Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/100318

Joint Model for Longitudinal and Time-To-Event Data: a Two-Stage Approach

*Srimanti Dutta and Arindom Chakraborty**

Department of Statistics, Visva-Bharati University, Santiniketan, India

Received: 22 Jan. 2020, Revised: 12 May 2020, Accepted: 20 Jun. 2020
Published online: 1 Nov. 2021

Abstract: In clinical and epidemiological studies, very often, observations are collected on more than one correlated processes. For example, in AIDS related studies, along with a longitudinal biomarker like CD4 cell count, data on time-to-death is also recorded. Modelling them separately may give bias estimates. This necessitates the concept of joint modelling where two or more processes are modelled together. To link these processes, the usual technique is to use the same or highly correlated subject-specific random-effects for all the sub-models. In this work, structural correlation based on the conditional distribution of time-to-event given longitudinal response is used. A computationally efficient two-stage method is used to find the estimates. At the first stage, longitudinal submodel is fitted using *nlme* package in **R**. In the second stage, to avoid the complexity of second order differentiation, we have used an adaptive gradient descent algorithm. The simulation study shows that this structural correlation is good enough to take care of the correlation between these two simultaneous processes. A rapid convergence is also achieved. The proposed method is finally applied to a data set related to AIDS studies.

Keywords: Non-linear mixed-effects, Two-stage approach, AFT model, Gradient Descent algorithm

1 Introduction

Medical and epidemiological studies often involve instances where we have to analyze repeated evaluations of outcomes of a particular characteristic of a patient along with event history data which may involve death, dropout or progress of a disease. Two well-known areas which include vast applications of these kinds of studies are AIDS research and cancer studies. A joint modelling approach is needed where we take into account two processes i.e., the longitudinal and the time-to-event simultaneously. These two processes in their very nature are interlinked with each other and separate analysis may give biased or inefficient estimates. A joint modelling approach thus comes into play. There are two types of joint models available in the literature. In the first type of models, our primary interest lies on the event outcome and we wish to examine the effect of longitudinal trajectories on it. In another type of models, we are interested in the effect of the survival process on the longitudinal outcome.

The modelling framework for the former situation is known as selection models (Little and Rubin [1], Little and Rubin [2]) and the latter is known as pattern-mixture models (Little [3]). Both of them signify the same joint distributions with different statistical interpretations. Fitzmaurice et al. [4], Molenberghs et al. [5], Molenberghs and Verbeke [6] in their works had studied in detail and compared the conclusions and interpretations based on these two types of models. Another model i.e., the latent variable model (or a shared random-effects model) has also been considered by some authors (Xu and Zeger [7], Henderson et al. [8], Verbeke et al. [9]) where the two processes are linked by means of shared random-effects. These models can be distinguished by the method of factorization chosen in their resulting joint likelihood.

Hence, the primary focus of joint modelling lies in acquiring reliable estimates of the parameters in the model based on where our interest is focused. For this purpose, both frequentist and Bayesian approach had been explored extensively in literature where sometimes these two approaches have also been combined for the reason of taking the privileges of strengths offered by both the situations. Unlike earlier researches on joint modelling involving univariate framework (consideration of single longitudinal and time-to-event outcome), works in recent times focus extensively on multiple longitudinal outcomes (Ibrahim et al. [10]) along with an event of interest or competing events of interest. Rizopoulos and

* Corresponding author e-mail: arindom.chakraborty@visva-bharati.ac.in

Ghosh [11] in their work had developed multivariate joint model considering multiple longitudinal outcomes along with a time-to-event based on data on 407 patients with chronic kidney disease who underwent a renal transplantation and had three separate biomarker measurements recorded repeatedly. Here, a bayesian approach was adopted using a spline-based semi-parametric setup with the latent terms being formulated with Dirichlet process allowing for general shapes of their distribution.

Further, Andrinopoulou et al. [12] developed an extended joint model that handled one continuous and one ordinal longitudinal outcome, as well as allowing for a competing risk setting. In recent times, Bayesian estimation has been facilitated by the development of packages like *JMbayes* (Rizopoulos [13]) which can fit a wide range of joint models for longitudinal outcomes in both continuous and ordinal form. Hickey et al. [14] considered a joint model with multivariate linear mixed sub-model for longitudinal process and Cox proportional hazards regression model with time-varying covariates for the survival process with a zero-mean multivariate latent Gaussian process introduced to bind the two sub-models. A practical algorithm for fitting the models had been proposed and the method was demonstrated using *joineRML* in **R**. *JM* package (Rizopoulos [15]) had also been developed for joint models to be implemented in a frequentist perspective.

In this context, there is another important issue to be considered regarding the model assumptions for the longitudinal growth trajectories which are usually fitted through linear mixed-effects models. These models though take into account the inherent correlation introduced by the repeated measurements along with the covariates suffers from a drawback in the cases where lack of linearity is observed in the data.

However, sometimes the likelihood function may depend on the parameters in a non-linear fashion where we are compelled to use non-linear approach where there is much greater scope of obtaining a more interpretable model involving a reduced number of parameters. Moreover, it is quite common in HIV studies that the CD4 cell trajectories of infected patients display a pattern which exhibits skewness or a departure from linearity. Oncological studies also present with some situations where longitudinal measurements of patients though found to be linear before the detection of cancer started to show non-linear pattern nearer to the time point when the cancer was detected. Thus, it is clear that non-linear mixed-effects models can be particularly useful in these instances as they are based on underlying causes or mechanisms which generates such data. Li et al. [16] had used a quasi-likelihood type approximation for non-linear variables and had transformed their model into a multivariate linear mixed model framework with an algorithm being developed via an extension to EM approach.

However, the classical approach suffers from some serious drawbacks as current maximum likelihood methods for joint modelling are only advantageous when the dimension of random-effects is not high. Further, for multivariate linear mixed-effects models, the model fitting and analysis becomes intractable in cases where the dimension of variables are greater than four due to the increase of parameters in the covariance matrix for random-effects. For the non-linear framework, high-dimensional integration is more troublesome to handle in the presence of multiple random-effects where Monte Carlo or Gauss-Hermite quadrature fails to serve the purpose (Fieuws et al. [17]).

Hence, two-stage estimation has been explored in literature which guarantees to solve those problems to some extent. Two-stage approaches have been investigated previously by Bycott and Taylor [18], Self and Pawitan [19], Tsiatis et al. [20], and Dafni and Tsiatis [21]. In joint models, this approach had been considered by Albert and Shih [22] where they addressed to the problem of accommodating a large number of longitudinal biomarkers and high dimensional random-effects by proposing a two-stage regression calibration approach (also known as two-stage multilevel method). Sweeting and Thompson [23] considered fitting the longitudinal process separately in the first stage and then utilized the maximum likelihood estimates and best linear unbiased predictors of random-effects in the second stage to deal with the survival process which was treated with standard survival analysis software. Ye and Wu [24] had compared the efficacy of joint likelihood and two-step methods for various joint models. Sayers et al. [25] compared three two-stage approaches where they summarized the information on the longitudinal observations for each of them in the first stage and considered a simple approach, an individual regression approach and multilevel model for longitudinal trajectories in the second stage. Donnelly et al. [26] constructed a two-stage approach where they used a Coxian phase-type distribution in the second stage by considering random-effects (incorporated as covariates) as a proxy measure in the linear mixed-effects model in the first stage. Huong et al. [27] proposed ordinary, full likelihood and modified two-stage approaches and had compared their performances.

In this work, we have adopted a two-stage approach to our proposed joint model with $(n+1)$ components. Here, the first n components describe the longitudinal process which is viewed in the non-linear framework. The $(n+1)^{th}$ component describes the time-to-event and our main objective is to examine how the time-to-event depends on the longitudinal process. The association between the two processes has been defined through Bartlett decomposition of the covariance matrix. We had considered an accelerated failure time model under a selection model framework. We had considered a linear mixed-effects model in the first stage for the longitudinal process and had used the information

obtained from that stage to estimate the parameters in the survival process. This has been achieved by adopting a gradient descent algorithm with an adaptive learning rate.

The article is organized as follows. In Section 2, we describe the two-stage approach for proposed joint model. We have also considered frequentist approach to estimate parameters. In Section 3, we examine the performance of our two-stage estimation approach on our proposed model with the help of a simulation study. Section 4 applies the two-stage approach on a popular data set on AIDS study and some conclusions are discussed in Section 5.

2 Joint Models and Estimation

2.1 Model and Methodology

Here, we denote y_{ij} as the j^{th} $(j = 1,2,..,n_i)$ longitudinal trajectory for the i^{th} $(i = 1,2,..,m)$ subject and T_i as the time-to-event outcome for i^{th} individual with δ_i as the censoring indicator. We model the joint distribution of the longitudinal and time-to-event data with given subject-specific random-effects $\boldsymbol{b}_i$ by the (n_i+1) variate normal distribution

$$(\boldsymbol{y}_i', \log T_i)' \sim N(\boldsymbol{g}(\boldsymbol{x}_i\boldsymbol{\beta}, \boldsymbol{z}_i\boldsymbol{b}_i), \boldsymbol{\Sigma}_i),$$

$$\boldsymbol{V}_i = \begin{pmatrix} \boldsymbol{y}_i \\ \log T_i \end{pmatrix} = \boldsymbol{g}(\boldsymbol{x}_i\boldsymbol{\beta}, \boldsymbol{z}_i\boldsymbol{b}_i) + \boldsymbol{\varepsilon}_i \tag{1}$$

where $\boldsymbol{g}(\cdot)$ depicts a vector of non-linear functions which is assumed to be monotonic, continuously differentiable and whose first order derivative is uniformly bounded. Here, $\boldsymbol{x}_i$ denotes the design matrix for the fixed-effects, $\boldsymbol{z}_i$ denotes the design matrix for the random-effects and $\boldsymbol{\beta}$ is the vector of regression parameters. Again,

$$\boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_i) \text{ with } \boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{y_i} & \boldsymbol{\sigma}_{1i} \\ \boldsymbol{\sigma}_{1i}' & \sigma_T^2 \end{pmatrix} \tag{2}$$

where the vector $\boldsymbol{\sigma}_{1i}$ captures the structural association between $\boldsymbol{y}_i$ and $\log T_i$. This association between two processes in the joint model framework had been previously captured through the subject specific random-effects $\boldsymbol{b}_i$ in literature. In our proposed modelling structure we have captured mainly the longitudinal correlation by $\boldsymbol{b}_i$ which indicates that the dependency between the two processes can still be captured using the conditional distribution even if these two processes do not share common $\boldsymbol{b}_i$.

Owing to difficulties due to positive definiteness constraints and high-dimensional complexites it is cumbersome to model the entire covariance matrix for each subject. This issue can be addressed by factorization of the joint distribution of $(\boldsymbol{y}_i, \log T_i)$. In our proposed modelling framework, we factor the joint distribution of $\boldsymbol{y}_i$ and $\log T_i$ into two components: a marginal non-linear model for $\boldsymbol{y}_i$ and a correlated regression model for $\log T_i$ given $\boldsymbol{y}_i$. Similar type of setting has been used by Fitzmaurice and Laird [28] and Gueorguieva and Agresti [29] in this context. In presence of subject specific random-effects $\boldsymbol{b}_i$, we let

$$\boldsymbol{x}_i = \begin{pmatrix} \boldsymbol{x}_{i1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{x}_{i2} \end{pmatrix}, \boldsymbol{z}_i = \begin{pmatrix} \boldsymbol{z}_{i1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{z}_{i2} \end{pmatrix} \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$$

Then, by implementing the Bartlett decomposition of a covariance matrix, the new models can be expressed as:

$$\boldsymbol{y}_i|\boldsymbol{b}_i = \boldsymbol{g}(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1, \boldsymbol{z}_{i1}\boldsymbol{b}_i) + \boldsymbol{\varepsilon}_{i1} \tag{3}$$

and

$$\log T_i|\boldsymbol{y}_i, \boldsymbol{b}_i = \boldsymbol{x}_{i2}\boldsymbol{\beta}_2 + \boldsymbol{z}_{i2}\boldsymbol{b}_i + \boldsymbol{B}_i(\boldsymbol{y}_i - \boldsymbol{g}(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1, \boldsymbol{z}_{i1}\boldsymbol{b}_i)) + \varepsilon_{i2} \tag{4}$$

where $\boldsymbol{B}_i = \boldsymbol{\sigma}_{1i}'\boldsymbol{\Sigma}_{y_i}$ is the vector reflecting structural association between these two processes for the i^{th} individual where $\boldsymbol{\sigma}_{1i} = \sigma_{cov}^2\boldsymbol{1}$. Here we also capture the local dependency through non-zero $\boldsymbol{z}_{i1}$ and $\boldsymbol{z}_{i2}$. Further let us assume, $\boldsymbol{\varepsilon}_{i1} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{y_i})$ and $\varepsilon_{i2} \sim N(0, \sigma_T^2)$. To capture longitudinal correlation, we have assumed $\boldsymbol{b}_i$ to be a p component vector following $N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_b)$. The covariance matrix $\boldsymbol{\Sigma}_b$ is structured as the AR(1) process i.e

$$\boldsymbol{\Sigma}_b = \sigma_b^2 \begin{pmatrix} 1 & \rho & .. & \rho^{p-1} \\ \rho & 1 & .. & \rho^{p-2} \\ . & . & .. & . \\ . & . & .. & . \\ \rho^{p-1} & \rho^{p-2} & .. & 1 \end{pmatrix}$$

Censoring information is incorporated through indicator variable δ_i such that $\delta_i = 1$ if the i^{th} individual is not censored and zero otherwise. The time-to-event observation T_i^* is $min(\log T_i, C_i)$ where C_i denotes censored time point for the i^{th} individual and T_i denotes the time to event observations of the i^{th} individual. Thus the contribution of the time-to-event for the likelihood can be expressed as:

$$f(T_i^*|\boldsymbol{y}_i, \boldsymbol{b}_i) = h(T_i^*|\boldsymbol{b}_i)^{\delta_i} S(T_i^*|\boldsymbol{b}_i)$$

where $h(T_i^*|\boldsymbol{b}_i)$ is the conditional hazard and $S(T_i^*|\boldsymbol{b}_i)$ is the conditional survival function. Under the log-normal assumption, we have

$$S(T_i^*|\boldsymbol{y}_i, \boldsymbol{b}_i) = 1 - \Phi\left(\frac{T_i^* - \lambda(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{b}_i)}{\sigma_A}\right)$$

where

$$\lambda(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{b}_i) = \boldsymbol{x}_{i2}\boldsymbol{\beta}_2 + \boldsymbol{z}_{i2}\boldsymbol{b}_i + \boldsymbol{B}_i(\boldsymbol{y}_i - \boldsymbol{g}(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1, \boldsymbol{z}_{i1}\boldsymbol{b}_i))$$

denotes the conditional mean of T_i^* structurally dependent on the longitudinal trajectory and the subject-specific random-effects and

$$\sigma_A^2 = \sigma_T^2 - \boldsymbol{\sigma}_{1i}{}' \boldsymbol{\Sigma}_{y_i}^{-1} \boldsymbol{\sigma}_{1i}$$

denotes the conditioned variability.

It is a well-known fact that the AFT (accelerated failure time) structure in joint modelling is troublesome to deal with compared to Cox model, since $f(T_i^*|\boldsymbol{y}_i, \boldsymbol{b}_i)$ is more complicated and unlike the Cox model the baseline function involves unknown quantities. As a result, it is not possible to use the point mass function with masses assigned to all uncensored survival times T_i for the baseline hazard function.
Hence, the complete data likelihood for the i^{th} individual :

$$L_i = \left(\prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{b}_i)\right) f(T_i^*|\boldsymbol{y}_i, \boldsymbol{b}_i) f(\boldsymbol{b}_i|\sigma_b^2, \rho)$$

Assuming independence among subjects, we can take $\boldsymbol{\Sigma}_{y_i} = \sigma_y^2 \boldsymbol{I}$, where $\boldsymbol{I}$ is a $n_i \times n_i$ matrix. For notational simplicity, we let $\boldsymbol{D} = \{y_{ij}\} \cup \{T_i^*\} \cup \{\delta_i\}$ be the observed data and $\boldsymbol{\Psi} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \sigma_y^2, \sigma_b^2, \rho, \sigma_T^2, \sigma_{cov}^2, \boldsymbol{\nu}')'$ be the parameter vector. Here $\boldsymbol{\nu}$ denotes the coefficient vector of random-effects in the time-to-event trajectory.
Here, we have,

$$T_i^*|\boldsymbol{y}_i, \boldsymbol{b}_i \sim N(\lambda(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{b}_i), \sigma_A^2)$$

The contribution owing to the accelerated failure time structure is

$$f_S = \phi^*(W)^{\delta}(1 - \Phi(W))^{1-\delta}$$

where

$$W_i = \frac{T_i^* - \lambda(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{b}_i)}{\sigma_A}$$

denotes the standard normal variate relating to the time-to-event contribution for the i^{th} subject, $\phi^*(W)$ denotes the probability density function (pdf) and $\Phi(W)$ denotes the cumulative distribution function (cdf) of the standard normal variate W.

Combining all submodels, the likelihood for the i^{th} individual is given by,

$$L_i(\boldsymbol{\psi}|\boldsymbol{D}) \propto (\exp(-\frac{(T_i^* - \lambda(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{b}_i))^2}{2\sigma_A^2}))^{\delta_i}(1 - \int \exp(-\frac{(T_i^* - \lambda(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{b}_i))^2}{2\sigma_A^2}))^{1-\delta_i}$$
$$\exp(-(\boldsymbol{y_i} - \boldsymbol{g}(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1, \boldsymbol{z}_{i1}\boldsymbol{b}_i))' \boldsymbol{\Sigma}_{y_i}^{-1} (\boldsymbol{y_i} - \boldsymbol{g}(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1, \boldsymbol{z}_{i1}\boldsymbol{b}_i))) \exp(-\boldsymbol{b}_i' \boldsymbol{\Sigma}_b^{-1} \boldsymbol{b}_i)$$

2.2 Estimation

We need to estimate the parameters involved in the two models denoted by $\boldsymbol{\Psi} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \sigma^2_y, \sigma^2_b, \rho, \sigma^2_T, \sigma^2_{cov}, \boldsymbol{\nu}')'$. Due to limitations in existing classical approaches for the parametric estimation under non-linear framework, we have proposed a two-stage approach where we fit the non-linear longitudinal submodel at the first stage. We use the estimates obtained from this method and plug them in to the survival model. This estimation of the longitudinal process separately constitutes the first stage of our two-stage estimation. The next stage involves using of the estimated longitudinal parameters and the best linear unbiased estimators (BLUP) of the subject-specific random-effects in the survival process and then estimating the survival parameters by gradient descent algorithm with adaptive learning rate.

Two-Stage Approach

Stage-I : A non-linear mixed-effects model has been considered for the longitudinal process where we model the longitudinal trajectories as

$$\widetilde{\boldsymbol{y}}_{NL_i}|\boldsymbol{b}_i = \boldsymbol{g}(\boldsymbol{x}_{i11}\boldsymbol{\beta}_{11}, \boldsymbol{z}_{i11}\boldsymbol{b}_i) + \boldsymbol{\varepsilon}_{i11}$$

where $\boldsymbol{x}_{i11}$ and $\boldsymbol{z}_{i11}$ respectively are the covariate vectors of the fixed-effects parameter vector $\boldsymbol{\beta}_{11}$ and the subject specific random-effects $\boldsymbol{b}_i$ constituting the components of non-linear vector functions $\boldsymbol{g}(\cdot)$. Here $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_b)$ with $\boldsymbol{\Sigma}_b$ as a AR(1) process.

The model had been fitted through *lme* function in *nlme*() package (Pinheiro and Bates [30], Pinheiro et al. [31]) in R and the corresponding output has been utilized for obtaining the estimated longitudinal submodel. It can be represented as:

$$\widetilde{\boldsymbol{y}}_{NL_i}|\boldsymbol{b}_i = \boldsymbol{g}(\boldsymbol{x}_{i11}\boldsymbol{\beta}^*_{11}, \boldsymbol{z}_{i11}\boldsymbol{b}^*_i) + \boldsymbol{\varepsilon}_{i11}$$

where $\boldsymbol{\beta}^*_{11}$ and $\boldsymbol{b}^*_i$ correspond to the estimated fixed-effects parameters and best linear unbiased predictors respectively obtained from the fitted model summary.

In this work we have also considered a linear set up along with the non-linear set up as

$$\widetilde{\boldsymbol{y}}_{L_i}|\boldsymbol{b}_i = \boldsymbol{x}_{i21}\boldsymbol{\beta}_{21} + \boldsymbol{z}_{i21}\boldsymbol{b}_i + \boldsymbol{\varepsilon}_{i21}$$

where $\boldsymbol{x}_{i21}$ is the covariate vector of the fixed-effects for the i^{th} trajectory. Here, $\boldsymbol{\beta}_{21} = (\beta_I, \beta_S)$, where β_I and β_S represents the fixed effect intercept and the slope of the longitudinal profile for the i^{th} patient. $\boldsymbol{z}_{i21}$ represents the covariate vector incorporating both the random-effects intercept and slope for the i^{th} longitudinal trajectory. The fitted longitudinal model can be represented in the form:

$$\widetilde{\boldsymbol{y}}_{L_i}|\boldsymbol{b}_i = \boldsymbol{x}_{i21}\boldsymbol{\beta}^*_{21} + \boldsymbol{z}_{i21}\boldsymbol{b}^*_i + \boldsymbol{\varepsilon}_{i21}$$

where $\boldsymbol{\beta}^*_{21}$ and $\boldsymbol{b}^*_i$ respectively correspond to the estimated parameter values of the fixed-effects and the best linear unbiased predictors obtained from the submodel fit.

Stage-II: For the non-linear framework, the information from the first step of the procedure has thus been implemented as plug-in estimates in the survival part of our model which can be expressed as:

$$T^*_i|\widetilde{\boldsymbol{y}}_{NL_i}, \boldsymbol{b}^*_i = \boldsymbol{x}_{i12}\boldsymbol{\beta}_{12} + \boldsymbol{z}_{i12}\boldsymbol{b}^*_i + \boldsymbol{B}_{i1}(\widetilde{\boldsymbol{y}}_{NL_i} - \boldsymbol{g}(\boldsymbol{x}_{i11}\boldsymbol{\beta}^*_{11}, \boldsymbol{z}_{i11}\boldsymbol{b}^*_i)) + \varepsilon_{i12}$$

where $\boldsymbol{B}_{i1} = \boldsymbol{\sigma}'_{1i}\boldsymbol{\Sigma}^*_{y_{NL_i}}$

For the linear extension, the fitted values of the parameters in the *lme* model and the empirical Bayes estimates of the subject-specific parameters thus obtained are used as plug-in estimates in the survival part of our model where

$$T^*_i|\widetilde{\boldsymbol{y}}_{L_i}, \boldsymbol{b}^*_i = \boldsymbol{x}_{i22}\boldsymbol{\beta}_{22} + \boldsymbol{z}_{i22}\boldsymbol{b}^*_i + \boldsymbol{B}_{i2}(\widetilde{\boldsymbol{y}}_{L_i} - \boldsymbol{x}_{i21}\boldsymbol{\beta}^*_{21} - \boldsymbol{z}_{i21}\boldsymbol{b}^*_i) + \varepsilon_{i22}$$

where $\boldsymbol{B}_{i2} = \boldsymbol{\sigma}'_{1i}\boldsymbol{\Sigma}^*_{y_{L_i}}$.

Hence, we are left with the task of estimation of the parameters for the time-to-event process only which we estimate by adopting a gradient-descent optimization algorithm with an adaptive learning rate G where the parameter vector at $(h+1)^{th}$ step can be expressed as

$$\boldsymbol{\Theta}^{(h+1)} = \boldsymbol{\Theta}^{(h)} - G^{(h)}(\boldsymbol{\Theta}) * \boldsymbol{U}(\boldsymbol{\Theta}^{(h)})$$

where $\boldsymbol{U}(\boldsymbol{\Theta}^{(h)})$ denotes the score vector of the parameter at the h^{th} step.

The choice of the learning rate is very important as a large G can lead to skipping the optimal solution whereas a small G may result in needing many iterations to converge to best values. So, we adapt our learning rate at each iteration step depending on how closer or farther we are to the true solution. The first-order partial derivatives comprising the elements of the score vectors under the non-linear framework can be expressed as:

$$\begin{aligned}
\frac{\partial lnL_i}{\partial \boldsymbol{\beta}_{12}} &= \frac{\zeta}{\sigma_A}\boldsymbol{x}_{i12} \\
\frac{\partial lnL_i}{\partial \sigma_T^2} &= \frac{\zeta}{2\sigma_A^3}(T_i^* - \boldsymbol{x}_{i12}\boldsymbol{\beta}_{12} - \boldsymbol{z}_{i12}\boldsymbol{b}_i^* - \boldsymbol{B}_{i1}(\widetilde{\boldsymbol{y}}_{NL_i} - \boldsymbol{g}(\boldsymbol{x}_{i11}\boldsymbol{\beta}_{11}^*, \boldsymbol{z}_{i11}\boldsymbol{b}_i^*))) \\
\frac{\partial lnL_i}{\partial \sigma_{cov}^2} &= \zeta[\frac{\mathbf{1}(\widetilde{\boldsymbol{y}}_{NL_i} - \boldsymbol{g}(\boldsymbol{x}_{i11}\boldsymbol{\beta}_{11}^*, \boldsymbol{z}_{i11}\boldsymbol{b}_i^*))}{\sigma_A} - \frac{n_i\sigma_{cov}^2}{\sigma_A^3}(T_i^* - \boldsymbol{x}_{i12}\boldsymbol{\beta}_{12} - \boldsymbol{z}_{i12}\boldsymbol{b}_i^* - \boldsymbol{B}_{i1}(\widetilde{\boldsymbol{y}}_{NL_i} - \boldsymbol{g}(\boldsymbol{x}_{i11}\boldsymbol{\beta}_{11}^*, \boldsymbol{z}_{i11}\boldsymbol{b}_i^*)))] \\
\frac{\partial lnL_i}{\partial \nu_s} &= \zeta\,\frac{b_{i(s)}}{\sigma_A}
\end{aligned}$$

where $\zeta = \left(\delta_i W_i + (1-\delta_i)\,\dfrac{\phi^*(W_i)}{(1-\Phi_i(W_i))}\right)$, ($W_i$ defined in previous section)

$$\begin{aligned}
\boldsymbol{z}_{i12} &= \boldsymbol{\nu} = (\nu_1, \nu_2, ..., \nu_s, ...)', \\
G^{(h)}(\boldsymbol{\Theta}) &= \frac{(\boldsymbol{\Theta}^{(h-1)} - \boldsymbol{\Theta}^{(h-2)})(\boldsymbol{U}(\boldsymbol{\Theta}^{(h-1)}) - \boldsymbol{U}(\boldsymbol{\Theta}^{(h-2)}))}{(||\boldsymbol{U}(\boldsymbol{\Theta}^{(h-1)}) - \boldsymbol{U}(\boldsymbol{\Theta}^{(h-2)})||)^2}, \\
\boldsymbol{\sigma}_1 &= \sigma_{cov}^2\mathbf{1}, \;\; n_i = \text{number of time points for the } i^{th} \text{ individual}
\end{aligned}$$

with $G^{(h)}$ denoting the learning rate for the h^{th} parameter step and U being the score function.

Apart from this, we have considered two separate joint models namely JM_1 and JM_2 under our proposed non-linear mixed-effects model where the former has been modelled in a way that the processes would be linked with both subject-specific random-effects and structural association i.e., $\boldsymbol{\sigma}_1$ and the latter links the two processes solely by structural association. This approach has also been adopted under the linear framework and thus the performances of the estimation procedure in the survival model can be examined. Further, it gives an interesting view about how the nature of the association between the two processes affects their estimates.

3 Simulation Study

We attempt to investigate the efficiency of this two-stage estimation method by conducting a simulation study considering 200 individuals each with varying number of longitudinal measurements generated from $Uniform(4, 10)$. A non-linear model is considered for longitudinal data and subsequently, a linear model has been considered for the sake of comparison. To examine importance of structural dependency between the two processes, we have considered two joint models i.e., JM_1 and JM_2. The former has the two processes linked by means of both subject-specific random-effects and structural dependency by means of conditional distribution and the latter has two processes linked solely by structural dependency.

We had thus generated 100 datasets under the non-linear framework for each of the joint models with the following model assumptions. Continuous covariates for fixed and random-effects in the longitudinal submodel were generated from

normal distributions. The subject-specific random-effects for each individual, in this case, has been assumed to follow a bivariate normal distribution with $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_b)$, where $\boldsymbol{\Sigma}_b$ follows the AR(1) structure with ρ being the longitudinal correlation. Again, the longitudinal trajectories are assumed to follow n_i variate normal distribution where $y_{ij} = \alpha + \beta x_{ij} + \beta_1 x_{ij}^2 + b_{i1} + b_{i2} t_{ij} + \varepsilon_{ij}$ denotes the i^{th} patient's longitudinal profile at the j^{th} time point where $\boldsymbol{\varepsilon}^{'}s$ are generated from $N(\boldsymbol{0}, \sigma_y^2 \boldsymbol{I})$. For the sake of identifiability we assume $\sigma_y^2 = 1$. Here, (α, β, β_1) denotes the fixed-effects parameter vector for the non-linear longitudinal trajectory, x_{ij} and t_{ij} respectively being the covariate vector for the fixed-effects and random-effects respectively. $x_{ij}^{'}s$ are generated from N(3, 0.5) and t_{ij} represents the time point of the j^{th} longitudinal observation for the i^{th} patient. The inclusion of t_{ij} ensures that the effect of time is considered as covariate in the longitudinal trajectory.

Time-to-event observations i.e $\log T_i^{'s}$ are generated from $N(\mu_{T_{1i}}, \sigma_{T_C})$ where $\mu_{T_{1i}} = \beta_0 + \nu_1 b_{i1} + \nu_2 b_{i2} + \sigma_{cov}^2 \sum_{j=1}^{n_i} (y_{ij} - \alpha - \beta x_{ij} - \beta_1 x_{ij}^2 - b_{i1} - b_{i2} t_{ij})$ and $\sigma_{T_C} = (\sigma_T^2 - n_i \sigma_{cov}^4)$. β_0 denotes the intercept of the time-to-event trajectory where (ν_1, ν_2) is the coefficient vector for the subject-specific random-effects. Here, $\sigma_{cov}^2 \boldsymbol{1} = Cov(\log T_i, \boldsymbol{y}_i)$ reflects the structural correlation between the two submodels.

The censoring time associated with the i^{th} individual i.e., $C_i \sim exp(0.5)$ where we compute $T_i^* = min(\log T_i, C_i)$ and $\delta_i = 1$ if $\log T_i \leq C_i$ and 0 otherwise. Thus (T_i^*, δ_i) contains the censoring information with the censoring percentage observed to be around 35%-40%. We also considered a setup with 20% censoring and found no remarkable difference in the results. We assumed the true parameter values to be $\alpha = 1.85$, $\beta_0 = 1$, $\beta = 1.3$, $\beta_1 = 0.5$, $\sigma_{cov}^2 = 0.1$, $\sigma_T^2 = 0.8$, $\sigma_b^2 = 0.8$, $\rho = 0.3$, $\nu_1 = -0.2$, $\nu_2 = 0.8$.

Further, as discussed previously in Section 1, for the sake of establishing the importance of the structural dependency in our proposed model we have carved out two joint models. JM_1 with two kinds of association between the two processes (structural and through subject-specific random-effects) and JM_2 or the joint model with local independence (structural association only) where $\nu_1 = \nu_2 = 0$.

We have simultaneously considered a linear mixed-effects model for longitudinal data using similar set up.

Here, the longitudinal observation for the i^{th} patient at the j^{th} time point can be expressed as $y_{ij} = \alpha + \beta x_{ij} + b_{i1} + b_{i2} t_{ij} + \varepsilon_{ij}$. Time to event observations i.e $\log T_i^{'s}$ have been generated from $N(\mu_{T_{2i}}, \sigma_{T_C})$ where $\mu_{T_{2i}} = \beta_0 + \nu_1 b_{i1} + \nu_2 b_{i2} + \sigma_{cov}^2 \sum_{j=1}^{n_i} (y_{ij} - \alpha - \beta x_{ij} - b_{i1} - b_{i2} t_{ij})$ and $\sigma_{T_C} = (\sigma_T^2 - n_i \sigma_{cov}^4)$ expresses the conditional mean of the survival observations directly dependent on both the longitudinal trajectories and the subject-specific random-effects.

Under this model segments the parameter space can be defined as:

$$\boldsymbol{\Theta}_1 = (\alpha, \beta_0, \beta, \beta_1, \sigma_T^2, \sigma_{cov}^2, \sigma_b^2, \rho, \nu_1, \nu_2)$$
$$\boldsymbol{\Theta}_2 = (\alpha, \beta_0, \beta, \beta_1, \sigma_T^2, \sigma_{cov}^2, \sigma_b^2, \rho)$$
$$\boldsymbol{\Theta}_3 = (\alpha, \beta_0, \beta, \sigma_T^2, \sigma_{cov}^2, \sigma_b^2, \rho, \nu_1, \nu_2)$$
$$\boldsymbol{\Theta}_4 = (\alpha, \beta_0, \beta, \sigma_T^2, \sigma_{cov}^2, \sigma_b^2, \rho)$$

where $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ are the parameter space for JM_1 and JM_2 for non-linear framework and $\boldsymbol{\Theta}_3$ and $\boldsymbol{\Theta}_4$ respectively are the parameter space JM_1 and JM_2 under the linear framework.

Hence we obtain $(\boldsymbol{x}_i, \boldsymbol{y}_i, T_i^*, \delta_i)$ as our dataset for the purpose of the simulation study and model comparison. In the non-linear framework, we fit the data thus obtained in the form of a quadratic mixed-effects model through *lme*() function available under *nlme* package. The corresponding bias, Standard Error (SE), Mean Square Error (MSE) and coverage probabilities (CP) for 95% equal tailed confidence intervals for all the parameters under the sub-models has been displayed in Table 1. Estimates for the intercept, linear effect parameter and quadratic effect parameter under the two sub-models have been found to be $\hat{\alpha} = 1.5096$ with Standard error (se)=0.1152 for JM_1, $\hat{\alpha} = 1.5100$ with se=0.1111 for JM_2, $\hat{\beta} = 1.2934$ with se=0.0766 for JM_1, $\hat{\beta} = 1.2948$ with se=0.0800 for JM_2, $\hat{\beta}_1 = 0.4887$ with se=0.0127 for JM_1, $\hat{\beta}_1 = 0.4999$ with se=0.0122 for JM_2. The longitudinal correlation ρ has been estimated to be $\hat{\rho} = 0.7131$ with se=0.1364 for JM_1 and $\hat{\rho} = 0.7120$ with se=0.2961 for JM_2. The estimated variability of the longitudinal component for both JM_1 and JM_2 is $\hat{\sigma}_b^2 = 0.6201$ with se=0.0001.

Moreover, for each of the joint models the best linear unbiased predictors (BLUP) for the subject-specific random-effects i.e., $\hat{b}_{i1}$ and $\hat{b}_{i2}$ have also been obtained and thus the estimated parameters along with the BLUPs are plugged in

the survival process of the model where we estimate the parameters by adopting gradient-descent optimization algorithm. As for the parameters in the survival process, the intercept for the time-to-event trajectory has been estimated to be $\hat{\beta}_0 = 0.9001$ with se=0.0009 for JM_1 and $\hat{\beta}_0 = 1.1000$ with se=0.0001 for JM_2. The structural association parameter between the two processes has been estimated to be $\hat{\sigma}^2_{cov} = 0.0302$ with se=0.2382 for JM_1 and $\hat{\sigma}^2_{cov} = 0.0040$ with se=0.0007 for JM_2. The unconditional variability for the time-to-event observations is estimated to be $\hat{\sigma}^2_T = 0.7000$ with se=0.0010 for JM_1 and $\hat{\sigma}^2_T = 0.9005$ with se=0.1000 for JM_2. The estimated coefficients of the components of the subject-specific random-effects in the time-to-event process have been found to be $\hat{v}_1 = -0.1900$ with se=0.0011 and $\hat{v}_2 = 0.7000$ with se=0.0008 for JM_1. Both the joint models in non-linear framework performed equally well with comparable Bias, low standard error and mean square errors and the findings has been displayed in Table 1.

As for the linear model for longitudinal observations we fit it with a linear mixed-effects by *lme()* function available in the *nlme* package in **R**. For the JM_1 and JM_2 sub-models, the summary provides us with the estimate of the intercept and slope of the fixed-effects i.e., $\hat{\alpha} = 1.8021$ with se=0.0265 for JM_1 and $\hat{\alpha} = 1.8342$ with se=0.0287 for JM_2, $\hat{\beta} = 1.3126$ with se=0.0077 for JM_1 and $\hat{\beta} = 1.2936$ with se=0.1847 for JM_2. σ^2_b has been estimated to be $\hat{\sigma^2_b} = 0.6287$ with se=0.1446 for JM_1 and $\hat{\sigma^2_b} = 0.6389$ with se=0.1226 for JM_2.

The longitudinal correlation for JM_1 and JM_2 are respectively $\hat{\rho} = 0.4457$ with se=0.1183 and $\hat{\rho} = 0.3510$ with se=0.1334. Further, the best linear unbiased estimates (BLUP) of the slope and intercept components of the subject-specific random-effects i.e., $\hat{b}_{i1}$ and $\hat{b}_{i2}$ for each of the submodels are obtained and these along with these estimates are utilized as plug-in estimates in the survival part of the model. Hence we are only left with the task of estimation of the time-to-event parameters which we have achieved by adopting the gradient-descent optimization algorithm with an adaptive learning rate as done with our proposed non-linear framework.

The estimated intercept of the time-to-event trajectory is $\hat{\beta}_0 = 1.0836$ with se=0.0163 for JM_1 and $\hat{\beta}_0 = 1.0999$ with se=0.0001 for JM_2. The structural association between the two components of the joint models i.e σ^2_{cov} is $\hat{\sigma}^2_{cov} = 0.2002$ with se=0.0001 for JM_1 and $\hat{\sigma}^2_{cov} = 0.0003$ with se=0.0001 for JM_2. The unconditional variability for for the time-to-event is $\hat{\sigma}^2_T = 0.9716$ with se=0.0008 for JM_1 and $\hat{\sigma}^2_T = 0.7568$ with se=0.0001 for JM_2. The coefficients for the intercept and slope for the subject-specific random-effects in the time-to-event trajectory were estimated to be $\hat{v}_1 = -0.1997$ with se=0.0016 and $\hat{v}_2 = 0.6378$ with se=0.0018 and it is seen that JM_1 and JM_2 perform reasonably well with small bias, standard error (se) and mean square error (MSE) with satisfactory coverage probabilities for all the parameters.

It can be easily seen from Table 1 that structural correlation alone (JM_2) can give almost similar kind of results compared to JM_1 in terms of biases. Moreover, it can be noted that the estimates for linear and non-linear under JM_1 and JM_2, in most cases, are found to be close enough. This shows that the two-stage method performs well in a different setup.

3.1 Robustness study

It must be noted that we have adopted a fully parametric model for our analysis and have modelled the survival process by parametric accelerated failure time model with a lognormal distributional assumption. To examine the robustness of our model under model misspecification, we specify the survival distribution as generalized logistic distribution with probability density function $f(x;\mu,\varsigma) = \frac{e^{-\frac{x-\mu}{\varsigma}}}{\varsigma(1+e^{-\frac{x-\mu}{\varsigma}})^2}$ where $-\infty < x, \mu < \infty$ and $\varsigma > 0$. The results are summarized in Table 2. It displays the results under both the linear and non-linear framework and the subsequent joint models under each of them with bias, standard error, mean square error and the coverage probabilities of the 95% confidence intervals. The results are satisfactory with low bias, SE, MSE and desirable coverage probabilities for both the linear and non-linear approaches.

In terms of absolute bias, both the models JM_1 and JM_2 under linear framework performs better for longitudinal parameters than the non-linear framework, whereas the coverage probabilities for α, σ^2_b and ρ under non-linear framework are found to be on the slightly higher side. Again, the time-to-event parameters β_0, σ^2_{cov} amd σ^2_T under JM_1 reveals sightly lower absolute bias for the non-linear framework and the coverage probabilities for the time-to-event submodel parameters are observed to have better performance under the non-linear framework. JM_2 exhibits comparable performance in terms of absolute bias for the time-to-event submodel parameters with better coverage probability of β_0 and σ^2_T for non-linear framework.

4 DATA Analysis

4.1 AIDS Data

In an multicenter open-label trial 467 patients each of whom had either previously received zidovudine (ZDV) and had 300 or fewer CD4 cells per cubic millimetre of blood or were diagnosed with acquired immunodeficiency syndrome (AIDS) were being randomly assigned for being treated with either (ddI) didanosine (500mg per day) or (ddC) zalcitabine (2.25 mg per day). CD4 blood cells are a part of the infection-fighting system of the body produced in the spleen, lymph nodes and thymus gland. Thus the absolute number of CD4 lymphocyte cells per cubic millimetre of blood is an effective indicator for progression of AIDS and so it has been measured repeatedly over intervals of 2, 6, 12 and 18 months. CD4 cell count over time serves as a biomarker and is an indication of enhanced risk to infection. Both of the drugs didanosine and zalcitabine are commonly used to treat patients infected by human immunodeficiency virus (HIV) who were either unable to tolerate zidovudine treatment or were showing signs of disease progression despite it.

This data was collected from the experiment carried out by the Terry Beirn Community Programs for Clinical Research on AIDS. A detailed overview can be found in Abrams et al. [32]. The goal of this study was to compare and judge the safety and relative efficacy of these two alternative antiretroviral drugs in curing patients suffering from the HIV. The study revealed that about 66% (309) patients had experienced disease progression or death, and nearly about 40% (188) had died after a median follow-up of 15.6 months. About 35% (164) had become study drug intolerant, and about 57% (143) of the still living and monitored had been discontinued permanently from the original study drug. Only 1% (4) patients' vital status could not be known and about 7% (31) patients were not participating in the study anymore.

In this two-step method, in the first stage, we have incorporated the covariate Observation time (Obstime) which constitutes follow up periods of the patients under study. The square root of the CD4 cell counts (square root transformation adopted to reduce the right skewness in the CD4 data) serves as the repeated longitudinal observations.

According to our non-linear framework the CD4 cell counts in the data may be expressed in the form:

$$y_{ij} = \alpha + \beta \text{Obstime}_{ij} + \beta_1 \text{Obstime}_{ij}^2 + \text{Obstime}_{ij}\text{Drug}_i + b_{i1} + \text{Obstime}_{ij} b_{i2} + \varepsilon_{11ij},$$

where α, β and β_1 are respectively the intercept, linear effect parameter and the quadratic effect parameter for the CD4 cell trajectory with the subject-specific random-effects distribution assumed to follow a bivariate normal with zero mean vector and $\boldsymbol{\Sigma}_b = \sigma_b^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ as the dispersion matrix.

We obtain the estimates of the parameters using *lme()* function available in *nlme* package in **R**. These estimates and the best linear unbiased predictors of the subject-specific random-effects thus obtained from the first step are utilized as plug-in estimates in the survival part of the model. Thus the time-to-event outcome variable i.e., death due to acquired immunodeficiency syndrome can be expressed as:

$$\begin{aligned} \log T_i = {} & \beta_0 + \nu_1 \hat{b}_{i1} + \nu_2 \hat{b}_{i2} + \sigma_{cov}^2 \sum_{j=1}^{n_i} (y_{ij} - \alpha - \beta \text{Obstime}_{ij} \\ & - \beta_1 \text{Obstime}_{ij}^2 - \text{Obstime}_{ij}\text{Drug}_i - b_{i1} - \text{Obstime}_{ij} b_{i2}) + \varepsilon_{12ij} \end{aligned}$$

where β_0 is the intercept, ν_1 and ν_2 respectively are the coefficients of the subject-specific random-effects of the time-to-event trajectory i.e., $\log T_i$. Time-to-event observations indicate the time up to which i^{th} patient has lived with this disease (in case of death) or time at which the censoring took place (the patient if alive till the end of the study). σ_{cov}^2 captures the structural association prevalent between the CD4 cell count observations of the affected patient and the time to death or disease progression. We carry out the estimation procedure by adopting a gradient-descent optimization procedure for estimating the parameters of the survival process. We have also compared two submodels under the linear framework namely JM_1 and JM_2 $(\nu_1 = \nu_2 = 0)$ where, in the former, the two processes are associated by both subject-specific random-effects and structural association and the latter has solely structural association binding the two processes.

For the linear set up the following model is used:

$$y_{ij} = \alpha + \beta\ \text{Obstime}_{ij} + \text{Obstime}_{ij}\text{Drug}_i + b_{i1} + b_{i2}\text{Obstime}_{ij} + \varepsilon_{21ij},$$

where α and β denotes the intercept and slope component of the CD4 cell trajectory with $\boldsymbol{b}_i = (b_{i1}, b_{i2})$ following a bivariate normal distribution with $\mathbf{0}$ mean and $\sigma_b^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ as the dispersion matrix. The time-to-event outcome variable

can be expressed as:

$$\log T_i = \beta_0 + \nu_1 \hat{b}_{i1} + \nu_2 \hat{b}_{i2} + \sigma^2_{cov} \sum_{j=1}^{n_i} (y_{ij} - \alpha - \beta \text{Obstime}_{ij} - \text{Obstime}_{ij}\text{Drug}_i - b_{i1} - \text{Obstime}_{ij} b_{i2}) + \varepsilon_{22ij}$$

The results for both the approaches have been summarized in Table 3 with estimated longitudinal intercept for linear and non-linear set-up being $\hat{\alpha} = 2.5118$ and $\hat{\alpha} = 2.5204$ respectively. The slope parameters for linear and non-linear models were estimated to be $\hat{\beta} = -0.0375$ and $\hat{\beta} = -0.0456$ respectively for linear effect and $\hat{\beta}_1 = 0.0007$ for quadratic effect. The estimated longitudinal correlation is $\hat{\rho} = 0.0007$ for linear model and $\hat{\rho} = 0.1058$ in the case of non-linear counterpart. The longitudinal variability is estimated to be $\hat{\sigma^2_b} = -0.1500$ for linear and $\hat{\sigma^2_b} = 0.1058$ for non-linear framework. The estimated survival parameters by gradient-descent algorithm for linear and non-linear models respectively are $\hat{\beta}_0 = 1.0990$ and $\hat{\beta}_0 = 0.6999$ (intercept), $\hat{\sigma}^2_{cov} = 0.0012$ and $\hat{\sigma}^2_{cov} = 0.0013$ (structural association between the two processes), $\hat{\sigma}^2_T = 1.7000$ and $\hat{\sigma}^2_T = 1.5096$ (unconditional variability for the time-to-event observations), $\hat{\nu}_1 = -0.1999$ and $\hat{\nu}_1 = 0.1000$ (coefficient of b_{i1}), $\hat{\nu}_2 = -0.7999$ and $\hat{\nu}_2 = 0.6000$ (coefficient of b_{i2}).

Four models are comparable in this data set. However, JM_1 under linear mixed-effects model gives the best fit in terms of AIC value as displayed in Table 4. It may be noted that from this data linear mixed-effects model be it JM_1 or JM_2 fits better compared to quadratic mixed-effects model. All models converged rapidly within 10 iterations (Table 5).

Table 1: Summary statistics for parameter estimation of the simulated data for the models

Parameter	True	Linear mixed-effects								Quadratic mixed-effects							
		JM_1				JM_2				JM_1				JM_2			
		Abs. BIAS	SE	MSE	CP	Abs. BIAS	SE	MSE	CP	Abs. BIAS	SE	MSE	CP	Abs. BIAS	SE	MSE	CP
longitudinal																	
α	1.8500	0.0479	0.0265	0.0029	1.0000	0.0158	0.0287	0.0011	1.0000	0.3404	0.1152	0.1291	0.9700	0.3400	0.1111	0.1236	0.9700
β	1.3000	0.0126	0.0077	0.0002	0.9900	0.0064	0.1847	0.0034	1.0000	0.0066	0.0766	0.0059	0.9900	0.0052	0.0800	0.0059	0.9900
β_1	0.5000	-	-	-	-	-	-	-	-	0.0113	0.0127	0.0003	1.0000	0.0001	0.0122	0.0003	1.0000
σ^2_b	0.8000	0.1713	0.1446	0.0502	0.9900	0.1611	0.1226	0.0409	0.9900	0.1799	0.0001	0.6282	0.9700	0.1799	0.0001	0.6282	0.9700
ρ	0.3000	0.0580	0.1183	0.0174	1.0000	0.0510	0.1334	0.0204	1.0000	0.4131	0.1364	0.1893	0.9800	0.4120	0.2961	0.2134	0.9800
time-to-event																	
β_0	1.0000	0.0836	0.0163	0.0072	0.9600	0.0999	0.0001	0.0099	0.9600	0.0999	0.0009	0.0099	1.0000	0.1000	0.0001	0.0100	1.0000
σ^2_{cov}	0.1000	0.1002	0.0001	0.0100	0.9800	0.1003	0.0001	0.0101	0.9800	0.0698	0.2382	0.0616	0.9600	0.0960	0.0007	0.0092	0.9700
σ^2_T	0.8000	0.1716	0.0008	0.0294	0.9900	0.0432	0.0001	0.0018	0.9900	0.1000	0.0010	0.0100	0.9900	0.1005	0.1000	0.0201	0.9900
ν_1	-0.2000	0.0003	0.0016	0.0001	0.9900	-	-	-	-	0.0100	0.0011	0.0001	1.0000	-	-	-	-
ν_2	0.8000	0.1622	0.0018	0.0263	0.9900	-	-	-	-	0.1000	0.0008	0.0100	0.9700	-	-	-	-

5 Discussion

To avoid complexity in joint models where the longitudinal submodel is subject to non-linear mixed-effects model, a computationally more efficient two-stage process has been considered. It may be noted that the proposed method performs

Table 2: Results of fitting lognormal AFT model when data are simulated from generalized logistic distribution

Parameter	True	Linear mixed-effects								Quadratic mixed-effects							
		JM_1				JM_2				JM_1				JM_2			
		Abs. BIAS	SE	MSE	CP	Abs. BIAS	SE	MSE	CP	Abs. BIAS	SE	MSE	CP	Abs. BIAS	SE	MSE	CP
longitudinal																	
α	1.8500	0.0475	0.2345	0.0561	0.9700	0.0177	0.2043	0.0404	0.9700	0.3781	0.9935	1.1299	0.9800	0.1225	0.5283	0.2941	0.9900
β	1.3000	0.0195	0.0749	0.0059	0.9900	0.0116	0.0687	0.0049	0.9900	0.8435	0.4081	0.8780	0.9600	0.1525	0.2728	0.0977	0.9700
β_1	0.5000	-	-	-	-	-	-	-	-	0.0563	0.0442	0.0051	1.0000	0.0826	0.0606	0.0105	1.0000
σ^2_b	0.8000	0.1882	0.1688	0.0633	0.9600	0.2169	0.1145	0.0602	0.9600	0.7900	0.0006	0.6241	0.9700	0.7778	0.0134	0.6051	0.9700
ρ	0.3000	0.0327	0.1085	0.0126	0.9700	0.0248	0.1121	0.0138	0.9700	0.5131	0.0364	0.2643	0.9800	0.4120	0.2961	0.2134	0.9800
time to event																	
β_0	1.0000	0.0999	0.0001	0.0099	0.9700	0.0999	0.0001	0.0099	0.9700	0.0999	0.0001	0.0099	0.9900	0.0980	0.0001	0.0096	0.9900
σ^2_{cov}	0.1000	0.0988	0.0056	0.0098	0.9600	0.0994	0.0027	0.0099	0.9600	0.0098	0.0001	0.0001	0.9800	0.2144	0.0459	0.0481	0.9700
σ^2_T	0.8000	0.1000	0.0242	0.0106	0.9800	0.1000	0.0013	0.0100	0.9800	0.0644	0.0205	0.0046	0.9900	0.0048	0.0502	0.0024	0.9900
ν_1	-0.2000	0.0001	0.0001	0.0001	0.9700	-	-	-	-	0.3981	0.0801	0.1649	0.9700	-	-	-	-
ν_2	0.8000	0.0001	0.0001	0.0001	0.9700	-	-	-	-	0.1976	0.0007	0.0390	0.9700	-	-	-	-

Table 3: AIDS data study

Parameter	Linear mixed-effects		Quadratic mixed-effects	
	Est_{JM_1}	Est_{JM_2}	Est_{JM_1}	Est_{JM_2}
longitudinal				
α	2.5118	2.5118	2.5204	2.5204
β	-0.0375	-0.0375	-0.0456	-0.0456
β_1	-	-	0.0007	0.0007
σ_b^2	0.0298	0.0298	0.0004	0.0004
ρ	-0.1500	-0.1500	0.1058	0.1058
time-to-event				
β_0	1.0990	1.0999	0.6999	0.6999
σ_{cov}^2	0.0012	0.0013	0.0013	0.0013
σ_T^2	1.7000	1.7000	1.5096	1.5096
ν_1	-0.1999	-	0.1000	-
ν_2	-0.7999	-	0.6000	-

Table 4: Model comparison

AIC	Model			
	Linear mixed-effects		Quadratic mixed-effects	
	JM_1	JM_2	JM_1	JM_2
Aids Data Study	2728.330	2731.654	2737.666	2744.343

Table 5: Number of iterations needed for achieving convergence

Model			
Linear mixed-effects		Quadratic mixed-effects	
JM_1	JM_2	JM_1	JM_2
6-7	5-7	5-6	7-8

well in both linear and non-linear setup. From simulation studies, it may be noted that structural association alone, can take care of the association between these two processes. The proposed method is found to be robust against time-to-event model misspecification. Effect of influential observations both in response and covariates are out of the scope of this paper. This will be addressed in our future work for which work has already been started. Sensitivity analysis may be carried out with different structures of the variance-covariance matrix of $\boldsymbol{b}_i$. A bayesian analogous to this two-stage method may be of future interest.

Acknowledgments

We would like to thank an anonymous reviewer for providing us with valuable suggestions which enhanced the quality of work.

Conflict of Interest

The authors have declared no conflict of interest.

6 References

[1] R. J. Little and D. B. Rubin, The analysis of social science data with missing values, Sociological Methods & Research, 18(2-3), 292–326 (1989).

[2] R. J. Little and D. B. Rubin, Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annual review of public health, 21(1):121–145 (2000).

[3] R. J. Little, Pattern-mixture models for multivariate incomplete data. Journal of the American Statistical Association, 88(421):125–134 (1993). [4] G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, Advances in longitudinal data analysis: an historical perspective. In Longitudinal data analysis, Chapman and Hall/CRC, pages 17–42 (2008).

[5] G. Molenberghs, H. Thijs, M. G. Kenward, and G. Verbeke, Sensitivity analysis of continuous incomplete longitudinal outcomes, Statistica Neerlandica, 57(1):112–135 (2003).

[6] G. Molenberghs and G. Verbeke, Missing data concepts, Models for Discrete Longitudinal Data, pages 481–488 (2005). [7] J. Xu and S. L. Zeger. Joint analysis of longitudinal data comprising repeated measures and times to events, Journal of the Royal Statistical Society: Series C (Applied Statistics), 50(3):375–387 (2001).

[8] R. Henderson, P. Diggle and A. Dobson, Joint modelling of longitudinal measurements and event time data, Biostatistics, 1(4):465–480 (2000).

[9] G. Verbeke, G. Molenberghs and D. Rizopoulos, Random effects models for longitudinal data. In Longitudinal research with latent variables, Springer, pages 37–96, (2010).

[10] J. G. Ibrahim, M. H. Chen and D. Sinha, Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials, Statistica Sinica, pages 863–883, (2004).

[11] D. Rizopoulos and P. Ghosh. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event, Statistics in medicine, 30(12):1366–1380, (2011).

[12] E. R Andrinopoulou, D. Rizopoulos, J. J Takkenberg and E. Lesaffre, Joint modeling of two longitudinal outcomes and competing risk data, Statistics in medicine, 33(18):3167–3178, (2014).

[13] D. Rizopoulos, The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. arXiv preprint arXiv:1404.7625 (2014).

[14] G. L. Hickey, P. Philipson, A. Jorgensen and R. Kolamunnage-Dona, joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. BMC medical research methodology, 18(1), 1-14 (2018).

[15] D. Rizopoulos, JM: An R package for the joint modelling of longitudinal and time-to-event data. Journal of Statistical Software (Online), 35(9):1–33 (2010).

[16] H. Li, Y. Zhan, R. J. Carroll, S. K. Keadle, J. N. Sampson and C. E. Matthews. A joint modeling and estimation method for multivariate longitudinal data with mixed types of responses to analyze physical activity data generated by accelerometers, Statistics in medicine, 36(25):4028–4040 (2017).

[17] S. Fieuws, G. Verbeke, B. Maes and Y. Vanrenterghem. Predicting renal graft failure using multivariate longitudinal profiles. Biostatistics, 9(3):419–431 (2007).

[18] P. Bycott and J. Taylor, A comparison of smoothing techniques for cd4 data measured with error in a time dependent cox proportional hazards model, Statistics in medicine, 17(18):2061–2077 (1998).

[19] S. Self and Y. Pawitan, Modeling a marker of disease progression and onset of disease. In AIDS epidemiology, Springer, pages 231–255 (1992).

[20] A. A Tsiatis, V. Degruttola and M. S. Wulfsohn, Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and cd4 counts in patients with aids, Journal of the American statistical association, 90(429):27–37 (1995).

[21] U. G. Dafni and A. A. Tsiatis, Evaluating surrogate markers of clinical outcome when measured with error, Biometrics, pages 1445–1462 (1998).

[22] P. S. Albert and J. H. Shih, On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. Biometrics, 66(3):983–987 (2010).

[23] M. J. Sweeting and S. G. Thompson, Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture, Biometrical Journal, 53(5):750–763 (2011).

[24] Q. Ye and L. Wu, Two-step and likelihood methods for joint models of longitudinal and survival data, Communications in Statistics-Simulation and Computation, 46(8):6019–6033 (2017).

[25] A. Sayers, J. Heron, A. D. Smith, C. Macdonald-Wallis, M. Gilthorpe, F. Steele and K. Tilling, Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: a simulation study of childhood growth and bp, Statistical methods in medical research, 26(1):437–452 (2017).

[26] C. Donnelly, L. M. McFetridge, A. H. Marshall and H. J. Mitchell, A two-stage approach to the joint analysis of longitudinal and survival data utilising the coxian phase-type distribution, Statistical methods in medical research, 27(12):3577–3594 (2018).

[27] P. T. T Huong, D. Nur, H. Pham and A. Branford, A modified two-stage approach for joint modelling of longitudinal and time-to-event data, Journal of Statistical Computation and Simulation, 88(17):3379–3398 (2018).

[28] G. M. Fitzmaurice and N. M. Laird, Regression models for a bivariate discrete and continuous outcome with clustering, Journal of the American statistical Association, 90(431):845–852 (1995).

[29] R. V. Gueorguieva and A. Agresti, A correlated probit model for joint modeling of clustered binary and continuous responses, Journal of the American Statistical Association, 96(455):1102–1112 (2001).

[30] J. Pinheiro and D. Bates, Mixed-effects models in S and S-PLUS. Springer Science & Business Media (2006).

[31] J. Pinheiro, D. Bates, S. DebRoy and D. Sarkar, R core team, Nlme: Linear and nonlinear mixed effects models, R Package version, 3(57), 1-89 (2007).

[32] D. I. Abrams, A. I. Goldman, C. Launer, J. A. Korvick, J. D. Neaton, L. R. Crane, M. Grodesky, S. Wakefield, K. Muth, S. Kornegay, A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection, New England Journal of Medicine, 330(10):657–662 (1994).