Bayesian Joint Models for Longitudinal and Survival Data, with Application to HIV Data

by

Harper Xiaolan Cheng

B.Sc., University of Toronto, 2019

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Department of Statistics)

The University of British Columbia
(Vancouver)

April 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the project entitled:

Bayesian Joint Models for Longitudinal and Survival Data, with Application to HIV Data

submitted by **Harper Xiaolan Cheng** in partial fulfillment of the requirements for the degree of **Master of Science** in **Department of Statistics**.

Examining Committee:

Dr. Lang Wu, Department of Statistics
Supervisor

Abstract

Joint models for longitudinal and survival data have gained great popularity in biostatistics research in that these models allow for simultaneous evaluations on longitudinal and time-to-event data. In this project, we first give a brief review on models that are used for longitudinal and survival data analysis. And then some common approaches for evaluating joint models are discussed, with a focus on the Bayesian method. The R package `JMbayes` developed by Dr.Dimitris Rizopoulos at Erasmus University Rotterdam is employed to demonstrate the application of joint models using the HIV data. We are interested in the association between biomarker trajectories during HIV treatment and characteristics of viral rebounds after treatment termination. The longitudinal and survival processes are linked through four different association structures and the corresponding statistical inferences are derived based on the Bayesian method. Sensitivity analyses are conducted to ensure that parameter estimates are robust to the choice of priors, and joint model diagnostics are performed to certify the validity of such models. In addition to the Bayesian method, joint models are evaluated using the likelihood method with the R package `JM` for comparison purposes. Estimates for association structures produced by the likelihood method are comparable to those of Bayesian method when flat priors are used. For the HIV data, it is concluded that the shared random effects model is the most appropriate specification of the relationship between the longitudinal and survival process. The biomarker level during treatment is unrelated to the subsequent characteristics of viral rebound at 5% significance level.

Table of Contents

Abstract . . . . **iii**

Table of Contents . . . . **iv**

List of Tables . . . . **vii**

List of Figures . . . . **viii**

1 Introduction **1**

2 Review of Models for Longitudinal and Survival Data **3**

2.1 Models for Longitudinal Data 3

2.1.1 Linear Mixed Effects Model 4

2.1.2 Other Types of Mixed Effects Model 8

2.1.3 Model Diagnostics 9

2.2 Models for Survival Data 9

2.2.1 Hazard and Survival Function 10

2.2.2 Cox Proportional Hazard Model 10

3 Joint Modelling of Longitudinal and Survival Data 12

3.1 Introduction to Joint Model 12

3.2 Model Specification 13

3.2.1 Association Structures 14

3.3 Statistical Inference 16

3.3.1 Likelihood Method 17

3.3.2 Bayesian Method 17

3.4 Bayesian Parameter Estimation 24

3.4.1 Posterior Mean and Standard Deviation 25

3.4.2 Posterior Standard Error 25

3.4.3 Individual Predictions 26

4 Analysis of HIV Data with Bayesian Joint Modelling 29

4.1 Background 29

4.2 Data Description 30

4.3 Separate Models for Longitudinal Data and Survival Data 34

4.3.1 Linear Mixed Effects Model 34

4.3.2 Cox PH Model 41

4.4 Bayesian Joint Modeling for Longitudinal and Survival Data . . . 44

4.4.1 Joint Model with Shared Random Effects 44

4.4.2 Joint Model with Shared True Longitudinal Values $\eta_i(t)$. 49

4.4.3 Joint Model with Shared True Longitudinal Value $\eta_i(t)$ and Corresponding Slope $\eta_i'(t)$ 52

4.4.4 Joint Model with Shared Cumulative Effects 55

4.5 Discussion on the Choice of Association Structure 58

4.5.1 Model Comparison 58

4.5.2 Joint Model Diagnostics 60

4.6 Individual Predictions 68

5 Analysis of HIV Data using Joint Models Based on Likelihood Method 71

5.1 Motivation 71

5.2 Theoretical Framework 71

5.3 Model Specification and Results 73

5.4 Discussion on MLE and Bayesian Estimates 76

6 Conclusion 78

Bibliography 80

7 Appendix 83

7.1 R code 83

7.2 LME Diagnostic Plots 83

7.3 Joint Model Diagnostic Plots 87

7.3.1 Trace Plot for Analysis 2-4 87

7.3.2 ACF Plot for Analysis 2-4 91

7.3.3 Trace Plot for Analysis 14-16 95

7.3.4 ACF Plot for Analysis 14-16 97

List of Tables

Table 4.1 Data Description of HIV Dataset 30

Table 4.2 Posterior Estimates on Association Structures for Analysis 1-4 46

Table 4.3 Posterior Estimates on Association Structures for Sensitivity Analysis 1 and 2 48

Table 4.4 Posterior Estimates on Association Structures for Analysis 5-8 and Sensitivity Analysis 3-4 51

Table 4.5 Posterior Estimates on Association Structures for Analysis 9-12 and Sensitivity Analysis 5-6 53

Table 4.6 Posterior Estimates on Association Structures for Analysis 13-16 and Sensitivity Analysis 7-9 57

Table 4.7 DIC for Joint Models with Different Association Structures . . 60

Table 5.1 Likelihood and Bayesian Estimates on Association Structure . 75

List of Figures

Figure 2.1 An Example of LME Models with Random Intercept (Weekly Growth of Lab Rats) 6

Figure 2.2 An Example of LME Models with Both Random Intercepts and Random Slopes (Female Annual Income) 7

Figure 3.1 PDF of inverse Gamma Distribution with Different Shape and Scale Parameters 24

Figure 4.1 Individual Longitudinal Trajectories of (log10) RNA Viral Loads for the Duration of the Study 32

Figure 4.2 Individual Longitudinal Trajectories of (log) CD4 Cell Counts for the Duration of the Study 33

Figure 4.3 Individual RNA Viral Load (on a log10 scale) Trajectories of Eligible Patients during First Six Months of ART Treatment . 35

Figure 4.4 Individual CD4 Count (on a natural scale) Trajectories of Eligible Patients during First Six Months of ART Treatment . . . 36

Figure 4.5 Predicted and Observed RNA Levels on Selected Patients. Observed values are shown in circles. 39

Figure 4.6 Predicted and Observed CD4 Levels on Selected Patients. Observed values are shown in circles. 40

Figure 4.7 Number of People Experiencing RNA Viral Rebound 42

Figure 4.8 Number of People Experiencing RNA Viral Peak Point 43

Figure 4.9 Trace Plot for Intercept Association Structure (Analysis 1) . . 61

Figure 4.10 Trace Plot for Slope Association Structure (Analysis 1) 62

Figure 4.11 Trace Plot for Cumulative Association Structure (Analysis 13) 63

Figure 4.12 ACF Plot for Intercept Association Structure (Analysis 1) . . . 65

Figure 4.13 ACF Plot for Slope Association Structure (Analysis 1) 66

Figure 4.14 ACF Plot for Slope Association Structure (Analysis 13) 67

Figure 4.15 Estimated Conditional Survival Probabilities on Viral Rebound for Patient 1 69

Figure 4.16 Estimated Conditional RNA Viral Load (log10-scale) for Patient 1 70

Figure 7.1 Residual Plot for RNA Viral Loads 84

Figure 7.2 Residual Plot for CD4 Cell Counts 84

Figure 7.3 QQ Plot of Residuals for RNA Viral Loads 85

Figure 7.4 QQ Plot of Residuals for CD4 Cell Counts 85

Figure 7.5 QQ Plot of Random Effects for RNA Viral Loads 86

Figure 7.6 QQ Plot of Random Effects for CD4 Cell Counts 86

Figure 7.7 Trace Plot for Intercept Association Structure (Analysis 2) . . 87

Figure 7.8 Trace Plot for Slope Association Structure (Analysis 2) 88

Figure 7.9 Trace Plot for Intercept Association Structure (Analysis 3) . . 88

Figure 7.10 Trace Plot for Slope Association Structure (Analysis 3) 89

Figure 7.11 Trace Plot for Intercept Association Structure (Analysis 4) . . 89

Figure 7.12 Trace Plot for Slope Association Structure (Analysis 4) 90

Figure 7.13 ACF Plot for Intercept Association Structure (Analysis 2) . . . 91

Figure 7.14 ACF Plot for Slope Association Structure (Analysis 2) 92

Figure 7.15 ACF Plot for Intercept Association Structure (Analysis 3) . . . 92

Figure 7.16 ACF Plot for Slope Association Structure (Analysis 3) 93

Figure 7.17 ACF Plot for Intercept Association Structure (Analysis 4) . . . 93

Figure 7.18 ACF Plot for Slope Association Structure (Analysis 4) 94

Figure 7.19 Trace Plot for Cumulative Association Structure (Analysis 14) 95

Figure 7.20 Trace Plot for Cumulative Association Structure (Analysis 15) 96

Figure 7.21 Trace Plot for Cumulative Association Structure (Analysis 16) 96

Figure 7.22 ACF Plot for Cumulative Association Structure (Analysis 14) 97

Figure 7.23 ACF Plot for Cumulative Association Structure (Analysis 15) 98

Figure 7.24 ACF Plot for Cumulative Association Structure (Analysis 16) 98

Chapter 1

Introduction

In biostatistics studies, it is often of interest to model the time-to-event data together with repeated measurements which usually suffer from non-ignorable dropouts, missingness, and measurement errors. The survival and longitudinal process can be modelled jointly where the survival analysis handles informative dropouts while the mixed effects model attends to missingness and measurement errors. The joint likelihood allows for estimations on all parameters at the same time such that a more accurate inference and unbiased parameter estimates can be obtained. Individual predictions on survival probabilities and longitudinal outcomes are readily available from the joint model inference as well.

In joint models, the longitudinal and survival process can be linked in different ways through association structures. For example, the two processes can be linked through shared random effects, through current biomarker level, or through history of biomarker level, etc. The choice of association structures depends on underlying biological process. If the relationship between the two processes are unclear, statistical models with different association structures can be compared using deviance information criterion (DIC) value. A model with the smallest DIC is believed to fit the given data the best, and the corresponding association structure is likely to be the most appropriate specification of the relationship.

To derive statistical inferences of joint models, joint likelihood functions can be evaluated by either the likelihood or the Bayesian method. For the likelihood method, maximum likelihood estimates (MLE) are obtained by numerical integration techniques such as Gaussian quadrature and Monte Carlo. To achieve high efficiency on parameter estimation, Laplace approximation and EM algorithm have been developed to approximate parameter values. The likelihood method is computationally challenging and the approximation can be slow at convergence.

Another approach is the Bayesian method which has the advantage of borrowing information from other studies as a way of improving efficiency and reducing bias in parameter estimation, especially in studies with small sample size. One of the major challenges for Bayesian inference is that the posterior distribution can be computationally intensive and may not have an analytical expression for high dimensional problems. In that case, estimations of posterior distributions can be realized by Markov chain Monte Carlo (MCMC) where random samples are drawn from a proposed distribution that is constantly updated through each iteration based on the Markov chain mechanism. Posterior expectation, standard deviation, and standard error of the parameter can be summarized from the resulting Bayesian posterior distribution.

This report reviews joint modelling of longitudinal and survival data. The focus is on Bayesian joint models, but the likelihood method is briefly discussed to supplement the main body of this report. More specifically, we review joint model specifications, different association structures, the choice of prior distributions, the estimation of posterior distributions and model parameters, and the interpretation of joint models. The implementation of Bayesian joint models is illustrated with an analysis on the HIV dataset using the R package `JMbayes`. The joint model inference is further derived using the likelihood method with the R package `JM`. MLE and Bayesian estimates are compared and contrasted.

Chapter 2

Review of Models for Longitudinal and Survival Data

2.1 Models for Longitudinal Data

It is common to have repeated measurements in clinical data. Measurements that are taken over time are usually correlated within the same patient, while vary to some extent between different patients. Clinical data are frequently accompanied with missingness, measurement errors, dropouts and other characteristics that could complicate the data structure. To analyze longitudinal data, the statistical model should be able to accommodate within-individual correlation and between-individual variation while allowing for missingness and measurement errors. Some common methods on modelling the longitudinal data include mixed effects models, GEE models, transitional models, etc. The focus of this report is on mixed effects models.

Mixed effects models can be regarded as an extension of classic regression models for cross sectional data where random effects are introduced onto each individual to account for both the within individual correlations and between individual variations. The random effects provide flexibility of incorporating individual departures from population average, allowing for individual-specific as well as population-level inferences. In addition, it is impractical to have every patient's measurements taken at same time points because patients may have varying follow-up times and different times of visit. Mixed effects models can model data with unbalanced structure meaning that missingness in the response variable is allowed as long as they are missing at random.

2.1.1 Linear Mixed Effects Model

We first assume that the response is continuous and normally distributed which prompts the choice of linear mixed effects (LME) models. To elucidate the connection between LME models and regular regression models, random effects in the LME models can be viewed as extra terms that are incorporated to appropriate parameters of the regression model to address individual characteristics during the longitudinal process. Let a simple regression model with one covariate be

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n_i \tag{2.1}$$

where the response y_{ij} are for individual i at time point j, covariate x_{ij} is the measurement, β_0 is the intercept and β_1 is the parameter coefficient for x_{ij}. Individual departures from population average can be captured by including random intercepts or/and random slopes to the regression model. Random intercepts allow individuals to start with different response values while random slopes support different rates of progression. Examples of LME models with random intercepts and LME models with both random intercepts and random slopes can be found in the following subsections.

To express LME models in matrix formats, let $\boldsymbol{y_i} = (y_{i1}, \ldots, y_{in_i})^T$ be n_i repeated measurements taken at different time points for individual i where $i = 1, ..., n$. Since unbalanced data can be handled by mixed effects models, the number of measurements n_i can vary for different individuals. Suppose that the mixed effects model has p fixed effects and q random effects.

$$\boldsymbol{y_i} = X_i \boldsymbol{\beta} + Z_i \boldsymbol{b_i} + \boldsymbol{e_i} \tag{2.2}$$

$$\boldsymbol{b_i} \sim N(0, D), \quad \boldsymbol{e_i} \sim N(0, R_i) \tag{2.3}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ represents the population parameters or the fixed-effects such as gender, age at recruitment, treatment arm, etc, $\boldsymbol{b_i} = (b_{i1}, \ldots, b_{iq})^T$ is the random effects that are used to model individual departures from the mean measurement on the population level, X_i which is of dimension $n_i \times p$ and Z_i which is of dimension $n_i \times q$ are corresponding design matrices for the fixed and random effects, respectively, $\boldsymbol{e_i} = (e_{i1}, \ldots, e_{in_i})^T$ are random errors of repeated measurements for the i-th individual, D which is of dimension $q \times q$ and R_i which is of dimension $n_i \times n_i$ are variance-covariance matrices for random effects b_i and random errors e_i. Assumptions for the above model state that individuals are independent of each other; and that random effects $\boldsymbol{b_i}$ and random errors $\boldsymbol{e_i}$ are independently and normally distributed with mean zero and variance-covariance matrix D and R_i.

LME Models with Random Intercepts

To give an example of when to use LME models with random intercepts, the weekly weight growths of lab rats are examined. Figure 2.1 depicts weekly growth of rats measured by their weights with each rat represented by a grey line and the average weight represented by a red line. As can be seen from the plot, the initial weight for each rat varies at the beginning of the experiment, but the growth rate appears to be similar across all rats. In this case, individual deviations can be characterized by adding random intercepts to β_0 in Equation 2.1 such that $\beta_{0i} = \beta_0 + b_i$. On the other hand, random slopes seem unnecessary considering that they all grow at comparable rates. The LME model with random intercepts is written as

$$\begin{aligned} y_{ij} &= \beta_{0i} + \beta_1 x_{ij} + e_{ij} \\ &= (\beta_0 + b_i) + \beta_1 x_{ij} + e_{ij} \\ &= (\beta_0 + \beta_1 x_{ij}) + b_i + e_{ij} \\ &= X_i \boldsymbol{\beta} + Z_i \boldsymbol{b_i} + \boldsymbol{e_i} \qquad (2.4) \\ &= \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{ij} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} b_i + \begin{pmatrix} e_{i1} \\ \vdots \\ e_{ij} \end{pmatrix} \end{aligned}$$

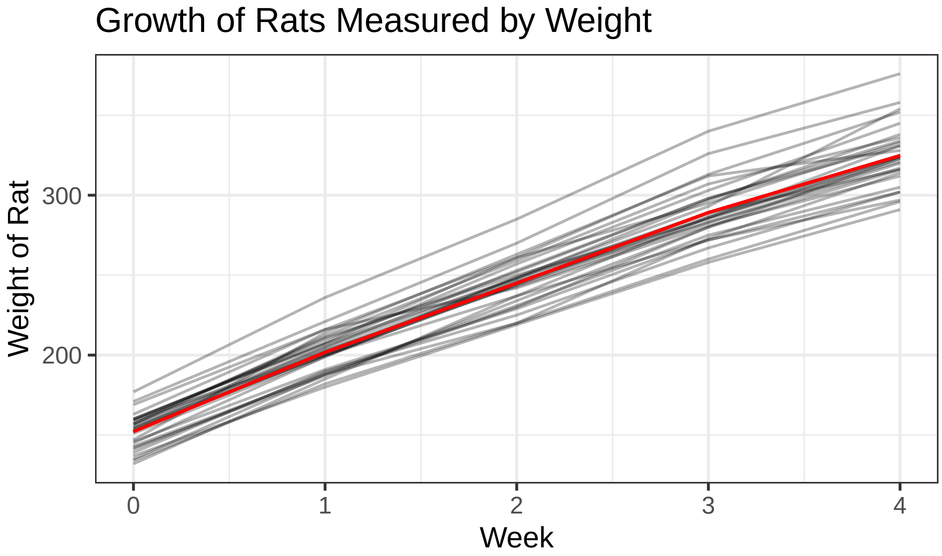

Figure 2.1: An Example of LME Models with Random Intercept (Weekly Growth of Lab Rats)

LME with Random Slopes and Random Intercepts

Another type of situation is shown in Figure 2.2 where it demonstrates the change of annual income for selected American female workers from 1968 to 1990. In this example, individuals started off earning different income in 1968. Over the duration of the study, they have experienced drastically different changes in terms of their salary.

Thus, besides from adding random intercepts $\beta_{0i} = \beta_0 + b_{0i}$, it is appropriate to include random slopes as well, that is $\beta_{1i} = \beta_1 + b_{1i}$. The LME model with both random intercepts and random slopes can be written as

$$\begin{aligned}
y_{ij} &= \beta_{0i} + \beta_{1i}x_{ij} + e_{ij} \\
&= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})x_{ij} + e_{ij} \\
&= (\beta_0 + \beta_1 x_{ij}) + (b_{0i} + b_{1i}x_{ij}) + e_{ij} \\
&= X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{e_i} \qquad (2.5) \\
&= \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{ij} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{ij} \end{pmatrix} \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ \vdots \\ e_{ij} \end{pmatrix}
\end{aligned}$$

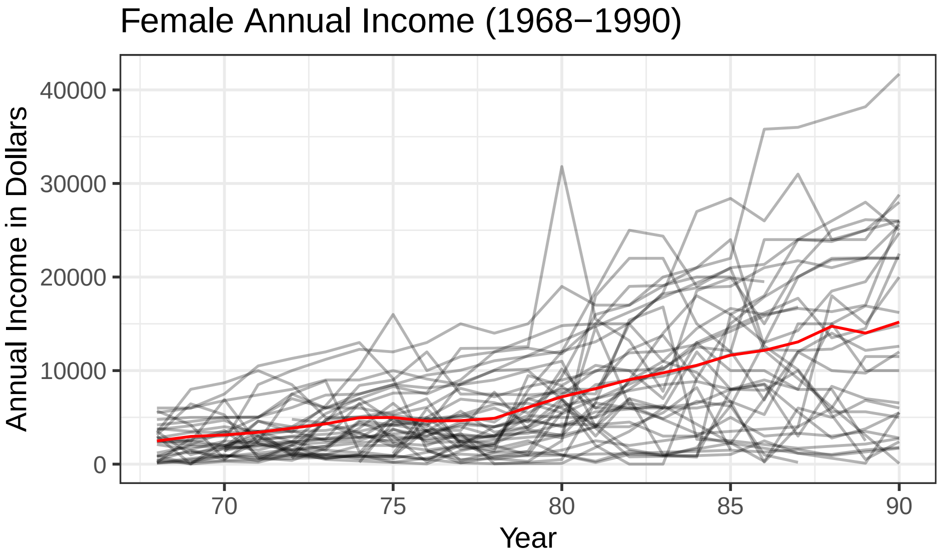

Figure 2.2: An Example of LME Models with Both Random Intercepts and Random Slopes (Female Annual Income)

2.1.2 Other Types of Mixed Effects Model

Generalized Linear Mixed Effects Models

The response variable is not always continuous and normally distributed, but rather it could be binary or count data. In these cases, the response should be modelled with binomial or Poisson distribution. A generalized linear mixed model (GLMM) is proposed to model responses whose distributions belong to the exponential family which includes but not limited to normal, gamma, Wishart, binomial, Poisson distribution. The GLMM is expressed as

$$g(E(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b_i})) = \eta_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} \tag{2.6}$$

$$\boldsymbol{b_i} \sim N(0, D) \tag{2.7}$$

where the link function $g(.)$ is monotonic and differentiable and it links the conditional expectation of the response to the linear predictor η_i.

Non-linear Mixed Effects Model

Linear models that are discussed above fit the data without having to understand the underlying scientific justifications. In some other cases, for example, in pharmacokinetics studies, the biological process has been well-studied and the relationship between the response and covariates has been established. Typically, the relationship which is based on the biological mechanism is nonlinear. This type of model called nonlinear mixed effects model (NLME model) not only provides a better fit to current data, but also produces more accurate predictions than LME models. An NLME model is formulated as follows,

$$\boldsymbol{y_i} = g(\boldsymbol{t_i}, \boldsymbol{\beta_i}) + \boldsymbol{e_i} \tag{2.8}$$

$$\boldsymbol{\beta_i} = h(\boldsymbol{x_i}, \boldsymbol{\beta}, \boldsymbol{b_i}), \quad \boldsymbol{b_i} \sim N(0, D), \quad \boldsymbol{e_i} \sim N(0, R_i) \tag{2.9}$$

where $g(.)$ is a known nonlinear function, $h(.)$ is usually a linear function that links covariates $\boldsymbol{x_i}$, fixed effects $\boldsymbol{\beta}$, and random effects $\boldsymbol{b_i}$. Model assumptions are the same as that of linear mixed effects models.

2.1.3 Model Diagnostics

Assumptions for LME models include homogeneity of residual variance, normality of random effects and normality of residuals. A residual plot showing no observable patterns suggests variance homoscadasticity. Quantile-quantile plot (QQ plot) is used to check for normality. If the points lie roughly on a straight line, then there may be no violation of the normality assumption. A bonus from the QQ plot is that individuals with unusual measurements (i.e., outliers) can be identified visually.

On the other hand, diagnostics for GLMM is a bit different. Unlike a linear model, for logistic and Poisson distribution, variance of the response is not constant and is bounded by a mean-variance relationship. Thus, the Pearson or deviance residuals instead of regular residual is used. Additionally, it is necessary to check for over-dispersion problems where empirical variance is larger than its theoretical value. In the presence of over-dispersion (and in rare cases, under-dispersion), a dispersion parameter can be introduced to correct the mean-variance relationship and a quasi-binomial or quasi-Poisson distribution is used instead.

2.2 Models for Survival Data

To identify the relationship between survival times and covariates, it is often of interest to model the time-to-event data. For example, we can study if the dropout is associated with a certain treatment or side effect to determine whether the dropout is informative; or we can study if time to death is related to the level of a certain biomarker. The event of interest may be death, relapse, dropout, and other events of interest.

One of the main characteristics of survival data is that it is usually censored meaning that the occurrence of such an event is not observed for some individuals. If individuals have not yet experienced such an event before the termination of the study or before them dropping out, it is called right-censored. In medical research, it is common to have right censoring which should be accounted for in statistical analyses. Another important feature of survival data is that it is often right skewed because the majority of participants are expected to experience the event at an earlier time with a few outliers whose survival times are longer than average. In addition, the follow-up times likely vary among individuals since people enter and leave the study at different time points. To accommodate such features, regression models for modelling survival data, such as Cox proportional hazard model (Cox

PH model) and accelerated failure time model (AFT model), are commonly used.

2.2.1 Hazard and Survival Function

For n individuals, survival data is arranged as $\{(t_i, \delta_i), i = 1, ..., n\}$ where t_i is the smaller of the survival time and censoring time, δ_i is the censoring indicator. $\delta_i = 1$ means that the event is observed for person i while $\delta_i = 0$ means that the event has yet to be observed or it is right censored.

Unlike regular regression models where the mean of the response is modelled, in survival analyses, the hazard or the risk of an event is modelled instead. This is because researchers are usually more interested in the probability of someone surviving past a certain time point rather than the expected time of an event happening. The hazard function

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t}, \quad t > 0 \tag{2.10}$$

is defined as the probability of someone experiencing an event at time t given that they have survived up to that time point.

Alternatively, the survival process can be modelled by the survival function

$$S(t) = P(T \geq t), \quad t > 0 \tag{2.11}$$

which describes the probability of someone surviving up to time point t. It is easy to show that the relationship between the hazard function and survival function is

$$h(t) = -\frac{d\text{log}S(t)}{dt}, \quad \text{or} \quad S(t) = \text{exp}[-\int_0^t h(s)ds] \tag{2.12}$$

2.2.2 Cox Proportional Hazard Model

Cox PH models are commonly used for modelling the hazard function described in Equation 2.10.

Let $h_i(t)$ be the hazard function for individual i, and $h_0(t)$ be the unspecified baseline hazard function that describes the hazard when all covariates are set to zero. $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^T, \quad i = 1, \ldots, n$ is a vector of covariates and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is the corresponding parameter. A Cox PH model is written as

$$\begin{aligned} h_i(t) &= h_0(t)\text{exp}(\boldsymbol{x_i}^T \boldsymbol{\beta}) \\ &= h_0(t)\text{exp}(\beta_1 x_{i1} + \cdots + \beta_p x_{ip}) \end{aligned} \tag{2.13}$$

where exp$(\boldsymbol{x_i^T \beta})$ can be interpreted as the relative risk associated with covariate $\boldsymbol{x_i}$. Since $h_i(t)$ and $h_0(t)$ are taken to be nonparametric, no distributional assumptions are imposed on the data itself. In other words, Cox PH models provide the flexibility of accommodating survival data of any distribution.

One assumption for Cox PH models is that the hazard ratio $\frac{h_i(t)}{h_0(t)}$ is constant with respect to time, therefore, the so-called proportional hazard assumption. That is, the increase or decrease in the risk of an event happening is the same at all times. For model diagnostics, the proportional hazard assumption is examined with Schoenfeld residual plot. A plot showing no clear patterns of Schoenfeld residual against time is an indication of no violation to the model assumption. However, if the hazard is not proportional, alternative models such as AFT models should be used instead.

Chapter 3

Joint Modelling of Longitudinal and Survival Data

3.1 Introduction to Joint Model

In research studies, repeated measurements often arise together with time-to-event data. However, statistical inferences could be invalid if the longitudinal and survival process are analyzed separately. In particular, when the primary focus is on survival probabilities, the longitudinal outcome is regarded as a time-dependent variable which is directly affected by the occurrence of the event. However, traditional methods for analyzing survival data requires the time-dependent variables to be independent of the survival process. Modelling the two outcomes separately may lead to invalid inferences. Another scenario is when the interest is on the longitudinal outcome. Since no measurements are available after the occurrence of the event, dropouts are dependent on the survival process rendering the measurements to be missing not at random. Inferences of mixed effects models are unreliable due to the fact that the longitudinal model is unable to handle data structures when the mechanism of missingness is missing not at random.

To account for the above-mentioned issues and to obtain valid inferences, the survival and longitudinal process can be modelled jointly where missing measurements and measurement errors are addressed.

3.2 Model Specification

Joint models are designed to model the longitudinal and survival process altogether. Generalized linear mixed effects models described in Equation 2.6

$$g[E\{y_i(t)|\boldsymbol{b_i(t)}\}] = \eta_i(t) = \boldsymbol{x_i}(t)^T\boldsymbol{\beta} + \boldsymbol{z_i}(t)^T\boldsymbol{b_i} \tag{3.1}$$

are used to represent the longitudinal process where $\boldsymbol{z_i}(t)$ is the observed value of time-dependent variable at time point t with possible measurement errors. The above model approximates the true measurement at time t such that the longitudinal outcome can be linked to the survival process in the joint model.

Cox PH models Equation 2.13 are used to model the survival process

$$h_i(t) = h_0(t)\text{exp}(\boldsymbol{\gamma}^T\boldsymbol{w_i}) \tag{3.2}$$

where $\boldsymbol{w_i} = (w_{i1}, \dots, w_{ir})^T$ represents r baseline covariates with regression coefficients $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_r)^T$.

The joint model is built upon the above Cox PH model where the longitudinal component is incorporated into the relative risk term exp$(\boldsymbol{\gamma}^T\boldsymbol{w_i})$ such that

$$h_i(t) = h_0(t)\text{exp}[\boldsymbol{\gamma}^T\boldsymbol{w_i} + f\{\eta_i(t), \boldsymbol{b_i}, \boldsymbol{\alpha}\}] \tag{3.3}$$

where $f(.)$ is called the association structure that links the longitudinal and survival process (different choices of association structure can be found in Section 3.2.1), $\boldsymbol{\alpha}$ measures the association between the level of time-dependent variable, such as biomarkers, and the hazard of an event. The joint model suggests that the hazard for individual i at time t is associated with the linear predictor $\eta_i(t)$ as well as the strength of association $\boldsymbol{\alpha}$.

The non-parametric baseline hazard function $h_0(t)$ may be modelled by B-splines which are consisted of piecewise polynomial functions joined at knots [1]. Specifically, the logarithm baseline hazard function is constructed from a linear combination of B-splines

$$\text{log}h_0(t) = \gamma_{h_0,0} + \sum_{q=1}^{Q} \gamma_{h_0,q}B_q(t, \boldsymbol{v})$$

where $B_q(t, \boldsymbol{v})$ is the q-th basis function of a B-spline with knots $\boldsymbol{v}$. The number of knots should be determined strategically to avoid overfitting or underfitting problems. It has been shown that for a dataset of sample size n, a regression model with number of knots that are between $\frac{n}{10}$ and $\frac{n}{20}$ provides adequate fit to the observed data as well as reliable extrapolation of future data [2]. Locations of knots are determined by percentiles of survival times or censoring times.

3.2.1 Association Structures

The function $f(.)$ describes how the longitudinal process may be associated with the risk of the event. While the true association might be complicated, there are several common and plausible association structures that help address the relationship. The longitudinal and survival process are linked through association structures in several different ways which are dependent on the underlying biological mechanism.

Association through Shared Random Effects

When the risk of an event happening is related to individual trajectories such as individual intercepts or/and individual slopes, the longitudinal and survival process may be linked through the random effects which characterize the longitudinal process and the resulting joint model is called a *shared parameter model*. Suppose that $\boldsymbol{b_i} = (b_{1i}, \ldots, b_{mi})$ represents a collection of m random effects for the i-th individual in the mixed effects model used to model the longitudinal process, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^T$ quantifies the possible association between each random effect and the risk of the event. The association structure can be expressed as

$$\begin{aligned} f &= \boldsymbol{\alpha}^T \boldsymbol{b_i} \\ &= \alpha_1 b_{1i} + \cdots + \alpha_m b_{mi}, \quad i = 1, \ldots, n. \end{aligned} \tag{3.4}$$

Alternatively, the longitudinal process can be linked with the survival process through both random effects and their corresponding fixed effects $\boldsymbol{\beta_i} = (\beta_1, \ldots, \beta_m)$ such that

$$\begin{aligned} f &= \boldsymbol{\alpha}^T (\boldsymbol{\beta_b} + \boldsymbol{b_i}) \\ &= \alpha_1(\beta_1 + b_{1i}) + \cdots + \alpha_m(\beta_m + b_{mi}), \quad i = 1, \ldots, n. \end{aligned} \tag{3.5}$$

Equation 3.4 are preferred when there are reasons to believe that besides other time-independent covariates, the survival process is associated with individual trajectory characteristics such as individual intercepts and slopes which are captured by random effects.

Association through the Linear Predictor $\eta_i(t)$

Sometimes the risk of an event may be directly associated with the current value, or the current unobserved true value in case of measurement error, of the longitudinal process. In this case, it may be reasonable to assume the following association structure,

$$f = \alpha\eta_i(t) \tag{3.6}$$

where α measures the association between the risk of an event for individual i at time t, and the current true longitudinal measurements.

This specification is preferable when the survival probability is best reflected by the current level of time-dependent variables.

Association through the Current Value $\eta_i(t)$ and/or the Current Rate of Change $\eta_i'(t)$

The third option incorporates both the current value $\eta_i(t)$ and the rate of change $\eta_i'(t)$ into the association structure such that

$$f = \alpha_1\eta_i(t) + \alpha_2\eta_i'(t) \tag{3.7}$$

where α_1 represents the association between the hazard and current true value of longitudinal measurements, and α_2 represents the association between the hazard and its rate of change at time t.

This specification is more suitable in describing scenarios where we believe that the risk of an event may be explained by both the current level of time-dependent variables and its rate of change.

Association through the History of the Longitudinal Process

The last option links the risk with cumulative effects of the longitudinal outcome such that

$$f = \alpha \int_0^t \eta_i(s) ds \tag{3.8}$$

where α quantifies the association between the hazard and cumulative measurements of time-dependent variables.

The first and the third options are similar in a way that they both believe that the hazard is associated with how fast the measurement level changes. The difference lies in that the first option uses the baseline measurement while the second option concerns with the current measurement level at time t. However, these settings may be inadequate for more complicated cases where the hazard is also related to previous influences of the longitudinal measurements up to time t. The third option which considers the history of longitudinal process is more relevant in addressing such a scenario.

As mentioned earlier, the true association between the longitudinal and survival process, if exists, might be highly complicated. However, the foregoing possible association structures are both mathematically simple and scientifically reasonable. Naturally, there could be other association structures which might not be easily represented by an explicit mathematical expression. For the purpose of this project, we will mostly focus on the foregoing association structures for analyzing the HIV dataset in later chapters.

3.3 Statistical Inference

Several statistical techniques can be used to make inferences of the joint model given by Equation 3.3. For this project, we will focus on two methods, namely the likelihood method and the Bayesian method.

3.3.1 Likelihood Method

The likelihood method maximizes the joint likelihood function

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(t_i, \delta_i | \boldsymbol{b_i}, \boldsymbol{\theta}) f(\boldsymbol{y_i} | \boldsymbol{b_i}, \boldsymbol{\theta}) f(\boldsymbol{b_i} | D) d\boldsymbol{b_i} \tag{3.9}$$

and finds the maximum likelihood estimates for unknown parameters $\boldsymbol{\theta}$. By estimating all parameters simultaneously, more accurate inferences can be derived. The major challenge for the likelihood method concerns the computational difficulties as the joint likelihood function might be intractable, especially for high-dimensional problems. Numerical integration techniques such as Gaussian quadrature and Monte Carlo methods can be used to approximate the integral when the above likelihood function does not have an analytical solution. However, computation may be inefficient if a large number of parameters are involved in the model. Another approach is to apply the EM algorithm where the random effects are considered as latent variables. But we might encounter convergence problems.

On a side note, joint model inferences can also be derived using the two-stage method where the longitudinal and survival process are modelled separately. Estimations on the true level of time-dependent variables obtained from the mixed effects model are used to fit the survival model. The advantage of this method lies in its implementation simplicity. The pitfall is that it may produce biased estimates. What is more, since the longitudinal and survival process are analyzed separately, the relationship between survival probabilities and longitudinal measurements cannot be established [3].

3.3.2 Bayesian Method

Apart from the likelihood method and two-stage method mentioned above, Bayesian approach with prior knowledge incorporated can estimate parameters in a more efficient way.With information from both the prior knowledge and the data, posterior distributions on unknown parameters can be derived. The essence of Bayesian inference is based on the Bayes' Rules where parameters $\boldsymbol{\theta}$ are treated as random variables and the corresponding posterior distributions $f(\boldsymbol{\theta}|\boldsymbol{x})$ depend on both the prior distributions $f(\boldsymbol{\theta}|\boldsymbol{\theta_0})$, which is determined by hyperparameters $\boldsymbol{\theta_0}$, and the likelihood function $f(\boldsymbol{x}|\boldsymbol{\theta})$. More specifically, the Bayes' Rules state that

$$f(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{3.10}$$

where the normalizing constant $f(\boldsymbol{x}) = \int f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the information provided by the data and it is independent of parameters $\boldsymbol{\theta}$. Thus, the posterior distribution is proportional to the expression on the numerator, that is

$$f(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}). \tag{3.11}$$

Bayesian Inference for Joint Models

For Bayesian inference of the joint model, the validity of the posterior distribution relies on the following assumptions. Study subjects are assumed to be independent of each other, and given the random effects, both the longitudinal and survival process are believed to be independent. Formally, for joint models with unknown parameters $\boldsymbol{\theta}$ and random effects $\boldsymbol{b_i}$, the posterior distribution of parameter space is written as

$$f(\boldsymbol{\theta}, \boldsymbol{b}) \propto \prod_{i=1}^{n} f(t_i, \delta_i|\boldsymbol{b_i}, \boldsymbol{\theta})f(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\theta})f(\boldsymbol{b_i}|D)f(\boldsymbol{\theta}). \tag{3.12}$$

The first term describes the survival process which is modelled by

$$f(t_i, \delta_i|\boldsymbol{b_i}, \boldsymbol{\theta}) = h_i(t_i)^{\delta_i}S(t) = h_i(t_i)^{\delta_i}\text{exp}\{-\int_0^{t_i} h_i(s)ds\} \tag{3.13}$$

where $h_i(.)$ is given by Equation 3.3 and it describes the survival function for the i-th individual. Since the above integral may be intractable, numerical integration such as the Gauss-Kronrod and Gauss-Legendre quadrature rule can be used for approximation.

The second term describes the longitudinal process which is modelled by

$$f(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\theta}) = \prod_{l=1}^{n_i} f(y_{il}|\boldsymbol{b_i}, \boldsymbol{\theta}) = \prod_{l=1}^{n_i} \text{exp}(\frac{y_{il}\psi_{il}(\boldsymbol{b_i}) - c[\psi_{il}(\boldsymbol{b_i})]}{a(\varphi)} - d(y_{il}, \varphi)) \tag{3.14}$$

for some known function $c(.)$, $a(.)$, and $d(.)$. The above specification is the general form of probabilistic distributions for response variables that are in the exponential family with $\psi_{il}(\boldsymbol{b_i})$ and φ being the canonical and dispersion parameter, respectively.

Markov Chain Monte Carlo

When evaluating the posterior distribution, computational difficulty lies in the calculation of the normalizing constant as it may not have a closed-form solution if the parameter space $\boldsymbol{\theta}$ is high dimensional. Additionally, point estimates such as expectation of the parameter space and marginal distribution of a certain parameter θ_j might be intractable as well. To evaluate intractable integrals, Markov chain Monte Carlo (MCMC) can be used to numerically approximate high-dimensional integrals.

MCMC is a technique where samples are drawn from a proposal distribution that is constantly updated through each iteration based on the Markov chain mechanism. After a sufficient amount of iterations, the Markov chain converges to a stationary distribution that can be used to approximate the true posterior distribution [4]. Samples from the first few iterations, or the so-called burn-in period, are usually discarded to minimize the influence of starting values. The Monte Carlo samples $x^{(i)}$ behave in the same way as if they were drawn from the true posterior distribution provided that large number of iterations are achieved. Thus, Monte Carlo samples sampled from the post-convergence distribution can be used for deriving Bayesian inference and the calculations such as posterior means, posterior standard deviations, posterior standard errors, Bayesian credible intervals, and other summary statistics are made possible [5].

As the name suggests, MCMC is composed of Monte Carlo sampling and Markov chain.

Monte Carlo simulation is the process for random selections of independent and identical samples from a target distribution. It allows us to obtain parameter estimation without having to undergo deterministic integration which may be intractable for high-dimensional problems. In the Bayesian inference context, the posterior distribution of unknown parameters is the target distribution that we are interested in sampling from. Suppose N i.i.d samples $\{x^{(i)}\}_{i=1}^{N}$ are obtained from the posterior distribution by Monte Carlo sampling. When N is large enough, point estimates for $E(\theta)$ can be approximated empirically by the sample mean, that is

$$E(\theta) = \frac{1}{N}\sum_{i=1}^{N} x^{(i)}. \tag{3.15}$$

Similarly, the intractable integrals $I(f)$ is approximated with a tractable sum

$$I(f) = \frac{1}{N}\sum_{i=1}^{N} f(x^{(i)}) \tag{3.16}$$

for some function of interest f. However, the posterior distribution remains unknown at the time of sampling meaning that it is impossible to sample directly from the posterior. Instead, a proposal distribution which is believed to resemble the posterior distribution can be used to draw Monte Carlo samples from. Such a compromise leads to another problem in that the proposal distribution does not equate to the posterior distribution rendering the estimation inaccurate. Thus, a Markov chain is introduced to the Monte Carlo sampling process such that it allows for ongoing updates on unknown parameters by either accepting or rejecting the next Monte Carlo sample. Specifically, a Markov chain is a stochastic process where the probability of the next event depends solely on the state of the current event, that is $p(x^{(i)}|x^{(i-1)}, ..., x^{(1)}) = p(x^{(i)}|x^{(i-1)})$. The stationary distribution of the Markov chain would be a good approximation of the posterior distribution and the Monte Carlo samples would imitate samples drawn from the true posterior distribution.

Metropolis-Hasting algorithm

The decision of whether a chain moves to the next state is based on the MCMC algorithm which determines whether the new sample should be accepted or rejected. One of the most popular algorithms is the Metropolis-Hastings algorithm which is described as follows (we will use a one-dimensional parameter space for demonstration):

1. Select starting value θ_0
2. For $i = 1, ..., n$, repeat
    (a) Sample $u \sim \text{Unif}[0, 1]$

(b) Sample $\theta^* \sim q(\theta^*|\theta_{i-1})$

(c) If $u < \min\{1, \frac{g(\theta^*)q(\theta_{i-1}|\theta^*)}{g(\theta_{i-1})q(\theta^*|\theta_{i-1})}\}$, $\theta_i = \theta^*$ else $\theta^i = \theta^{(i-1)}$

Let $g(\theta) = f(x|\theta)f(\theta)$ be the target distribution up to a normalizing constant which means that $g(\theta)$ is proportional to the posterior distribution $f(\theta|x)$. The starting value for the parameter is selected as θ_0. A new value θ^* is drawn from the proposal distribution $q(\theta^*|\theta_{i-1})$ whose probabilistic distribution is conditioned on the previous value θ_{i-1} for the i-th iteration. The term $\frac{g(\theta^*)q(\theta_{i-1}|\theta^*)}{g(\theta_{i-1})q(\theta^*|\theta_{i-1})}$ is the acceptance probability. If the numerator is greater than the denominator meaning that the newly sampled value θ^* is more likely to happen (or closer to the true value of θ) than the current value θ_{i-1}, then the newly sampled value θ^* is accepted and the proposal distribution is updated accordingly. On the other hand, if the numerator is smaller than the denominator meaning that the current value θ_{i-1} is more likely to happen, the newly sampled value is only accepted if this ratio is greater than u where u is drawn from a uniform distribution defined on the interval 0 to 1. Otherwise, θ^* is rejected and θ_i takes the same value as that of θ_{i-1}. Under this mechanism, a better approximation to the posterior distribution is gained after each iteration. Given a sufficient number of iterations, the Markov chain will eventually converge to a stationary distribution that can be used to approximate the posterior distribution.

The choice of the proposal distribution $q(\theta)$ requires careful consideration. An ideal $q(\theta)$ should allow for fast convergence while maintaining a reasonable level of sample rejection rate. In random walk Metropolis-Hasting, the proposal distribution is a normal distribution centered on the parameter value of previous iteration, θ_{i-1}, that is $q(\theta^*|\theta_{i-1}) \sim N(\theta_{i-1}, \sigma^2 I)$ where σ^2 is assigned to an appropriate value such that the proposal distribution possesses the desirable properties discussed previously.

One Long Chain vs Multiple Shorter Chains

When running the MCMC algorithm, one can invest all computing powers to elongate one single chain and obtain a stationary distribution. However, there has been concerns on whether a single chain will be stuck on one mode and never gets the chance of randomly walking to the next mode [6]. This is because by the Metropolis-Hastings algorithm, the chain would spend more time on regions with higher likelihoods while regions with low likelihoods are not explored resulting the chain not being able to move to the next region with higher likelihood. This issue would lead to an inaccurate estimation of the posterior distribution. An alternative solution to ”one long chain” method is to run multiple shorter chains with different starting values in parallel in the hope that all modes can be explored within at least one chain [7]. The downside of running multiple chains is that the length of each chain is shortened due to limited computing resources. In addition, some chains may converge relatively slowly given a certain set of starting points.

Prior Distribution and Default Prior for Joint Model

Prior distributions represent prior scientific knowledge or researchers' own belief about the parameter before data collection or statistical analyses. The resulting posterior distribution is under the influence of both the data and the prior distribution. With large sample size, effects of the prior on the posterior are attenuated meaning that the posterior is primarily determined by the data. In other cases, however, different choices of the prior could affect the posterior and the parameter estimation in a major way. To avoid distorted conclusions derived under unsuitable priors and to check if the posterior is sensitive to different prior specifications, posterior inferences under different priors can be compared using a sensitivity analysis [8].

Some popular choices of the prior include non-informative/diffuse/flat prior which assigns approximately equal weights to every region of the likelihood function and it is used when no prior knowledge is available or preferred. Apart from flat priors, there is a family of prior distributions called the conjugate prior which provides computational convenience. The posterior distribution under the conjugate prior is in the same probabilistic distribution family as the prior, and it ensures that the posterior can be computed analytically.

For prior distributions used in Bayesian inferences for joint models, default choices of prior distributions for fixed effects coefficient $\boldsymbol{\beta}$, regression coefficient of the survival model $\boldsymbol{\gamma}$, association parameter $\boldsymbol{\alpha_1}$, slope association parameter $\boldsymbol{\alpha_2}$, and spline coefficient of baseline hazard function $\boldsymbol{\gamma_{h_0}}$ are taken to be normal distributions. The mean value of the normal distribution is centered at zero meaning that the prior assumes that the explanatory variables are not associated with the response. Specifically, default choices for the above parameters are as follows,

$$\boldsymbol{\beta} \sim N\left(\mathbf{0}, a_{\beta} I\right), \quad \boldsymbol{\gamma} \sim N\left(\mathbf{0}, a_{\gamma} I\right) \tag{3.17}$$

$$\boldsymbol{\alpha_1} \sim N\left(\mathbf{0}, a_{\alpha_1} I\right), \quad \boldsymbol{\alpha_2} \sim N\left(\mathbf{0}, a_{\alpha_2} I\right) \tag{3.18}$$

$$\boldsymbol{\gamma_{h_0}} \sim N\left(\mathbf{0}, a_{h_0} I\right) \tag{3.19}$$

where a_β, a_γ, a_{α_1}, a_{α_2}, and a_{h_0} are constant of one's own choice. They can be set to large values leading to large variance if researchers prefer not to impose prior knowledge on the estimation. Such priors are called non-informative or flat priors as discussed previously. Here, non-informative means that the variance of such prior distributions is considerably large which gives relatively flat-shaped distributions, favoring no particular regions of the likelihood function.

For the variance of error terms, we use an inverse Gamma prior

$$f_x(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{x}\right)^{\alpha+1}\exp\left(-\frac{\beta}{x}\right)$$

where α and β are the shape and scale parameter, respectively, and $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ is the Gamma function. The inverse Gamma distribution is commonly used as a conjugate prior for the variance structure of a normal distribution [9]. Let the residual standard deviation be σ_r, the values for the shape and scale parameter be $\alpha = \beta = \frac{1}{a_\sigma \sigma_r^2}$ where a_{σ_r} is a constant of one's choice. The prior for the variance of error terms is

$$R_i \sim IG\left(\frac{1}{a_\sigma\sigma_r^2}, \frac{1}{a_\sigma\sigma_r^2}\right). \tag{3.20}$$

Figure 3.1 illustrates the density plot of inverse Gamma Distribution with different choices of shape and scale parameter. When small values are assigned to these parameters, the density plot gives a relatively flat shape. That means a_{σ_r} should be set to a comparably large value if a non-informative prior is preferred.

Inverse Wishart distribution is the generalization of the inverse Gamma distribution for the multivariate cases. Thus, for the variance-covariance matrix of random effects, we use an inverse Wishart prior $\mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$ which is the conjugate prior for the covariance matrix of a multivariate normal distribution [10]. The inverse Wishart distribution has the density function

$$f_{\boldsymbol{x}}(\boldsymbol{x};\boldsymbol{\Psi},\nu) = \frac{|\boldsymbol{\Psi}|^{\frac{\nu}{2}}}{2^{\nu p/2}\Gamma_p(\frac{\nu}{2})}|\boldsymbol{x}|^{\frac{-(\nu+p+1)}{2}}e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}\boldsymbol{x}^{-1})}$$

where $\boldsymbol{x}$ and the scale matrix $\boldsymbol{\Psi}$ are $p \times p$ positive definite matrices, ν is the degree of freedom. Since random effects $\boldsymbol{b_i}$ for $i = 1, ..., n$ are assumed to follow a multivariate normal distribution, $\boldsymbol{b_i} \sim N(\boldsymbol{0}, \boldsymbol{D})$, and the prior for variance covariance matrix $\boldsymbol{D}$ follows a Wishart distribution $\mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$, the posterior distribution $f(\boldsymbol{D}|\boldsymbol{b_i})$ would then follow an inverse Wishart distribution $\mathcal{W}^{-1}(\boldsymbol{b_i}\boldsymbol{b_i^T} + \boldsymbol{\Psi}, n+\nu)$.

Let the number of random effects (including random intercept and random slopes)

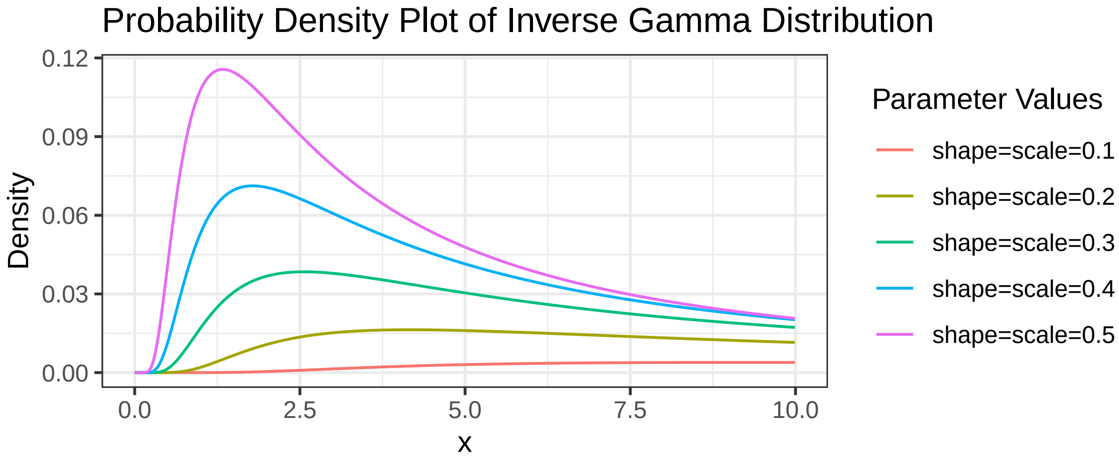

Figure 3.1: PDF of inverse Gamma Distribution with Different Shape and Scale Parameters

be m, the variance-covariance matrix of the random effects calculated from the longitudinal process be $\boldsymbol{d}$, and $\boldsymbol{d}^{-1}$ be the inverse of matrix $\boldsymbol{d}$. For hyperparameters of the inverse Wishart distribution, the degree of freedom ν is set as m, and the precision parameter $\boldsymbol{\Psi}$ is set to be $m\boldsymbol{d}^{-1}$. Thus, the prior for variance-covariance matrix of random effects $\boldsymbol{D}$ is distributed as

$$\boldsymbol{D} \sim \mathcal{W}^{-1}(m\boldsymbol{d}^{-1}, m) \tag{3.21}$$

As stated earlier, for the purpose of diminishing subjectivity in choosing priors, sensitivity analyses can be conducted to test whether the obtained results are robust to prior distributions with different hyperparameters.

3.4 Bayesian Parameter Estimation

Standard deviations are commonly used in frequentist statistics as a way of quantifying uncertainty of an estimator. The concept can be extended to Bayesian inferences where posterior standard deviations can be used to compare accuracy of Bayesian posterior expectations under different prior distributions[11]. On the other hand, posterior standard errors are used to assess how accurate an estimate is compared to its true value. Thus, posterior standard deviations and posterior standard errors provide some insights on selecting priors that can give rise to relatively accurate parameter estimations, though they are not the only criteria for choosing the prior. The evaluation of prior distributions also depends on the consistency of posterior distribution, the rate of convergence, etc.

3.4.1 Posterior Mean and Standard Deviation

While posterior mean can be computed directly from Bayes rules, the integrals involved in posterior distributions and posterior expectations may be intractable in high-dimensional problems. Alternatively, the posterior distribution and expectation can be estimated by MCMC samples.

To illustrate the calculation, suppose that θ is the parameter of interest, and $\theta_1, ..., \theta_n$ are MCMC samples drawn from proposal distributions. The posterior expectation and standard deviation are estimated as follows,

$$E(\theta|Y) = \hat{\theta} = \overline{\theta_n} = \frac{\sum_{i=1}^{n} \theta_i}{n} \tag{3.22}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (\theta_i - \hat{\theta})^2}{n-1}} \tag{3.23}$$

Bayesian credible intervals can be calculated as well to assess if the parameter of interest is a significant predictor to the response. The definition and calculation of Bayesian credible intervals are comparable to frequentist confidence interval. The lower and upper bound of a 95% credible interval are the 2.5th and 97.5th quantile of the posterior distribution. However, the Bayesian credible interval and the frequentist confidence interval differ interpretation-wise. A 95% credible interval means that given the observed data, there is a 95% chance that the true value of the parameter lies within this interval. Despite of the difference in interpretation, the credible interval can be used in the same manner as the confidence interval to determine whether an explanatory variable is statistically significant. If the value zero is excluded from the credible interval, then it provides evidence of this variable being significantly associated with the response [12].

3.4.2 Posterior Standard Error

Posterior standard errors are estimated with the help of the effective sample size using the time series methodology [13]. MCMC samples can be treated as data points aroused from time series where samples are autocorrelated within each chain. Dependency among samples would inflate the variance of parameters, and the effects of which is accounted for by the effective sample size, N_{eff} [14]. Recall that for independent and identical random samples, the standard error is $se = \frac{\hat{\sigma}}{\sqrt{n}}$ where n is the sample size. The MCMC analogue to this statement replaces sample size n with N_{eff} to adjust for autocorrelation. The effective sample size can be estimated with

$$\hat{N_{eff}} = \frac{M \times N}{\hat{\tau}} \tag{3.24}$$

where M is the number of Markov chains, N is the number of autocorrelated samples, and $\hat{\tau}$ is an expression involving estimated autocorrelation. Then, the posterior standard error is computed as

$$\hat{se} = \frac{\hat{\sigma}}{\sqrt{\hat{N_{eff}}}}. \tag{3.25}$$

The smaller the posterior standard error, the closer the parameter estimate is to its true value. Therefore, an estimate with a smaller posterior standard error is more accurate at parameter estimation.

3.4.3 Individual Predictions

Medical doctors are often interested in patients' prognoses in order to personalize their treatments. Joint models allow for individual-specific predictions on both the survival and longitudinal outcome. For the j-th patient, let the baseline covariates be $\boldsymbol{w}_j$ and the longitudinal measurements be $\mathcal{Y}_j(t) = \{y_j(t_{jl}); 0 \leq t_{jl} \leq t, l = 1, \ldots, n_j\}$ where t_{jl} records the time points when measurements were taken.

For the survival process, physicians are usually interested in predicting the probability of patient j surviving up to time point u where $u > t$, that is

$$\pi_j(u|t) = P(T_j^* \geq u | T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{w}_j, \mathcal{D}_n) \tag{3.26}$$

where T_j^* and T_j are the true event time and observed event time (recall that $T_j = \min(T_i^*, C_i)$), respectively, and $\mathcal{D}_n = \{T_i, \delta_i, \boldsymbol{y}_i; i = 1, \ldots, n\}$ denotes a set of sample from the target population upon which the joint model is built.

For the longitudinal process, the level of time-dependent variables $\omega_j(u|t)$ at time

u can also be predicted based on the fitted joint model, that is

$$\omega_j(u|t) = E\{y_j(u)|T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n\} \tag{3.27}$$

Under the Bayesian framework, the first-order estimate of $\pi_j(u|t)$ is

$$\begin{aligned}
\tilde{\pi}_j(u|t) &= \int P(T_j^* \geq u|T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D}_n)d\boldsymbol{\theta} \\
&= \int\int P(T_j^* \geq u|T_j^* > t, \boldsymbol{b}_j, \boldsymbol{\theta})p(\boldsymbol{b}_j|T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}_n})d\boldsymbol{b}_j d\boldsymbol{\theta} \\
&= \int\int \frac{S_j\{u|\mathcal{H}_j(u, \boldsymbol{b}_j, \boldsymbol{\theta})\}}{S_j\{t|\mathcal{H}_j(t, \boldsymbol{b}_j, \boldsymbol{\theta})\}} p(\boldsymbol{b}_j|T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}_n})d\boldsymbol{b}_j d\boldsymbol{\theta} \\
&\approx \int \frac{S_j\{u|\mathcal{H}_j(u, \hat{\boldsymbol{b}}_j, \hat{\boldsymbol{\theta}}); \hat{\boldsymbol{\theta}}\}}{S_j\{t|\mathcal{H}_j(t, \hat{\boldsymbol{b}}_j, \hat{\boldsymbol{\theta}}); \hat{\boldsymbol{\theta}}\}} p(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}_n})d\boldsymbol{\theta}
\end{aligned} \tag{3.28}$$

where $S_j(.)$ is the survival function at a certain time point, $\mathcal{H}_j(.)$ is the longitudinal history up to a certain time point, and $\boldsymbol{b}_j$ and $\boldsymbol{\theta}$ are the random effects and full parameter space which are both estimated by MCMC. To avoid complicated calculation involving high-dimensional integrals, $\tilde{\pi}_j(u|t)$ is estimated from Monte Carlo samples by the following scheme [15],

1. Draw $\boldsymbol{\theta}^{(l)}$ from its posterior distribution obtained by MCMC
2. Draw $\boldsymbol{b}_j^{(l)}$ from a multivariate $t_{(4)}$ distribution centered at Bayes estimates $\hat{\boldsymbol{b_i}}$ with scale matrix $\hat{Var}(\hat{\boldsymbol{b}}_j)$
3. Compute $\pi_j^{(l)}(u|t) = \frac{S_j\{u|\mathcal{H}_j(u, \boldsymbol{b}_j^{(l)}, \boldsymbol{\theta}^{(l)}); \boldsymbol{\theta}^{(l)}\}}{S_j\{t|\mathcal{H}_j(t, \boldsymbol{b}_j^{(l)}, \boldsymbol{\theta}^{(l)}); \boldsymbol{\theta}^{(l)}\}}$
4. Repeat Step 1-3 for $l = 1, \ldots, L$ times

The Monte Carlo estimate on the survival probability can be calculated as

$$\hat{\pi}_j(u|t) = \text{median}\{\pi_j^{(l)}(u|t), l = 1, \ldots, L\} \tag{3.29}$$

or

$$\hat{\pi}_j(u|t) = \text{mean}\{\pi_j^{(l)}(u|t), l = 1, \ldots, L\}. \tag{3.30}$$

Standard deviations, standard errors and credible intervals can be calculated in the same way as described in Section 3.4.

Similar to predictions of the survival probability, predictions for the longitudinal outcome $\omega_j(u|t)$ can be estimated as follows:

$$\begin{aligned}
\tilde{\omega}_j(u|t) &= \int E\{y_j(u)|T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}\} p(\boldsymbol{\theta}|\mathcal{D}_n) d\boldsymbol{\theta} \\
&= \int \int E\{y_j(u)|\boldsymbol{b}_j, \boldsymbol{\theta}\} p(\boldsymbol{b}_j|T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}_n) d\boldsymbol{b}_j d\boldsymbol{\theta} \\
&= \int [\boldsymbol{x}_j^T(u)\boldsymbol{\beta} + \boldsymbol{z}_j^T(u) \int \boldsymbol{b}_j p(\boldsymbol{b}_j|T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) d\boldsymbol{b}_j] p(\boldsymbol{\theta}|\mathcal{D}_n) d\boldsymbol{\theta} \\
&\approx \int [\boldsymbol{x}_j^T(u)\boldsymbol{\beta} + \boldsymbol{z}_j^T(u)\hat{\boldsymbol{b}}_{\boldsymbol{j}}] p(\boldsymbol{\theta}|\mathcal{D}_n) d\boldsymbol{\theta}.
\end{aligned} \tag{3.31}$$

The Monte Carlo estimate $\hat{\omega}_j(u|t)$ can be derived in the same manner as $\hat{\pi}_j(u|t)$ by applying the scheme described above.

Chapter 4

Analysis of HIV Data with Bayesian Joint Modelling

In this section, `JMbayes`, an R package that implements and evaluates joint models with the Bayesian method, is used to analyze the HIV dataset.

4.1 Background

Human immunodeficiency virus or HIV is a type of RNA retrovirus that deteriorates human's immune system. If the health condition is not controlled, the virus will eventually lead to acquired immunodeficiency syndrome or AIDS. One of the most promising treatment methods for AIDS is the antiretroviral therapy (ART). A significant drop in HIV viral loads is expected upon the administration of antiretroviral drugs. Once the ART is terminated, there is usually a rapid rise in viral loads to peak values followed by a gradual drop. The viral load would eventually stabilizes around set points. A key characteristic after treatment termination is viral loads rebound which is defined as the first rise of viral loads to a detectable level after ART interruption [16]. The reason for viral rebound is that HIV viruses are believed to harbour in a latent reservoir where the infected immune cells are not actively making new copies of HIV virus at the moment [17]. These latent reservoirs will start producing HIV viruses upon the cessation of medical interventions, giving rise to viral rebounds. It appears that characteristics of viral loads after treatment interruption, such as time to viral rebounds and time to reaching viral load peak points, vary among individuals. It is speculated that the variation is related to HIV viral decay and CD4 regeneration during the ART treatment. An analysis of the HIV dataset was conducted to elucidate possible relationship between individual trajectories during the ART treatment and characteristics of viral rebounds after treatment termination using joint models which was evaluated with the Bayesian method.

4.2 Data Description

Table 4.1: Data Description of HIV Dataset

Name of Variable	**Data Type**	**Description**
PATIENT	Factor	patient identifier, ranging from 1 to 76
GENDER	Factor	1=Male, 0=Female, 9=unknown
age	Numerical	age at seroconversion
CD4_V	Numerical	natural log of CD4 cell counts (cells/uL)
RNA_V	Numerical	log 10 of RNA viral loads (copies/ml)
treatment	Factor	1=receiving ART; 0=ART interruption
months_from_seroco	Numerical	specime date to seroconversion date
t1	Numerical	time from the start of ART to the time of measurement
t2	Numerical	time from the start of ART to the time of event

A total of 76 patients were recruited for the HIV study where all patients received ART treatment followed by treatment interruption. Because the treatment efficacy is best reflected by the change of biomarker levels (i.e., RNA viral loads and CD4 cell counts) during the early phase, the emphasis is placed on the first six months of the ART treatment for the longitudinal process. CD4 cell counts and RNA viral loads were measured at different time points for different individuals. If biomarker levels are below the detection limit of the microarray being used, the measurement is called *left censored*. Left censored viral loads are ignored except for the first censored value for each patient. By convention, the first left censored biomarker values are replaced with half the detection limit of the microarray being used.

For the longitudinal process, we are interested in RNA viral loads and CD4 cell counts trajectories during the ART treatment. Variable `t1` records time from the start of ART to the time when measurements were taken. For the survival process, two events of interest are RNA viral rebounds and the peak of RNA viral loads. Variable `t2` records time from the start of treatment to the time of that particular event of interest happening.

Before conducting the analysis, eligible patients were selected from the raw dataset. The HIV study involved a total of 76 patients among which there was one patient with missing information on all of the measurement times. This patient was subsequently excluded from the analysis. For the longitudinal process, the time frame of interest is first six months since the start of ART. The survival process concerns with time periods after ART termination. Patients who experienced both processes and had at least one measurement taken during treatment were selected which resulted in a total of 58 eligible patients.

The longitudinal trajectories of RNA viral loads and CD4 cell counts for eligible patients are shown in Figure 4.1 and Figure 4.2. As can be seen from the RNA viral loads trajectory plot, there is a rapid viral decay upon the start of ART treatment followed by an immediate viral rebound after treatment interruption. On the other hand, CD4 cell counts gradually increase during the ART treatment and slowly decrease after treatment termination. Individual trajectories seem to vary for different patients.

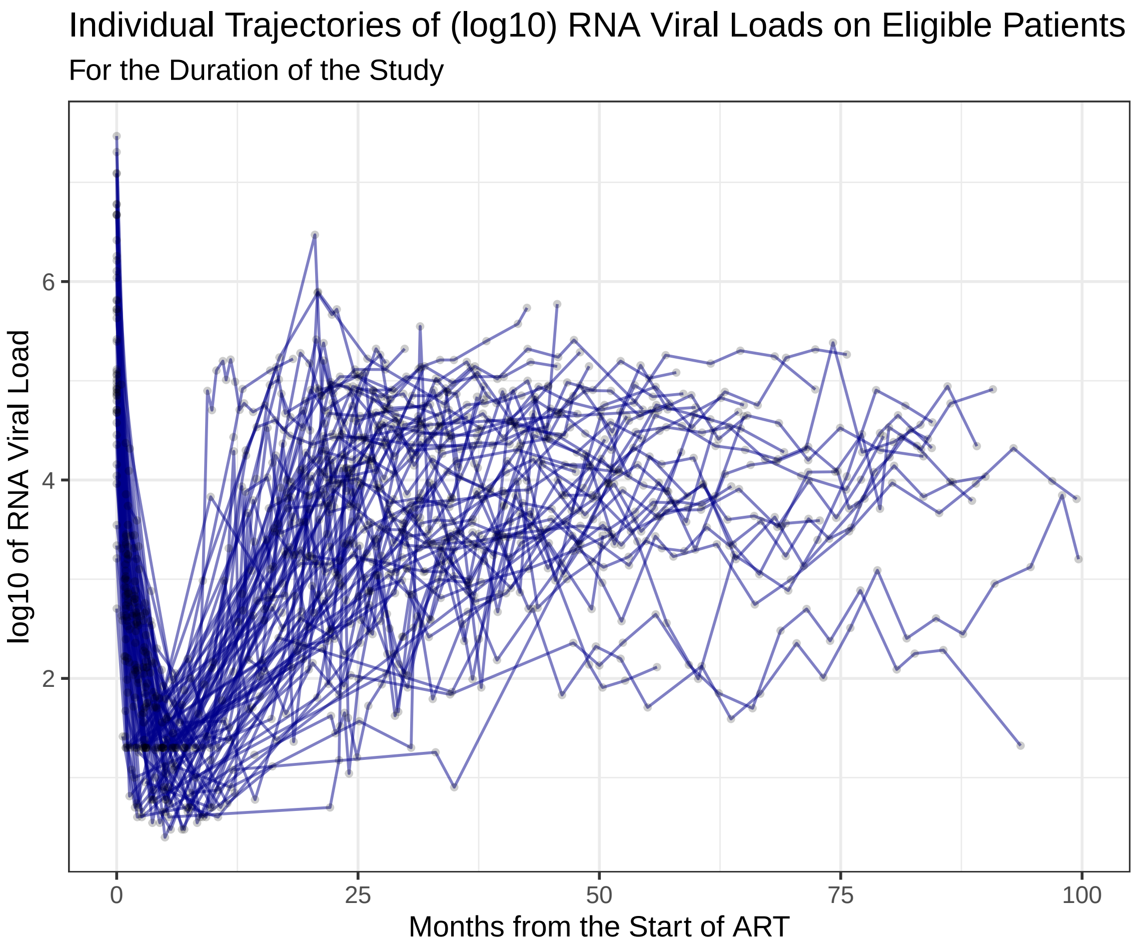

Figure 4.1: Individual Longitudinal Trajectories of (log10) RNA Viral Loads for the Duration of the Study

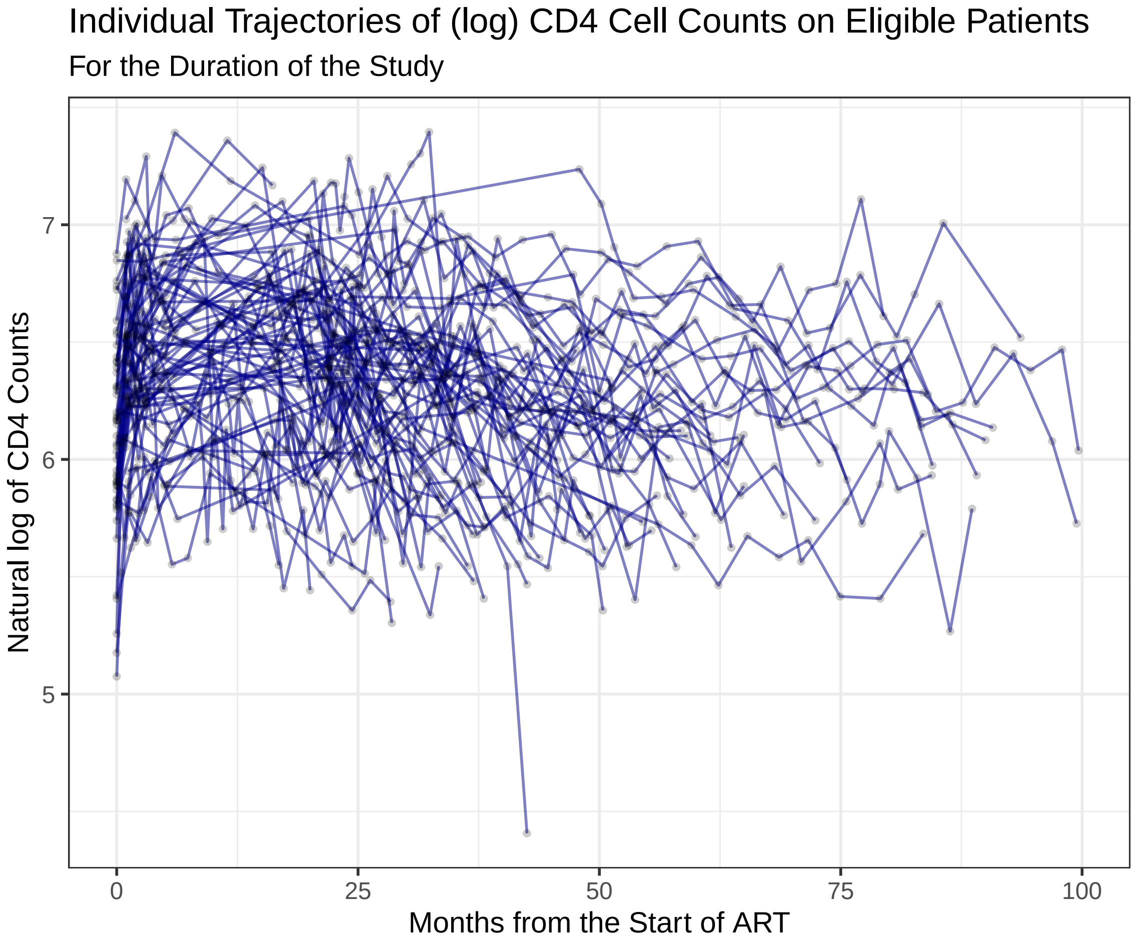

Figure 4.2: Individual Longitudinal Trajectories of (log) CD4 Cell Counts for the Duration of the Study

4.3 Separate Models for Longitudinal Data and Survival Data

To study the association between viral decay during the ART treatment and characteristics of viral rebounds after treatment termination, the LME model is used to model the longitudinal process of RNA viral decay and the Cox PH model is used to model the survival process. Furthermore, since HIV viruses attack and suppress immune systems by depleting CD4 cells, the CD4 cell counts are expected to increase once ART starts. It is likely that a rise in CD4 cell counts are related to RNA viral decay during ART treatment. To supplement the first analysis on the change of RNA viral loads, the CD4 cell counts are modelled separately with the LME model as well.

4.3.1 Linear Mixed Effects Model

Firstly, trajectories of RNA viral load decay and CD4 cell regeneration during the ART treatment were modelled with the LME model. Since LME models require that the response to be continuous and normally distributed, a natural logarithm was applied to CD4 cell counts to normalize the response. Similarly, a log10 transformation was applied to RNA viral loads. The logarithmic transformation is preferred over other types of transformation due to its simplicity in results interpretation.

Model Specification and Model Comparison

Before conducting a longitudinal analysis, it is important to decide whether random intercepts or/and random slopes should be incorporated into the model. This can be determined by plotting individual longitudinal trajectories. Figure 4.3 and Figure 4.4 depict RNA and CD4 trajectories during the first six months of ART treatment. It appears that patients started off with different levels of RNA viral loads and CD4 cell counts, and the rate of progression varied among individuals as well. Thus, it is appropriate to include both random intercepts and random slopes to correctly specify the LME model.

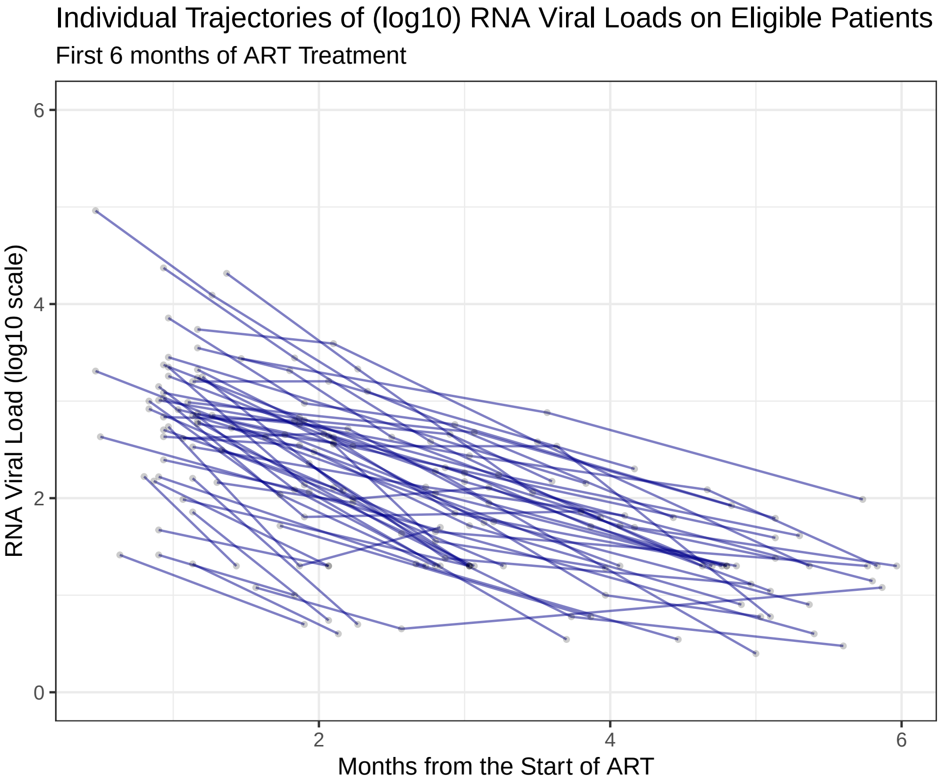

Figure 4.3: Individual RNA Viral Load (on a log10 scale) Trajectories of Eligible Patients during First Six Months of ART Treatment

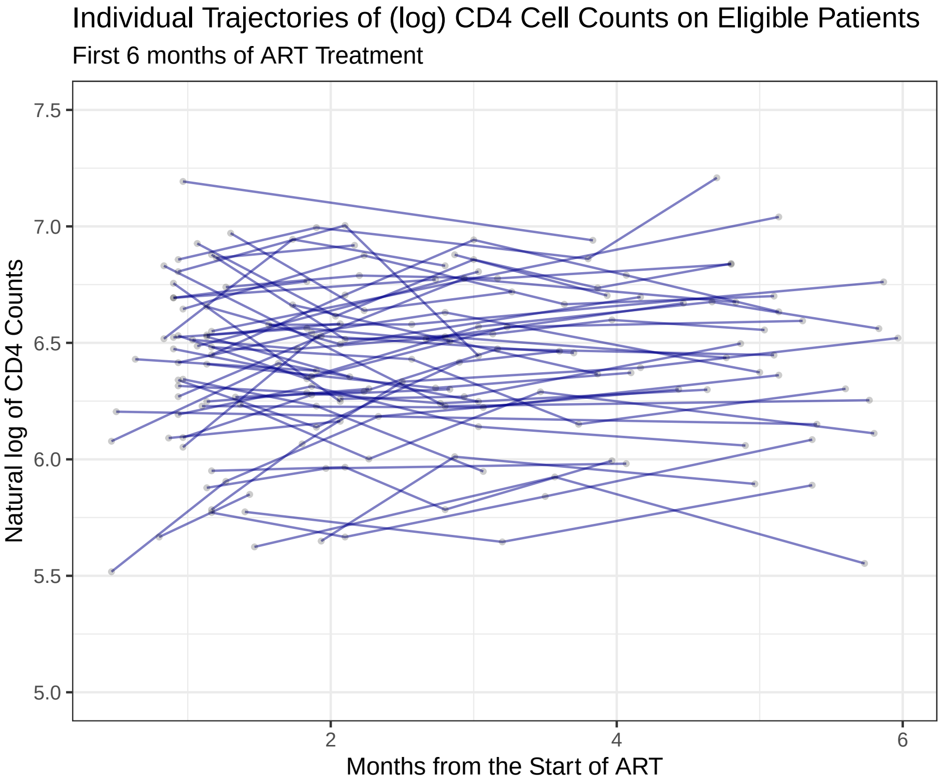

Figure 4.4: Individual CD4 Count (on a natural scale) Trajectories of Eligible Patients during First Six Months of ART Treatment

Let y_{ij} be the continuous response of RNA viral loads on a $log10$ scale, z_{ij} be the continuous response of CD4 cell counts on a natural log scale. And t_{ij} is the occasion of patient i at time point j where $i = 1, \dots, n$ and j may vary for different patients. Aside from t_{ij}, covariates of interest also include gender and age since people of different demographic characteristics are dissimilar physiologically and may respond differently to medical treatments. Model selections on nested models are performed in order to find the smallest model that fits the data almost as well as the full model [18]. The nested models for RNA viral loads are

$$y_{ij} = (\alpha_0 + a_{0i}) + (\alpha_1 + a_{1i})t_{ij} + \alpha_2 x_{\text{gender}} + \alpha_3 x_{\text{age}} + e_{ij} \tag{4.1}$$

$$y_{ij} = (\alpha_0 + a_{0i}) + (\alpha_1 + a_{1i})t_{ij} + e_{ij} \tag{4.2}$$

and the nested models for CD4 cell counts are

$$z_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \beta_2 x_{\text{gender}} + \beta_3 x_{\text{age}} + e_{ij}. \tag{4.3}$$

$$z_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + e_{ij} \tag{4.4}$$

Parameters of the above models are estimated with the maximum likelihood method which allows for comparison between nested models. Let SSR_{f} and SSR_{r} be the residual sum of squares for the full model and reduced model, respectively. The null hypothesis states that the smaller model is not significantly different from the larger model. To test this hypothesis, an F-test is adopted where the test statistic

$$F = \frac{(SSR_{\text{r}} - SSR_{\text{f}})/(\text{df}_{\text{r}} - \text{df}_{\text{f}})}{SSR_{\text{f}}/\text{df}_{\text{f}}} \sim F(\text{df}_{\text{r}}, \text{df}_{\text{f}})$$

follows an F-distribution with df_{r} and df_{f} being the residual degrees of freedom from the reduced and the full model, respectively. A resulting p-value less than the significance level provides evidence against the null hypothesis, which means that the smaller model does not give an adequate fit to the data compared to the larger model. On the other hand, if the null is not rejected, the smaller model is preferred.

The function `anova()` in R implements the F-test for nested model comparison. The result shows that both p-values are greater than the significance level. Thus, the reduced model with `t1` as the only predictor is adopted for modelling the longitudinal process of both RNA viral loads and CD4 cell counts.

The linear mixed effects models for $\log_{10} RNA\_V$ and $\log CD4\_V$ are written as follows:

$$y_{ij} = (\alpha_0 + a_{0i}) + (\alpha_1 + a_{1i})t_{ij} + e_{ij} \tag{4.5}$$

$$z_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + e_{ij} \tag{4.6}$$

where patients are treated as random effects with random intercepts a_{0i} and a random slopes a_{1i} when modelling RNA viral decay; and random intercepts b_{0i} and random slopes b_{1i} when modelling CD4 cell counts for the i-th patient.

Analysis and Results

Two linear mixed effects models described by Equation 4.5 and Equation 4.6 are fitted to the HIV data using the `lme()` function from the `nlme` package.

Firstly, the LME model was fitted to all 58 eligible patients previously selected. Model diagnostics were performed to detect the presence of outliers. Outliers can be identified from the residual plot where a data point is considered as an outlier if its standardized residual is more than 2 estimated standard deviation away from the mean. By this convention, patient 23 is an outlier since their average RNA viral load is considerably lower than that of others. Similarly, for CD4 cell counts, it is likely that patient 50 is an outlier because the mean measurement of this individual is substantially higher than that of others. The quantile-quantile plot (QQ plot) is used to check the normality assumption on error terms and random effects. As suggested by the plot, no violations are observed.

It is worth noting that outliers could be physiologically different from an average HIV patient in that they might respond extremely well or poorly to the ART treatment. These patients should be removed to obtain a more accurate inference on the population level. For the purpose of this study, patients with unusual measurements, namely, patient 23 and 50, were excluded. The total number of study subjects are thus 56.

Longitudinal analyses were conducted after the removal of the above-mentioned outliers. Results from the longitudinal submodel suggest that as treatment proceeds, RNA viral loads decrease while CD4 cell counts increase.

For LME model diagnostics, residual homoscedasticity and the normality of residuals and random effects are checked with the residual plot and QQ plot, respectively. Plots can be found in the Appendix under LME Diagnostic Plots. For the residual plot, the points randomly scatter around a horizontal line of zero indicating that variance of error term is constant. The QQ plot shows a roughly straight line which suggests that error terms and random effects are normally distributed. Model diagnostics reveal no violations to LME model assumptions.

To assess the goodness of fit, Figure 4.5 and Figure 4.6 depict individual fitted lines based on the LME model for six selected patients. The observed measurements are shown in circles. Observed values of both viral loads and CD4 cell counts fall roughly on the corresponding fitted line which indicates that the LME model fits the data fairly well.

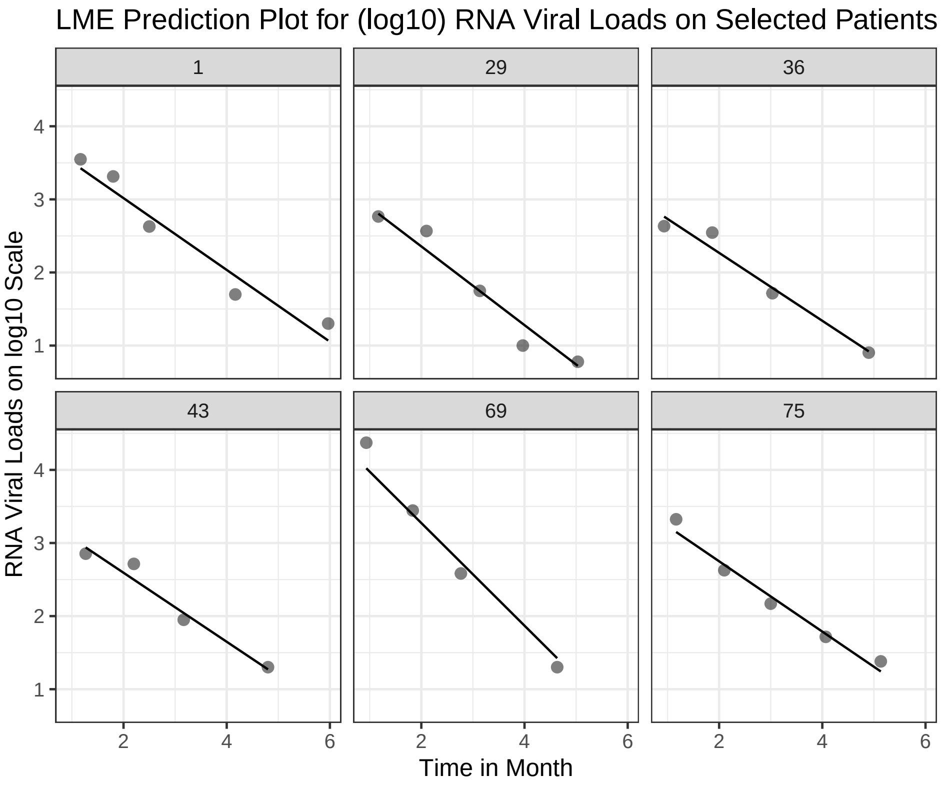

Figure 4.5: Predicted and Observed RNA Levels on Selected Patients. Observed values are shown in circles.

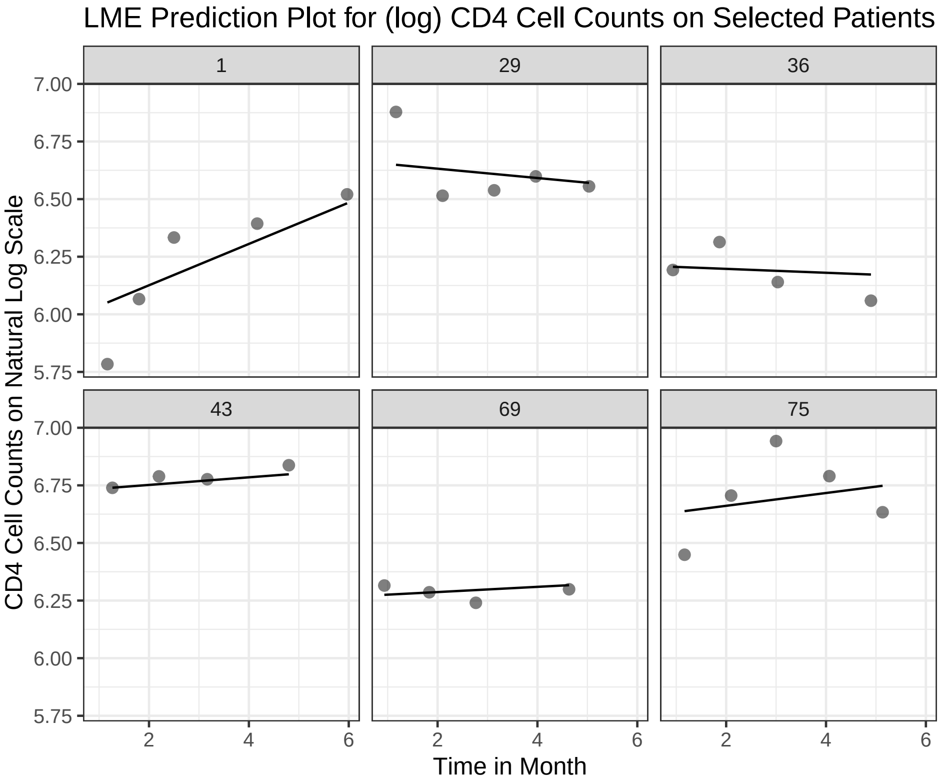

Figure 4.6: Predicted and Observed CD4 Levels on Selected Patients. Observed values are shown in circles.

4.3.2 Cox PH Model

Secondly, the Cox PH model was used to model the hazard of viral load rebound and the hazard of viral peak point after treatment termination, separately.

Model Specification and Model Comparison

Let $h_i(t)$ be the hazard of the event for individual i at time point t, and $\boldsymbol{w_i} = (w_{i,gender}, w_{i,age})$ be the demographic characteristics. Similar to the model selection process performed with the longitudinal model, an F-test can be used to compare nested survival models, namely, the full model and the null model

$$h_i(t) = h_0(t)\text{exp}(\gamma_1 w_{\text{gender}} + \gamma_2 w_{\text{age}}) \tag{4.7}$$

$$h_i(t) = h_0(t) \tag{4.8}$$

using the `anova()` function. P-values for both tests are greater than the significance level. Thus, the null model described by Equation 4.8 is preferred. Let $h_i^{(r)}(t)$ and $h_i^{(p)}(t)$ denote the hazard of experiencing viral rebound and viral peak point at time point t, respectively. $h_0^{(r)}(t)$ and $h_0^{(p)}(t)$ denote the corresponding baseline hazard function. Models for the two survival processes are written as follows

$$h_i^{(r)}(t) = h_0^{(r)}(t) \tag{4.9}$$

$$h_i^{(p)}(t) = h_0^{(p)}(t). \tag{4.10}$$

Analysis and Results

Cox PH models described by Equation 4.9 and Equation 4.10 were used to model the hazard of two events of interest using `survfit()` from the `survival` package. Figure 4.7 and Figure 4.8 provide visual examination of the number of people experiencing viral rebound and viral peak point at different time points. The majority of patients experienced viral load rebound within approximately 10 months after treatment termination, whereas patients experienced a peak in viral loads in much later periods. Most patients underwent viral peak point within 50 months after ART interruption.

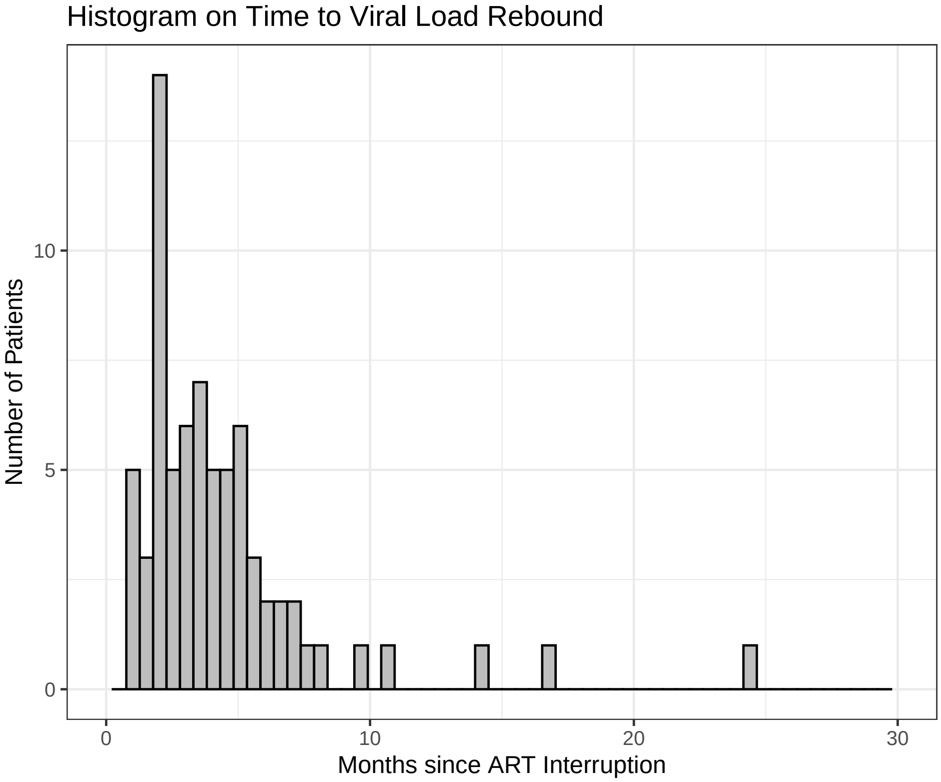

Figure 4.7: Number of People Experiencing RNA Viral Rebound

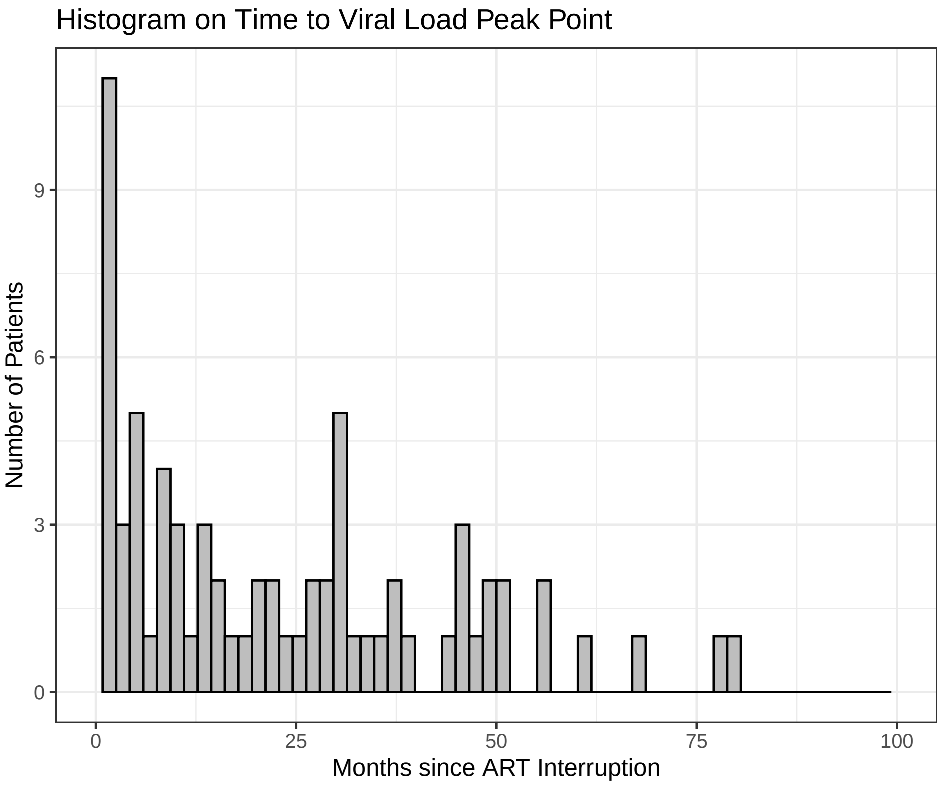

Figure 4.8: Number of People Experiencing RNA Viral Peak Point

4.4 Bayesian Joint Modeling for Longitudinal and Survival Data

The Bayesian joint model was implemented to study the association between longitudinal trajectories during the treatment and viral load characteristics after treatment termination. To that end, four joint models by pairwise combination were analyzed. To investigate the most appropriate specification of such an association, the longitudinal process and survival process were linked through four different association structures described in Section 3.2.1. Conclusions were based on the most reasonable association structure(s).

The `JMbayes` package provides a set of default settings for conducting the Bayesian inference on joint models. One of the settings that is of interest for alteration is the choice of hyperparameters. A Bayesian inference is sensitive to the prior distribution if there is a drastic change in the posterior distribution upon a minor alteration in the prior [19]. It is ideal that the posterior distribution is not sensitive to the choice of prior. Besides, misspecified hyperparameters could lead to distorted inferences. Therefore, the Bayesian inferences were obtained by using both default priors and priors with different hyperparameters. The results were compared to ensure that the resulting inference was reliable and robust to the choice of priors.

4.4.1 Joint Model with Shared Random Effects

Firstly, the longitudinal and the survival process were linked through shared random effects described by Equation 3.4. The random effects $a_i = \{a_{0i}, a_{1i}\}$ were used for modelling RNA viral loads, and $b_i = \{b_{0i}, b_{1i}\}$ were used for modelling CD4 cell counts. The random effects were linked to the Cox PH models through association structure $\alpha_a = \{\alpha_{a_0}^{(r)}, \alpha_{a_0}^{(p)}, \alpha_{a_1}^{(r)}, \alpha_{a_1}^{(p)}\}$, and $\alpha_b = \{\alpha_{b_0}^{(r)}, \alpha_{b_0}^{(p)}, \alpha_{b_1}^{(r)}, \alpha_{b_1}^{(p)}\}$, respectively. The superscripts (r) and (p) represent the event of viral rebound and viral peak point, respectively. Four resulting joint models are as follows:

- Analysis 1: Association between RNA viral rebound and RNA viral decay

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_{a_0}^{(r)} a_{0i} + \alpha_{a_1}^{(r)} a_{1i}\right) \tag{4.11}$$

- Analysis 2: Association between RNA viral peak point and RNA viral decay

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_{a_0}^{(p)} a_{0i} + \alpha_{a_1}^{(p)} a_{1i}\right) \tag{4.12}$$

- Analysis 3: Association between RNA viral rebound and CD4 cell counts

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_{b_0}^{(r)} b_{0i} + \alpha_{b_1}^{(r)} b_{1i}\right) \tag{4.13}$$

- Analysis 4: Association between RNA viral peak point and CD4 cell counts

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_{b_0}^{(p)} b_{0i} + \alpha_{b_1}^{(p)} b_{1i}\right) \tag{4.14}$$

The joint model inference was obtained by using the default priors specified in the `JMbayes` package. The prior distributions took on the following hyperparameters where $a_{\alpha_1} = 10$, $a_{\alpha_2} = 10$, $a_{h_0} = 10$, and $a_\sigma = 10$. That is,

$$\alpha_a \sim N(0, 10), \quad \alpha_b \sim N(0, 10) \tag{4.15}$$

$$\boldsymbol{\gamma_{h_0}} \sim N\left(\mathbf{0}, 10I\right), \quad R_i \sim IG\left(\frac{1}{10\sigma_r^2}, \frac{1}{10\sigma_r^2}\right) \tag{4.16}$$

The default MCMC sampling method is the Metropolis-Hastings algorithm for most posterior distributions. The exception is for the variance-covariance matrix of random effects which is known to follow an inverse Wishart distribution. `JMbayes` implements the “one long chain” method with the first 3000 iterations being the burn-in period followed by 20000 iterations. For the purpose of this analysis, the number of iteration after burn-in period was set to 50000 for more precise approximation. Because of the correlation between neighbouring chains, Markov chains are typically thinned where all samples are discarded except for every k-th observation as a way to reduce autocorrelation [20]. The default setting is to keep 2000 iterations for each parameter.

The Bayesian inferences on four joint models were drawn by using the above-mentioned settings. Resulting posterior estimates are summarized in Table 4.2.

The association structure is estimated by the posterior mean calculated from MCMC samples. Uncertainty of point estimators is measured by the posterior standard deviation which quantified the amount of variation of the parameter. Posterior standard errors can be used to assess the precision of point estimators. A small posterior standard error indicates that the posterior mean is an accurate representation of the

Table 4.2: Posterior Estimates on Association Structures for Analysis 1-4

	Posterior Mean	Standard Error	Standard Deviation	2.5%	97.5%
Analysis 1[1]					
$\alpha_{a_0}^{(r)}$	-0.1645	0.0210	0.3212	-0.7620	0.4604
$\alpha_{a_1}^{(r)}$	-0.1634	0.0387	0.7187	-1.8254	1.0864
Analysis 2[2]					
$\alpha_{a_0}^{(p)}$	0.5654	0.0166	0.3529	-0.0925	1.3106
$\alpha_{a_1}^{(p)}$	-1.1811	0.0287	0.8268	-2.9407	0.3185
Analysis 3[3]					
$\alpha_{b_0}^{(r)}$	-0.1298	0.0140	0.4267	-0.9547	0.6909
$\alpha_{b_1}^{(r)}$	1.8949	0.1073	1.8549	-2.0938	5.2723
Analysis 4[4]					
$\alpha_{b_0}^{(p)}$	-1.0176	0.0240	0.6119	-2.1849	0.1597
$\alpha_{b_1}^{(p)}$	-2.0962	0.0845	2.1154	-6.3284	1.9958

[1] Association between RNA Viral Rebound and RNA Viral Decay
[2] Association between RNA Viral Peak Point and RNA Viral Decay
[3] Association between RNA Viral Rebound and CD4 Cell Counts
[4] Association between RNA Viral Peak Point and CD4 Cell Counts

true parameter value. As can be seen from the results, posterior standard errors for all four analyses are relatively small with respect to the posterior mean. This indicates that the effective sample size is sufficiently large and that the obtained estimates are reliable.

The covariate is statistically significant if the corresponding Bayesian credible interval excludes zero. The results suggest that neither the initial value nor the rate of increase of RNA viral loads is significantly associated with the risk of the two survival processes. The conclusions are the same for CD4 cell counts. However, for Analysis 4 (Association between RNA viral peak point and CD4 cell counts), credible intervals for random intercepts and random slopes are (-2.1849, 0.1597) and (-6.3284, 1.9958), respectively, where majority of the values fall into the region below 0. This provides some evidence on CD4 level being negatively associated with viral peak point, although the association is not significant on a 5% significance level.

Sensitivity Analysis

The Bayesian inference is robust if results are insensitive to the choice of priors. For Analysis 1-4, the results suggest that neither RNA decay nor CD4 cell count is significantly associated with the risk of RNA viral rebound or viral peak point. It is likely that the baseline level and the rate of increase of CD4 cell counts are negatively associated with the risk of viral peak point, albeit rather weakly at a 5% significance level. For the sensitivity analysis, we are interested in evaluating the degree of association when priors are centered at zero with much larger and much smaller variance than the default variance. The first sensitivity analysis used priors with larger variance where hyperparameters were altered to be $a_{\alpha_1} = 100$, $a_{\alpha_2} = 100$. And $a_{\alpha_1} = 1$, $a_{\alpha_2} = 1$ were used for the second sensitivity analysis where the variance was set to be smaller than that of the default prior.

- Sensitivity Analysis 1: Large Variance for the Association Structures

$$\alpha_a \sim N(0, 100), \quad \alpha_b \sim N(0, 100) \tag{4.17}$$

- Sensitivity Analysis 2: Small Variance for the Association Structures

$$\alpha_a \sim N(0, 1), \quad \alpha_b \sim N(0, 1) \tag{4.18}$$

This first set of hyperparameters results in prior distributions that are less informative than the default priors due to their large variance. By adopting less informative priors, the posterior distributions are further influenced by the observed data while the impact of prior knowledge is attenuated.

The second set of priors centralizes around zero. It embodies the belief that the survival probability is unrelated to the longitudinal measurements. This information is applied to the Bayesian inference where more weights are given to neighbouring regions of the likelihood function around zero. Results for Sensitivity Analysis 1 and 2 are summarized in Table 4.3.

Table 4.3: Posterior Estimates on Association Structures for Sensitivity Analysis 1 and 2

	Posterior Mean	Standard Error	Standard Deviation	2.5%	97.5%
Sensitivity Analysis 1[1]					
Analysis 1					
$\alpha_{a_0}^{(r)}$	-0.1528	0.0262	0.3289	-0.7688	0.4645
$\alpha_{a_1}^{(r)}$	-0.2312	0.0521	0.8133	-1.9859	1.1387
Analysis 2					
$\alpha_{a_0}^{(p)}$	0.6049	0.0194	0.3687	-0.0907	1.3755
$\alpha_{a_1}^{(p)}$	-1.2833	0.0342	0.8715	-3.1336	0.2596
Analysis 3					
$\alpha_{b_0}^{(r)}$	-0.0634	0.0133	0.4420	-0.9387	0.8231
$\alpha_{b_1}^{(r)}$	2.4594	0.1645	2.3014	-2.3985	6.8177
Analysis 4					
$\alpha_{b_0}^{(p)}$	-1.1660	0.0275	0.6579	-2.3954	0.1300
$\alpha_{b_1}^{(p)}$	-3.9957	0.1610	2.9612	-10.0244	1.5257
Sensitivity Analysis 2[2]					
Analysis 1					
$\alpha_{a_0}^{(r)}$	-0.1285	0.0155	0.2853	-0.6773	0.4339
$\alpha_{a_1}^{(r)}$	-0.1280	0.0223	0.5606	-1.2880	0.9421
Analysis 2					
$\alpha_{a_0}^{(p)}$	0.5233	0.0136	0.3141	-0.0953	1.1780
$\alpha_{a_1}^{(p)}$	-0.7350	0.0164	0.6037	-1.9516	0.3878
Analysis 3					
$\alpha_{b_0}^{(r)}$	-0.1705	0.0260	0.4224	-0.9758	0.7086
$\alpha_{b_1}^{(r)}$	0.3623	0.0339	0.9472	-1.5781	2.1420
Analysis 4					
$\alpha_{b_0}^{(p)}$	-0.6551	0.0248	0.5130	-1.6705	0.3070
$\alpha_{b_1}^{(p)}$	-0.3235	0.0239	0.9314	-2.1121	1.5079

[1] Large Variance for the Association Structures ($N(0, 100)$)

[2] Small Variance for the Association Structures ($N(0, 1)$)

The relationship between CD4 cell counts and the risk of viral peak point is further inspected. Previous results from Analysis 4 suggest that there is a weak association between these two processes with 95% credible intervals for $\alpha_{b_0}^{(p)}$ and $\alpha_{b_1}^{(p)}$ being (-2.1849, 0.1597) and (-6.3284, 1.9958), respectively. When a flatter prior was used (Sensitivity Analysis 1), credible intervals cover more area in the negative region which provide stronger evidence on CD4 cell counts being negatively associated with viral peak point. On the other hand, when a more informative prior is utilized (Sensitivity Analysis 2), credible intervals cover less area in the negative region which provide even less evidence supporting the association compared to Analysis 4.

Regardless of the choice of priors, association between the longitudinal and survival process remains insignificant. However, moderate changes in posterior estimates are observed when priors with smaller variance are used. Therefore, joint model with shared random effects could be slightly sensitive to the choice of prior.

4.4.2 Joint Model with Shared True Longitudinal Values $\eta_i(t)$

Secondly, the longitudinal and survival process were linked through the true biomarker level $\eta_i(t)$. Since the longitudinal process was modelled with the LME model, the linear predictor $\eta_i(t)$ was essentially the expected value for patient i at time point t, i.e., $E(y_i(t))$ and $E(z_i(t))$ for RNA viral loads and CD4 cell counts, respectively. Let the association structure be $\alpha_\eta = \{\alpha_y^{(r)}, \alpha_y^{(p)}, \alpha_z^{(r)}, \alpha_z^{(p)}\}$ for the four above-mentioned analyses. The resulting joint models are as follows:

- Analysis 5: Association between RNA viral rebound and RNA viral decay

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_y^{(r)}\eta_i^{(r)}(t)\right) \tag{4.19}$$

- Analysis 6: Association between RNA viral peak point and RNA viral decay

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_y^{(p)}\eta_i^{(p)}(t)\right) \tag{4.20}$$

- Analysis 7: Association between RNA viral rebound and CD4 cell counts

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_z^{(r)}\eta_i^{(r)}(t)\right) \tag{4.21}$$

- Analysis 8: Association between RNA viral peak point and CD4 cell counts

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_z^{(p)}\eta_i^{(p)}(t)\right) \tag{4.22}$$

For the default analyses, joint model inferences were obtained under the default priors. In particular, the default prior distribution for the association structure α_η was

$$\alpha_\eta \sim N(0, 10). \tag{4.23}$$

Similar to the sensitivity analysis conducted for the shared random effects model, hyperparameters of the association structure were altered for Analysis 5-8. The same two sets of hyperparameters, namely, $\alpha = 100$ and $\alpha = 1$ were used for Sensitivity Analysis 3 and 4, respectively.

- Sensitivity Analysis 3: Large Variance for the Association Structures

$$\alpha_\eta \sim N(0, 100) \tag{4.24}$$

- Sensitivity Analysis 4: Small Variance for the Association Structures

$$\alpha_\eta \sim N(0, 1) \tag{4.25}$$

Posterior estimates for the default analyses and sensitivity analyses are summarized in Table 4.4 with significant results highlighted in red. For joint model inferences based on default priors, true levels of CD4 cell counts are significantly associated with RNA viral peak point. The negative posterior mean indicates that a high CD4 level contributes negatively towards the risk of experiencing viral peak point. For one unit increase in $log(CD4\_V)$, the risk of experiencing viral peak point decreases by $1 - e^{-0.2675} = 23.5\%$. 95% credible intervals for Analysis 5-7 include 0 which suggests that the association between true longitudinal values and survival probabilities is insignificant at 5% significance level.

For the sensitivity analyses, posterior means for association structures are comparable regardless of the choice of priors. One thing to note is that the association between the true level of CD4 cell counts and the risk of viral peak point is significant when the prior for association structure took on smaller variance; however, the association is insignificant under larger variance. In addition, the relationship between the true level of CD4 cell counts and the risk of viral rebound is significant under the prior with variance 100; however, this association is not significant under priors with smaller variance. The above results suggest that joint models with shared true longitudinal measurements are somewhat sensitive to the choice of prior making the obtained Bayesian inference subject to uncertainty to some extent.

Table 4.4: Posterior Estimates on Association Structures for Analysis 5-8 and Sensitivity Analysis 3-4

	Posterior Mean	Standard Error	Standard Deviation	2.5%	97.5%
Default Analyses [1,2,3,4]					
$\alpha_y^{(r)}$	-0.1708	0.0141	0.0867	-0.3473	0.0006
$\alpha_y^{(p)}$	-0.0706	0.0051	0.0448	-0.1733	0.0009
$\alpha_z^{(r)}$	0.1833	0.0103	0.1001	-0.0188	0.3868
$\alpha_z^{(p)}$	-0.2675	0.0178	0.1311	-0.5770	-0.0643
Sensitivity Analysis 3 [5]					
$\alpha_y^{(r)}$	-0.1632	0.0131	0.0888	-0.3365	0.0092
$\alpha_y^{(p)}$	-0.0706	0.0051	0.0448	-0.1733	0.0009
$\alpha_z^{(r)}$	0.1792	0.0073	0.0855	0.0106	0.3474
$\alpha_z^{(p)}$	-0.2458	0.0246	0.1437	-0.5648	0.0037
Sensitivity Analysis 4 [6]					
$\alpha_y^{(r)}$	-0.1795	0.0087	0.0748	-0.3176	-0.0119
$\alpha_y^{(p)}$	-0.0676	0.0044	0.0432	-0.1785	-0.0004
$\alpha_z^{(r)}$	0.1706	0.0097	0.0950	-0.0467	0.3567
$\alpha_z^{(p)}$	-0.2503	0.0198	0.1397	-0.5965	-0.0193

[1] Analysis 5: Association between RNA Viral Rebound and RNA Viral Decay

[2] Analysis 6: Association between RNA Viral Peak Point and RNA Viral Decay

[3] Analysis 7: Association between RNA Viral Rebound and CD4 Cell Counts

[4] Analysis 8: Association between RNA Viral Peak Point and CD4 Cell Counts

[5] Large Variance for the Association Structures $(N(0, 100))$

[6] Small Variance for the Association Structures $(N(0, 1))$

4.4.3 Joint Model with Shared True Longitudinal Value $\eta_i(t)$ and Corresponding Slope $\eta_i'(t)$

The third option for association structures is based on the assumption that the survival probability is related to both true biomarker levels $\eta_i(t)$ and their rate of change $\eta_i'(t)$. Let the association structure for $\eta_i(t)$ be $\alpha_{\eta_1} = \{\alpha_{y_1}^{(r)}, \alpha_{y_1}^{(p)}, \alpha_{z_1}^{(r)}, \alpha_{z_1}^{(p)}\}$, and the association structure for $\eta_i'(t)$ be $\alpha_{\eta_2} = \{\alpha_{y_2}^{(r)}, \alpha_{y_2}^{(p)}, \alpha_{z_2}^{(r)}, \alpha_{z_2}^{(p)}\}$. The four joint models are as follows:

- Analysis 9: Association between RNA viral rebound and RNA viral decay

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_{y_1}^{(r)}\eta_i^{(r)}(t) + \alpha_{y_2}^{(r)}\eta_i'^{(r)}(t)\right) \tag{4.26}$$

- Analysis 10: Association between RNA viral peak point and RNA viral decay

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_{y_1}^{(p)}\eta_i^{(p)}(t) + \alpha_{y_2}^{(p)}\eta_i'^{(p)}(t)\right) \tag{4.27}$$

- Analysis 11: Association between RNA viral rebound and CD4 cell counts

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_{z_1}^{(r)}\eta_i^{(r)}(t) + \alpha_{z_2}^{(r)}\eta_i'^{(r)}(t)\right) \tag{4.28}$$

- Analysis 12: Association between RNA viral peak point and CD4 cell counts

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_{z_1}^{(p)}\eta_i^{(p)}(t) + \alpha_{z_2}^{(p)}\eta_i'^{(p)}(t)\right) \tag{4.29}$$

The above joint models were evaluated using default priors and priors with larger and smaller variance for the association structure.

- Default Priors

$$\alpha_{\eta_1} \sim N(0, 10), \quad \alpha_{\eta_2} \sim N(0, 10) \tag{4.30}$$

- Sensitivity Analysis 5: Large Variance for the Association Structures

$$\alpha_{\eta_1} \sim N(0, 100), \quad \alpha_{\eta_2} \sim N(0, 100) \tag{4.31}$$

- Sensitivity Analysis 6: Small Variance for the Association Structures

$$\alpha_{\eta_1} \sim N(0, 1), \quad \alpha_{\eta_2} \sim N(0, 1) \tag{4.32}$$

The resulting Bayesian posterior estimates are summarized in Table 4.5 with significant results highlighted in red and conflicting results in light blue. For model inferences derived under the default priors, the risk of viral rebound is significantly associated with both the true level of RNA viral loads and their rate of change. Specifically, patients with a low level of RNA viral loads are less likely to experience viral rebound while patients with a faster viral load decay are more prone to viral rebound. Furthermore, the true level of CD4 cell counts are negatively associated with the risk of experiencing viral peak point, that is patients with higher level of CD4 cells are less susceptible to the peak of viral loads.

As for the sensitivity analyses, Bayesian estimates of Sensitivity Analysis 5 are in accordance with results obtained using default priors. However, the association between RNA viral decay and viral rebounds is insignificant under the influence of more informative priors. Moreover, the rate of CD4 cell counts change is positively associated with viral rebounds under priors with large and default variance; however, the direction of association reverses under the influence of more informative priors. Thus, Bayesian inferences for joint models with shared true longitudinal measurements $\eta_i(t)$ and its corresponding slopes $\eta_i'(t)$ is sensitive to the choice of prior.

Table 4.5: Posterior Estimates on Association Structures for Analysis 9-12 and Sensitivity Analysis 5-6

	Posterior Mean	Standard Error	Standard Deviation	2.5%	97.5%
Default Analyses					
Analysis 9[1]					
$\alpha_{y_1}^{(r)}$	-0.4212	0.0117	0.1095	-0.6341	-0.2081
$\alpha_{y_2}^{(r)}$	4.1779	0.1769	1.7917	0.6385	7.6965
Analysis 10[2]					
$\alpha_{y_1}^{(p)}$	-0.0593	0.0050	0.0500	-0.1680	0.0330
$\alpha_{y_2}^{(p)}$	-0.5258	0.0940	1.1840	-2.7824	1.9526
Analysis 11[3]					
$\alpha_{z_1}^{(r)}$	0.1328	0.0385	0.1862	-0.2540	0.4515
$\alpha_{z_2}^{(r)}$	0.0696	0.2855	2.5013	-4.5760	5.2198

[1] Association between RNA viral Rebound and RNA Viral Decay
[2] Association between RNA viral Peak Point and RNA Viral Decay
[3] Association between RNA viral Rebound and CD4 Cell Counts

Analysis 12[4]					
$\alpha_{z_1}^{(p)}$	-0.2903	0.0275	0.1280	-0.5865	-0.0475
$\alpha_{z_2}^{(p)}$	0.9575	0.1326	2.5306	-4.0773	5.7214
Sensitivity Analysis 5[5]					
Analysis 9					
$\alpha_{y_1}^{(r)}$	-0.4766	0.0258	0.1666	-0.8246	-0.1303
$\alpha_{y_2}^{(r)}$	5.3105	0.4308	2.7127	0.6654	11.4860
Analysis 10					
$\alpha_{y_1}^{(p)}$	-0.0550	0.0064	0.0587	-0.1855	0.0427
$\alpha_{y_2}^{(p)}$	-0.6450	0.1052	1.3279	-3.0598	2.2293
Analysis 11					
$\alpha_{z_1}^{(r)}$	0.1213	0.0255	0.1507	-0.1832	0.4074
$\alpha_{z_2}^{(r)}$	0.8958	0.4072	3.0243	-4.9443	6.8292
Analysis 12					
$\alpha_{z_1}^{(p)}$	-0.2822	0.0256	0.1279	-0.5384	-0.0151
$\alpha_{z_2}^{(p)}$	2.2022	0.2580	3.6893	-4.6722	9.6410
Sensitivity Analysis 6[6]					
Analysis 9					
$\alpha_{y_1}^{(r)}$	-0.2050	0.0151	0.1024	-0.3869	0.0038
$\alpha_{y_2}^{(r)}$	0.7790	0.0726	0.9462	-1.0859	2.6840
Analysis 10					
$\alpha_{y_1}^{(p)}$	-0.0635	0.0046	0.0487	-0.1697	0.0215
$\alpha_{y_2}^{(p)}$	-0.2979	0.0466	0.7915	-1.8170	1.3071
Analysis 11					
$\alpha_{z_1}^{(r)}$	0.1750	0.0148	0.1044	-0.0612	0.3561
$\alpha_{z_2}^{(r)}$	-0.0968	0.0479	0.9409	-1.9276	1.7174
Analysis 12					
$\alpha_{z_1}^{(p)}$	-0.2690	0.0431	0.1421	-0.5585	-0.0132
$\alpha_{z_2}^{(p)}$	0.1148	0.0249	0.9617	-1.8072	2.0530

[4] Association between RNA viral Peak Point and CD4 Cell Counts

[5] Large Variance for the Association Structures ($N(0, 100)$)

[6] Small Variance for the Association Structures ($N(0, 1)$)

4.4.4 Joint Model with Shared Cumulative Effects

Lastly, the risk of an event at time t was linked to cumulative effects of longitudinal measurements which can be expressed as the integral of biomarker levels from baseline to time t. Let the association structure for the cumulative effects be $\alpha_\eta = \{\alpha_y^{(r)}, \alpha_y^{(p)}, \alpha_z^{(r)}, \alpha_z^{(p)}\}$. The four joint models are as follows:

- Analysis 13: Association between RNA viral rebound and RNA viral decay

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_y^{(r)} \int_0^t \eta_i^{(r)}(s)ds\right) \tag{4.33}$$

- Analysis 14: Association between RNA viral peak point and RNA viral decay

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_y^{(p)} \int_0^t \eta_i^{(p)}(s)ds\right) \tag{4.34}$$

- Analysis 15: Association between RNA viral rebound and CD4 cell counts

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_z^{(r)} \int_0^t \eta_i^{(r)}(s)ds\right) \tag{4.35}$$

- Analysis 16: Association between RNA viral peak point and CD4 cell counts

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_z^{(p)} \int_0^t \eta_i^{(p)}(s)ds\right) \tag{4.36}$$

The above joint models were evaluated using default priors and priors with different variance for the association structure.

- Default Priors:

$$\alpha_\eta \sim N(0, 10) \tag{4.37}$$

- Sensitivity Analysis 7: Large Variance for the Association Structures

$$\alpha_\eta \sim N(0, 100) \tag{4.38}$$

- Sensitivity Analysis 8: Small Variance for the Association Structures

$$\alpha_\eta \sim N(0, 0.1) \tag{4.39}$$

- Sensitivity Analysis 9: Smaller Variance for the Association Structures

$$\alpha_{\eta} \sim N(0, 0.01) \tag{4.40}$$

Posterior estimates are summarized in Table 4.6. Analyses with default prior suggest that the risk of viral rebound is positively and significantly associated with the cumulative effect of CD4 cell counts. The susceptibility to viral rebound is greater in patients with higher cumulative levels of CD4 cells. Results from the sensitivity analyses are consistent with those from the default analyses. In particular, the estimates are almost invariant under priors $\alpha_{\eta} \sim N(0, 100)$, $\alpha_{\eta} \sim N(0, 10)$, and $\alpha_{\eta} \sim N(0, 0.1)$. Thus, joint models with shared cumulative effects are robust to the choice of prior.

Table 4.6: Posterior Estimates on Association Structures for Analysis 13-16 and Sensitivity Analysis 7-9

	Posterior Mean	Standard Error	Standard Deviation	2.5%	97.5%
Default Analyses [1,2,3,4]					
$\alpha_y^{(r)}$	-0.0015	0.0006	0.0046	-0.0166	0.0036
$\alpha_y^{(p)}$	-0.0003	0.0001	0.0012	-0.0029	0.0015
$\alpha_z^{(r)}$	0.0263	0.0011	0.0077	0.0118	0.0418
$\alpha_z^{(p)}$	-0.0061	0.0005	0.0036	-0.0126	0.0031
Sensitivity Analysis 7 [5]					
$\alpha_y^{(r)}$	-0.0015	0.0006	0.0046	-0.0166	0.0036
$\alpha_y^{(p)}$	-0.0003	0.0001	0.0012	-0.0029	0.0015
$\alpha_z^{(r)}$	0.0263	0.0011	0.0077	0.0118	0.0418
$\alpha_z^{(p)}$	-0.0061	0.0005	0.0036	-0.0126	0.0031
Sensitivity Analysis 8 [6]					
$\alpha_y^{(r)}$	-0.0015	0.0006	0.0046	-0.0166	0.0036
$\alpha_y^{(p)}$	-0.0003	0.0001	0.0012	-0.0029	0.0015
$\alpha_z^{(r)}$	0.0267	0.0017	0.0098	0.0102	0.0495
$\alpha_z^{(p)}$	-0.0061	0.0005	0.0036	-0.0126	0.0031
Sensitivity Analysis 9 [7]					
$\alpha_y^{(r)}$	-0.0025	0.0012	0.0061	-0.0215	0.0034
$\alpha_y^{(p)}$	-0.0003	0.0001	0.0012	-0.0033	0.0013
$\alpha_z^{(r)}$	0.0275	0.0019	0.0088	0.0109	0.0446
$\alpha_z^{(p)}$	-0.0057	0.0005	0.0036	-0.0135	0.0021

[1] Analysis 13: Association between RNA Viral Rebound and RNA Viral Decay

[2] Analysis 14: Association between RNA Viral Peak Point and RNA Viral Decay

[3] Analysis 15: Association between RNA Viral Rebound and CD4 Cell Counts

[4] Analysis 16: Association between RNA Viral Peak Point and CD4 Cell Counts

[5] Large Variance for the Association Structure ($N(0, 100)$)

[6] Small Variance for the Association Structure ($N(0, 0.1)$)

[7] Smaller Variance for the Association Structure ($N(0, 0.01)$)

4.5 Discussion on the Choice of Association Structure

In the previous section, the survival and longitudinal process were linked through four different association structures. The models were then examined in detail to elucidate the most reasonable specification. The ideal model should produce similar results under different prior distributions and the interpretation should be clinically meaningful as well.

Joint models are insensitive to the choice of prior when the two processes are linked through shared cumulative effects making this association structure appropriate in addressing the relationship. However, statistical significance does not imply clinical significance. One can argue that the posterior estimates under this specification are insubstantial to produce a clinically significant interpretation, that is, one unit increase in the cumulative level of $logCD4$ cell counts would only result in a $e^{0.0263} = 1.03$ fold increase in the risk of experiencing viral rebound. On the other hand, joint models with shared random effects are slightly sensitive to priors with small variance. But, the resulting posterior estimates are larger and can therefore provide more meaningful insights in clinical practice.

In this section, we compared and evaluated joint models with different association structures in terms of their DIC values and model diagnostics to provide some insight on choosing the most appropriate association structure.

4.5.1 Model Comparison

The deviance information criterion (DIC) is typically used as a model selection technique for Bayesian hierarchical modelling whose posterior estimations are based on the MCMC method. Let the deviance be

$$D(\theta) = -2log(f(y|\theta)) + C \tag{4.41}$$

where y denotes the data, θ denotes the parameters, $f(y|\theta)$ is the likelihood function, and C is a constant that will be cancelled out when models are compared. The effective number of parameters p_D is defined as

$$p_D = \overline{D(\theta)} - D(\bar{\theta}). \tag{4.42}$$

DIC considers the input from both deviance and effective number of parameters such that

$$DIC = p_D + \overline{D(\theta)}. \tag{4.43}$$

which can be easily calculated from the MCMC samples. An ideal model would be the one with the least possible number of parameters but still has a good fit to the data. DIC penalizes both the deviance and effective number of parameters such that models with smaller DIC values are the ones that are small in size but adequate to give an acceptable fit to the given data [21].

To determine the most appropriate specification of the association between the longitudinal and survival process, DIC was used to compare four joint models with different association structures [22]. The resulting DIC values are summarized in Table 4.7.

The relationship between the longitudinal measurements and survival events are as follows:

- **Relation 1:** RNA Viral Rebound and RNA Viral Loads
- **Relation 2:** RNA Viral Peak Point and RNA Viral Loads
- **Relation 3:** RNA Viral Rebound and CD4 Cell Counts
- **Relation 4:** RNA Viral Peak and CD4 Cell Counts

and the association structures are as follows:

- **Association Structure 1:** Linked through Shared Random Effects
- **Association Structure 2:** Linked through True Biomarker Levels
- **Association Structure 3:** Linked through True Biomarker Levels and their Slopes
- **Association Structure 4:** Linked through Shared Cumulative Effects

It is worth noting that the difference in DIC between different association structures is not substantial. Take joint model comparison of **Relation 2** for example, the percentage difference between the largest and smallest DIC values is $\frac{811.5247-802.0023}{802.0023} =$

Table 4.7: DIC for Joint Models with Different Association Structures

	Relation 1	Relation 2	Relation 3	Relation 4
Association Structure 1	818.7051	802.0023	653.5890	644.4220
Association Structure 2	819.3705	807.0684	650.0760	639.2527
Association Structure 3	818.6841	807.0234	649.9863	638.8735
Association Structure 4	816.4797	811.5247	645.4156	643.9661

1.2%. For other relations, the percentage difference is even smaller. It is fairly likely that joint models with all four association structures perform equally well at fitting to the given HIV dataset.

4.5.2 Joint Model Diagnostics

As discussed at the beginning of this section, joint models with shared random effects and joint models with shared cumulative effects appear to be insensitive to the choice of prior. Additionally, DIC values for these two models are comparable to those with other association structures, that is, DIC values do not provide strong evidence against choosing these two models. To further investigate the model validity, diagnostics were conducted on joint models with shared random effects and joint models with shared cumulative effects.

When deriving joint model inferences using the Bayesian method with MCMC, the posterior or the target distribution is approximated by Monte Carlo samples drawn from the proposal distribution. It is important to ensure that the stationary distribution of the Markov chain provides an accurate approximation to the true posterior distribution. In addition, neighbouring MCMC samples are correlated to each other and the autocorrelation decreases as two samples become further apart. If autocorrelation decreases slowly, the empirical distribution of MCMC samples would be noisy and imprecise when approximating the target distribution. This is because a slow decay in autocorrelation would result in a small effective sample size which, as described in Section 3.4.2, would contribute to a large posterior standard error rendering the posterior estimates inaccurate. MCMC diagnostics plots such as the trace plot and the autocorrelation plot can help evaluate the quality of such an approximation.

Trace Plot

To assess how quickly the chain converges, a trace plot was used where parameter values generated from the Markov chain were plotted against the number of iterations. The chain is believed to be mixed well and the approximation is accurate if sampled values move up and down within a narrow range with no apparent trend.

For joint models with shared random effects, Figure 4.9 and Figure 4.10 depict the movement of sampled values for association structure $\alpha_{a_0}^{(r)}$ and $\alpha_{a_1}^{(r)}$ from **Analysis 1** (Association between RNA Viral Loads and Viral Rebound). Neither of the trace plot shows apparent trends indicating that the Markov chain converges fairly quickly.

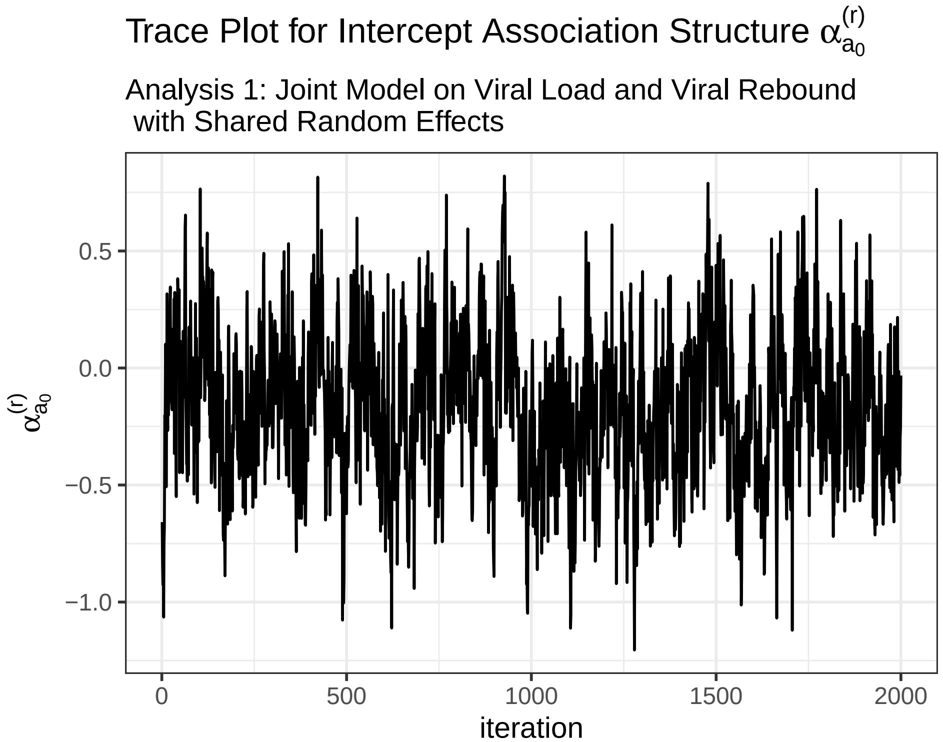

Figure 4.9: Trace Plot for Intercept Association Structure (Analysis 1)

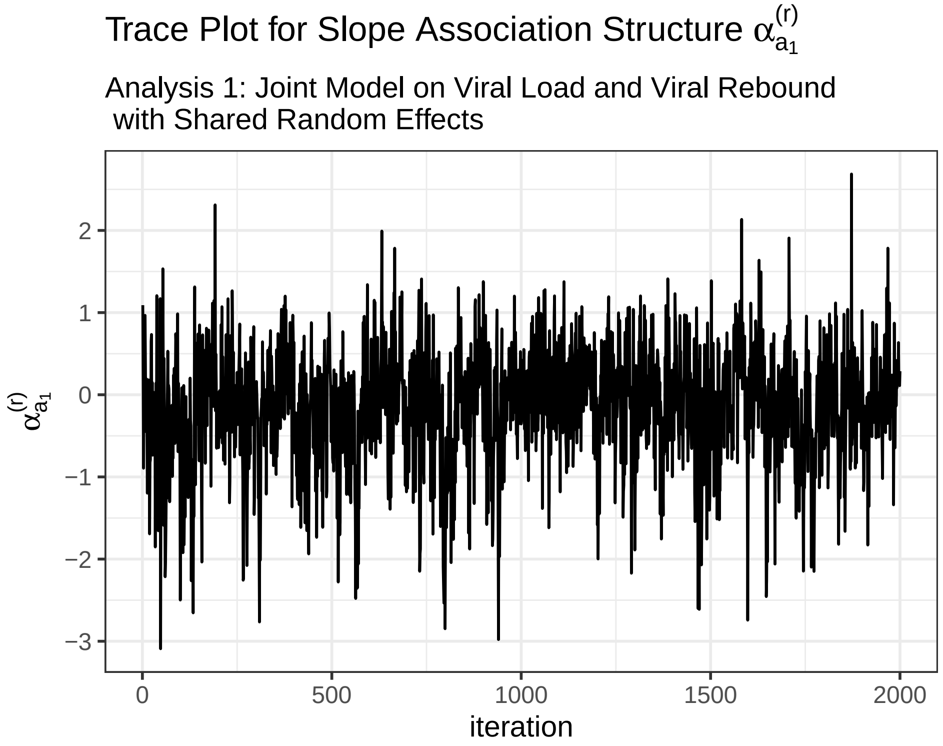

Figure 4.10: Trace Plot for Slope Association Structure (Analysis 1)

For joint models with shared cumulative effects, Figure 4.11 depicts the movement of sampled values for association structure $\alpha_y^{(r)}$ from **Analysis 13** (Association between RNA Viral Loads and Viral Rebound). The trace plot for the cumulative association structure shows that the chain is unstable at some regions indicating that the posterior distribution is noisy and that the Markov chain might converge relatively slowly.

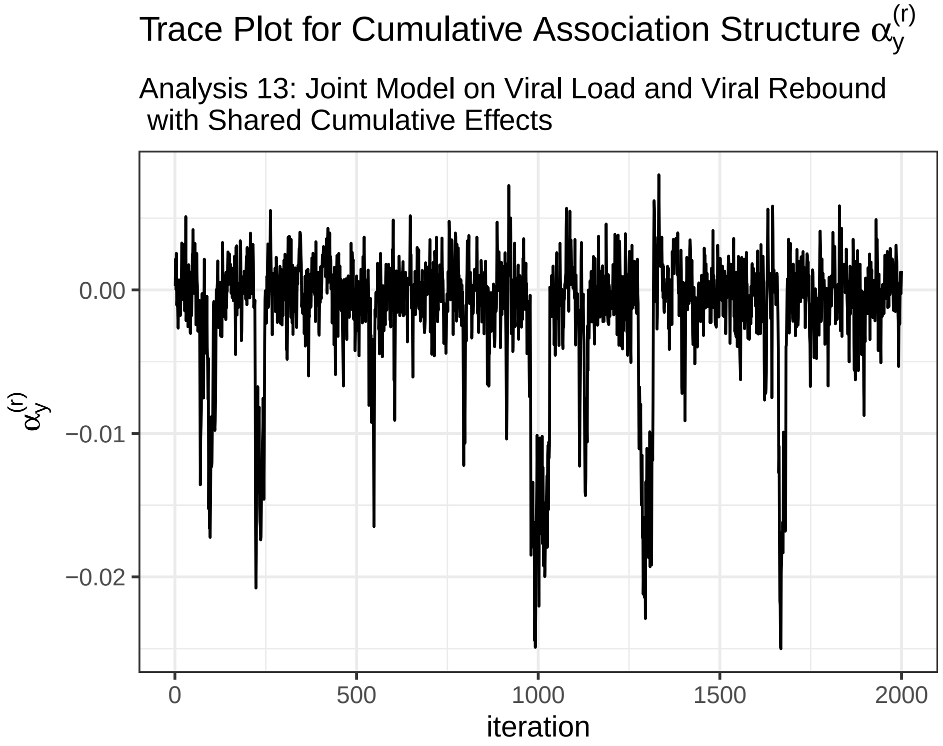

Figure 4.11: Trace Plot for Cumulative Association Structure (Analysis 13)

ACF Plot

In addition to trace plots which assess the convergence of Markov chains, autocorrelation function (ACF) plots can be used to examine the correlation between lagged observations where "lag" is defined as a fixed passing time in time series. In R, ACF plots can be generated with function `acf()` and the default maximum lag is $10log_{10}(N/m)$ where N is the number of observations and m is the number of series. For MCMC samples used in this analysis, N is the number of iterations and m is 1 because the "one long chain" method was implemented. To ensure a reasonably small posterior standard error, it is ideal that MCMC samples are not highly autocorrelated at longer lags and the correlation between lagged observations should decrease quickly.

For joint models with shared random effects, Figure 4.12 and Figure 4.13 depict autocorrelation for association structure α_{a_0} and α_{a_1} from **Analysis 1** (Association between RNA Viral Loads and Viral Rebound). MCMC samples for α_{a_0} are slightly correlated at shorter lags, but the autocorrelation stabilizes around zero fairly quickly. The autocorrelation for α_{a_1} decreases rapidly to zero after a few lags. ACF plots for the two association structures suggest that MCMC samples are not highly correlated at longer lags which ensured a sufficient effective sample size for parameter estimation.

For joint models with shared cumulative effects, Figure 4.14 depicts the autocorrelation for association structure $\alpha_y^{(r)}$ from Analysis 13 (Association between RNA Viral Loads and Viral Rebound). The plot indicates that autocorrelation is large at short lags and the correlation decreases at a relatively slow rate. This is an indication that MCMC samples are highly correlated even at longer lags and that the resulting approximation might be imprecise.

Trace plots and ACF plots for association structures used in Analysis 2-4 and Analysis 14-16 are attached in the Appendix Section 7.3.

Trace plots and ACF plots for joint models with shared random effects reveal no apparent problems regarding chain convergence and insufficient effective sample size. On the other hand, diagnostics for joint model with shared cumulative effects suggest that the empirical distribution of the MCMC samples may not be an accurate representation of the true posterior distribution. Even though Markov chains for both models converge as no errors were produced by the analytical software, it is possible that Markov chains for joint models with shared random effects converged more quickly and the resulting posterior distribution was less noisy compared to joint models with shared cumulative effects. Thus, shared random effects models are the most appropriate specification for the relationship between the longitudinal and survival process.

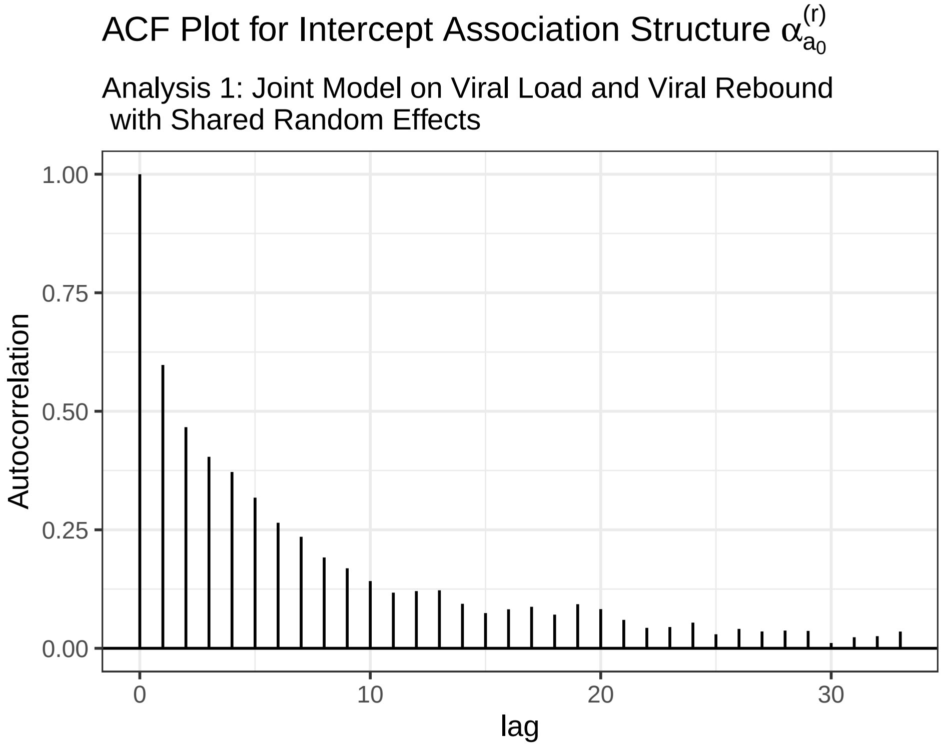

Figure 4.12: ACF Plot for Intercept Association Structure (Analysis 1)

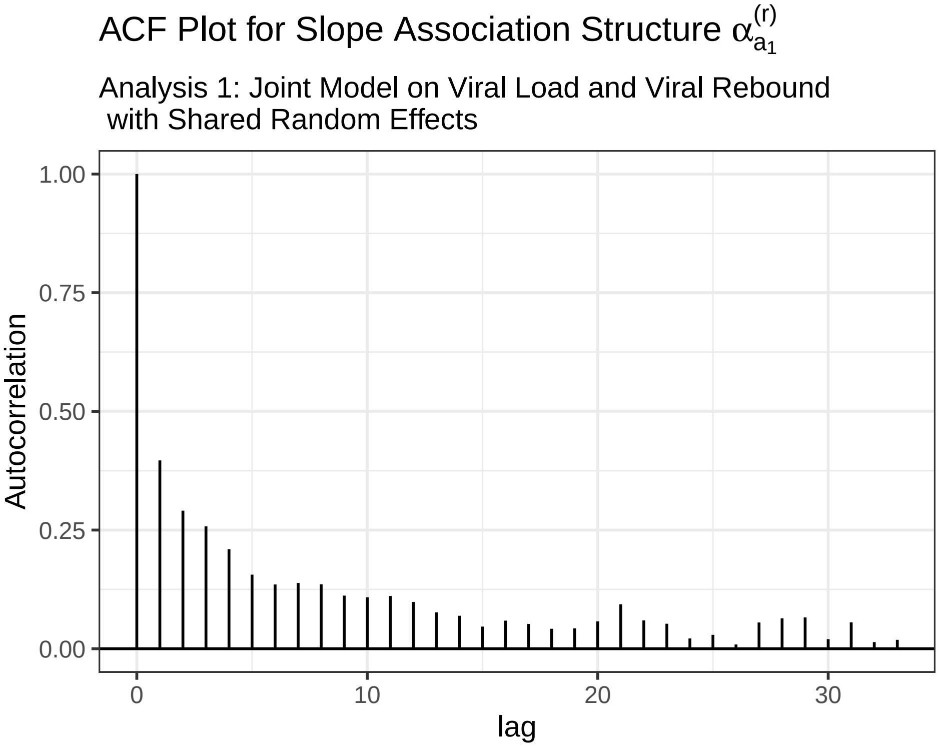

Figure 4.13: ACF Plot for Slope Association Structure (Analysis 1)

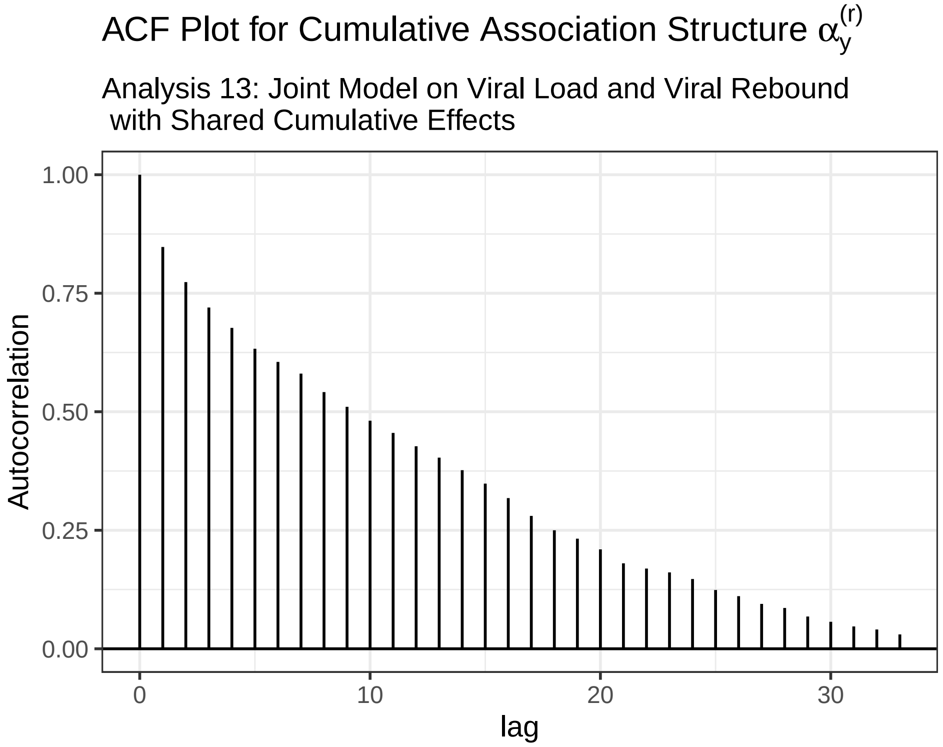

Figure 4.14: ACF Plot for Slope Association Structure (Analysis 13)

4.6 Individual Predictions

Now that the most optimal model has been determined, we can now use joint models with shared random effects to make predictions on individual patients. In the `JMbayes` package, function `survfitJM()` and `predict()` implement the algorithms described in Section 3.4.3 such that individual-specific predictions on survival probabilities and longitudinal measurements can be obtained. By default, estimates for survival and longitudinal outcomes are based on 200 Monte Carlo samples at time points $\{u : u > t_l, l = 1, \ldots, 35\}$. The empirical mean of Monte Carlo samples is used to estimate the survival probability and longitudinal outcome. 95% pointwise credible intervals can be obtained as well. Figure 4.15 and Figure 4.16 are visual illustrations on conditional estimates of the survival and longitudinal outcome for patient 1 based on the fitted model of **Analysis 1** (Association between RNA Viral Loads and Viral Rebound).

To the left of the vertical dotted line of Figure 4.15 is log10 RNA viral loads after treatment interruption with the vertical dotted line representing the last measurement date. For the purpose of making survival predictions, the last measurement date is selected as the one before the occurrence of viral load rebound such that the event is yet to be observed. The survival probability of viral rebound is plotted to the right of the dotted line with the shaded area being the 95% confidence region. The plot suggests that the survival probability decreases over time meaning that there is an increasing probability of experiencing viral rebound as time progresses.

Figure 4.16 illustrates predictions for the longitudinal outcome. The vertical dotted line represents the last measurement date. The red line to the right of the dotted line is the longitudinal prediction on log10 RNA viral loads for patient 1. 95% confidence regions are bounded within the dashed lines. The plot suggests that there is a decreasing trend in RNA viral loads throughout the course of the ART treatment. Note that values below zero are not physiologically possible and these predictions should be interpreted carefully based on physicians' own knowledge.

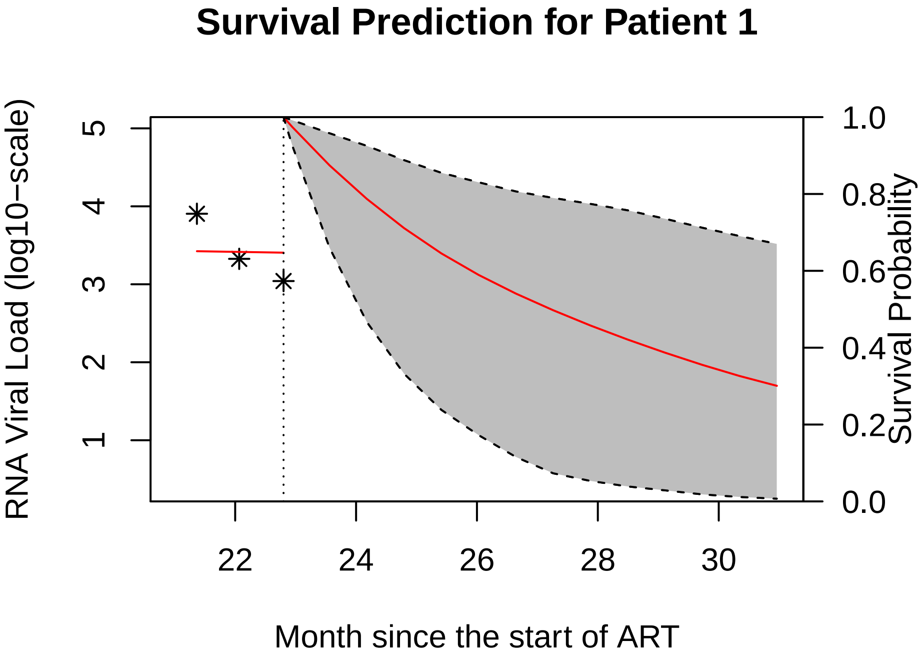

Figure 4.15: Estimated Conditional Survival Probabilities on Viral Rebound for Patient 1

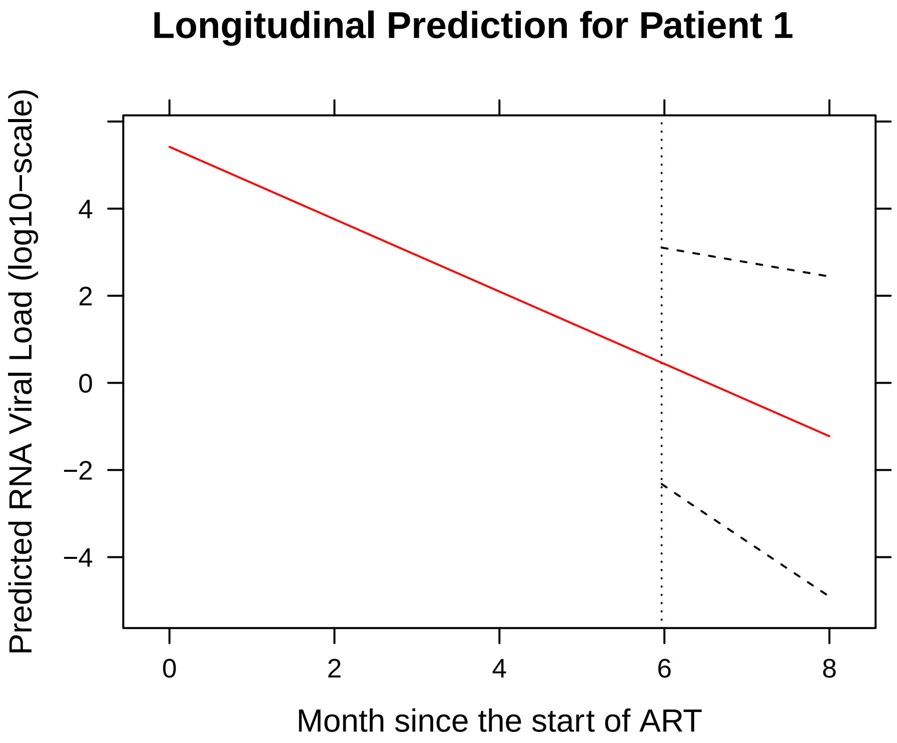

Figure 4.16: Estimated Conditional RNA Viral Load (log10-scale) for Patient 1

Chapter 5

Analysis of HIV Data using Joint Models Based on Likelihood Method

5.1 Motivation

Apart from the Bayesian method detailed in the previous chapter, statistical inferences of the joint model can be made based on the likelihood method. When a non-informative prior is used to derive the Bayesian inference, no prior knowledge on parameters is integrated into the calculation. Therefore, Bayesian estimates under a flat prior should be comparable to the maximum likelihood estimates.

In this chapter, the HIV dataset was used to demonstrate an analysis of joint models using the likelihood method which is implemented by the R package `JM`. The MLE and Bayesian estimates with non-informative priors were compared to see if they agreed with each other.

5.2 Theoretical Framework

The maximum likelihood estimates on random effects coefficient $\boldsymbol{b_i}$ and all the other unknown parameters $\boldsymbol{\theta}$ are obtained by finding the values that maximize the likelihood function

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(t_i, \delta_i | \boldsymbol{b_i}; \boldsymbol{\theta}, \boldsymbol{\beta}) f(\boldsymbol{y_i} | \boldsymbol{b_i}, \boldsymbol{\theta}) f(\boldsymbol{b_i} | D) d\boldsymbol{b_i} \tag{5.1}$$

where δ_i is the survival indicator defined in Section 2.2.1 (Hazard and Survival Function), $\boldsymbol{\beta}$ is the fixed effects coefficient, $\boldsymbol{y_i}$ is the longitudinal outcome for i-th individual, and D is the variance-covariance matrix for random effects $\boldsymbol{b_i}$. The survival process can be expressed as

$$f(t_i, \delta_i | \boldsymbol{b_i}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \{h_i(t_i | \mathcal{H}_i(t_i); \boldsymbol{\theta}, \boldsymbol{\beta})\}^{\delta_i} \mathcal{S}(t_i | \mathcal{H}_i(t_i); \boldsymbol{\theta}, \boldsymbol{\beta}) \tag{5.2}$$

where $\mathcal{H}_i(t)$ denotes the longitudinal history up to time t, $h(.)$ is given by Equation 3.3

$$h_i(t) = h_0(t)\text{exp}[\boldsymbol{\gamma}^T \boldsymbol{w_i} + f\{\eta_i(t), \boldsymbol{b_i}, \boldsymbol{\alpha}\}] \tag{5.3}$$

and the survival function $\mathcal{S}(.)$ is defined as

$$\begin{aligned} \mathcal{S}(t | \mathcal{H}_i(t); \boldsymbol{\theta}, \boldsymbol{\beta}) &= P(T_i^* > t | \mathcal{H}_i(t); \boldsymbol{\theta}, \boldsymbol{\beta}) \\ &= \text{exp}\{- \int_0^t h_i(s | \mathcal{H}_i(t); \boldsymbol{\theta}, \boldsymbol{\beta}) ds\} \end{aligned} \tag{5.4}$$

Since the joint likelihood given by Equation 5.1 might be intractable for high dimensional parameter space, to maximize the likelihood function, numerical integration techniques such as Gaussian quadrature and Monte Carlo can be used to approximate the above integral. In addition, Laplace approximation and EM algorithm are frequently used to obtain approximated estimations on model parameters [23].

The package `JM` implements the Gauss-Hermite quadrature method for approximating integrals in Equation 5.1. Integrals in Equation 5.4 are approximated with the Gauss–Kronrod quadrature method. The EM algorithm is used to maximize the likelihood function. If EM encounters non-convergence issues, a quasi-Newton algorithm would replace EM as the optimization method until convergence is achieved. As for the baseline hazard function $h_0(t)$ in Equation 3.3, `JM` provides several model specifications and numerical integration options. Two of the methods used for the subsequent analysis are

- Piecewise-constant baseline hazard function

$$h_0(t) = \sum_{q=1}^{Q} \xi_q I(v_{q-1} < t \leq v_q)$$

  where v denotes the time-scale, and xi_q is the hazard of the interval $(v_{q-1}, v_q]$.

- Spline-approximated baseline hazard function

$$\text{log} h_0(t) = \gamma_{h_0,0} + \sum_{q=1}^{Q} \gamma_{h_0,q} B_q(t, v)$$

  where $B_q(t, v)$ is the q-th basis function of a B-spline with knot v.

The above two baseline hazard functions are approximated with the Gauss-Hermite integration rule.

5.3 Model Specification and Results

The R package `JM` supports two association structures $f(.)$. The first one links the longitudinal and survival process through $\eta_i(t)$, that is,

$$f = \alpha\eta_i(t).$$

The second option links the two processes through $\eta_i(t)$ and its corresponding slope $\eta_i'(t)$, that is,

$$f = \alpha_1\eta_i(t) + \alpha_2\eta_i'(t).$$

These two specifications of the association structure correspond to Analysis 5-8: Joint Model with Shared True Longitudinal Values $\eta_i(t)$ and Analysis 9-12: Joint Model with Shared True Longitudinal Value $\eta_i(t)$ and Corresponding Slope $\eta_i'(t)$. Thus, Analysis 5-8 and 9-12 evaluated with association structures using non-informative priors were re-examined with the likelihood method. To recap on the model specifications of the above-mentioned analyses, models of interest are listed as follows,

1. Joint Model with Shared True Longitudinal Values $\eta_i(t)$

- Analysis 5: Association between viral rebound and viral decay

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_y^{(r)}\eta_i^{(r)}(t)\right) \tag{5.5}$$

- Analysis 6: Association between viral peak point and viral decay

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_y^{(p)}\eta_i^{(p)}(t)\right) \tag{5.6}$$

- Analysis 7: Association between viral rebound and CD4 cell counts

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_z^{(r)}\eta_i^{(r)}(t)\right) \tag{5.7}$$

- Analysis 8: Association between viral peak point and CD4 cell counts

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_z^{(p)}\eta_i^{(p)}(t)\right) \tag{5.8}$$

2. Joint Model with Shared True Longitudinal Value $\eta_i(t)$ and Corresponding Slope $\eta_i'(t)$

- Analysis 9: Association between viral rebound and viral decay

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_{y_1}^{(r)}\eta_i^{(r)}(t) + \alpha_{y_2}^{(r)}\eta_i'^{(r)}(t)\right) \tag{5.9}$$

- Analysis 10: Association between viral peak point and viral decay

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_{y_1}^{(p)}\eta_i^{(p)}(t) + \alpha_{y_2}^{(p)}\eta_i'^{(p)}(t)\right) \tag{5.10}$$

- Analysis 11: Association between viral rebound and CD4 cell counts

$$h_i^{(r)}(t) = h_0^{(r)}(t)\text{exp}\left(\alpha_{z_1}^{(r)}\eta_i^{(r)}(t) + \alpha_{z_2}^{(r)}\eta_i'^{(r)}(t)\right) \tag{5.11}$$

- Analysis 12: Association between viral peak point and CD4 cell counts

$$h_i^{(p)}(t) = h_0^{(p)}(t)\text{exp}\left(\alpha_{z_1}^{(p)}\eta_i^{(p)}(t) + \alpha_{z_2}^{(p)}\eta_i'^{(p)}(t)\right) \tag{5.12}$$

Let $\alpha_\eta = \{\alpha_y^{(r)}, \alpha_y^{(p)}, \alpha_z^{(r)}, \alpha_z^{(p)}\}$, $\alpha_{\eta_1} = \{\alpha_{y_1}^{(r)}, \alpha_{y_1}^{(p)}, \alpha_{z_1}^{(r)}, \alpha_{z_1}^{(p)}\}$, and $\alpha_{\eta_2} = \{\alpha_{y_2}^{(r)}, \alpha_{y_2}^{(p)}, \alpha_{z_2}^{(r)}, \alpha_{z_2}^{(p)}\}$. For the Bayesian method, prior distributions for these association structures take on large variance, specifically,

$$\alpha_\eta \sim N(0, 100), \quad \alpha_{\eta_1} \sim N(0, 100), \quad \alpha_{\eta_2} \sim N(0, 100) \tag{5.13}$$

such that the priors can be regarded as relatively non-informative. Maximum likelihood estimates and Bayesian estimates on the association structure were compared and summarized in Table 5.1.

Table 5.1: Likelihood and Bayesian Estimates on Association Structure

	Likelihood			Bayesian		
	Estimate	S.E.	p-value [1]	Estimate	S.E	Credible Interval [1]
Shared $\eta_i(t)$						
$\alpha_y^{(r)}$	-0.433	0.139	0.002	-0.163	0.013	(-0.337, 0.009)
$\alpha_y^{(p)}$	-0.101	0.076	0.184	-0.071	0.005	(-0.173, 0.001)
$\alpha_z^{(r)}$	0.422	0.345	0.221	0.179	0.007	(0.011, 0.347)
$\alpha_z^{(p)}$	-0.547	0.599	0.361	-0.246	0.025	(-0.565, 0.004)
Shared $\eta_i(t)$ and $\eta_i'(t)$						
$\alpha_{y_1}^{(r)}$	-0.677	0.194	< 0.001	-0.477	0.026	(-0.825, -0.130)
$\alpha_{y_2}^{(r)}$	5.968	3.140	0.057	5.311	0.431	(0.665, 11.486)
$\alpha_{y_1}^{(p)}$	-0.100	0.096	0.295	-0.055	0.006	(-0.186, 0.043)
$\alpha_{y_2}^{(p)}$	-2.687	3.240	0.407	-0.645	0.105	(-3.060, 2.229)
$\alpha_{z_1}^{(r)}$	0.299	0.424	0.480	0.121	0.026	(-0.183, 0.407)
$\alpha_{z_2}^{(r)}$	1.854	7.240	0.798	0.896	0.407	(-4.944, 6.829)
$\alpha_{z_1}^{(p)}$	-1.314	0.703	0.062	-0.282	0.026	(-0.538, -0.015)
$\alpha_{z_2}^{(p)}$	-0.757	21.513	0.972	2.202	0.258	(-4.672, 9.641)

[1] Significant results are highlighted in red

5.4 Discussion on MLE and Bayesian Estimates

The results have shown that, in general, the likelihood and Bayesian method produce similar results in terms of direction of association, although some estimates differ in statistical significance and the scale of estimated value. The likelihood and Bayesian results should be compared and contrasted with the following issues taken into account.

Bayesian estimates should resemble MLE results when all of the unknown parameters follow non-informative priors. However, priors of certain parameters did not take on flat priors in the sensitivity analyses conducted earlier. For the variance of the error terms, its prior distribution is described in Equation 3.20. Similarly, the prior distribution for variance-covariance matrix of random effects is described in Equation 3.21. Priors for the above-mentioned parameters remained unchanged for sensitivity analyses. Thus, Bayesian estimates obtained from sensitivity analyses do not equate to MLE results because not all prior distributions for unknown parameters are necessarily non-informative.

Secondly, the standard errors for MLE and Bayesian estimates differ to some extent. This is due to the difference in definition of standard errors under the frequentist and Bayesian framework. The standard error for MLE is defined as the standard deviation of the sampling distribution, that is

$$SE = \frac{\sigma}{\sqrt{N}}$$

where σ is the sample standard deviation and N is the sample size; whereas the posterior standard error is defined as

$$SE = \frac{\sigma}{\sqrt{N_{eff}}}$$

where σ is the posterior standard deviation obtained from MCMC samples, and N_eff is the effective sample size defined in Section 3.4.2. It is worth noting that the standard error for MLE were considerably large with respect to its estimated value. Large standard error is an indicator of the estimated parameter value being inaccurate and unreliable.

Lastly, the previous chapter has shown that MCMC diagnostics for joint models with shared $\eta_i(t)$ and joint models with shared $\eta_i(t)$ and $\eta_i'(t)$ reveal some problems on convergence and insufficient effective sample size. Even though the software did not produce any non-convergence errors when running the program, the fact that MCMC converges slowly should raise some concerns over the quality of the Bayesian approximation. Unsatisfactory diagnostics on the joint model are often evidence of the empirical distribution being a poor approximation to the true posterior distribution meaning that the Bayesian estimates may not be an accurate representation of the true parameter value.

Therefore, conflicting results obtained from MLE and the Bayesian estimates might be attributable to the use of non-informative priors on selected parameters, different definition of the standard error, large standard error of MLE and poor MCMC diagnostics of the Bayesian method.

Chapter 6

Conclusion

The relationship between viral loads/CD4 cell counts and characteristics of viral rebound/peak point is first explored with joint modelling using the Bayesian method. Based on results from the sensitivity analysis, model comparison with DIC values and MCMC diagnostics, joint models with shared random effects is the most appropriate specification for modelling the relationship between the longitudinal and survival process. Results suggest that neither the initial level nor the rate of change of RNA viral loads or CD4 cell counts is significantly associated with the risk of viral rebound or viral peak point. One thing worth noting is that, despite of an insignificant association, there is some evidence supporting that the initial CD4 cell counts and its rate of change are negatively associated with viral peak point. In other words, patients with higher initial levels of CD4 cell counts and patients with faster CD4 regeneration rate are less likely to experience viral peak point after treatment termination. Individual-specific predictions on the survival probability and longitudinal outcome are made possible by joint modelling. These above-mentioned results can help physicians personalize treatment plans for patients based on their own estimated prognoses.

To supplement the Bayesian method, inferences of the joint model are obtained by the likelihood method. Results show that in general, maximum likelihood estimates are comparable to the Bayesian estimates when non-informative priors are used. However, the standard error for MLE is considerably larger than that of the Bayesian estimates. This could be caused by the different definition of standard errors under the frequentist and Bayesian framework. Additionally, standard errors for MLE are substantially large with respect to their estimates. Large standard errors are a sign that the estimated value being inaccurate. The reason for large standard errors is not clear at this moment, but it could be due to inaccuracy of the numerical integration.

Future analyses can focus on modelling the longitudinal process with nonlinear mixed effects (NLME) models which better capture the HIV viral load dynamics over the entire period of study. The HIV dynamics have been well established which makes it convenient when modelling the viral loads with the NLME model. Another advantage of using NLME models is that model specification is based on the underlying data-generation mechanism meaning that it provides more reliable predictions than linear mixed effects model.

Bibliography

[1] Brian D. Marx Paul H. C. Eilers. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–121, 1996. → page 13

[2] F Harrell. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* Springer-Verlag, 2001. → page 14

[3] Veyrat-Follet C Guedj J Desmée S, Mentré F. Nonlinear mixed-effect models for prostate-specific antigen kinetics and link with survival in the context of metastatic prostate cancer: A comparison by simulation of two-stage and joint approaches. *AAPS J*, 17(3):691–699, 2015. → page 17

[4] de Freitas N.-Doucet A. et al Andrieu, C. An introduction to mcmc for machine learning. *Machine Learning*, 50:5–43, 2003. → page 19

[5] Matteson D.S Ruppert D. *Bayesian Data Analysis and MCMC In Statistics and Data Analysis for Financial Engineering*. Springer Texts in Statistics, 2015. → page 19

[6] Jennifer A. Hoeting Geof H. Givens. *Computational Statistics*. John Wiley Sons, October 2012. → page 21

[7] Simon Jackman. *Bayesian Analysis for the Social Sciences*. John Wiley Sons, October 2009. → page 22

[8] Andrew Gelman. Prior distribution. *Encyclopedia of Environmetrics*, 3:1634–1637, 2002. → page 22

[9] Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer Science Business Media, June 2009. → page 23

[10] Dimitris Korobilis Gary Koop. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358, 2010. → page 23

[11] B Efron. Frequentist accuracy of bayesian estimates. *J R Stat Soc Series B Stat Methodol*, 77(3):617–646, June 2015. → page 24

[12] Costa LM Saragiotto BT Hespanhol L, Vallio CS. Understanding and interpreting confidence and credible intervals around effect estimates. *Braz J Phys Ther.*, 23(4):290–301, 2019. → page 25

[13] Dimitris Rizopoulos. *The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data using MCMC.* Erasmus Medical Center Rotterdam. → page 25

[14] Stan Development Team. *Stan Reference Manual*, 2.23 edition. → page 26

[15] Dimitris Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819–829, 2011. → page 27

[16] Emily Land. Undetectable to viral rebound: When and why? → page 29

[17] Pierson T et al. Finzi D, Hermankova M. Identification of a reservoir for hiv-1 in patients on highly active antiretroviral therapy. *Science*, 278(5341):1295–1300, 1997. → page 29

[18] Jeffrey N. Rouder, Christopher R. Engelhardt, Simon McCabe, and Richard D. Morey. Model comparison in anova. *Psychonomic Bulletin & Review*, 23(6):1779–1786, 2016. → page 37

[19] Małgorzata Roos, Thiago G. Martins, Leonhard Held, and Håvard Rue. Sensitivity analysis for bayesian hierarchical models. *Bayesian Analysis*, 10(2):321–349, Jun 2015. → page 44

[20] Mitchell J. Eaton William A. Link. On thinning of chains in mcmc. *Methods in Ecology and Evolution*, 3(1):112–115, February 2012. → page 45

[21] Angelika van der Linde. Dic in variable selection. *Statistica Neerlandica*, 59(1), Feb. → page 59

[22] Marta García-Fiñana Ruwanthi Kolamunnage-Dona Maha Alsefri, Maria Sudell. Bayesian joint modelling of longitudinal and time to event data: a methodological review. *BMC Medical Research Methodology*, April 2020. → page 59

[23] Dimitris Rizopoulos. Jm: An r package for the joint modelling of longitudinal and time-to-event data. → page 72

Chapter 7

Appendix

7.1 R code

R code for data processing, visualization, and data analysis on the HIV dataset can be found on my GitHub repository:

https://github.com/harpercheng91/UBC-Master-s-Project

7.2 LME Diagnostic Plots

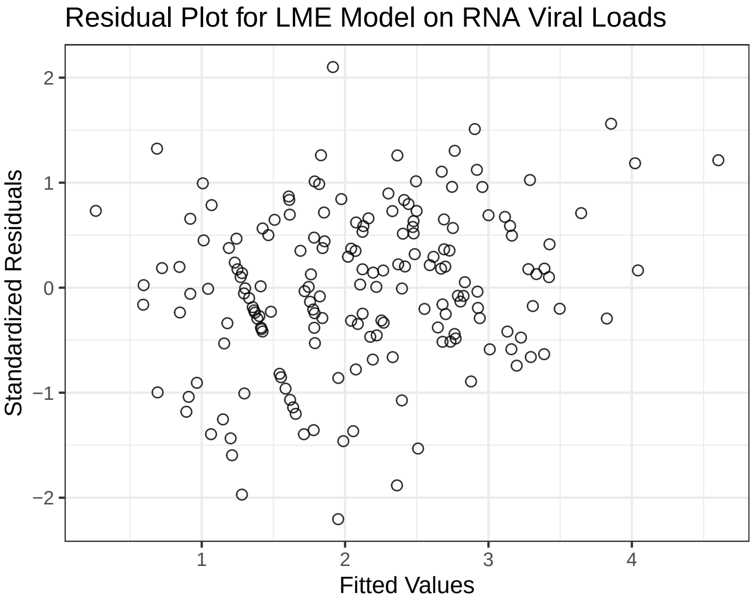

Figure 7.1: Residual Plot for RNA Viral Loads

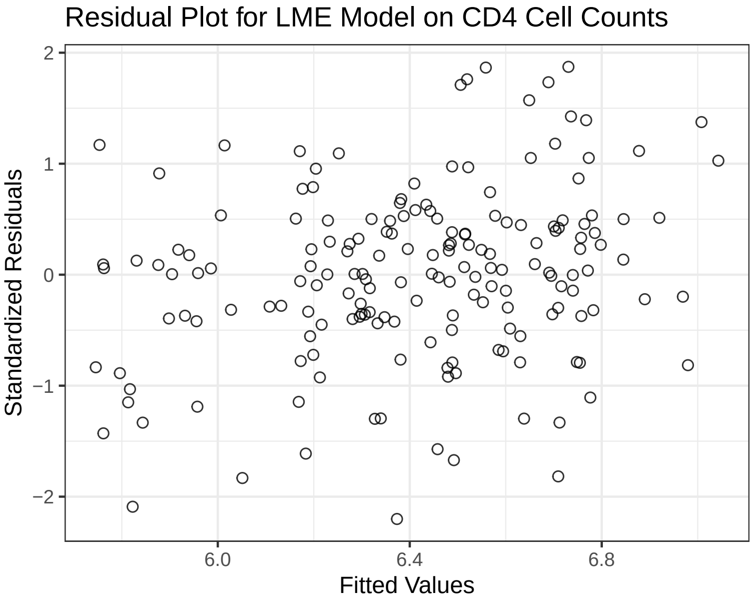

Figure 7.2: Residual Plot for CD4 Cell Counts

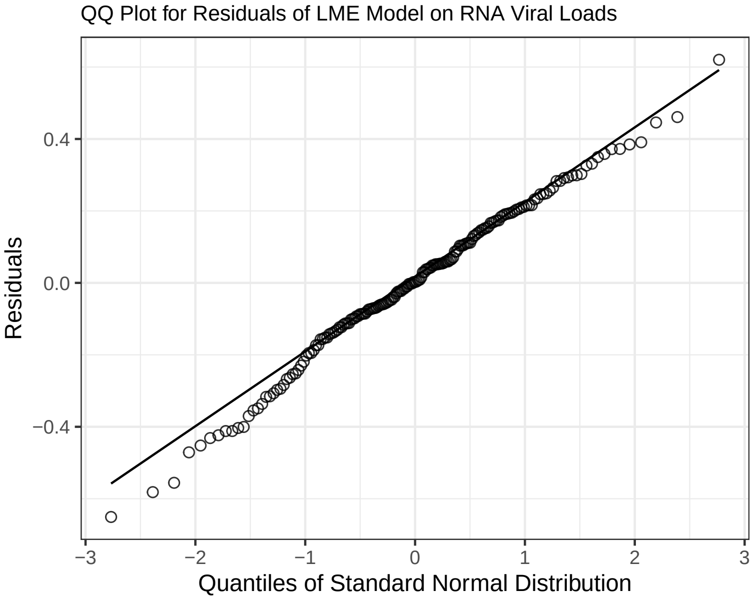

Figure 7.3: QQ Plot of Residuals for RNA Viral Loads

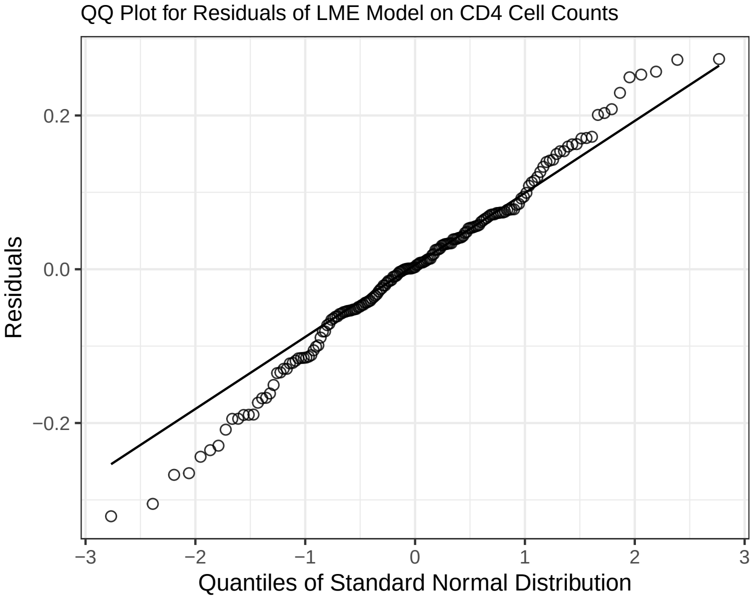

Figure 7.4: QQ Plot of Residuals for CD4 Cell Counts

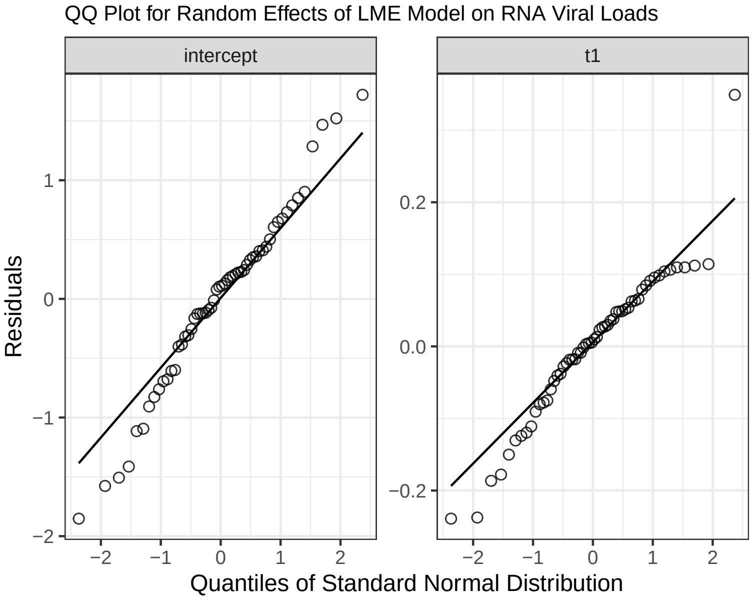

Figure 7.5: QQ Plot of Random Effects for RNA Viral Loads

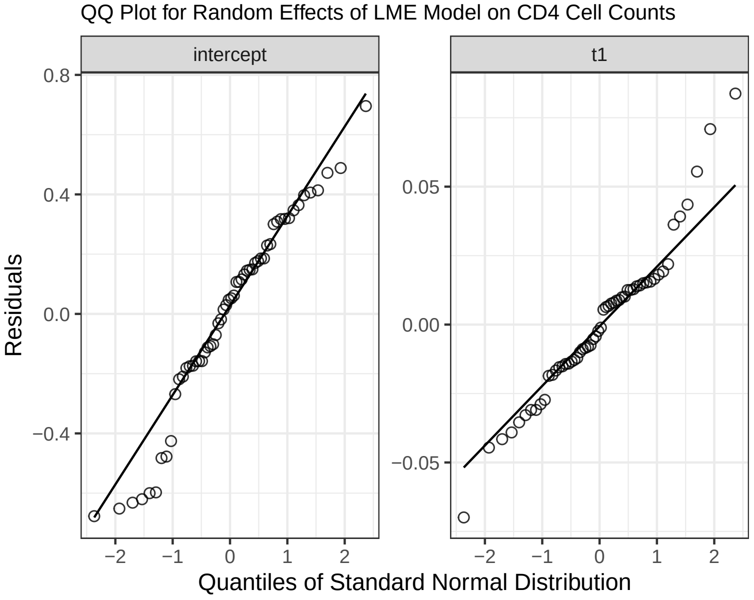

Figure 7.6: QQ Plot of Random Effects for CD4 Cell Counts

7.3 Joint Model Diagnostic Plots

7.3.1 Trace Plot for Analysis 2-4

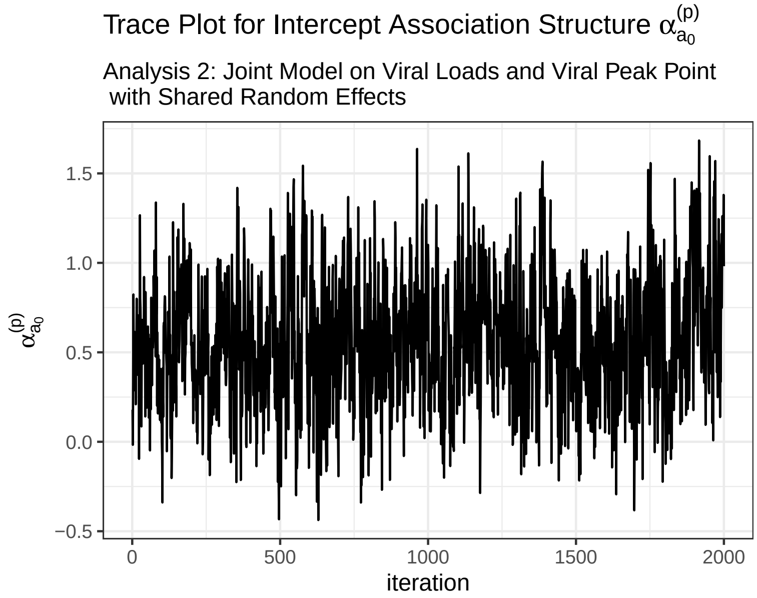

Figure 7.7: Trace Plot for Intercept Association Structure (Analysis 2)

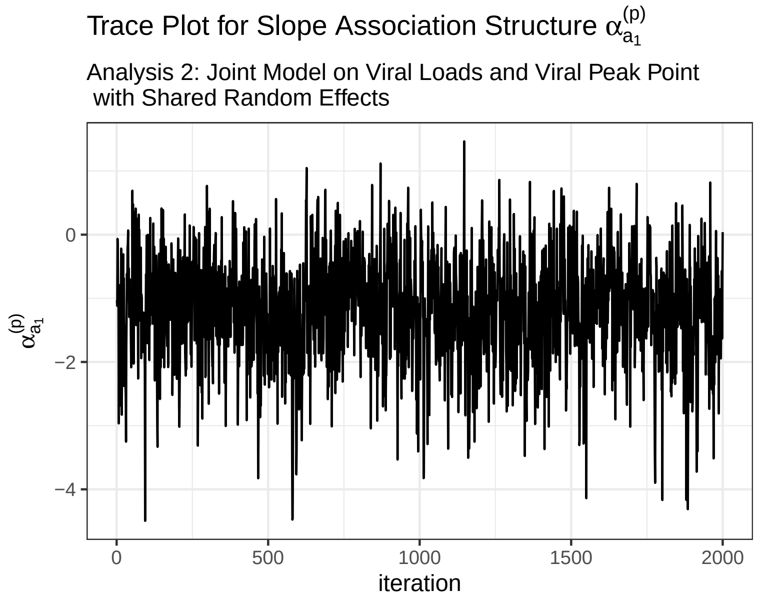

Figure 7.8: Trace Plot for Slope Association Structure (Analysis 2)

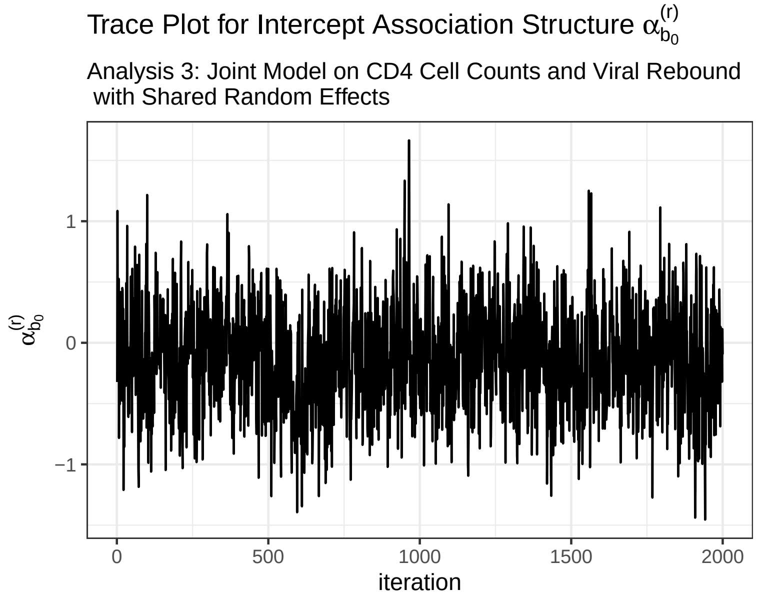

Figure 7.9: Trace Plot for Intercept Association Structure (Analysis 3)

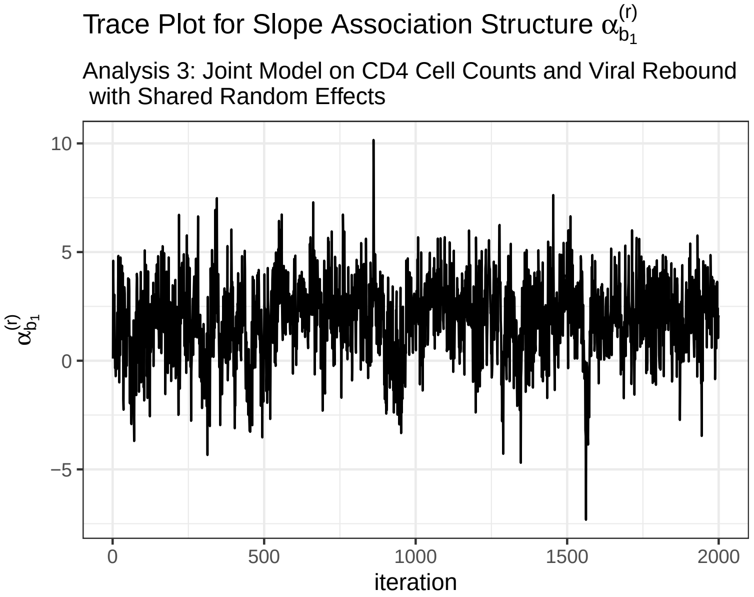

Figure 7.10: Trace Plot for Slope Association Structure (Analysis 3)

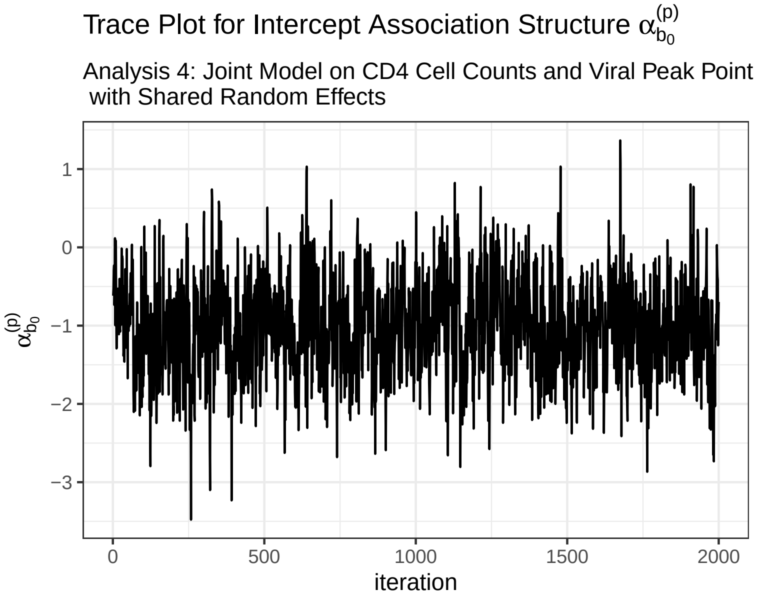

Figure 7.11: Trace Plot for Intercept Association Structure (Analysis 4)

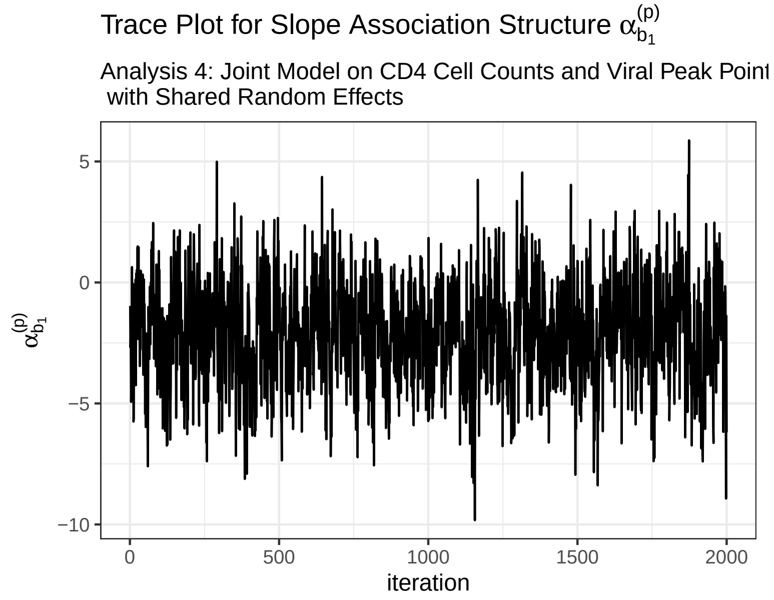

Figure 7.12: Trace Plot for Slope Association Structure (Analysis 4)

7.3.2 ACF Plot for Analysis 2-4

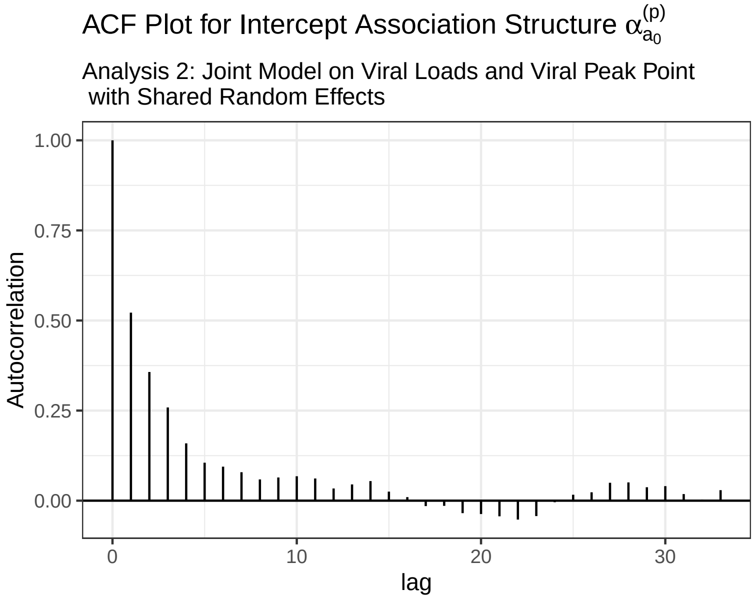

Figure 7.13: ACF Plot for Intercept Association Structure (Analysis 2)

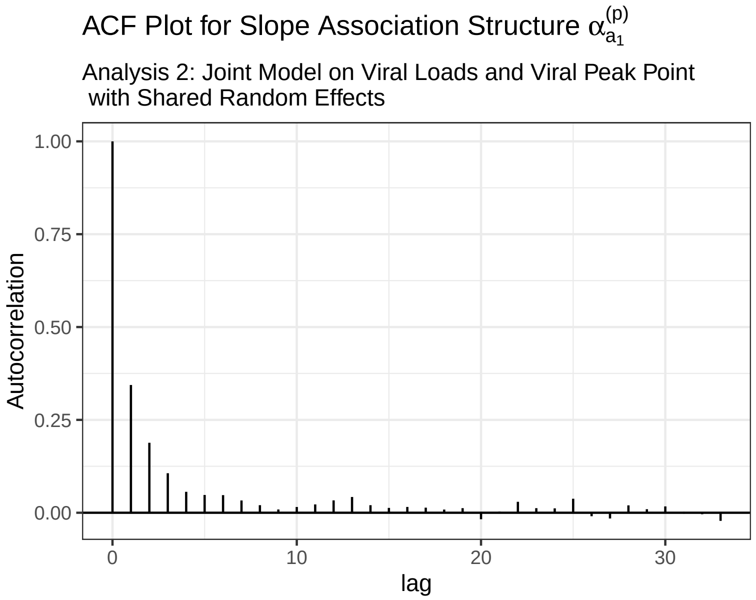

Figure 7.14: ACF Plot for Slope Association Structure (Analysis 2)

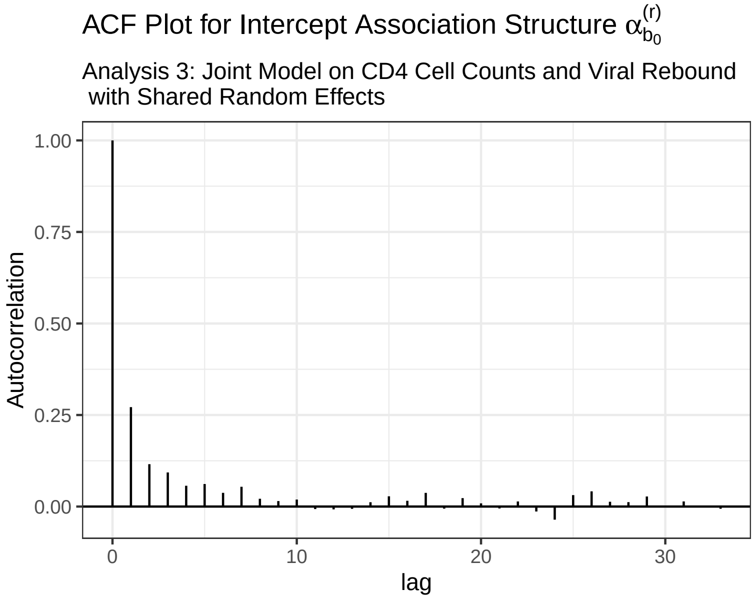

Figure 7.15: ACF Plot for Intercept Association Structure (Analysis 3)

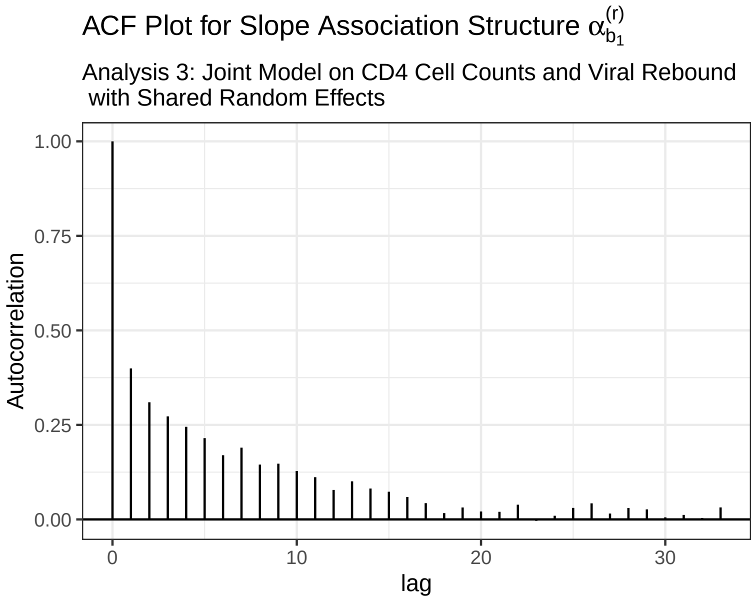

Figure 7.16: ACF Plot for Slope Association Structure (Analysis 3)

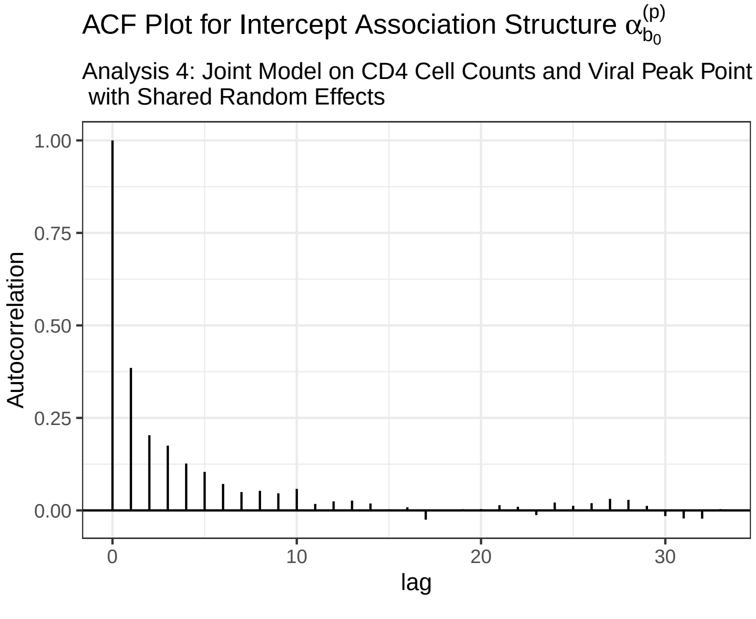

Figure 7.17: ACF Plot for Intercept Association Structure (Analysis 4)

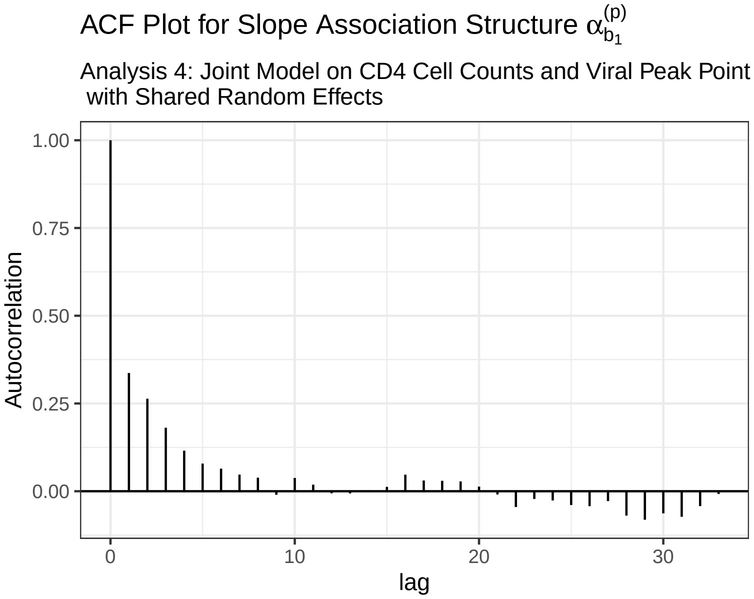

Figure 7.18: ACF Plot for Slope Association Structure (Analysis 4)

7.3.3 Trace Plot for Analysis 14-16

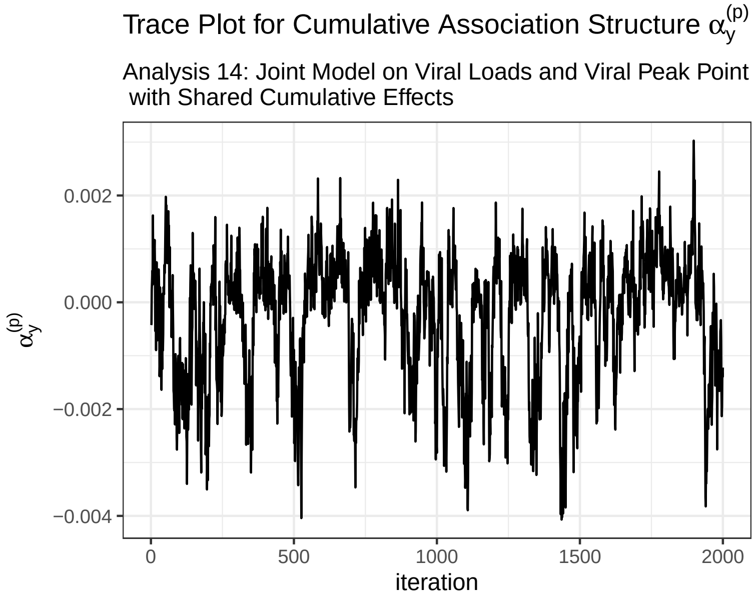

Figure 7.19: Trace Plot for Cumulative Association Structure (Analysis 14)

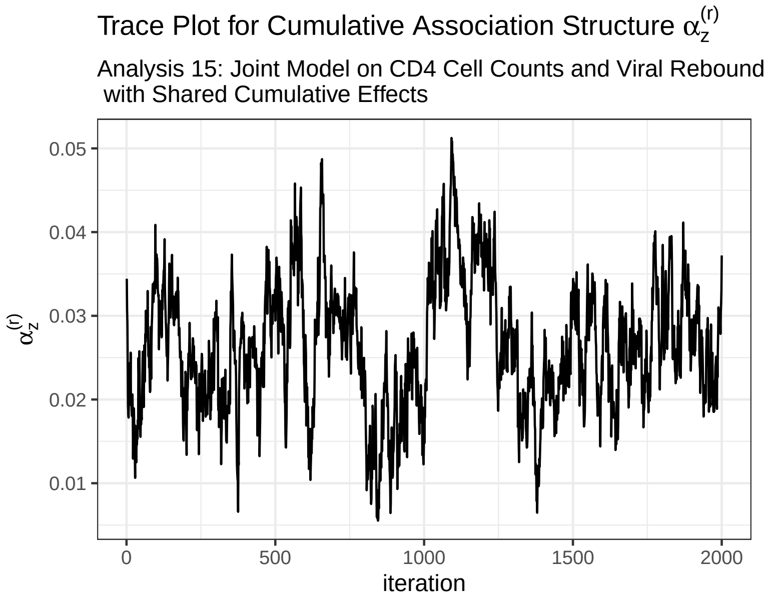

Figure 7.20: Trace Plot for Cumulative Association Structure (Analysis 15)

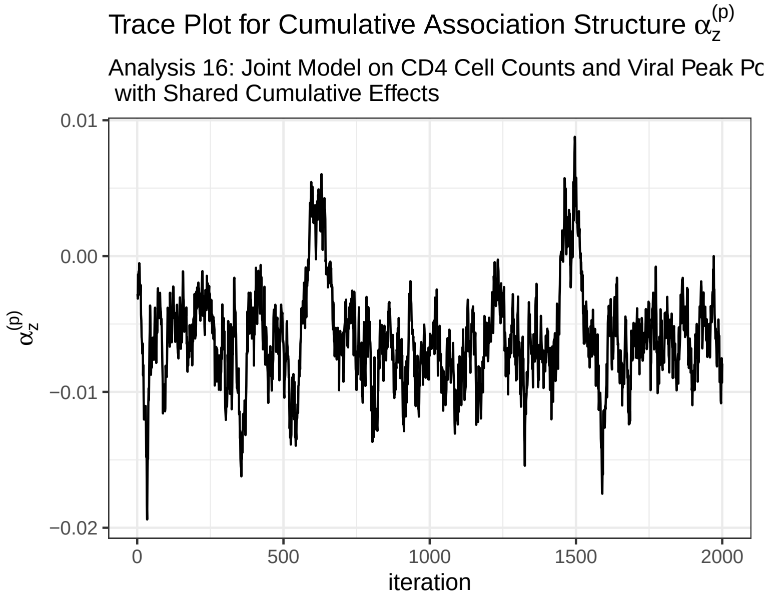

Figure 7.21: Trace Plot for Cumulative Association Structure (Analysis 16)

7.3.4 ACF Plot for Analysis 14-16

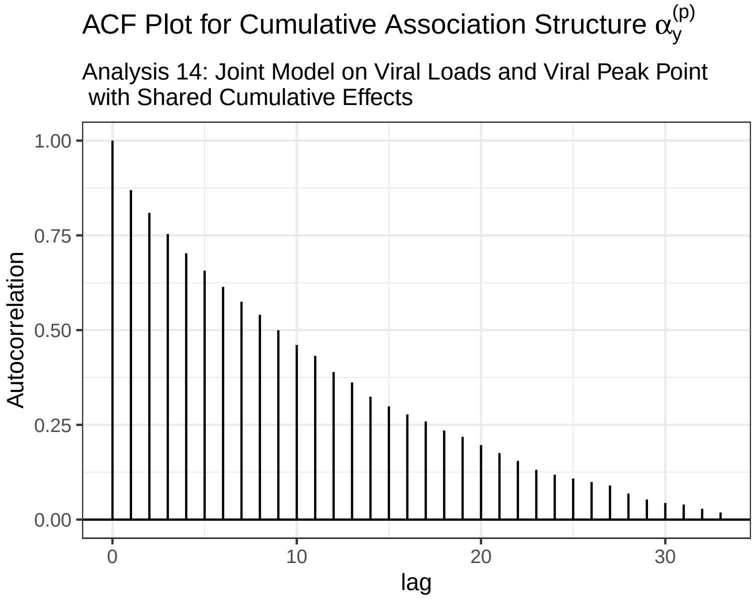

Figure 7.22: ACF Plot for Cumulative Association Structure (Analysis 14)

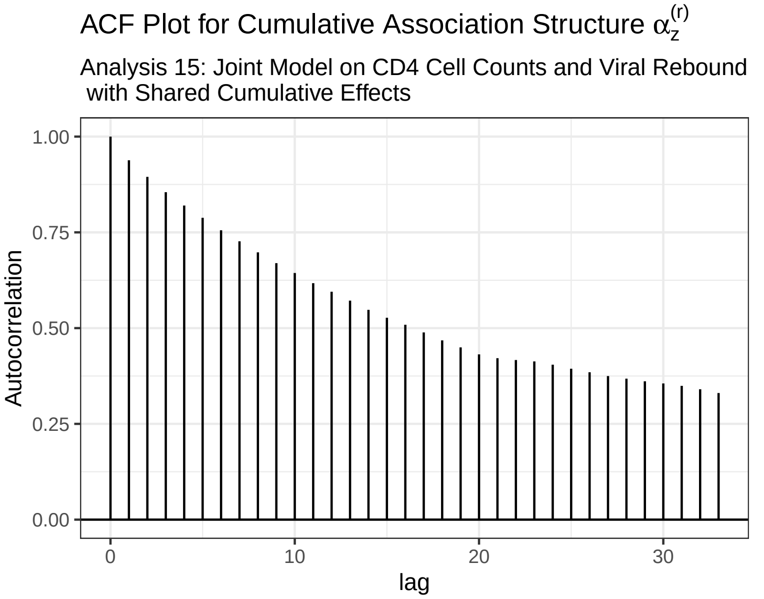

Figure 7.23: ACF Plot for Cumulative Association Structure (Analysis 15)

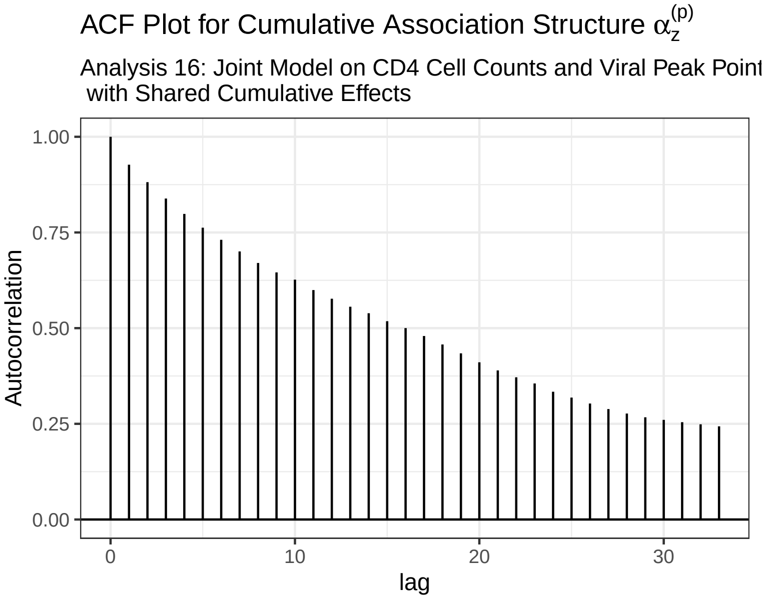

Figure 7.24: ACF Plot for Cumulative Association Structure (Analysis 16)