Journal of Statistical Research *https://doi.org/10.3329/jsr.v58i1.75414*
2024, Vol. 58, No. 1, pp. 75-96 *ISSN 0256 - 422 X*

JOINT MODELS FOR LONGITUDINAL DATA

LANG WU[*]

Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Email: lang@stat.ubc.ca

SIHAOYU GAO

Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Email: sihaoyu.gao@stat.ubc.ca

QIAN YE

Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Email: qian.ye@stat.ubc.ca

SUMMARY

In a longitudinal study, data on different types of variables are often collected repeatedly over time. Some variables may be continuous, and some variables may be binary or times to an event of interest. Even for a single variable, data may be collected at different phases of the study with different characteristics. These different types of variables are typically associated or correlated, since they are measurements on the same individuals in the study. Analysis of data on each of these variables separately, ignoring other variables, may be inefficient and may also lead to biased results. Standard multivariate models with several correlated responses may not be easy to specify for different types of variables or when the models are nonlinear. Jointly modelling these variables simultaneously not only may be more efficient but may also reduce biases in parameter estimation. Statistical inference can then be based on the joint likelihood for all observed data. In this article, we briefly review several different types of joint models for longitudinal data. We focus on mixed effects models and likelihood methods for inference. We illustrate these joint models with datasets from HIV/AIDS studies.

Keywords and phrases: Generalized linear mixed model, joint model, likelihood method, measurement error, nonlinear mixed effects model.

AMS Classification: 62-07

[*] Corresponding author

1 Introduction

In a longitudinal study, data on more than one variables are often collected repeatedly over time. These variables may be of different types, such as some being continuous and some being binary. Because data on these variables are collected on the *same* individuals in the study, they are often correlated or associated. In other words, not only the repeated measurements of a variable may be correlated, data on different variables at a given time point may also be correlated. In data analyses, the correlations or associations among these variables should be incorporated in a statistical model in order to make more efficient statistical inference and avoid potential biases. Analyses for longitudinal data on each variable separately may be inefficient and biased. Standard multivariate models for several correlated responses, such as multivariate linear mixed effects models, may not be appropriate when the variables are of different types or the models are *nonlinear*. Thus, jointly modelling longitudinal data on several variables simultaneously are desirable. In this article, we provide a review of joint models for longitudinal data.

As an example, we consider an HIV/AIDS longitudinal study. Figure 1 shows the longitudinal data collected on two variables (CD4 cell count and viral load). Here viral loads may drop below a detection limit (left censored), while CD4 is known to be measured with errors. As we can see, data on these two variables appear to be negatively correlated. While both variables may be viewed as continuous variables, we are sometimes interested in investigating whether viral load being left censored (a binary variable) may be associated with (true) CD4 values, or we may be interested in investigating whether CD4 below 200 (a binary variable, where 200 is a critical threshold value for CD4 data) may be associated with (true) viral load values. Moreover, sometimes CD4 may be viewed as count data. Thus, two different types of longitudinal variables may be associated. To incorporate this association, joint models may be desirable. We will discuss more details in later sections.

Mixed effects models are widely used in the analysis of longitudinal data, since they allow for both individual-specific inference and population-average inference. Moreover, joint models may be natually specified for mixed effects models. Thus, we focus on mixed effects models in this article. A mixed effects model for longitudinal data can be obtained from the corresponding regression models for cross-sectional data by introducing random effects in the model. These random effects incorporate the association among the repeated measurements since each individual shares the same random effects which may be interpreted as latent characteristics of the individual. Moreover, these random effects reflect the variations of the longitudinal data across individuals.

There are different types of mixed effects models, depending on the types of the response variable as well as its relationship with covariates. In this article, we focus on generalized linear mixed models (GLMMs) and nonlinear mixed effects (NLME) models, since GLMMs are useful for non-normal responses, such as binary or count responses, and NLME models are nonlinear models. In both cases, multivariate models, which also incorporate the association among the response variables, are difficult to specify, making joint modelling an attractive approach. Moreover, both GLMMs and NLME models include linear mixed effects (LME) models as a special case. Semi-parametric or nonparametric mixed effects models may also be approximated by LME models (Wu, 2009). Therefore, GLMM and NLME models cover a wide variety of mixed effects models for

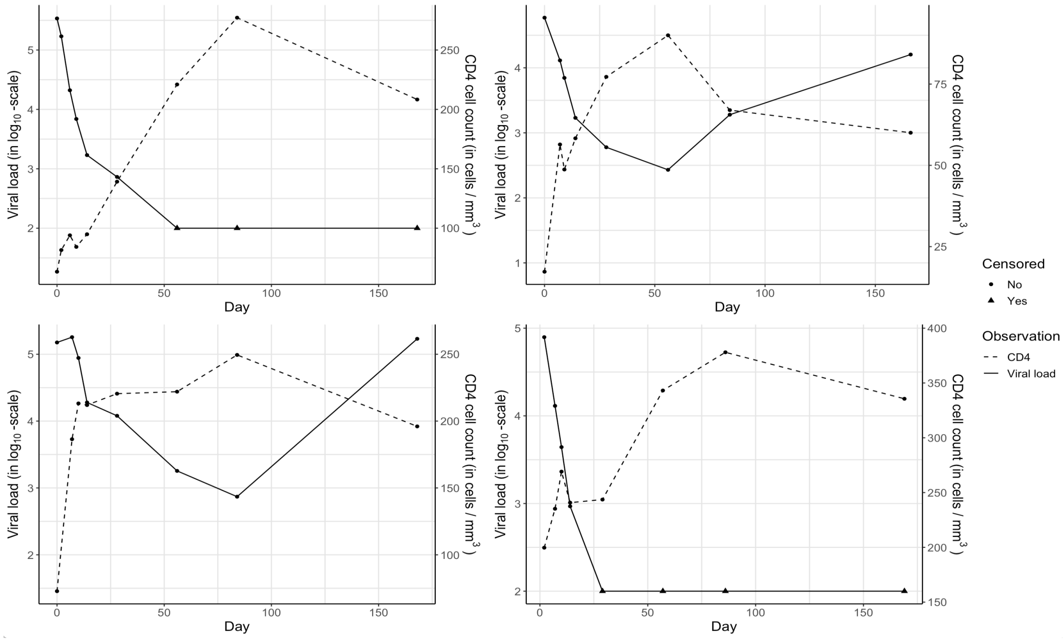

Figure 1: Longitudinal data on CD4 and viral loads (in $\log_{10}$ scale) for four randomly selected subjects. Viral load trajectories are represented by the solid lines, while CD4 trajectories are represented by dashed lines. Observed values are denoted by solid dots, while left-censored viral load values are imputed by the detection limit and are denoted by triangular dots.

longitudinal data.

We first consider a few simple examples where joint models may be useful. For simplicity, here we focus on LME models, though later we consider more general GLMM and NLME models. Let y_{ij} and x_{ij} be two continuous variables for individual i measured at time t_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$. Joint models are desirable in the following situations:

- *Example 1: Covariate measurement error problem.* Suppose that y is a response and x is a covariate with measurement errors. To address measurement errors, we may consider the following joint model

$$\begin{aligned} y_{ij} &= \beta_0 + b_{0i} + \beta_1 x^*_{ij} + \epsilon_{ij}, \\ x_{ij} &= \alpha_0 + a_{0i} + \alpha_1 t_{ij} + e_{ij} = x^*_{ij} + e_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \end{aligned} \tag{1.1}$$

where β_0, β_1, α_0, and α_1 are fixed effects, b_{0i} and a_{0i} are the random effects, x^*_{ij} is the true but unobserved covariate value of x_{ij}, ϵ_{ij} is random error, and e_{ij} denotes measurement errors. The above two models are joint or associated since they share the same variable x^*. Similar joint models may be obtained if the LME response model is replaced by a GLMM or NLME model.

- *Example 2: Modelling two associated longitudinal processes.* Suppose that we are interested in the association between two longitudinal variables x and y in a longitudinal study, such as CD4 and viral load. We may then consider the following models

  $$\begin{aligned} y_{ij} &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})t_{ij} + \epsilon_{ij}, \\ x_{ij} &= \alpha_0 + a_{0i} + (\alpha_1 + a_{1i})t_{ij} + e_{ij}, \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i. \end{aligned}$$

  To incorporate the association between x and y, we may assume that the random effect b_{1i} and a_{1i} in the two models are correlated, since these random effects may represent latent characteristics of the same individuals. Then, joint models are desirable to make simultaneous inference. The above two LME models may also be replaced by GLMM or NLME models.

- *Example 3: Joint models for longitudinal and survival data.* In a longitudinal study, we may also be interested in certain events such as dropouts. When modeling the longitudinal data with informative dropouts, we may consider the following joint model

  $$\begin{aligned} y_{ij} &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})x_{ij} + \epsilon_{ij}, \\ h_i(t) &= h_0(t)\text{exp}(\beta_2 b_{0i} + \beta_3 b_{1i}), \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i, \end{aligned}$$

  where $h_i(t)$ represents the hazard function of the event for individual i at time t, $h_0(t)$ is an unspecified baseline hazard function, and β_j's are fixed effects. Here the random effects in the longitudinal model, (b_{0i}, b_{1i}), are viewed "covariates" in the survival model. Alternatively, we may consider a survival model with measurement errors in a time-dependent covariate y_{ij}.

In Section 2, we will discuss more general joint models.

For statistical inference of joint models, a commonly used so-called *two-step method* proceeds as follows:

- Step 1: Estimate the shared variables or parameters in one model based on the observed data, ignoring the other model.

- Step 2: Estimate the parameters in the other model separately, substituting the shared variables or parameters with the estimated values from Step 1.

While the two-step method is straightforward to implement, it may under-estimate the standard errors of parameter estimates since it ignores uncertainty in estimation in Step 1, and it may even lead to biased estimation in some cases. Moreover, the two-step estimates may not be fully efficient. Therefore, it is desirable to conduct statistical inference for joint models based on the joint likelihood of all the observed data. The maximum likelihood estimates (MLEs) of all model parameters can be simultaneously obtained by maximizing the joint likelihood. A potential challenge of the joint likelihood approach lies in computational complexity, as joint likelihoods often involve high-dimensional and intractable integrals.

There is a large literature on the analysis of longitudinal data and joint models, especially joint models for longitudinal and survival data. Longitudinal data analysis based on mixed effects models are reviewed in Davidian and Giltinan (1995), Fitzmaurice et al. (2012), McCulloch and Searle (2004), Vonesh (2014), Pinheiro and Bates (2006), Wu (2009), and Lavielle (2014), among others. Joint modeling for longitudinal data and survival data is reviewed in Rizopoulos (2012), Elashoff et al. (2016), among others. Since there has been sufficient reviews of joint models for longitudinal and survival data, in this article we will focus on joint models for longitudinal data. References will be given in the corresponding sections. The article is organized as follows. In Section 2, we will describe several common situations where joint models are desirable. In Section 3, we present several examples to illustrate the joint models. We conclude the article with some discussion in Section 4.

2 Joint Models for Longitudinal Data

In this section, we briefly review several situations where joint models are desirable. As noted in Section 1, we focus on GLMM and NLME models since they include LME models as special cases and the usual multivariate models may not be easy to specify.

2.1 GLMM with covariate measurement errors

Let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$ be the n_i repeated measurements of the response y for individual i, $i = 1, 2, \ldots, n$. We consider a time-dependent covariate w, which is measured with errors. Let $\mathbf{w}_i = (w_{i1}, w_{i2}, \ldots, w_{in_i})^T$ denote the repeated measurements of the error-prone covariate w for individual i. In addition to $\mathbf{w}_i$, we may also consider other covariates, denoted as $\mathbf{x}_i$ and $\mathbf{z}_i$, which can be either baseline or time-dependent variables without measurement errors. A general GLMM for the response y can be written as

$$\begin{aligned} h(E(y_{ij})) &= \beta_w w_{ij}^* + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}_i, \\ \mathbf{b}_i &\sim N(0, B), \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i, \end{aligned} \tag{2.1}$$

where $h(\cdot)$ is a known link function, such as the logit link for binary responses, w_{ij}^* denotes the true but unobserved value of covariate w_{ij}, β_w and $\boldsymbol{\beta}$ are fixed parameters, $\mathbf{b}_i$ represents random effects, and B is a covariance matrix. We assume that, conditioning on the random effects $\mathbf{b}_i$, the repeated measurements $y_{i1}, y_{i2}, \ldots, y_{in_i}$ are independent and each follows a distribution in the exponential family, which includes binomial distributions, Poisson distribution, and normal distributions. Note that, when the link function $h(y) = y$, the above GLMM reduces to an LME model.

To address measurement errors, we consider the following NLME model for covariate w_{ij}, which may also be viewed as a classic measurement error model (Carroll et al., 2006; Yi, 2017),

$$\begin{aligned} & w_{ij} = g(t_{ij}, \boldsymbol{\alpha}, \mathbf{a}_i) + e_{ij} \equiv w_{ij}^* + e_{ij}, \\ & \mathbf{a}_i \sim N(0, A), \quad \mathbf{e}_i \sim N(0, \sigma^2 I_{n_i}), \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i, \end{aligned} \tag{2.2}$$

where $g(\cdot)$ is a linear or nonlinear function, $\boldsymbol{\alpha}$ is a vector of fixed parameters, $\mathbf{a}_i$ is a vector of random effects, $w_{ij}^* = g(t_{ij}, \boldsymbol{\alpha}, \mathbf{a}_i)$ is the unobserved true covariate value, and $\mathbf{e}_i = (e_{i1}, \ldots, e_{in_i})^T$

are the covariate measurement errors. We assume that $\mathbf{a}_i$ and $\mathbf{e}_i$ are independent. Note that when the function $g(\cdot)$ is linear in the random effects $\mathbf{a}_i$, the above NLME model reduces to an LME model. In some applications, an NLME model may be derived based on the underlying data generation mechanisms, such as HIV viral dynamic models (Wu and Ding, 1999; Wang et al., 2020). Such an NLME covariate measurement error model is more desirable than the usual empirical LME covariate measurement error model, since the NLME model is a mechanistic model which may provide better predictions of the unobserved true covariate values than an empirical LME model. In such situations, the advantages of an NLME model are more obvious when the repeated measurements are infrequent, since the NLME model should hold even when data are not available.

The above two models (2.1) and (2.2) are linked through the shared (unobserved) variable w_{ij}^*, so they may be viewed as a joint model. We consider a (joint) likelihood method for estimating all model parameters simultaneously. Let $f(\cdot)$ denote a generic density function, and let $\boldsymbol{\theta}$ be the collection of all parameters in models (2.1) and (2.2). Then, the joint likelihood for all observed data is given by

$$L_o(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int \int f(\mathbf{y}_i|\mathbf{w}_i^*, \mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{w}_i|\mathbf{a}_i, \boldsymbol{\theta}) f(\mathbf{a}_i|\boldsymbol{\theta}) f(\mathbf{b}_i|\boldsymbol{\theta})\; d\mathbf{a}_i\; d\mathbf{b}_i,$$

which is often intractable since it involves a high-dimensional integral and generally does not have a closed-form expression. Thus, a major challenge in likelihood inference is the evaluation of the intractable likelihood $L_o(\boldsymbol{\theta})$. This computational issue is shared by other joint models. So we will discuss the common computational issues in a later section.

2.2 Joint models for several longitudinal processes

In a longitudinal study, suppose that we are interested in two or more possibly associated longitudinal processes and the nature of the association. For example, the longitudinal processes may arise from (i) repeated measurements on two or more correlated variables such as CD4 and viral load, or (ii) repeated measurements on a single variable at different phases of the study, such as viral load measurements during an antiviral treatment and after treatment interruption. For simplicity, here we focus on two longitudinal variables. Suppose that one variable may be modelled by a GLMM (2.1), and another variable may be modelled by an NLME model (2.2). To incorporate the association between these two processes, we may consider the following approaches: (a) the two models share the same random effects since the longitudinal measurements on the two variables are made on the same individuals in the study; (b) the random effects from the two models are correlated, which is less restrictive than (a); and (c) at each time point t_{ij}, the two models share a latent variable γ_j, which induces the association between the two processes. As an illustration, in the following we consider case (ii) and two continuous variables. Extensions to other cases are conceptually straightforward. In case (ii), we may be interested in investigating, say, whether the individual viral decay rates during treatment are associated with the individual viral rebound rates after treatment interruption.

Let y_{ij} be a continuous longitudinal response value for individual i at time t_{ij} in an *early* time period, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$. Let w_{il} be a continuous longitudinal response value for the same individual i at a *later* time u_{il}, $i = 1, 2, \ldots, n$, $l = 1, 2, \ldots, m_i$, and $u_{il} > t_{i,n_i}$ for all l's. The first NLME model for the earlier time period is given by

$$\begin{aligned} & y_{ij} = g(t_{ij}, \boldsymbol{\beta}, \mathbf{b}_i) + e_{ij}, \\ & \mathbf{b}_i \sim N(0, B), \quad \mathbf{e}_i \sim N(0, \sigma_1^2 I_{n_i}), \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i, \end{aligned} \tag{2.3}$$

where $g(\cdot)$ is a linear or nonlinear function, $\boldsymbol{\beta}$ is a vector of fixed parameters, $\mathbf{b}_i$ is a vector of random effects, and $\mathbf{e}_i = (e_{i1}, \ldots, e_{in_i})^T$ are the random errors of within-individual measurements. The second NLME model for the later time period is given by

$$\begin{aligned} & w_{il} = h(u_{il}, \boldsymbol{\alpha}, \mathbf{a}_i) + \epsilon_{il}, \\ & \mathbf{a}_i \sim N(0, A), \quad \boldsymbol{\epsilon}_i \sim N(0, \sigma_2^2 I_{m_i}), \qquad i = 1, 2, \ldots, n, \quad l = 1, 2, \ldots, m_i, \end{aligned} \tag{2.4}$$

where $h(\cdot)$ is a linear or nonlinear function and may be different from $g(\cdot)$, $\boldsymbol{\alpha}$ is a vector of fixed parameters, $\mathbf{a}_i$ is a vector of random effects, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$ are the random errors of within-individual measurements. Since the two NLME models are assumed for the same individuals, the individual-specific characteristics of the two longitudinal processes may be associated. For example, the viral decay rates during a treatment may be associated with the viral rebound rates after treatment interruption. Thus, we may assume that the two NLME models (2.3) and (2.4) are linked through correlated random effects, i.e., we assume that

$$\begin{pmatrix} \mathbf{a}_i \\ \mathbf{b}_i \end{pmatrix} \sim N \left(0, \begin{pmatrix} A & C^T \\ C & B \end{pmatrix} \right), \tag{2.5}$$

where the covariances of interest are contained in the matrix C, i.e., $C \neq 0$ implies that the random effects $\mathbf{a}_i$ and $\mathbf{b}_i$ may be correlated. Alternatively, we may use the random effects $\mathbf{b}_i$ in the first NLME model as "covariates" in the second NLME model. In both cases, the two NLME models may be viewed as a joint model.

For the above joint model, we can estimate all parameters simultaneously based on the joint likelihood for all observed data. Let $\boldsymbol{\theta}$ be the collection of all model parameters. The observed-data joint likelihood for the joint model can be written as

$$L_o(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int f(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{w}_i | \mathbf{a}_i, \boldsymbol{\theta}) f(\mathbf{a}_i, \mathbf{b}_i | \boldsymbol{\theta}) d(\mathbf{a}_i, \mathbf{b}_i), \tag{2.6}$$

which is again intractable, so it offers similar computational issues as noted earlier. Note that the above joint model can be extended to other cases, such as two joint GLMMs or a joint GLMM and NLME model. Again, a GLMM or NLME model includes LME models as special cases.

2.3 Joint models with shared latent process

In a longitudinal study, it may be reasonable to assume an (unobserved) latent process which governs several observed longitudinal processes, and thus induces the association among these processes.

For example, in a vaccine study, each individual may receive multiple vaccinations over time, say, every six months. After each vaccination, longitudinal data on several immune response biomarkers are observed, which may be governed by individuals' (unobserved) true immune statuses. An individual's true immune status, called "hidden state" here, may change from "low" to "high" and then from "high" to "low" before next vaccination. This process may be assumed to follow a first-order Markov process since the current state only depends on the immediate previous state. Thus we may consider a hidden Markov mixed effects model for the "hidden states". That is, the assumed Markov process is a latent process which governs the observed immune status data. Hidden Markov mixed effects models are reviewed in Altman (2007) and Bartolucci and Farcomeni (2019), among others.

Specifically, let z_{ij} be the hidden state for subject i at time t_{ij}, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, n_i$, and let $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{in_i})^T$. We assume that there are $K + 1$ hidden states, i.e., $z_{ij} = 0, 1, \ldots, K$. For simplicity, here we focus on $K = 1$, i.e., two immune statuses "high" or "low". We also focus on two observed longitudinal processes. Let $y^{(p)}$ be the immune response biomarker p, $p = 1, 2$. We may consider an NLME model for $y^{(p)}$. Given the hidden state $z_{ij} = k$ at time t_{ij}, we assume

$$y_{ij}^{(p)} = g_p(t_{ij}, \boldsymbol{\beta}_k^{(p)}, \mathbf{b}_i^{(p)}) + e_{ij}^{(p)}, \qquad i = 1, 2, \ldots, n; j = 1, 2, \ldots, n_i,\ p = 1, 2, \tag{2.7}$$

where the parameters $\boldsymbol{\beta}_k^{(p)}$ depend on the hidden state k, and other notation is similar as in previous sections. For individual i, we assume that the hidden states $z_{i1}, \ldots, z_{in_i}$ follow a first-order Markov chain:

$$P(\mathbf{z}_i) = P(z_{i0}) \prod_{j=1}^{n_i} P(z_{ij} | z_{i,j-1}),$$

where z_{i0} denotes the initial state. Let

$$p_{ij}(kl) = P(z_{ij} = l | z_{i,j-1} = k)$$

be the transition probability from state k at time $t_{i,j-1}$ to state l at time t_{ij}, and let $p_{i0} = P(z_{i0})$ be the initial distribution. Note that the transition probability $p_{ij}(kl)$ may vary across individuals and may also depend on covariates such as age and immunocompromised status. Thus, we assume the following generalized linear mixed model (GLMM):

$$\text{logit}(p_{ij}(kl)) = \alpha_{kl0} + \mathbf{w}_{1i}^T \boldsymbol{\alpha}_{kl} + \mathbf{w}_{2i}^T \mathbf{b}_i, \quad i = 1, 2, \ldots, n; j = 1, 2, \ldots, n_i,\ k, l = 0, 1, \ldots, K, \tag{2.8}$$

where $\mathbf{w}_{1i}, \mathbf{w}_{2i}$ are vectors of covariates, and $\alpha_{kl0}, \boldsymbol{\alpha}_{kl}$ are vectors of fixed effect parameters. In model (2.8), the transition probability $p_{ij}(kl)$ is assumed to depend on the random effects $\mathbf{b}_i = (\mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})$ from the two NLME models. This is because the random effects $\mathbf{b}_i$ represent individual-specific characteristics of the longitudinal biomarker process which may influence the transition probabilities.

The (joint) likelihood for all observed data is given by

$$L_o(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int \Big\{ \sum_{\mathbf{z}_i} f(\mathbf{y}_i^{(1)} | \mathbf{z}_i, \mathbf{b}_i^{(1)}, \boldsymbol{\theta}) f(\mathbf{y}_i^{(2)} | \mathbf{z}_i, \mathbf{b}_i^{(2)}, \boldsymbol{\theta}) f(\mathbf{z}_i | \boldsymbol{\theta}) f(\mathbf{b}_i | \boldsymbol{\theta}) \Big\}\, d\mathbf{b}_i,$$

where $\boldsymbol{\theta}$ denotes collection of all parameters. Again, it involves an intractable integration.

2.4 Joint models for mean and variance of longitudinal data

For many longitudinal data in practice, a major feature is that the *within individual* repeated measurements exhibit significant variations and these variations appear to change over time. It is important to understand the nature of the within-individual systematic and random variations, since they allow us to conduct more efficient inferences, such as obtaining narrower confidence intervals for key parameters of interest and detecting significant treatment effects which may otherwise be masked by the large variations or noises. To understand the nature of the within-individual systematic and random variations, we may model the within-individual variations, together with the model for the mean. In other words, we may jointly model the mean and the variance of the longitudinal data. Such a joint model also allows us to conduct robust inferences against outliers, as described below. Related literature in this direction may be found at Lin et al. (1997), Pourahmadi (1999), Hedeker et al. (2008), German et al. (2022), and Ye and Wu (2024).

We again consider an NLME model

$$\begin{aligned} & y_{ij} = g(t_{ij}, \boldsymbol{\beta}, \mathbf{b}_i) + e_{ij}, \qquad \text{or } \ E(y_{ij}|\mathbf{b}_i) = g(t_{ij}, \boldsymbol{\beta}, \mathbf{b}_i), \\ & \mathbf{b}_i \sim N(0, B), \qquad e_{ij} \sim N(0, \sigma_{ij}^2), \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i, \end{aligned} \tag{2.9}$$

where the notation is similar to those in previous sections. In NLME model (2.9), we assume a more flexible error distribution $e_{ij} \sim N(0, \sigma_{ij}^2)$, where the within-individual longitudinal data variance σ_{ij}^2 is allowed to be individual-specific and time-dependent. To better understand the nature of the within-individual variation σ_{ij}^2, we can model the variance process σ_{ij}^2 and investigate whether within-individual repeated measurements variation may be partially explained by time-varying covariates (say) $\mathbf{x}_{ij}$. We can also introduce an additional random effect to the variance model for σ_{ij}^2 to incorporate possible correlations of the variances over time and the between-individual variabilities. That is, we consider the following model for the variance of the within-individual repeated measurements σ_{ij}^2:

$$\log(\sigma_{ij}^2) = \alpha_1 + \boldsymbol{\alpha}_2^T \mathbf{x}_{ij} + \eta a_i, \qquad i = 1, \ldots, n, \quad j = 1, 2, \ldots, n_i, \tag{2.10}$$

where vector $\boldsymbol{\alpha} = (\alpha_1, \boldsymbol{\alpha}_2)$ and η are fixed parameters, and a_i is a random effect. We may assume that $\exp(a_i)$ follows an inverse gamma distribution or $a_i \sim N(0, 1)$. In the variance model (2.10), if we assume that

$$\exp(a_i) \sim k/\chi_k^2,$$

i.e., the inverse χ_k^2 distribution, which is a special case of an inverse gamma distribution, then the random errors e_{ij} in the NLME model (2.9) follow a $t(k)$-distribution with degrees of freedom k. Since a $t(k)$-distribution has heavier tails than the standard normal distribution, the joint model (2.9) and (2.10) may be used for *robust* inference against *outliers* in the within-individual repeated measurements.

Likelihood inference for all parameters, denoted by $\boldsymbol{\theta}$, can be based on the following (joint) likelihood

$$L_o(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int \int f(\mathbf{y}_i|\ \mathbf{b}_i, \boldsymbol{\theta}) f(\boldsymbol{\sigma}_i^2|\boldsymbol{\alpha}, a_i) f(\mathbf{b}_i, a_i|\boldsymbol{\theta})\ d(a_i, \mathbf{b}_i),$$

which again involves a high-dimensional intractable integration. With a second model for variance, the likelihood computation may involve other challenges such as convergence issues. Moreover, for unbiased estimation of the variance, we may consider restricted MLEs (REML) of the parameters. The computationally efficient h-likelihood method (Lee et al., 2018) allows one to obtain REML easily.

2.5 Joint models for longitudinal and survival data

In the analysis of longitudinal data, we are often also interested in certain events, such as dropouts or infections of disease or deaths. We may then consider a survival model for the time to event of interest. In this case, joint modelling for the longitudinal data and survival data is necessary. There is a large literature on joint models for longitudinal and survival data, as reviewed in Rizopoulos (2012) and Elashoff et al. (2016), among others. In the following, we briefly illustrate such joint models in the context of survival analysis with measurement errors in a time-dependent covariate. Similar joint models can also be used in the analysis of longitudinal data with informative dropouts and other applications.

For individual i, let s_i be the survival time or event time, subject to right censoring, $i = 1, 2, \ldots, n$. We assume that the censoring is non-informative. Let c_i be the censoring time. Due to censoring, we only observe $t_i = \min\{s_i, c_i\}$. Let $\delta_i = I(s_i \leq c_i)$ be the censoring indicator such that $\delta_i = 0$ if the survival time for individual i is right censored and $\delta_i = 1$ otherwise. Let $w_i(t)$ be an error prone time-dependent covariate whose unobserved true value is $w_i^*(t)$. We consider the following survival model

$$\lambda_i(t) = \lambda_0(t) \exp(w_i^*(t)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2), \qquad i = 1, \ldots, n, \tag{2.11}$$

where $\lambda_i(t)$ is the hazard function for individual i, $\lambda_0(t)$ is the baseline hazard, $\mathbf{x}_i$ contains other covariates, and $\boldsymbol{\beta} = (\beta_1, \boldsymbol{\beta}_2)$ are unknown regression parameters. We may consider NLME model (2.2) as a measurement error model for $w_i(t)$, which includes LME models as a special case. The joint likelihood based on observed longitudinal data and survival data is given by

$$\begin{aligned} L_o(\boldsymbol{\theta}) = \prod_{i=1}^{n} \int \Bigg\{ & \left[\lambda_0(t_i) \exp\{w_i^*(t_i)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2\}\right]^{\delta_i} \\ & \times \exp\left[-\int_0^{t_i} \lambda_0(u) \exp\{w_i^*(u)\beta_1 + \mathbf{x}_i^T \boldsymbol{\beta}_2\} du\right] \\ & \times f(\mathbf{w}_i|\mathbf{a}_i, \boldsymbol{\alpha}) f(\mathbf{a}_i|A) \Bigg\} \, d\,\mathbf{a}_i. \end{aligned}$$

The computation here can be more challenging than the joint models in previous sections since the baseline hazard is often unspecified.

Since there is a large literature on joint models for longitudinal and survival data, here we omit the details. Interested readers may find reviews in Rizopoulos (2012), Wu et al. (2012), and Elashoff et al. (2016).

2.6 Computation issues

The observed-data likelihoods for the joint models in previous sections all involve high-dimensional and intractable integration, especially when the models are nonlinear in the (unobservable) random effects. In the joint model literature, much attention has been on the computational side for likelihood inference. We briefly review some of these methods as follows:

- *Numerical integration methods.* Numerical integration methods such as the Gaussian Hermite quadrature method may work well when the dimension of the integration is not high (say, no more than two dimensions). When the dimension of the integration is high, numerical integration methods can be tedious and computationally very intensive.
- *EM algorithms*. The EM algorithm is often used for likelihood inference of GLMM and NLME models, treating the random effects as "missing data". Since GLMM and NLME models are nonlinear in the random effects, the E-step of the EM algorithm again involves an intractable integration. Monte Carlo methods or other approximate methods have been used in the E-step of the EM algorithm, leading to different types of EM algorithms such as Monte Carlo EM (MCEM) algorithm (Wei and Tanner, 1990; Wu, 2009) and Stochastic approximation EM (SAEM) algorithm (Delyon et al., 1999; Lavielle, 2014; Comets et al., 2017). The SAEM is computationally more efficient than MCEM and is implemented in software "Monolix" – see more details below.
- *Approximate methods*. Various computationally efficient approximate methods have also been proposed in the literature, such as the linearization method (Lindstrom and Bates, 1990) and h-likelihood method (Lee et al., 2018). These methods avoid evaluating the intractable integration based on either Laplace approximations or Taylor series approximations. These approximate methods are computationally much more efficient than numerical integration methods and EM algorithms, but they may be less accurate, since numerical integration methods and EM algorithms may be made more accurate by increasing (say) quadrature points or numbers of Monte Carl samples (in the expense of computational time) while this may not be possible for the approximate methods.

As an example, we briefly describe the SAEM for the GLMM (2.1) with covariate measurement error model (2.2) as described earlier. By treating the random effects as "missing data", we have "complete data" $\{(\mathbf{y}_i, \mathbf{w}_i, \mathbf{a}_i, \mathbf{b}_i), i = 1, 2, \ldots, n\}$. The log-likelihood of this "complete data" can be written as

$$\begin{aligned} l_c(\boldsymbol{\theta}) &\equiv \sum_{i=1}^{n} l_c(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{w}_i, \mathbf{b}_i, \mathbf{a}_i) \\ &= \sum_{i=1}^{n} \big\{ \log f(\mathbf{y}_i|\mathbf{w}_i, \mathbf{b}_i, \boldsymbol{\theta}) + \log f(\mathbf{w}_i|\mathbf{a}_i, \boldsymbol{\theta}) + \log f(\mathbf{b}_i|\boldsymbol{\theta}) + \log f(\mathbf{a}_i|\boldsymbol{\theta}) \big\}. \end{aligned}$$

The E-step is to compute the conditional expectation of the complete data log-likelihood given the observed data and the current parameter estimates. Beginning with some starting values, assuming that the current parameter estimate is $\boldsymbol{\theta}^{(k)}$ at the k-th EM iteration, where $k = 1, 2, \ldots$, the conditional expectation at the $(k+1)$-th EM iteration can be written as:

$$\begin{aligned}Q_k(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= \sum_{i=1}^{n} E\Big[l_c(\boldsymbol{\theta}; \mathbf{y}_i, \mathbf{w}_i, \mathbf{b}_i, \mathbf{a}_i) \mid \mathbf{y}_i, \mathbf{w}_i, \boldsymbol{\theta}^{(k)}\Big] \\ &= \sum_{i=1}^{n} \int\int \Big\{ \log f(\mathbf{y}_i|\mathbf{w}_i, \mathbf{b}_i, \boldsymbol{\theta}^{(k)}) + \log f(\mathbf{w}_i|\mathbf{a}_i, \boldsymbol{\theta}^{(k)}) + \log f(\mathbf{b}_i|\boldsymbol{\theta}^{(k)}) + \\ &\qquad \log f(\mathbf{a}_i|\boldsymbol{\theta}^{(k)})\Big\} \times f(\mathbf{b}_i, \mathbf{a}_i|\mathbf{w}_i, \mathbf{y}_i, \boldsymbol{\theta}^{(k)})\, d\mathbf{b}_i\, d\mathbf{a}_i,\end{aligned}$$

which usually does not have a closed-form expression.

There are different approaches to evaluate the above E-step in the literature. The MCEM is to use Monte Carlo simulations to approximate the E-step $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$, which requires large Monte Carlo samples from the conditional distribution $f(\mathbf{b}_i, \mathbf{a}_i|\mathbf{w}_i, \mathbf{y}_i, \boldsymbol{\theta}^{(k)})$ for the approximation to work well. Thus, MCEM can be computationally very intensive. The SAEM, on the other hand, evaluates $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ by a stochastic approximation procedure, which only samples a small number m_k (often $m_k = 1$) from the conditional distribution $f(\mathbf{b}_i, \mathbf{a}_i|\mathbf{w}_i, \mathbf{y}_i, \boldsymbol{\theta}^{(k)})$ in the E-step. Then $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ is updated based on the following approximation:

$$\begin{aligned}Q_k(\boldsymbol{\theta}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) \approx Q_{k-1}(\boldsymbol{\theta}) + \gamma_k \Big[\frac{1}{m_k}\sum_{j=1}^{m_k} \big\{ \log f(\mathbf{y}_i|\mathbf{w}_i, \mathbf{b}_i^{(j)}, \boldsymbol{\theta}^{(k)}) + \log f(\mathbf{w}_i|\mathbf{a}_i^{(j)}, \boldsymbol{\theta}^{(k)}) \\ + \log f(\mathbf{b}_i^{(j)}|\boldsymbol{\theta}^{(k)}) + \log f(\mathbf{a}_i^{(j)}|\boldsymbol{\theta}^{(k)})\big\} - Q_{k-1}(\boldsymbol{\theta})\Big],\end{aligned}$$

where $\{\gamma_k\}_{k\geq 1}$ is a sequence of positive step size and $(\mathbf{a}_i^{(j)}, \mathbf{b}_i^{(j)})$ is the j-th sample from the conditional distribution $f(\mathbf{b}_i, \mathbf{a}_i|\mathbf{w}_i, \mathbf{y}_i, \boldsymbol{\theta}^{(k)})$. The M-step is to maximize $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$ to obtain updated parameter estimates $\boldsymbol{\theta}^{(k+1)}$, which can be based on a standard optimization method. Delyon et al. (1999) shows that, under standard regularity conditions, the SAEM algorithm converges to a (local) maximum of the likelihood. The SAEM is implemented in software **Monolix** (`https://monolix.lixoft.com/`), with the corresponding R package **saemix**. See a sample implementation at `https://github.com/Sihaoyu1220/JointModel`.

Based on our experience, when the dimensions of the random effects in the joint models are not low (say, more than two), approximate methods and SAEM may be more desirable.

3 Examples

In the following, we present several examples to illustrate joint models based on real datasets from HIV/AIDS studies.

3.1 Example I: A joint model for two longitudinal processes

In this example, we consider a longitudinal dataset collected at two different phases of the study: we model viral decay during treatment and viral rebound after treatment interruption and study their association. This allows us to possibly predict viral rebound characteristics after treatment based on viral decay during treatment, so that early intervention may be considered. This dataset consists of 76 patients followed over time, with the longest follow-up period being 3233 days, the shortest being 503 days, and the mean being 1582 days. The number of repeated measurements through the entire study period on each patient varies, with a minimum of 10 measurements, a maximum of 48 measurements, and a mean of 25 measurements.

Let y_{ij} be the $\log_{10}$-transformed viral load for patient i at time t_{ij} during treatment, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$. We consider the following NLME viral decay model during treatment (Wu and Ding, 1999):

$$\begin{aligned} y_{ij} &= \log_{10}\left(e^{P_{1i}-\lambda_{1i}t_{ij}} + e^{P_{2i}-\lambda_{2i}t_{ij}}\right) + e_{ij}, \\ P_{1i} &= P_1 + b_{1i}, \quad P_{2i} = P_2 + b_{2i}, \quad \lambda_{1i} = \lambda_1 + b_{3i}, \quad \lambda_{2i} = \lambda_2 + b_{4i}, \\ \mathbf{b}_i &\sim N(0, B), \quad e_{ij} \sim N(0, \sigma_1^2), \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i, \end{aligned} \tag{3.1}$$

where $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T$ are random effects, λ_1 is the first-phase viral decay rate, λ_2 is the second-phase viral decay rate, and the quantity $\log_{10}(e^{P_1} + e^{P_2})$ is typical viral load value at the start of treatment.

For viral rebound after treatment interruption, let w_{il} be the $\log_{10}$-transformed viral load for the same patient i at a later time u_{il} since treatment interruption, $i = 1, 2, \ldots, n$, $l = 1, 2, \ldots, m_i$. We consider the NLME viral rebound model by Wang et al. (2020) with some modifications. Since viral rebound after ART interruption is very complicated, it may be more flexible to allow the individual-specific rate of viral decline β_{5i} to be time-dependent, say β_{5il} for individual i at time u_{il}, and then β_{5il} may be modelled nonparametrically. In addition, random effects are needed for this parameter due to the large between-individual variations. Therefore, we consider the following semiparametric NLME viral rebound model:

$$\begin{aligned} w_{il} &= \frac{\beta_{1i}u_{il}}{u_{il} + \exp(\beta_{2i} - \beta_{3i}u_{il})} + \frac{\beta_{4i}}{1 + \exp(\beta_{5il}u_{il})} + \epsilon_{il}, \\ \beta_{5il} &= \beta_5 z_{il}^* + r(u_{il}) + v_i(u_{il}), \quad v_i(t) \sim GP(0, \gamma), \\ \beta_{ki} &= \beta_k + \tau_{ki}, \quad i = 1, 2, \ldots, n, \quad l = 1, 2, \ldots, m_i, \quad k = 1, \ldots, 5, \end{aligned} \tag{3.2}$$

where z_{il}^* is the true (but unobservable and possibly mis-measured) CD4 count, $\tau_i = (\tau_{1i}, \ldots, \tau_{5i})$'s are random effects, $\epsilon_{il} \sim N(0, \sigma_2^2)$ contains within-individual random error, and $r(u_{il})$ and $v_i(u_{il})$ are unknown smooth fixed and random functions. The fixed effects β_1 represents setpoint after rebound, β_2 controls the timing of viral rise, β_3 characterizes the rate of viral rebound, and β_4 denotes initial viral load value at the start of rebound. Since CD4 reflects immune status and is known to be measured with substantial errors, it may be useful to use the true CD4 value z_{il}^* to partially explain the large variations in β_{5il}.

Since the viral decay trajectories *during* ART and the viral rebound trajectories *after* ART interruption may be associated, we assume that the NLME viral decay model (3.1) and the viral rebound model (3.2) are linked through correlated random effects, i.e., we assume that

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\tau}_i \end{pmatrix} \sim N \left(0, \begin{pmatrix} B & C^T \\ C & \Sigma \end{pmatrix} \right), \tag{3.3}$$

where $C \neq 0$ contains the covariance between the random effects.

Table 1 presents the resulting parameter estimates and standard errors, obtained using the joint likelihood and SAEM algorithm. Table 2 summarizes the estimated correlations between the random effects of the two models. We find that b_{4i} is negatively associated with τ_{3i} with a correlation of -0.645, suggesting that a faster second phase viral decay during treatment may be associated with a slower viral rebound rate following treatment interruption. In addition, the initial viral decay rates during treatment appears to be negatively associated with the viral setpoints following treatment interruption with a correlation of -0.637, suggesting that the faster the viral decay after the start of treatment, the lower the setpoints following treatment interruption. These results agree with the findings in Gao et al. (2022), where the random effects in the viral decay model are treated as "covariates" in the viral rebound model.

Table 1: Parameter estimates for the joint model in Example II.

Parameter	Estimate	SE	z-value	p-value
P_1	17.479	0.428	40.844	0.000
λ_1	4.082	0.288	14.194	0.000
P_2	2.967	0.419	7.077	0.000
λ_2	0.058	0.026	2.218	0.027
β_1	3.379	0.103	32.889	0.000
β_2	8.615	1.036	8.319	0.000
β_3	3.428	0.405	8.458	0.000
β_4	1.272	0.107	11.91	0.000
β_5	-0.104	0.002	-64.209	0.000

3.2 Example II: GLMM with covariate measurement error

The dataset consists of 46 patients who were followed over time, both during an anti-HIV treatment and after treatment interruption. Viral loads and CD4 counts were repeatedly measured on these patients, with the longest follow-up period being 196 days and the shortest being 51 days. As shown in Figure 1, viral load declines during treatment and then rebounds after treatment interruption. The number of viral load measurements on each patient varies, with a minimum of 2 measurements and a maximum of 9 measurements. On average, each patient has 6.74 repeated viral load measurements.

Table 2: Estimated correlations between random effects of the joint models (3.1) and (3.2).

Random Effects	b_{3i}	b_{4i}	τ_{1i}	τ_{3i}
b_{3i}	1.000	0.416	-0.637	-0.353
b_{4i}	0.416	1.000	-0.038	-0.645
τ_{1i}	**-0.637**	-0.038	1.000	0.257
τ_{3i}	-0.353	**-0.645**	0.257	1.000

Regarding CD4 data, out of the 361 observations, 254 observations (70.36%) exhibit observed CD4 counts above 200 cells/mm^3, which is a commonly used threshold. Both viral load and CD4 are typically measured with errors.

Let y_{ij}^* be the observed CD4 value for individual i at time t_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$. We define a new binary response $y_{ij} = 0$ if the observed CD4 is less than 200, and $y_{ij} = 1$ otherwise. The objective is to investigate the relationship between the dichotomized CD4 value y and viral load w, which is subject to measurement errors, both during and after treatment. For simplicity, here we ignore left censoring in viral load. We treat y as a response and w as a time-dependent covariate and consider the following GLMM (Gao and Wu, 2024)

$$\log\left(\frac{P(y_{ij}=1)}{P(y_{ij}=0)}\right) = (\beta_0 + b_{0i}) + \beta_w w_{ij}^*, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i, \tag{3.4}$$

where w_{ij}^* represents the (unobserved) true viral load value for individual i at time t_{ij}. The key parameter β_w measures the strength of the association between viral load and CD4. To estimate the true viral load w_{ij}^*, we consider an NLME model for viral decay during treatment (Wu and Ding, 1999) and an NLME model for viral rebound after treatment interruption (Wang et al., 2020). Both NLME models are derived based on the underlying data-generation mechanisms and thus allow us to better address measurement errors than empirical LME models. We then consider a unified NLME model as follow:

$$\begin{aligned} w_{ij} &= \left[\log_{10}\left(e^{\alpha_{1i}-\alpha_{2i}t_{ij}} + e^{\alpha_{3i}}\right)\right] I(t_{ij} < T_i) + \left[\frac{\alpha_{4i}t_{ij}^*}{t_{ij}^* + \exp(\alpha_{5i} - \alpha_{6i}t_{ij}^*)} + \alpha_{7i}\right] I(t_{ij} \geq T_i) + e_{ij} \\ &\equiv w_{ij}^* + e_{ij}, \qquad (3.5) \\ \alpha_{ki} &= \alpha_k + a_{ki}, \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n_i, \quad k = 1, \ldots, 7, \end{aligned}$$

where $I(\cdot)$ is an indicator function, the vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_7)^T$ consists of fixed effect parameters, and the random effects are represented by $\mathbf{a}_i = (a_{1i}, \ldots, a_{7i})^T \sim N(0, A)$, with A being a covariance matrix. The term e_{ij} denotes measurement error, and T_i represents the (known) treatment interruption time for individual i, while $t_{ij}^* = t_{ij} - T_i$ indicates the measurement time since treatment interruption. We assume that the random effects $\mathbf{a}_i$ and the measurement error e_{ij} are independent, and e_{ij} are i.i.d. $\sim N(0, \sigma^2)$, given the random effects.

We consider likelihood inference for the above joint model based on a SAEM algorithm, implemented in software "monolix". Table 3 presents the estimated parameters. The estimated value of the parameter of interest, $\beta_w = -1.773$, indicates a negative association between CD4 and viral load (p-value < 0.001). For example, a subject with viral load of 50 has an estimated probability of 0.996 for the corresponding CD4 count to exceed 200. We see that estimate of the key parameter β_w suggests that lower viral loads are significantly associated with higher probabilities of CD4 counts exceeding the critical threshold of 200 cells/mm^3.

Table 3: Parameter estimates for joint models (3.4) and (3.5).

Parameter	Estimate	Standard error	z-value	p-value
α_1	11.497	0.190	60.642	0.000
α_2	0.272	0.021	12.720	0.000
α_3	5.536	0.285	19.455	0.000
α_4	1.274	0.244	5.229	0.000
α_5	14.161	4.103	3.452	0.001
α_6	0.237	0.065	3.628	0.000
α_7	2.656	0.168	15.796	0.000
β_0	9.125	1.092	8.356	0.000
β_w	-1.892	0.225	-8.426	0.000

3.3 Example III: Joint models for mean and variance

This dataset includes 1791 individuals, with the median follow-up being 320 days and the range between 73 and 365 days. The median number of repeated measurements is 6 (ranging from 4 to 15). We consider a joint model for the mean and variance and compare the effectiveness of different regimens. To model the viral dynamics in the early stage of treatment, we consider the following one-compartment exponential viral decay NLME model

$$y_{ij} = \beta_1 + (\beta_2 + u_{2i})e^{-(\beta_3 + u_{3i})t_{ij}} + e_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, n_i, \tag{3.6}$$

where y_{ij} is the $\log_{10}$-transformed viral load measured at time t_{ij} for individual i, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ contains fixed parameters, $\mathbf{u}_i = (u_{2i}, u_{3i})^T$ denotes random effects, and e_{ij}'s are the within-individual random errors. We allowed the variation of the within-individual repeated measurements to be time-dependent and individual-specific, i.e., we assumed

$$e_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2).$$

We also assumed $\mathbf{u}_i \sim N(0, D)$. To avoid very large/small estimates, which may be unstable, the time variable t_{ij} was re-scaled to be between 0 and 1.

To understand if the large and time-varying variations σ_{ij}^2 of the within-individual viral load repeated measurements may be partially explained by time-varying CD4 counts, which may lead to more efficient inference and thus more accurate estimate of the viral decay rates, we modeled the variance as follows

$$\log(\sigma_{ij}^2) = \alpha_0 + \alpha_1 z_{ij}^* + a_i, \tag{3.7}$$

where z_{ij}^* denotes the "true" but unobservable (square-root-transformed) CD4 cell count at time t_{ij} for individual i, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^T$ are fixed parameters and a_i is a random effect. Note that the CD4 cell count values were square-root-transformed to achieve a roughly symmetric distribution. We assumed $\exp(a_i) \sim k/\chi_k^2$ to address potential outliers in the viral load data. The degrees of freedom k was pre-specified to be a small number, i.e., $k = 3$, to allow for relatively heavy tails of the t-distribution. To address measurement errors in CD4 cell count values, we modeled the CD4 cell count process based on the following empirical LME model

$$z_{ij} \equiv z_{ij}^* + \varepsilon_{ij} = (\gamma_0 + b_{0i}) + (\gamma_1 + b_{1i})t_{ij} + (\gamma_2 + b_{2i})t_{ij}^2 + \varepsilon_{ij}, \tag{3.8}$$

for $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, where the true CD4 cell count was assumed to be $z_{ij}^* = (\gamma_0 + b_{0i}) + (\gamma_1 + b_{1i})t_{ij} + (\gamma_2 + b_{2i})t_{ij}^2$, $\mathbf{b}_i = (b_{0i}, b_{1i}, b_{2i})^T$ denote the random effects in the CD4 cell count model, and ε_{ij}'s are the measurement errors. We assumed that $\mathbf{b}_i \sim \mathcal{N}(0, B)$ and $\varepsilon_{ij} \sim_{iid} \mathcal{N}(0, \omega^2)$ are independent.

We consider both the joint model (JM) and the two-step (TS) method using the h-likelihood to estimate parameters. For comparison purposes, we also consider a commonly used linearization method for NLME models based on Lindstrom and Bates (1990), denoted by LB, without modeling the variance σ_{ij}^2. The objective is to evaluate treatment effectiveness. Thus we focus on estimating the initial decay rate β_3. The results based on one of the treatment group are shown in Table 4. Overall, the estimates based on the JM and TS methods were close to each other. However, the LB method produced quite different results, especially for the parameter of primary interest β_3. More importantly, the standard errors (SE) of $\hat{\beta}_3$ differ substantially across the three methods. The LB method, which did not model the within-individual variances, produced the largest SE of $\hat{\beta}_3$. The JM and TS methods, on the other hand, produced smaller SEs (i.e., more efficient) and thus smaller p-values. These results demonstrated the advantages of modeling the within-individual variations. We see that, by modeling the variance of the within-individual errors, the JM and TS methods lead to more efficient estimates (i.e., smaller SE's) of the parameters than the LB method which does not model the variance. Moreover, estimates of the parameters in the variance model allowed us to better understand the systematic component in the within-individual variation, since we see that CD4 cell count values partially explained this variation and this variation decreased with CD4 cell count values over time.

3.4 Example V: A Comparison of Joint Model with Separate Models

We show a simple example of efficiency gain from a joint model for longitudinal and survival data, compared with separate models, based on Zhang and Wu (2019). The dataset is from an HIV/AIDS study, and the details of the dataset and more complete data analysis can be found in Zhang and Wu

Table 4: *Parameter estimates for regimen NNRTIs*

Method		β_1	β_2	β_3	α_0	α_1	γ_0	γ_1	γ_2
LB	Est	1.46	1.77	**4.71**					
	SE	0.013	0.063	**0.263**					
TS	Est	1.54	2.21	**9.99**	3.82	-1.16	5.89	0.69	-0.37
	SE	0.003	0.050	**0.147**	0.254	0.041	0.035	0.092	0.087
JM	Est	1.54	2.22	**10.03**	4.44	-1.27	5.87	0.78	-0.43
	SE	0.003	0.050	**0.149**	0.250	0.041	0.032	0.056	0.053

Table 5: Efficiency gain from the joint model over separate analysis.

Parameter	Separate analysis		Joint model	
	Estimate	Standard error	Estimate	Standard error
α_1	5.45	0.23	5.34	0.21
α_2	8.64	0.28	8.62	0.25
α_3	49.72	4.68	47.14	4.51

(2019). We consider the covariate measurement error problem for a time-dependent covariate (viral load) in a Cox survival model for time to first CD4:CD8 decline.

Specifically, we consider the following survival model for time to first CD4:CD8 decline

$$\lambda_i(t) = \lambda_0(t)\exp(\beta_1 z_i + \beta_2 x_i^*(t)),$$

where $x_i^*(t)$ is the true viral load value at time t, z_i is treatment indicator such that $z_i = 1$ for arm I and $z_i = 0$ for arm II. The observed viral load data for individual i at time t_{ij} is denoted by x_{ij} ($\log_{10}$ transformed), which is subject to measurement error, $i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n_i$. To address measurement errors in viral loads, we consider the following NLME model

$$x_{ij} = \log_{10}\{e^{\alpha_1 + a_{1i}} + e^{(\alpha_2 + a_{2i}) - (\alpha_3 + a_{3i}) * t_{ij}}\} + e_{ij} \equiv x_{ij}^* + e_{ij},$$
$$\mathbf{a}_i = (a_{1i}, a_{2i}, a_{3i}) \sim N(0, A), \quad e_{ij} \sim N(0, \sigma^2),$$

where $(\alpha_1, \alpha_2, \alpha_3)^T$ is a vector of fixed effects, a_{ki}'s are random effects, and e_{ij} represents measurement error. Part of the data analysis results is shown in Table 5. We see that estimates based on the joint model have smaller standard errors than those based on separate analysis, showing efficiency gain from the joint model over separate analysis.

4 Discussion

In the analysis of longitudinal data, joint models are desirable when two or more different types of variables are associated. Joint models are also desirable when the longitudinal models for several variables are nonlinear in the random effects. In both cases, multivariate models such as multivariate LME models may be difficult to specify. Ignoring the associations between the variables may lead to less efficient statistical inference and possible biased estimation, compared to jointly modeling these variables simultaneously. The associations between the variables may be incorporated in different ways, such as shared or correlated random effects from different models or treat one variable as a covariate in the model for the other variable. When jointly modelling different types of variables, mixed effects models are natural choices for the longitudinal data. The maximum likelihood method is a standard approach for statistical inference of joint models, but the computation can be challenging since the likelihoods often involve a high-dimensional and intractable integration. Two step methods are computationally simpler and may use existing software, but they may lead to less efficient and possibly biased inference. Various modified two-step methods have been proposed in the literature, such as using bootstrap methods to obtain standard errors of the estimates.

Outliers are common in longitudinal data, and they may lead to incorrect inference if not addressed (Sinha, 2004; Sinha and Sattar, 2014). It is known that MLEs are sensitive to outliers. In this case, one may consider M-estimators for robust inference or replace the assumed normal distributions by heavy-tail t-distributions. Longitudinal data may also be censored, such as data below detection limits, or may be missing. In these cases, models may be assumed for the censored data or missing data, leading to more complicated joint models (Yu et al., 2018).

Another issue is mis-specified distributions for the random effects and random errors (Sattar and Sinha, 2021). In this case, one may consider pseudo-likelihood methods or generalized estimating equation (GEE) methods. A main advantage of GEE methods is that they do not require distributional assumptions. We only need to correctly specify the mean structure, and use working correlations for repeated measurements. In the case of joint models, a challenge for GEE methods is to simultaneously specify the association between different types of variables and the association among the repeated measurements. This may lead to complicated GEE joint models.

Bayesian methods may also be considered for joint models. Bayesian methods may particularly be attractive when the sample size is small or prior information is available. Another attractive feature of Bayesian methods is that many existing software are available, such as WinBUGS and Integrated nested Laplace approximation (INLA).

Although there has extensive literature in joint models, much research remains to be done. For example, parametric mixed effects models may be extended to semiparametric or nonparametric mixed effects models for longitudinal data. Various survival models may also be considered, such as accelerated failure time models and competing risk models (Li and Yang, 2016).

References

Altman, R. M. (2007), "Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting," *Journal of the American Statistical Association*, 102, 201–210.

Bartolucci, F. and Farcomeni, A. (2019), "A shared-parameter continuous-time hidden Markov and survival model for longitudinal data with informative dropout," *Statistics in medicine*, 38, 1056–1073.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: a Modern Perspective*, Chapman and Hall/CRC, 2nd ed.

Comets, E., Lavenu, A., and Lavielle, M. (2017), "Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm," *Journal of Statistical Software*, 80, 1–41.

Davidian, M. and Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, 1st ed.

Delyon, B., Lavielle, M., and Moulines, E. (1999), "Convergence of a stochastic approximation version of the EM algorithm," *Annals of Statistics*, 27, 94–128.

Elashoff, R., li, G., and Li, N. (2016), *Joint Modeling of Longitudinal and Time-to-event Data*, London: Chapman and Hall/CRC press, 1st ed.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2012), *Applied longitudinal analysis*, John Wiley & Sons.

Gao, S. and Wu, L. (2024), "Generalized Linear Mixed Models with Censored Covariates and Measurement Errors, with Applications in HIV/AIDS Studies," *Technical report, UBC*.

Gao, S., Wu, L., Yu, T., Kouyos, R., Günthard, H. F., and Wang, R. (2022), "Nonlinear mixed-effects models for HIV viral load trajectories before and after antiretroviral therapy interruption, incorporating left censoring," *Statistical Communications in Infectious Diseases*, 14.

German, C. A., Sinsheimer, J. S., Zhou, J., and Zhou, H. (2022), "WiSER: Robust and scalable estimation and inference of within-subject variances from intensive longitudinal data," *Biometrics*, 78, 1313–1327.

Hedeker, D., Mermelstein, R. J., and Demirtas, H. (2008), "An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data," *Biometrics*, 64, 627–634.

Lavielle, M. (2014), *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*, Chapman and Hall/CRC press.

Lee, Y., Nelder, J. A., and Pawitan, Y. (2018), *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*, Chapman and Hall/CRC Press.

Li, G. and Yang, Q. (2016), "Joint inference for competing risks survival data," *Journal of the American Statistical Association*, 111, 1289–1300.

Lin, X., Raz, J., and Harlow, S. D. (1997), "Linear mixed models with heterogeneous within-cluster variances," *Biometrics*, 910–923.

Lindstrom, M. J. and Bates, D. M. (1990), "Nonlinear mixed effects models for repeated measures data," *Biometrics*, 46, 673–687.

McCulloch, C. E. and Searle, S. R. (2004), *Generalized, Linear, and Mixed models*, John Wiley & Sons.

Pinheiro, J. and Bates, D. (2006), *Mixed-effects Models in S and S-PLUS*, Springer science & business media.

Pourahmadi, M. (1999), "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation," *Biometrika*, 86, 677–690.

Rizopoulos, D. (2012), *Joint Models for Longitudinal and Time-to-event Data: With Applications in R*, CRC press.

Sattar, A. and Sinha, S. K. (2021), "Inference with Joint Models Under MIsspecified Random Effects Distributions," *J Stat Res*, 55, 187–205.

Sinha, S. K. (2004), "Robust analysis of generalized linear mixed models," *Journal of the American Statistical Association*, 99, 451–460.

Sinha, S. K. and Sattar, A. (2014), "Analysis of incomplete longitudinal data with informative dropout and outliers," *Canadian Journal of Statistics*, 42, 670–695.

Vonesh, E. F. (2014), *Generalized Linear and Nonlinear Models for Correlated Data: Theory and Applications Using SAS*, SAS Institute.

Wang, R., Bing, A., Wang, C., Hu, Y., Bosch, R. J., and DeGruttola, V. (2020), "A flexible nonlinear mixed effects model for HIV viral load rebound after treatment interruption," *Statistics in Medicine*, 39, 2051–2066.

Wei, G. C. and Tanner, M. A. (1990), "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," *Journal of the American statistical Association*, 85, 699–704.

Wu, H. and Ding, A. A. (1999), "Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials," *Biometrics*, 55, 410–418.

Wu, L. (2009), *Mixed Effects Models for Complex Data*, Chapman and Hall/CRC Press.

Wu, L., Liu, W., Yi, G. Y., Huang, Y., et al. (2012), "Analysis of longitudinal and survival data: joint modeling, inference methods, and issues," *Journal of Probability and Statistics*, 2012.

Ye, Q. and Wu, L. (2024), “Jointly Modeling Means and Variances for Nonlinear Mixed Effects Models with Measurement Errors and Outliers,” *Technical report*.

Yi, G. Y. (2017), *Statistical analysis with measurement error or misclassification: strategy, method and application*, Springer.

Yu, T., Wu, L., and Gilbert, P. B. (2018), “A joint model for mixed and truncated longitudinal data and survival data, with application to HIV vaccine studies,” *Biostatistics*, 19, 374–390.

Zhang, H. and Wu, L. (2019), “Joint model of accelerated failure time and mechanistic nonlinear model for censored covariates, with application in HIV/AIDS,” *The Annals of Applied Statistics*, 13, 2140–2157.

Received: February 3, 2024

Accepted: May 5, 2024