7</td>
<td>26.9</td>
<td>34.9</td>
<td>15.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LCI</td>
<td>-19.8</td>
<td>3.1</td>
<td>15.6</td>
<td>3.9</td>
<td>4.14</td>
<td>5.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UCI</td>
<td>-8.4</td>
<td>49.9</td>
<td>59.4</td>
<td>52.0</td>
<td>58.8</td>
<td>24.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>3.1</td>
<td>12.2</td>
<td>12.5</td>
<td>11.8</td>
<td>15.5</td>
<td>5.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td></td>
<td>PM</td>
<td>-13.6</td>
<td>27.4</td>
<td>45.6</td>
<td>30.6</td>
<td>34.0</td>
<td>16.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LCI</td>
<td>-18.2</td>
<td>6.6</td>
<td>18.1</td>
<td>6.6</td>
<td>5.2</td>
<td>6.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UCI</td>
<td>-8.8</td>
<td>50.3</td>
<td>58.9</td>
<td>55.9</td>
<td>58.6</td>
<td>24.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>3.34</td>
<td>12.1</td>
<td>11.8</td>
<td>12.0</td>
<td>14.9</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td></td>
<td>PM</td>
<td>-14.6</td>
<td>25.1</td>
<td>44.8</td>
<td>31.8</td>
<td>21.2</td>
<td>13.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LCI</td>
<td>-19.8</td>
<td>1.35</td>
<td>14.1</td>
<td>2.30</td>
<td>0.80</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UCI</td>
<td>-4.2</td>
<td>56.3</td>
<td>59.5</td>
<td>58.4</td>
<td>55.0</td>
<td>24.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>4.20</td>
<td>15.5</td>
<td>12.3</td>
<td>16.5</td>
<td>15.5</td>
<td>5.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td>PM</td>
<td>-14.2</td>
<td>24.8</td>
<td>47.6</td>
<td>29.4</td>
<td>34.5</td>
<td>16.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LCI</td>
<td>-19.4</td>
<td>3.0</td>
<td>26.2</td>
<td>4.8</td>
<td>4.5</td>
<td>6.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UCI</td>
<td>-7.9</td>
<td>50.0</td>
<td>60.0</td>
<td>54.5</td>
<td>58.1</td>
<td>24.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>3.18</td>
<td>12.6</td>
<td>10.0</td>
<td>13.5</td>
<td>15.2</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td>PM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LCI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UCI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td></td>
<td>PM</td>
<td>-15.3</td>
<td>27.7</td>
<td>44.6</td>
<td>31.1</td>
<td>33.5</td>
<td>16.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LCI</td>
<td>-19.7</td>
<td>3.7</td>
<td>14.5</td>
<td>5.8</td>
<td>3.0</td>
<td>6.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>UCI</td>
<td>-7.4</td>
<td>53.2</td>
<td>59.6</td>
<td>54.6</td>
<td>58.3</td>
<td>25.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>3.3</td>
<td>12.4</td>
<td>12.7</td>
<td>12.3</td>
<td>15.9</td>
<td>5.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6: Posterior mean (PM), standard deviation (SD) and credible interval (CI) for regression coefficients of covariates in the model for $\beta_3$, and corresponding standard deviation (SD), lower limit ($L_{CI}$) and upper limit ($U_{CI}$) of 95% equal-tail credible interval (CI).

<table>
<thead>
<tr>
<th>Model</th>
<th>log(VL)</th>
<th>HAAR</th>
<th>N2NN1</th>
<th>N2P2</th>
<th>N2P1</th>
<th>N3</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>PM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>PM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>PM</td>
<td>-0.0010</td>
<td>-0.0125</td>
<td>0.0190</td>
<td>-0.0477</td>
<td>-0.0684</td>
<td>-0.0255</td>
</tr>
<tr>
<td></td>
<td>LCI</td>
<td>-0.0110</td>
<td>-0.0505</td>
<td>-0.0217</td>
<td>-0.0975</td>
<td>-0.1264</td>
<td>-0.0547</td>
</tr>
<tr>
<td></td>
<td>UCI</td>
<td>0.0105</td>
<td>0.0569</td>
<td>0.0569</td>
<td>0.0032</td>
<td>0.0060</td>
<td>0.0041</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.0056</td>
<td>0.0203</td>
<td>0.0202</td>
<td>0.0257</td>
<td>0.0308</td>
<td>0.0149</td>
</tr>
<tr>
<td>IV</td>
<td>PM</td>
<td>0.0118</td>
<td>-0.03491</td>
<td>-0.0774</td>
<td>-0.0349</td>
<td>-0.0897</td>
<td>-0.0366</td>
</tr>
<tr>
<td></td>
<td>LCI</td>
<td>0.0032</td>
<td>-0.06386</td>
<td>-0.1192</td>
<td>-0.0638</td>
<td>-0.1455</td>
<td>-0.0658</td>
</tr>
<tr>
<td></td>
<td>UCI</td>
<td>0.0208</td>
<td>-0.00295</td>
<td>-0.0355</td>
<td>-0.0030</td>
<td>-0.0351</td>
<td>-0.0078</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.0044</td>
<td>0.0150</td>
<td>0.0219</td>
<td>0.0150</td>
<td>0.0279</td>
<td>0.0143</td>
</tr>
<tr>
<td>V</td>
<td>PM</td>
<td>0.0118</td>
<td>-0.03491</td>
<td>-0.0774</td>
<td>-0.0349</td>
<td>-0.0897</td>
<td>-0.0366</td>
</tr>
<tr>
<td></td>
<td>LCI</td>
<td>0.0032</td>
<td>-0.06386</td>
<td>-0.1192</td>
<td>-0.0638</td>
<td>-0.1455</td>
<td>-0.0658</td>
</tr>
<tr>
<td></td>
<td>UCI</td>
<td>0.0208</td>
<td>-0.00295</td>
<td>-0.0355</td>
<td>-0.0030</td>
<td>-0.0351</td>
<td>-0.0078</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.0044</td>
<td>0.0150</td>
<td>0.0219</td>
<td>0.0150</td>
<td>0.0279</td>
<td>0.0143</td>
</tr>
</tbody>
</table>

In summary, our results suggest that among six candidate models Model VI is slightly better fit model based on all model selection criteria. In the following section we further report our results for model VI below.

3.7 Further results from model VI

One objective of this data analysis is to identify and classify all individuals into clinically sensible groups based on their CD4 trajectories. Our empirical analysis suggested 4 classes of CD4 trajectories: (i) stable, (ii) steady increase, (iii) increase first and remaining stable, and (iv) increase followed by decrease. Although subclasses are possible, these four classes represent the majority
of CD4 trajectories observed in this sample of patients. The mixture modeling enables us to obtain a summary of class membership at both the population and individual levels. At population level, the MCMC procedure yields samples from the posterior distribution of \( \pi = (\pi_1, \pi_2, \pi_3, \pi_4)^T \), the population proportion of individuals in each class as given in (2.14). The estimates and their 95% equal-tail CIs are shown in Table 7. It can be seen that class 3 (increase first and remaining stable) had the largest proportion (35.29%) followed by class 2 (steady increase) (31.45%) and class 4 (increase followed by a decrease) (24.89%), and class 1 had the lowest proportion (8.28%) (stable).

At individual level, the posterior probability, \( p_{ik} = E[I(c_i = k)] \), of individual \( i \) belonging to the \( k \)th \( (k = 1, 2, 3, 4) \) class, can be approximated by \( \frac{1}{M} \sum_{m=1}^{M} I(c_i^{(m)} = k) \), over the sample \( c^{(m)}_i \) of class membership of individual \( i \) drawn from the posterior distribution (2.12), \( (m = 1, ..., M) \). Individual classification probabilities are in barplot (Figure 10 for 30 randomly selected individuals. These individual classification probabilities help clinical diagnosis of disease progression stage. In addition to Figure 9, Table 8 shows posterior classification probabilities for the four patients shown in Figure 3(a). The classification probabilities and the trajectory class membership suggested in Figure 3(a) appear to match well. The trajectory of patient 102 appears to be that of class 1 because the CD4 counts fluctuated moderately; meanwhile the model-based classification yielded a probability 95% of being in class 1. The CD4 counts of the patient 143 increased steadily through the course of the follow-up, as a result the model classifies the trajectory into class 2 with a probability of 99%. Likewise, the CD4 counts of the patient 328 was classified into class 3 with a probability 100%. Finally, patient 970 was classified into class 4 with a probability of 94% because his/her CD4 counts increased at the beginning but decreased later, with probability being 94%.

<table>
<thead>
<tr>
<th>Class</th>
<th>Proportion(( \pi ))</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.28%</td>
<td>(6.25, 10.64%)</td>
</tr>
<tr>
<td>2</td>
<td>31.45%</td>
<td>(29.07, 33.75%)</td>
</tr>
<tr>
<td>3</td>
<td>35.29%</td>
<td>(33.73, 37.47%)</td>
</tr>
<tr>
<td>4</td>
<td>24.98%</td>
<td>(22.83, 26.56%)</td>
</tr>
</tbody>
</table>

The mathematical form of Model VI can be constructed for population mean using the estimates of the model parameters given in Table 4–6. In particular, the peak of CD4 trajectory is given by
Figure 10: Posterior probabilities of belonging to 4 trajectory classes for 30 patients (from the 51st to the 80th patient).

Table 8: Individual posterior probabilities of belonging to 4 trajectory classes for four representative patients.

<table>
<thead>
<tr>
<th>Class</th>
<th>Patient ID</th>
<th>$p_{i1}$</th>
<th>$p_{i2}$</th>
<th>$p_{i3}$</th>
<th>$p_{i4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>102</td>
<td>95%</td>
<td>0%</td>
<td>1%</td>
<td>4%</td>
</tr>
<tr>
<td>2</td>
<td>143</td>
<td>0%</td>
<td>99%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>328</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>970</td>
<td>0%</td>
<td>0%</td>
<td>6%</td>
<td>94%</td>
</tr>
</tbody>
</table>
the equation by

\[ \beta_2 = 76.4 - 15.3 \times \log(\text{viral load}) + 27.7 \times N_2 N_1 + 44.6 \times N_2 P_2 + 31.1 \times N_2 P_1 + 33.5 \times N_3 + 16.2 \times Others + 9.3 \times \log(CD_4_0) \]

, (only applicable to classes 2, 3 and 4), where viral load is a time-varying covariate and \( CD_4_0 \) is the baseline CD4 value. At the population level, CD4 trajectories are represented by the following for the four classes:

1. Class 1: stable

\[ \hat{CD}4_1(t) = \beta_l + 22.5, \quad (3.9) \]

2. Class 2: steady increase

\[ \hat{CD}4_2(t) = \beta_l + 22.5 + \frac{\beta_2 t^{1.77}}{0.47 + t^{0.84}}, \quad (3.10) \]

3. Class 3: increase first and remaining stable

\[ \hat{CD}4_3(t) = \beta_l + 22.5 + \frac{\beta_2 t^{0.84}}{0.47 + t^{0.84}}, \quad (3.11) \]

4. Class 4: increase followed by decrease

\[ \hat{CD}4_4(t) = \beta_l + 22.5 + \frac{\beta_2 t^{0.84}}{0.47 + t^{0.84}}. \quad (3.12) \]

### 3.8 Class membership prediction

Given a reasonably well fit model, one potential application is prediction of future class of trajectory (forecasting) given current status including current class of CD4 trajectory. Our mixture model allows for simultaneous model fitting and membership classification. This can be done using the multinomial logistic regression model for the class probability discussed in Section 2.3. The allows for dependence of posterior membership probability on selected predictors. It is of clinical interest to predict a patient’s CD4 trajectory class as an aid for disease staging and prognosis. For example, it would be interesting at the end of one year to forecast the CD4 trajectory within two year utilizing knowledge of the current trajectory and HAART regimens being used, etc. In other words, we would like to estimate a patients class membership probability at the end of the second year, with
only information on the patient by the first year. In this section we discuss the process of class membership prediction. We illustrate the process using year one results to predict year two outcome. The approach can be easily extended setting of different time points or intervals.

3.8.1 Current class membership probability

To predict future CD4 trajectory using currently available clinical information of the patients, a summary of current CD4 trajectory is highly useful. To this end, it is necessary to summarize a patient’s current trajectory. For this purpose, the whole data set was truncated at the end of the first year, and the mixture model was fitted to the year one data. We obtain the class membership probabilities $(p_{11i}, p_{12i}, p_{13i}, p_{14i})$ $(i = 1, \cdots, n)$. Clinically the four classes of trajectory somewhat reflect four stages of the disease progression in terms CD4. Among HAART naive patients, stable (class 1) may be viewed as an earlier stage where the immune system has not seen a meaningful decline; upon initial treatment using HAART, initial response is typically a boost of CD4 in the patient so CD4 counts increase steadily over a certain period of time (class 2); upon loosing the initial drug effects, improvement in CD4 may stagnated, resulting in class 3; finally, when patient immune system is not responding to treatments, the trajectory of class 4 is likely to follow. Hence the four classes may roughly echo progressively deteriorating stages of disease prognosis, we therefore assign scores 1-4 to classes 1, 2, 3 and 4 respectively.

$$s_{1i} = \sum_k kp_{1ki}, (k = 1, \cdots, 4),$$

where $p_{1ki}$ is current posterior membership probability for the $i$th patient. Figure 11 shows the class membership probabilities based on year 1 data for the first 30 patients.

3.8.2 Class membership probability based on truncated year 2 data

Similarly, we can fit the mixture model for the data truncated at the end of year 2 and obtain class membership probabilities. Figure 12 shows class membership probability based on truncated year 2 data.

With the estimated classification probabilities using the year two data, a multinomial logistic regression described in section 2.3 can be fitted. The covariates in the multinomial logistic regression model include, among others, the class score $s_{1i}$ derived from year 1 data and year 1 model, number
Figure 11.: Class membership probabilities for the first 30 patients at the end of year one.

Figure 12.: Class membership probability based truncated year two data.
of HAARTs in the first year, and the CD4 baseline at the beginning of the first year. Table 9 summarizes the estimated posterior mean \((PM)\) for \(\gamma_1 - \gamma_3\), and corresponding 95% equal-tail credible interval (in the parenthesis) in membership prediction multinomial logistic regression.

**Table 9:** Posterior mean \((PM)\) for \(\gamma_1 - \gamma_3\) (95\% CI) in multinomial logistic predictive model for year two membership.

<table>
<thead>
<tr>
<th>Covariate vector</th>
<th>Intercept (\times 10^{-2})</th>
<th>Probability score (\times 10^{-2})</th>
<th>number of HAARTs</th>
<th>CD4 baseline (\times 10^{-4})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\gamma_1)</td>
<td>29.8(24.3,34.2)</td>
<td>2.9(2.2,3.5)</td>
<td>-11.3(-13.6,-9.8)</td>
<td>3.8(3.3, 4.3)</td>
</tr>
<tr>
<td>(\gamma_2)</td>
<td>30.2(27.1,33.9)</td>
<td>5.0(4.1,6.6)</td>
<td>-9.9(-11.7,-8.4)</td>
<td>3.6(3.1,4.2)</td>
</tr>
<tr>
<td>(\gamma_3)</td>
<td>12.2(10.6,14.5)</td>
<td>12.9(10.3,15.2)</td>
<td>-3.5(-4.2,-2.9)</td>
<td>0.5(0.4,0.6)</td>
</tr>
</tbody>
</table>

### 3.8.3 Predicting class membership at the end of year 2

Supposing there are new patients who only have year one data, we can fit a mixture model to obtain their membership probability, and, then, the membership score. Together with their numbers of HAART, and baseline CD4 information, based on the logistic results shown in Table 9, we can obtain the predicted class membership probabilities. Figure 13 shows the predicted class membership probabilities of the first 30 patients.

**Figure 13:** Predicted class membership probabilities at year 2 using truncated year one data and model.
3.8.4 Evaluation of prediction

The probabilities shown in Figure 12 are based on information from the first and second year, and the probabilities shown in Figure 13 are based on only information from the first year. We need to evaluate the agreement of two sets of probabilities. Let \( s_{1i} \) and \( s_{2i} \) be the membership score of the \( i \)th patient based on the year one and two data, respectively, and let \( s'_{2i} \) be the predicted membership score of the \( i \)th patient for the second year based on year one data. Similarly, let \( p_{1i} \) and \( p_{2i} \) be the membership probabilities of the \( i \)th patient based on the year one and two data, respectively, and let \( p'_{2i} \) be the predicted membership probabilities of the \( i \)th patient for the second year based on year one data. Towards this end, to evaluate the prediction accuracy, we need to evaluate the agreement between \( s_{2i} \) and \( s'_{2i} \) (\( i = 1, \cdots, n \)). Let \( c = (1, 2, 3, 4)^T \); define statistic \( R^2_i \) be,

\[
R^2_i = \frac{(s'_{2i} - s_{2i})^2}{V_{s'_{2i}}},
\]

in which \( V_{s'_{2i}} = c^T V_{p'_{2i}} c \). \( V_{p'_{2i}} \) is the variance covariance of \( p'_{2i} \) (\( i = 1, \cdots, n \)). In practice, \( V_{p'_{2i}} \) can be estimated based on all predicted probability, \( p'_{2i} \) (\( i = 1, \cdots, n \)), from the multinomial logistic regression. After some simplification, we obtain

\[
E(R^2_i) = 1 + \frac{[E(s'_{2i}) - s_{2i}]^2}{V_{s'_{2i}}}.
\]

Then, after assume \( s'_{2i} \) (\( i = 1, \cdots, n \)), are normally distributed and \( s_{2i} \) (\( i = 1, \cdots, n \)), are known constants, it can shown that \( R^2 = \sum_i R^2_i \) asymptotically follows a noncentral Chi-square distribution with a non-centrality parameter \( \lambda = \sum_i \frac{[E(s'_{2i}) - s_{2i}]^2}{V_{s'_{2i}}} \) and a degree of freedom \( n - m \), where \( n \) and \( m \) are number of patients and unknown parameters in the mixture model, respectively.

In this case,

\[
V_{p'_{2i}} = \begin{pmatrix}
0.090 & -0.035 & -0.028 & 0.044 \\
-0.035 & 0.152 & -0.002 & 0.071 \\
-0.028 & -0.002 & 0.111 & -0.007 \\
0.044 & 0.071 & -0.007 & 0.143
\end{pmatrix},
\]

and \( V_{s'_{2i}} = 4.97, R^2 = \sum_i R^2_i = 842.5, n = 810, \lambda = 32.5 \). It is not easy to determine the number of parameters, \( m \), in a complex mixture model, for example, \( m \) could be 25 or 14 if we do or do not take the variance-covariance parameters into consideration. The test results are shown in Table 10.
Table 10: Test results based on non-central Chisq distribution.

<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>DF</th>
<th>λ</th>
<th>$R^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>810</td>
<td>14</td>
<td>796</td>
<td>32.5</td>
<td>842.5</td>
<td>0.6376</td>
</tr>
<tr>
<td>810</td>
<td>25</td>
<td>785</td>
<td>32.5</td>
<td>842.5</td>
<td>0.7314</td>
</tr>
</tbody>
</table>

We fail to detect any statistically significant difference between the scores based on observed data and prediction, since p values are greater than 0.05.

### 3.8.5 Class membership transition matrices

For notation convenience, let $\mathbf{p}_{1i} = (p_{11i}, p_{12i}, p_{13i}, p_{14i})^T$ be the class membership probabilities of the $i$th patient, using the year 1 data, $\mathbf{p}_{2i} = (p_{21i}, p_{22i}, p_{23i}, p_{24i})^T$ be the class membership probabilities of the $i$th patient, using the year 2 data, and $\mathbf{p}'_{2i} = (p'_{21i}, p'_{22i}, p'_{23i}, p'_{24i})^T$ be the predicted class membership probabilities of the $i$th patient at the end of year 2, using the year 1 data. We define an individual probability distribution matrix between $\mathbf{p}_{1i}$ and $\mathbf{p}_{2i}$ as,

$$D_{(12)i} = \mathbf{p}_{1i} \mathbf{p}_{2i}^T.$$  \hfill (3.15)

Then, the mean probability distribution matrix is

$$D_{(12)} = n^{-1} \sum_{i=1}^{n} D_{(12)i}.$$  \hfill (3.16)

The population transition matrix, $\mathbf{T}_{(12)}$, is defined as

$$\mathbf{T}_{(12)} = \left( \left[ D_{(12)(1)}/\text{sum}(D_{(12)(1)}) \right]^T, \cdots, \left[ D_{(12)(4)}/\text{sum}(D_{(12)(4)}) \right]^T \right)^T,$$  \hfill (3.17)

in which $D_{(12)(k)}$, $(k = 1, \cdots, 4)$, is the $k$th row of $D_{(12)}$.

Similarly we can define probability distribution matrix between $\mathbf{p}_{1i}$ and $\mathbf{p}'_{2i}$ of the $i$th patient as

$$D_{(12')i} = \mathbf{p}_{1i} (\mathbf{p}'_{2i})^T.$$  \hfill (3.18)

Then, the mean probability distribution matrix is

$$D_{(12')} = n^{-1} \sum_{i=1}^{n} D_{(12')i}.$$  \hfill (3.19)

The population transition matrix, $\mathbf{T}_{(12')}$, is defined as

$$\mathbf{T}_{(12')} = \left( \left[ D_{(12')(1)}/\text{sum}(D_{(12')(1)}) \right]^T, \cdots, \left[ D_{(12')(4)}/\text{sum}(D_{(12')(4)}) \right]^T \right)^T,$$  \hfill (3.20)
in which \( D_{(12')k} \), \((k = 1, \ldots, 4)\), is the \( k \)th row of \( D_{(12')} \).

Similar to equations 3.15–3.20, we can define mean transition matrices \( T_{(23)} \), \( T_{(23')} \), \( T_{(34)} \) and \( T_{(34')} \), \( T_{(45)} \) and \( T_{(45')} \), and \( T_{(56)} \) and \( T_{(56')} \). It is obvious that the closer \( T_{(m(m+1))} \) and \( T_{(m(m+1)')} \), \((m = 1, \ldots, 5)\), are, the more accurate the prediction is. The mean transition matrices \( T_{(12)} \) and \( T_{(12')} \) are listed below.

\[
T_{(12)} = \begin{pmatrix}
0.31 & 0.30 & 0.29 & 0.10 \\
0 & 0.61 & 0.25 & 0.14 \\
0 & 0.25 & 0.53 & 0.18 \\
0 & 0.13 & 0.29 & 0.58 \\
\end{pmatrix}, \\
T_{(12')} = \begin{pmatrix}
0.30 & 0.31 & 0.28 & 0.11 \\
0 & 0.55 & 0.27 & 0.18 \\
0 & 0.30 & 0.47 & 0.23 \\
0 & 0.14 & 0.31 & 0.55 \\
\end{pmatrix}.
\]

We compare the determinants, traces and eigenvalues of \( T_{(12)} \) and \( T_{(12')} \) to examine who close they are to each other. \( \det(T_{(12)}) = 0.037 \) and \( \det(T_{(12)'}) = 0.022 \), in which \( \det \) denotes determinant; \( \text{trace}(T_{(12)}) = 2.03 \) and \( \text{trace}(T_{(12)'}) = 1.87 \); the maximum eigenvalues of \( T_{(12)} \) and \( T_{(12')} \) are 0.98 and 1 respectively, and the minimum eigenvalues of \( T_{(12)} \) and \( T_{(12')} \) are 0.28 and 0.18 respectively. Based on determinant, trace and eigenvalues, \( T_{(12)} \) and \( T_{(12')} \) are close to each other, which proves the accuracy of prediction.

**Table 11:** Closeness evaluation between mean transition matrices \( T_{(12)} \) and \( T_{(12')} \).

<table>
<thead>
<tr>
<th>Items</th>
<th>( T_{(12)} )</th>
<th>( T_{(12')} )</th>
<th>Difference percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determinant</td>
<td>0.037</td>
<td>0.022</td>
<td>40.5%</td>
</tr>
<tr>
<td>Trace</td>
<td>2.03</td>
<td>1.87</td>
<td>7.9%</td>
</tr>
<tr>
<td>Minimum eigenvalue</td>
<td>0.28</td>
<td>0.18</td>
<td>35.7%</td>
</tr>
<tr>
<td>Maximum eigenvalue</td>
<td>0.98</td>
<td>1.00</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

The process demonstrated in Section 3.8.1-3.8.3, can be repeated to predict class membership of the third year based year 2 data, and so on so forth. Table 12 show the results of the multinormial regressions from the third year to the sixth year for class membership predictions. Given a new patient, based on Table 12, can we predict the class membership probability.
Table 12: Summary of estimated posterior mean ($PM$) for $\gamma_1 - \gamma_3$, and corresponding 95% equal-tail credible interval (in the parenthesis) in membership prediction multinomial logistic regression.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Patients</th>
<th>Covariate vector</th>
<th>Intercept</th>
<th>Probability score($\times 10^{-2}$)</th>
<th>number of HAARTs</th>
<th>CD4 baseline($\times 10^{-2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>652</td>
<td>$\gamma_1$</td>
<td>33.8(28.7,38.0)</td>
<td>2.8(2.1,3.9)</td>
<td>-9.3(-11.6,-7.8)</td>
<td>3.4(3.5,4.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_2$</td>
<td>28.2(25.2,31.4)</td>
<td>5.3(4.7,6.9)</td>
<td>-9.8(-10.8,-8.7)</td>
<td>4.1(3.7,4.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_3$</td>
<td>15.2(13.5,17.8)</td>
<td>11.5(9.1,14.9)</td>
<td>-2.9(-3.9,-1.4)</td>
<td>0.3(0.1,0.5)</td>
</tr>
<tr>
<td>4</td>
<td>459</td>
<td>$\gamma_1$</td>
<td>32.4(26.5,37.1)</td>
<td>2.3(2.2,3.9)</td>
<td>-15.6(-17.3,-13.4)</td>
<td>3.1(3.3,4.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_2$</td>
<td>25.7(21.1,28.7)</td>
<td>5.2(4.7,6.4)</td>
<td>-7.1(-9.6,-5.1)</td>
<td>3.2(3.5,4.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_3$</td>
<td>15.1(13.9,17.3)</td>
<td>13.5(11.5,16.2)</td>
<td>-4.8(-6.0,-1.5)</td>
<td>0.7(0.5,0.9)</td>
</tr>
<tr>
<td>5</td>
<td>331</td>
<td>$\gamma_1$</td>
<td>35.1(30.5,39.4)</td>
<td>1.9(1.3,2.5)</td>
<td>-10.6(-11.2,-7.4)</td>
<td>3.2(3.8,4.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_2$</td>
<td>32.7(29.6,34.9)</td>
<td>4.0(3.1,5.4)</td>
<td>-8.9(-7.4,-10.2)</td>
<td>3.3(3.3,4.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_3$</td>
<td>14.0(12.2,16.7)</td>
<td>10.9(8.8,13.2)</td>
<td>-4.5(-3.1,-5.2)</td>
<td>0.4(0.2,0.6)</td>
</tr>
<tr>
<td>6</td>
<td>254</td>
<td>$\gamma_1$</td>
<td>27.3(22.5,32.1)</td>
<td>2.7(2.3,2.8)</td>
<td>-15.3(-17.5,-13.9)</td>
<td>3.6(3.3,4.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_2$</td>
<td>30.1(27.4,33.2)</td>
<td>5.4(4.3,6.0)</td>
<td>-7.9(-9.4,-6.1)</td>
<td>4.6(4.1,5.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma_3$</td>
<td>12.9(10.5,14.8)</td>
<td>14.9(12.8,17.4)</td>
<td>-1.5(-2.1,-0.8)</td>
<td>0.8(0.7,0.9)</td>
</tr>
</tbody>
</table>
Chapter 4
Viral load trajectories analysis

4.1 HIV viral load dynamic models

Three differential equations are used to describe corresponding parts in the process of HIV infection, which are target uninfected cell $T$, virus $V$ and infected cell $T^*$. 

\[
\begin{align*}
\frac{dT}{dt} &= \rho - dT - kVT \\
\frac{dT^*}{dt} &= kVT - \delta T^* \\
\frac{dV}{dt} &= \eta T^* - cV
\end{align*}
\] (4.1)

where $\rho$ and $d$ are producing and death rates of $T$, respectively; $V$ is removed from a body at rate $c$ and infects the target cells $T$ to $T^*$ at rate of $k$; $\delta$ and $\eta$ are the death rate and virus producing rate of $T^*$, respectively. (4.1) is a system of nonlinear ordinary differential equations (ODE). Even though (4.1) does have any closed form solution, we can derive various approximations.

When no $T$ is infected, it is noted that $V = 0$, $T^* = 0$ and uninfected cells $T$ are at equilibrium as $T = \rho/d$. Let $t = 0$ be the time of virus infection. Supposing the $T$ are infected with a certain amount of virus, the initial conditions are $T_0 = \rho/d$, $T^*_0 = 0$ and $V_0$. Whether the virus can grow and establish an infection or not depends on a crucial quantity called basic reproductive ratio $R$, which is defined as the number of newly infected cells arising from one infected cell when almost all cells are uninfected and $R = \frac{\rho k \eta}{d c}$. If $R < 1$, the virus would not spread, since every infected cell would not on average produce more than one new infected cell. When the infection starts with $N$ infected cells, it can be expected that, on average, there would be roughly $\ln N/\ln(1 - R)$ rounds of replications before the virus population dying out. On the other hand, when $R > 1$, every infected cell would produce more than one newly infected cell. It is guaranteed that an explosive multiplication of virus would be generated, as $V(t) = V_0 \exp(rt)$, in which $r$ is the exponential growth rate of the virus population and it is given by the larger root of the equation $r^2 + (\delta + c)R + \delta c(1 - r^2) = 0$, the
approximation of \( r = \delta(R - 1) \), which means each infected cell produces \( R \) newly infected cells before dying. Virus growth will not continue indefinitely because the supply of uninfected cells is limited.

The viral load decrease sharply, at the very beginning of initiation of HAART treatment. This change with time can be expressed by the differential equation as, \( dV/dt = P - \lambda V \), where \( P \) is the viral production rate, \( \lambda \) is the decay rate of viral load, and \( V \) is the HIV viral load in plasma. Assuming a pretreatment steady state exists, \( dV/dt = 0 \), and a perfect treatment effect that no new infection or new virion produced, the HIV dynamics can be expressed as a simple one-exponential equation [50]:

\[
V(t) = V(0) \exp(-\lambda t)
\]

where \( V(0) \) is the viral load at the baseline and \( V(t) \) is the viral load at time \( t \). Equation (4.2) can only reasonably describe the behavior of the viral dynamics during 1–2 weeks after the initialization of treatment.

Assuming no new infectious virions \( (V_I) \) but some noninfectious virions \( (V_{NI}) \) will still be produced, which means there is a perfect protease inhibitor treatment effect [13], the HIV dynamics can be expressed as the following system of ODE:

\[
\begin{align*}
\frac{dT^*}{dt} &= kV_I T - \delta T^* \\
\frac{dV_I}{dt} &= -cV_I \\
\frac{dV_{NI}}{dt} &= N\delta T^* - cV_{NI}
\end{align*}
\]

in which \( N \) is the number of new virions produced per infected cell during its life time. We can obtain a close form solution to the system of ODE (4.3), assuming constant supply of target cell \( T \) and quasi-steady state before treatment \( (dT^*/dt = 0 \) and \( dV/dt = 0 \)).

\[
V(t) = V_0 \exp(-\lambda t) + \frac{N_0}{\lambda - \delta} \times \left[ \frac{N_0}{\lambda - \delta} \{ \exp(-\delta t) - \exp(-\lambda t) \} - \delta t \exp(-\lambda t) \right]
\]

where \( V(t) = V_I(t) + V_{NI}(t) \). Equation (4.4) was applied to more frequent measured HIV–1 RNA data during the first week of treatment [13]. The estimated half-life of free virions and productively infected cells are about six hours and 1.6 days, respectively based on nonlinear least-squares regression.

The ODE (4.3) was later further extended [19], to include a longer period of treatment that a biphasical decay rate of plasma HIV-1 RNA was observed: the first phase, an initial rapid exponential decline of nearly 2-logs (first phase), and the second phase a slower exponential decline. In
this system, two more target cells are added: long-lived infected cells, macrophages (M), will be infected into $M^*$ with a rate of $k_M$, produce virions at rate of $p$ and die with a rate of $\mu_M$; and latently infected lymphocytes (L) will be produced by a rate constant $f_k$ and die at a rate of $\mu_L$.

The HIV dynamics can be expressed as:

\[
\begin{align*}
\frac{dT^*}{dt} &= kVT + \alpha L - \delta T^* \\
\frac{dL}{dt} &= f_k VT - \mu_L L \\
\frac{dM^*}{dt} &= k_M VM - \mu_M M^* \\
\frac{dV}{dt} &= N\delta T^* + pM^* - cV
\end{align*}
\]

(4.5)

where $\alpha$ is the rate of latent infected cells $L$ becoming productively infected cells. With the similar assumptions used for equation (4.4), a closed form solution to the system of ODE (4.5) can be expressed as,

\[
V(t) = V_0[A \exp(-\delta t) + B \exp(-\mu_L t) + C \exp(-\mu_M t) + (1 + A + B + C)]
\]

(4.6)

where A, B and C are functions of system parameters. 4.6 is too complicated to identify all parameters, even with additional peripheral blood mononuclear cells information. Therefore, some parameters in 4.6 are assumed to be replaced by the values from previous studies. The first six weeks since the treatment was used in equation (4.6) and the half-life of productively infected CD4 cells, long-lived infected cells and latently infected cells were estimated as 1.1 days, 14.1 days and 8.5 days, respectively.

Perfect treatment effect may not be a very reasonable assumption. Wu and Ding [20] proposed a system of ODE including a protease inhibitor efficacy parameter of $\gamma$, $0 \leq \gamma \leq 1$, while $\gamma = 0$ means the PI medications have no effect and $\gamma = 1$ means perfect effect. After some reasonable simplification, the system of ODE is:

\[
\begin{align*}
\frac{dT^*}{dt} &= kVT - \delta T^* \\
\frac{dV_I}{dt} &= (1 - \gamma)P - cV_I \\
\frac{dV_{NI}}{dt} &= \gamma P + P^* + N\delta T^* - cV_{NI}
\end{align*}
\]

(4.7)

where $P$ is the virus produced rate by productively infected cells, such as CD4 cell, $P^*$ accounts for virus produced from “mysterious” infected cells such as Langerhans cells and microglial cells, or long-lived infected cells such as macrophages and latent infected cells, and $k, T^*, \delta, V_I, V_{NI}, N$
and $c$ have the same meaning as ODE (4.3). The system of ODE (4.7) gives a closed form solution,

$$V(t) = \exp(P_1 - \lambda_1 t) + \exp(P_2 - \lambda_2 t) + (P_3 + P_4 t) \exp(-ct)$$  \hspace{1cm} (4.8)

where $V(t) = V_I(t) + V_{NI}(t)$, $\lambda_1 = \delta$ is the first-phase viral decay rate that may represent the minimum turnover rate of productively infected cells, such as CD4, $\lambda_2$ is a possibly compound clearance rate of long-lived and latently infected cells and the value depends on the infection rate and destroyed rate by HIV virus. Because $c$ has been estimated to be very rapid (less than 6 hours of half life), it can be negligible compared with other terms. Thus, the equation (4.8) can be further simplified as a two-exponential equation:

$$V(t) = \exp(p_1 - \lambda_1 t) + \exp(p_2 - \lambda_2 t)$$  \hspace{1cm} (4.9)

in which $p_1$ and $p_2$ are initial viral production rate from productively infected cells, long-lived and latently infected cells, respectively. NLME modeling can be used in the estimation of the parameters in equation (4.9).

Although the “cocktail” HAART treatment can suppress HIV in 60 to 90% of cases, 30 to 60% of patients will end up as being considered treatment failure eventually because of the viral load rebound [51]. However, the decay rates in all of the equations introduced are constant, which means they cannot be applied to model rebound viral load values. Several extensions have been developed in order to catch up viral load response that include rebound data, such as extending from two exponential (4.8) by replacing the second constant decay rate with a time varying decay rate function as [52]:

$$V(t) = \exp(P_1 - \lambda_1 t) + \exp(P_2 - \lambda_2(t)t).$$  \hspace{1cm} (4.10)

Both HIV viral load and CD4 cell count are surrogate biomarkers of HIV progress status. CD4 cell count is more often used as an endpoint for long follow-up trials or advanced patients population, but for trials with short follow-up periods, viral load is often used as a primary endpoint to quantify treatment effect, where CD4 cell count is viewed as a covariate to help predict virologic responses. In viral load trajectory analysis, CD4 cell count is viewed as a time-varying covariate in $\lambda_2(t)$. 

48
4.2 Motivating data set

The data set that motivated this research is from an AIDS clinical trial study (ACTG398), which is a randomized, double-blind, placebo-controlled, with an extension to more than 48 week study of saquinavir, indinavir, or nelfinavir added as second protease inhibitor to the 4-drug class regimen in patients with virologic failure defined by receiving saquinavir, nelfinavir, indinavir, or ritonavir[26]. This study consists of 481 HIV-1 infected patients. The plasma HIV-1 RNA (viral load) is repeatedly quantified at weeks 0, 2, 4, 8, 16, and every 8 weeks until the last patient on study. The number of viral load measurements for each individual varies from 2 to 13. Out of total 481 patients, 379 patients who had more than 2 measurements were included in data analysis. A log₁₀ transformation of viral load was used in the analysis in order to stabilize the variation of the measurement errors and to speed up estimation algorithm. CD4 cell counts were also measured throughout study on a similar scheme. 16.8% of 379 patients dropped out and never returned to the study, and 22% of CD4 measures were missing at scheduled time points (i.e., intermittently missing measurements or dropout patients may return to the study). The missing data or dropouts may be related to drug resistance or other clinical problems.

4.3 Mixture components specification and covariate model for missing values

4.3.1 Mixture components specification

Based on discussion in Section 4.1, three useful approximations of ODE solution, which can be used to capture viral load responses, have been proposed as follows.

\[
y(t) = \log_{10}(e^{p_1 - \lambda_1 t}), \quad (4.11)
\]

\[
y(t) = \log_{10}(e^{p_1 - \lambda_1 t} + e^{p_2 - \lambda_2 t}), \quad (4.12)
\]

\[
y(t) = \log_{10}(e^{p_1 - \lambda_1 t} + e^{p_2 - \lambda_2^* t}), \quad (4.13)
\]

where \( y(t) \) is the log₁₀ scaled plasma HIV-1 RNA levels at time \( t \), \( \lambda_1 \) and \( \lambda_2 \) are called the first- and second-phase viral decay rates, which may represent the minimum turnover rate of productively infected cells and that of latently or long-lived infected cells, respectively, \( \lambda_2^* \) is a time-varying decay rate depending on CD4 cell count. The parameters \( p_1 \) and \( p_2 \) are macro-parameters; \( e^{p_1} \) and \( e^{p_1} + e^{p_2} \) are the baseline viral load at time \( t = 0 \) in one- and two-compartment models, respectively.
It is generally assumed that $\lambda_1 > \lambda_2$, which assures that the model is identifiable and is appropriate for empirical studies [20]. Negative values of the decay rate $\lambda^*_i$ may correspond to viral increase and lead to viral rebound [20], suggesting that variation in the dynamic parameters may be partially associated with time-varying covariates such as repeated CD4 cell counts.

Based on discussion above, we consider one- and two-compartment models with constant decay rate(s) for trajectory classes 1 and 2 defined in Section 1.2, respectively, and a two-compartment model with a time-varying decay rate in the second compartment for trajectory class 3. Thus, the mean functions of $K = 3$ components in the mixture model are specified by

1. One-compartment model with a constant decay rate for class 1 trajectories

$$g_1(A_1, B_1, \beta_{1i}, X_i) = \log_{10}(e^{\beta_{1i1} - \beta_{1i2}t_{ij}}), \quad (4.14)$$

2. Two-compartment model with constant decay rates for class 2 trajectories

$$g_2(A_2, B_2, \beta_{2i}, X_i) = \log_{10}(e^{\beta_{2i1} - \beta_{2i2}t_{ij}} + e^{\beta_{3i3} - \beta_{3i4}t_{ij}}), \quad (4.15)$$

3. Two-compartment model with constant and time-varying decay rates for class 3 trajectories

$$g_3(A_3, B_3, \beta_{3i}, X_i) = \log_{10}(e^{\beta_{3i1} - \beta_{3i2}t_{ij}} + e^{\beta_{3i3} - \beta_{3i5}t_{ij}}), \quad (4.16)$$

Following the notation in Section (2.1), we can assign $\beta$, $b_i$, $A_k$, $B_k$, and $\beta_{ki}$ ($k = 1, 2, 3$), as follow,

$$\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)^T, \quad b_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4}, b_{i5}, b_{i6})^T,$$

$$\beta_{1i1} = \beta_{2i1} = \beta_{3i1} = \beta_1 + b_{i1},$$

$$\beta_{1i2} = \beta_{2i2} = \beta_{3i2} = \beta_2 + \beta_5 z_{i0} + b_{i2},$$

$$\beta_{2i3} = \beta_{3i3} = \beta_3 + b_{i3},$$

$$\beta_{2i4} = \beta_4 + b_{i4}, \quad \beta_{3i5} = \beta_4 + \beta_6 z_{i0}^* + b_{i4},$$

$$A_1 = \text{diag}(1, 1, 0, 0, 0, 0), \quad A_2 = \text{diag}(1, 1, 1, 1, 1),$$

$$A_3 = \text{diag}(1, 1, 1, 1, 1, 1), \quad B_1 = \text{diag}(1, 1, 0, 0, 0, 0),$$

$$B_2 = B_3 = \text{diag}(1, 1, 1, 1, 0, 0),$$

$$X_i = (x_{i1}, \cdots, x_{ij}, \cdots, x_{i_n}), \quad x_{ij} = (t_{ij}, z_{i0}, z_{ij}^*)^T,$$

where “diag” denotes a diagonal matrix with off-diagonal elements being 0, $z_{i0}$ is CD4 baseline, and $z_{ij}^*$ is true (but unobservable) value of CD4 at time $t_{ij}$ defined below in Section 4.3.2. It is noted
that (4.16) is a natural extension of (4.15) to employ a time-varying covariate CD4, \( z^*_{ij} \), to capture viral rebound in class 3 trajectories.

### 4.3.2 Covariate models with missing and mismeasured data

Various covariate mixed-effect models were investigated in the literature [23, 24, 53, 54]. This subsection briefly discusses (CD4) covariate measurement error models with missing observations. Let \( z_{ij} \) be the observed covariate values which may be missing because these covariate values may not be observed at the \( y_{ij} \) measurement time \( t_{ij} \), \( (i = 1, ..., n; j = 1, ..., n_i) \). Let \( z_i = (z_{mis,i}, z_{obs,i}) \), where \( z_{mis,i} \) and \( z_{obs,i} \) are the collections of missing and observed components of \( z_i \), respectively.

Let \( r_i = (r_{i1}, ..., r_{in_i})^T \) be a vector of missing covariate indicator such that \( r_{ij} = 1 \) if \( z_{ij} \) is missing and 0 otherwise.

In the presence of covariate measurement errors, we consider the following linear mixed-effects (LME) model to quantify the covariate process.

\[
\begin{align*}
  z_{ij} &= u_{ij}^T \alpha + v_{ij}^T a_i + \epsilon_{ij} (\equiv z^*_{ij} + \epsilon_{ij}), \\
  \epsilon_i &= (\epsilon_{i1}, ..., \epsilon_{in_i})^T \sim N_{n_i}(0, \sigma^2 I_{n_i}),
\end{align*}
\]

(4.18)

where \( z^*_{ij} = u_{ij}^T \alpha + v_{ij}^T a_i \) can be viewed as the true (but unobservable) covariate value at time \( t_{ij} \), \( u_{ij} = u_{ij}(t_{ij}) \) and \( v_{ij} = v_{ij}(t_{ij}) \) are \( l \times 1 \) design vectors, \( \alpha = (\alpha_1, ..., \alpha_l)^T \) and \( a_i = (a_{i1}, ..., a_{il})^T \) are unknown population (fixed-effects) and individual-specific (random-effects) parameter vectors, respectively. The random-effects vector \( a_i \) follows a multivariate normal distribution \( N(0, \Sigma_a) \), where \( \Sigma_a \) is an unrestricted variance-covariance matrix.

To allow for nonignorable missing mechanism in covariate, we need to assume a missing data model for the missing covariate mechanism. We consider following simple independent missing data model.

\[
  f(r|\eta) = \prod_{i=1}^n f(r_i|\eta) = \prod_{i=1}^n \prod_{j=1}^{n_i} [P(r_{ij} = 1|\eta)]^{r_{ij}} [1 - P(r_{ij} = 1|\eta)]^{1-r_{ij}},
\]

(4.19)

where \( \logit[P(r_{ij} = 1|\eta)] = \eta_0 + \eta_1 z_{ij} \), and \( \eta = (\eta_0, \eta_1)^T \) is a vector of unknown nuisance parameters. As we know, the nonignorable missing model (4.19) is not testable based on the observed data, it is important to carry out sensitivity analysis based on different missing data models. For example, we can consider the following alternative missing data models: \( \logit[P(r_{ij} = 1|\eta)] = \eta_0 + \eta_1 z_{i,j-1} + \eta_2 z_{ij} \) or \( \logit[P(r_{ij} = 1|\eta)] = \eta_0 + \eta_1 z_{i,j} + \eta_2 z_{i,j}^2 \). Similar to sensitivity analysis...
conducted by [55], the modeling results are robust against the missing (dropout) data models, we, therefore, focus on the simpler missing data model (4.19) to avoid too many nuisance parameters.

We assume that the second viral decay rate in mean function (4.16) of component 3, $\beta_{3i5}$, is related to the true CD4 values rather than the observed but possibly mismeasured CD4 values. To model the CD4 process, we consider empirical LME models and choose the best model (4.18) with quadratic ($l = 3$) based on AIC/BIC model selection criteria as suggested by [24] and [23]. Thus, we adopted the quadratic polynomial as the covariate model (4.18) with $u_{ij} = v_{ij} = (1, t_{ij}, t_{ij}^2)^T$ for the CD4 trajectory as follows:

$$z_{ij} = (\alpha_1 + a_{1i}) + (\alpha_2 + a_{2i})t_{ij} + (\alpha_2 + a_{2i})t_{ij}^2 + \epsilon_{ij}, \quad (4.20)$$

where true CD4 value $z_{ij}^{*} = (\alpha_1 + a_{1i}) + (\alpha_2 + a_{2i})t_{ij} + (\alpha_2 + a_{2i})t_{ij}^2$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)^T$ is a population (fixed-effects) parameter vector, $a_i = (a_{1i}, a_{2i}, a_{3i})^T$ is an individual-specific (random-effects) vector with distribution $N(0, \Sigma)$, and $\epsilon_i = (\epsilon_1, ..., \epsilon_{n_i})^T \sim N_{n_i}(0, \sigma^2 I_{n_i})$.

### 4.4 MCMC implementation

To carry out the Bayesian inference, we took weakly-informative prior distributions for the parameters in Models I, II and III. In particular, (i) fixed-effects were taken to be independent normal distribution $N(0, 100)$ for each element of the population parameter vectors $\beta$, $\alpha$ and $\eta$; (ii) we assume a noninformative inverse Gamma prior distribution $IG(0.01, 0.01)$, which has mean 1 and variance 100, for variance parameters $\sigma^2$ and $\sigma_2^2$; (iii) the priors for the variance-covariance matrices of the random-effects $\Sigma$ and $\Sigma_a$ were taken to be inverse Wishart distributions $IW(\Omega_1, \nu_1)$ and $IW(\Omega_2, \nu_2)$, where the diagonal elements for diagonal variance matrix $\Omega_1$ and $\Omega_2$ were 0.01, and $\nu_1 = \nu_2 = 4$; and (vi) finally, we set hyper-parameters of Dirichlet distribution in (2.10), $\phi_1 = \phi_2 = \phi_3 = 1$, assuming individuals have equal probabilities of coming from any one of three classes initially.

The MCMC sampler was implemented using WinBUGS software [42] interacted with a function called bugs in a package R2WinBUGS of R. When the MCMC procedure was applied to the actual clinical data, convergence of the generated samples was assessed using standard tools within WinBUGS software such as trace plots and Gelman-Rubin (GR) diagnostics [43]. Figure 14 shows the dynamic version of GR diagnostics based on Model I as obtained from the WinBUGS software.
for the representative parameters where the three curves are given: the middle and bottom curves below the dashed horizontal line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$) and average within-sample variance ($W$), respectively, and the top curve represents their ratio ($\hat{R}$). It is seen that $\hat{R}$ tends to 1, and $\hat{V}$ and $W$ will stabilize as the number of iterations increase indicating that the algorithm has approached convergence. Figures 15 and 16 show the trace plots and the histograms of six representative parameters, $\beta_1$–$\beta_6$. When the horizontal line on each plot of the trace plots is stable we consider convergence reached. This line is the cumulative average of the parameter estimate. The histograms show the distribution of each parameter and all look normally distributed.

**Figure 14.** Gelman-Rubin (GR) diagnostic plot based on the NLME mixture model with three Markov chains as obtained from the WinBUGS software for representative parameters.

With the convergence diagnostics observed, we proposed that, after an initial number of 50,000 burn-in iterations of three chains of length 100,000, every 50th MCMC sample was retained from the next 50,000 for each chain. Thus, we obtained a total of 3,000 samples of targeted posterior
Figure 15.: Traceplots of MCMC parameter samples ($\beta_1 - \beta_6$).

Figure 16.: Histograms of MCMC parameter samples ($\beta_1 - \beta_6$).
distributions of the unknown parameters for statistical inference.

4.5 Results

As mentioned above in Section 1.2, one of the primary objectives in this data analysis was to cluster all individuals into 3 classes of viral load trajectories: (i) decrease rapidly and constantly in a short-term period, (ii) decrease at the beginning and then maintain stable at a low level, and (iii) decrease at the beginning, but rebound later. Based on the mixture modeling, we are able to obtain a summary of class membership at both the population and individual levels. At population level, the MCMC procedure yields samples from the posterior distribution of \( \mathbf{\pi} = (\pi_1, \pi_2, \pi_3)^T \) in (2.14), the population proportion of individuals in each class. The estimates of population proportion and 95\% equal-tail CIs of \((\pi_1, \pi_2, \pi_3)\) are shown in Table 14. It can be seen that class 2 (decrease and maintain stable) has the largest proportion, 47.86\%, followed by class 3 (decrease and rebound later) with proportion, 42.86\%, and, then, class 1 (decrease all the time) with the lowest proportion, 9.28\%. At individual level, the posterior probability of individual \(i\) belonging to the \(k\)th \((k = 1, 2, 3)\) class, \(p_{ik} = E[I(c_i = k)]\), can be approximated by \(\frac{1}{M} \sum_{m=1}^{M} I(c_i^{(m)} = k)\), in which \(c_i^{(m)}\) is class membership of individual \(i\) drawn from the posterior distribution (2.12) in the \(m\)th MCMC iteration \((m = 1, \ldots, M)\), where \(M\) is total iteration number of posterior samples. Barplot shown in Figure 17 displays the probabilities for the 20 individuals. The probability corresponding to individual patient who is classified as either viral load rebound or not may help physicians to refine treatment strategy and to identify the reason of viral load rebound for such individual patient. Table 5 shows individual posterior probabilities for the six representative patients shown in Figure 3(a). The probabilities shown in Table 15 and the classes of trajectories shown in Figure 3(a) are matched quite well. The patients 31 and 105 belong to class 1 because their viral load decrease constantly in a early short-term period, with probabilities 56\% and 52\%, respectively; the viral loads of the patients 29 and 132 decrease and then maintain stable, and thus, they belong to class 2, with probabilities 99\% and 100\%, respectively; and finally, the patients 33 and 99 are in class 3 (viral load rebound), with both probabilities being 100\%.

The estimated population parameters presented in Table 13 indicate that the first-phase decay rate, and the second-phase decay rate without and with time-varying CD4 covariate may be approximated by \(\hat{\lambda}_1 = 91.2 + 6.5z_0\), \(\hat{\lambda}_2 = 0.44\) and \(\hat{\lambda}_2^*(t) = 0.44 + 39.6(-0.68 + 1.46t - 0.78t^2)\), respectively,
Table 13: Summary of the estimated posterior mean (PM) of population (fixed-effects), precision parameters, and corresponding standard deviation (SD), lower limit (L{C1}) and upper limit (U{C1}) of 95% equal-tail credible interval (CI).

<table>
<thead>
<tr>
<th></th>
<th>β_1</th>
<th>β_2</th>
<th>β_3</th>
<th>β_4</th>
<th>β_5</th>
<th>α_1</th>
<th>α_2</th>
<th>α_3</th>
<th>σ^2</th>
<th>σ^2_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>10.3</td>
<td>91.2</td>
<td>6.5</td>
<td>5.9</td>
<td>0.44</td>
<td>39.6</td>
<td>-0.68</td>
<td>1.46</td>
<td>-0.78</td>
<td>0.33</td>
</tr>
<tr>
<td>L_{C1}</td>
<td>10.1</td>
<td>81.8</td>
<td>0.52</td>
<td>5.7</td>
<td>-0.07</td>
<td>35.7</td>
<td>-0.76</td>
<td>1.14</td>
<td>-1.18</td>
<td>0.31</td>
</tr>
<tr>
<td>U_{C1}</td>
<td>10.4</td>
<td>98.9</td>
<td>12.5</td>
<td>6.1</td>
<td>0.84</td>
<td>43.5</td>
<td>-0.58</td>
<td>1.77</td>
<td>-0.33</td>
<td>0.36</td>
</tr>
<tr>
<td>SD</td>
<td>0.08</td>
<td>4.85</td>
<td>3.01</td>
<td>0.16</td>
<td>0.29</td>
<td>2.08</td>
<td>0.05</td>
<td>0.17</td>
<td>0.21</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 14: Summary of population proportion and 95% equal-tail credible interval (CI) for 3 trajectory classes.

<table>
<thead>
<tr>
<th>Class</th>
<th>Proportion(π)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.28 %</td>
<td>(7.25, 11.64%)</td>
</tr>
<tr>
<td>2</td>
<td>47.86%</td>
<td>(45.07, 50.75%)</td>
</tr>
<tr>
<td>3</td>
<td>42.86%</td>
<td>(39.83, 45.56%)</td>
</tr>
</tbody>
</table>

Table 15: Individual posterior probabilities of belonging to 3 trajectory classes for six representative patients.

<table>
<thead>
<tr>
<th>Class</th>
<th>Patient ID</th>
<th>p_{i1}</th>
<th>p_{i2}</th>
<th>p_{i3}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>57%</td>
<td>37%</td>
<td>6%</td>
</tr>
<tr>
<td>1</td>
<td>105</td>
<td>52%</td>
<td>41%</td>
<td>7%</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>0%</td>
<td>99%</td>
<td>1%</td>
</tr>
<tr>
<td>2</td>
<td>132</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>3</td>
<td>99</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Figure 17: Posterior probabilities of belonging to 3 trajectory classes for 20 patients (from the 21st to the 40th patient).
where \( z_0 \) is the baseline CD4 value. The population viral load processes, \( \hat{V} \), of 3 classes may be approximated by

\[
\hat{V}_1(t) = \exp\left\{ 10.3 - \hat{\lambda}_1 t \right\}, \quad \hat{V}_2(t) = \exp\left\{ 10.3 - \hat{\lambda}_1 t \right\} + \exp\left\{ 5.9 - \hat{\lambda}_2 t \right\}
\]

and

\[
\hat{V}_3(t) = \exp\left\{ 10.3 - \hat{\lambda}_1 t \right\} + \exp\left\{ 5.9 - \hat{\lambda}_2^* t \right\}
\]

Since the first-phase viral decay rate, \( \lambda_1 \), is significantly associated with the baseline CD4 (due to significant estimate of \( \beta_3 \)) and the second-phase viral decay rate \( \lambda_2^* \) in component 3 is significantly associated with the true (but unobserved) CD4 values, this suggests that the viral load \( V(t) \) may be significantly associated with both the baseline CD4 and the true CD4 values. This simple approximation considered here may provide a rough guidance and point to further research even though the true association described above may be more complicated.

### 4.5.1 The dependence of posterior probability on covariates

The finite mixture model in this article can be extended in most applications to model the probabilities of class membership as a function of covariates. This could be accomplished by using a logistic regression for the probability of belonging to class \( k \) with a reference to be the last class, \( K \),

\[
\ln(p_{ik}/p_{iK}) = x_i^T \gamma_k, \quad (k = 1, \ldots, K - 1)
\]  \hspace{1cm} (4.21)

where \( p_{ik} \) is posterior class membership probability defined in (2.12), \( x_i \) \((u \times 1)\) is a vector of intercept and other covariates for subject \( i \), and \( \gamma_k = (\gamma_{k1}, \ldots, \gamma_{ku})^T \), \((k = 1, \ldots, K - 1)\), are unknown parameter vectors. In practice, model (4.21) can be fitted for the \( m \)th iteration of MCMC procedure based on class membership indicators \( c_i^{(m)} \) \((m = 1, \ldots, M)\), and then we can make inference for \( \gamma_k \) based on all estimated \( \gamma_k^{(m)}, \quad (m = 1, \ldots, M) \), in which \( c_i^{(m)} \) is class membership indicator vector in the \( m \)th MCMC iteration and \( M \) is the number of total iterations. In the context of HIV dynamic example here, \( x_i \) is a \( 4 \times 1 \) vector of intercept and other three dummy treatment variables because there are four treatment groups, amprenavir, abacavir, efavirenz, and adefovir dipivoxil (see Section 4.2 for details). Other covariates, such as gender and age, could be incorporated but none of them are significant (results not shown). Unlike patients in class 3, those in classes 1 and 2 did not have viral load rebound indicating a successful therapy so that class 1 and 2 are combined when doing the logistic regression. The results of class membership logistic regression are summarized in Table 16. The estimated \( \gamma_k \) can be interpreted, for example, that the odds of having viral load rebound of patients in treatment 4 group, receiving adefovir dipivoxil, is \( \exp(0.46) = 1.58 \) times more than that in patients in treatment 1 group, receiving amprenavir. Based on results, patients
in treatment 1 group are most likely not to have viral load rebound, followed by treatment 3 and 2, and patients in treatment 4 group are most likely to have viral load rebound.

Table 16: Estimated parameters in class membership indicator logistic regression (Treatment 4 group is reference).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Treatment 1</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>-0.31</td>
<td>0.46</td>
<td>0.09</td>
<td>0.25</td>
</tr>
<tr>
<td>LCI</td>
<td>-0.65</td>
<td>0.10</td>
<td>-0.24</td>
<td>-0.041</td>
</tr>
<tr>
<td>UCI</td>
<td>-0.011</td>
<td>0.84</td>
<td>0.47</td>
<td>0.59</td>
</tr>
<tr>
<td>SD</td>
<td>0.16</td>
<td>0.17</td>
<td>0.18</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Chapter 5
Discussion

For CD4 and viral load trajectories with population heterogeneity characteristic, we have developed a Bayesian approach to finite mixture of NLME models that may be preferred over a single-class NLME model in the sense that all individuals do not have to variate around a common mean trajectory and can be clustered into several classes based the model. The proposed method may have a significant impact on AIDS research because, in the presence of heterogeneity in response, appropriate statistical inference for CD4 trajectories and HIV dynamics is important for making reliable clinical decisions. Besides, based on the multinomial logistic regression, we can make prediction about the class membership one year later based on the available data for CD4 trajectories. Although this topic is motivated by AIDS study, the basic concepts of the newly-developed mixture modeling approach have generally broader applications. Thus, our proposed method can be readily applied in other fields as long as the relevant technical specifications are met and longitudinal measurements are assumed to come from two or more identifiable subclasses within a population. Based on the class membership prediction, clinical physicians can change or modify the therapy strategy accordingly. For example, physicians may want to change the type of the HAART if a patient is predicted to be in the class 4 one year later.

One advantage of mixture modeling is its flexibility to handle longitudinal data with different characteristics (heterogeneity) and the results support that a mixture model provides a better fit to the data than a single class model[27]. Another advantage of mixture modeling is model-based probabilistic clustering to obtain class membership probabilities. The probabilities of belonging to any one of three classes can be obtained at both population and individual levels. This information may help physicians refine general therapeutic strategy and develop individualized treatment.

Finite mixture models are widely used in social statistics literature, but called growth mixture model (GMM) and latent curve model (LCM), which is a is a special type of GMM. A linear mean function is usually applied for all latent classes, and EM algorithm is often used to make
inference[27]. The proposed mixture model extended GMM in the sense that the mean function
can take a non-linear form. When the mixture model is extended to incorporate nonlinear mean
functions, the EM algorithm would become complex partially because the likelihood function does
not have a closed form anymore. In the E step of EM algorithm, we need to take an expectation of
the "complete" likelihood function. Since the "complete" likelihood function does not have a
closed form, we must either approximate the integration in the likelihood, using Taylor expansion, or
evaluate the integration using Monte Carlo method. Both approaches involve intensive computation
and great efforts of statistical programming. We proposed a Bayesian inference method for mixture
models based on nonlinear mean functions, which provides not only estimates of model parameters
but also model-based probabilities of belonging to any one of \( K \) classes, simultaneously. Deviance
information criterion (DIC) is currently widely used to compare candidate models under Bayesian
framework. WinBUGS has a built-in function to calculate DIC, but it does not calculate DIC for
mixture models due to the complexity of mixture models. As shown in Section 2.2.5, DIC for
mixture model in this case is calculated in \( \mathbb{R} \). The program can be used to calculate DIC in other
mixture model applications.

A fundamental problem for “traditional” Bayesian mixture model analysis, in which each compo-
nent has the same family of densities but with different sets of parameters, is label switching due to
the non-identifiability of mixture components under symmetric priors [57, 58]. The so-called “label
switching” problem originated from the fact, using the Stephens’ notation [57], that the likelihood
\[
L(\theta, x) = \prod_{i=1}^{n} \left\{ \pi_1 f(x_i; \phi_1, \eta) + \ldots + \pi_K f(x_i; \phi_K, \eta) \right\} ,
\]
is the same for all permutations of \( \theta \). This problem arises in traditional mixture models in which each
component has the same form of density function, but with component-specific sets of parameters.
However, there is no label switching problem in our case based on our understanding, since each
component \((k)\) has distinct mean function, \( g_k(\cdot) \), sharing the same set of parameters with different
dimensions of parameters for each component. With different mean functions with the same set of
parameters, the likelihood can be expressed as
\[
L(\theta, x) = \prod_{i=1}^{n} \left\{ \pi_1 f_1(x_i; g_1(\theta)) + \ldots + \pi_K f_K(x_i; g_K(\theta)) \right\} .
\]
Label switching problem does not happen because components correspond to different mean func-
tions \( g_k, (j = 1, \ldots, K) \), which are known and pre-specified. Instead of specifying different mean
functions for classes, alternatively, we can also specify an universal mean function for all classes and let data themselves determine how many clusters there are and shapes of trajectories. In that way the label switching problem may arise and we need to interpret the results differently.

Term “identifiability” means that for a model, any two distinct sets of parameters lead to different data distribution [56], which in turn leads to different likelihood function. As Pauler and Laird [29] pointed out, a finite mixture model is identifiable if each of component models is identifiable and distinguishable from each other. Hennig [56] gave some examples in which each component linear regression is identifiable but the whole mixture model is not identifiable. Based on our understanding, this identifiability problem does not happen in a longitudinal trajectory analysis case, because we know what outcomes are from the same subjects (In Henning’s case, the identifiability problem happened because there are more than one combinations of means of \( y \), see Section 2 Hennig’s paper for details), and each component has its specific mean function.
References


Medicine 17, 2463–2485.

Hoboken, New Jersey: Wiley.


measured with error, with application to AIDS studies. Journal of the American Statistical
Association 97, 955–964.


RT, Gulick RM, Valentine FT, Aberg JA, Rogers MD, Karol CN, Saah AJ, Lewis RH, Bessen
Dual vs single protease inhibitor therapy following antiretroviral treatment failure: a random-


Biostatistics 3, 459–475.


mixture modeling. Social and Personality Psychology Compass 2: 302–317


