Chapter 6

Penalized Spline Joint Models for Longitudinal and Time-To-Event Data

Huong Thi Thu Pham and Hoa Pham

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75975

Abstract

The joint models for longitudinal data and time-to-event data have recently received numerous attention in clinical and epidemiologic studies. Our interest is in modeling the relationship between event time outcomes and internal time-dependent covariates. In practice, the longitudinal responses often show non-linear and fluctuated curves. Therefore, the main aim of this chapter is to use penalized splines with a truncated polynomial basis to parameterize the non-linear longitudinal process. Then, the linear mixed effects model is applied to subject-specific curves and to control the smoothing. The association between the dropout process and longitudinal outcomes is modeled through a proportional hazard model. Two types of baseline risk functions are considered, namely a Gompertz distribution and a piecewise constant model. The resulting models are referred to as penalized spline joint models; an extension of the standard linear joint models.

Keywords: survival data, longitudinal data, joint models, time-dependent covariates, random effects

1. Introduction

The joint models for longitudinal data and time-to-event data are aimed to measure the association between the longitudinal marker level and the hazard rate for an event. The longitudinal data are collected repeatedly for several subjects. In this data, there are two types of covariates, namely, time-independent covariates and time-dependent covariates. Furthermore, there are also two different categories of time-dependent covariates, namely, external and internal covariates. In clinical studies, internal time-dependent longitudinal outcomes are often applied to monitor disease progression and failure time.

In modern survival analysis, Cox [1] has been considered as a very popular joint model to be used for time-independent covariates. These models measured the effect of time-independent covariates on the hazard rate for an event. Subsequently, the extended Cox model was developed for external time-dependent covariates. However, these latter models cannot handle longitudinal biomarkers. Therefore, Rizopoulos [2] introduced joint models for internal time-dependent covariates and the risk for an event based on linear mixed-effects models and relative risk models.

The basic assumption for the standard joint models proposed by Rizopoulos [2] is that the hazard rate at a given time of the dropout process is associated with the expected value of the longitudinal responses at the same time. The whole history of response has an influence on the survival function. Thus, it is crucial to obtain good estimates for the subject-specific trajectories in order to have an accurate estimation of the survival function. In addition, an important feature that we need to account for is that many observations in the sample often show non-linear and fluctuated longitudinal trajectories in time. Each observation has its own trajectory. Therefore, flexibility is needed for subject-specific longitudinal submodels in the joint models to improve the predictions.

There are several previous works to flexibly model the subject-specific longitudinal profiles in the joint models. Brown et al. [3] applied B-splines with multidimensional random effects. In particular, Brown et al. [3] assumed that both subject and population trajectories have the same number of basis functions. By doing this, the number of parameters in the longitudinal submodel is reasonably large. If we have to deal with the roughness of the fit for this model, the computational problems will increase especially when the dimension of the random effects vector is large. Ding and Wang [4] proposed the use of B-splines with a single multiplicative random effect to link the population mean function with the subject-specific profile. This simple model can gain an easy estimation for parameters, however may not be appropriate for many practical applications [5]. Rizopoulos [5] considered more flexible models using natural cubic splines with the expansion of the random effects vector. The roughness of the fit is still not mentioned in these models.

In this chapter, we present new approaches to model non-linear shapes of subjects-specific evolutions for joint models by extending the standard joint models of Rizopoulos [2]. In particular, we implement penalized splines using a truncated polynomial basis for the longitudinal submodel. Following this, the linear mixed-effects approach is applied to model the individual trajectories and impose smoothness over adjacent coefficients respectively. The ECM algorithm is used for parameter estimation. In addition, corresponding standard errors are calculated using the observed information matrix. However, as the matrices of random effects covariates in our models are different from the matrices of random effects covariates in the standard joint models, the JM package of Rizopoulos [6] cannot be used for our models. Therefore, a set of R codes are written for the penalized spline joint models to implement the proposed procedures on the simulated data and a case study respectively.

The chapter is organized as follows. Section 2 describes the penalized splines with truncated polynomial basis for the joint models. In this section, the two models are specified as penalized spline joint model with hazard rate at base line having Gompertz distribution (referred to as

http://dx.doi.org/10.5772/intechopen.75975

Model 1) and penalized spline joint model with a piecewise-constant baseline risk function (referred to as Model 2). The joint likelihood, score functions and the ECM algorithm for estimation are presented in Section 3. We then validate the proposed algorithm using extensive simulation studies and then apply it for AIDS data in Section 4. Finally, Section 5 gives concluding remarks.

2. The penalized spline joint models

In this section, we introduce the joint models using penalized spline with truncated polynomial basis. The proposed parametrization is based on the standard joint models of Rizopoulos [2] and the regression model of a longitudinal response using penalized spline.

Notations in this section are taken from Rizopoulos [2]. Let T_i^* be the true survival time and C_i be the censoring time for the i^{th} subject $(i = 1, \ldots, n)$. T_i denotes the observed failure time for the i^{th} subject $(i = 1, \ldots, n)$, which is defined as $T_i = min\left(T_i^*, C_i\right)$. If an i^{th} subject is not censored, this means that we have observed its survival time, we will have $T_i \leq C_i$. If an i^{th} subject is censored, this means that we lose its follow up, or the subject has died from other causes, we will have $T_i > C_i$. Furthermore, we define the event indicator as $\delta_i = I\left(T_i^* \leq C_i\right)$. The observed data for survival outcome are (T_i, δ_i), $i = 1, \ldots, n$.

For a longitudinal response, suppose that we have n subjects in the sample and the actual observed longitudinal data for each subject-i at time point t is $y_i(t)$. We measure the i^{th} subject at n_i time points. Thus, the longitudinal data consists of the measurements $y_{ij} = \left\{y_i\left(t_{ij}\right), j = 1, \ldots, n_i\right\}$ taken at time points t_{ij}. We denote the true and unobserved value of the longitudinal outcome at time t as $m_i(t)$. We assume the relation between $y_i(t)$ and $m_i(t)$ as

$$y_i(t) = m_i(t) + \varepsilon_i(t), \tag{1}$$

where $\varepsilon_i(t) \sim N\left(0, \sigma_\varepsilon^2\right)$.

When survival function $S(t)$ is assumed to have a specific parametric form associating with a longitudinal submodel, estimations for parameters of interest are usually based on the likelihood function [2]. In the maximum likelihood method, there are different treatments for different types of covariates in the longitudinal submodel. Here, we present the two different categories of time-dependent covariates and the estimation techniques for these covariates will be introduced in the following sections. We let the time-dependent covariate for the i^{th} subject at time t be denoted by $y_i(t)$. We let $\mathcal{Y}_i(t) = \left\{y_i(s), 0 \leq s < t\right\}$ denote the covariate history of the i^{th} subject up to time t. According to Kalbfleisch and Prentice [7], the exogenous covariates are the covariates satisfying the condition:

$$\Pr(s \leq T_i < s + ds | T_i \geq s, \mathcal{Y}_i(s)) = \Pr(s \leq T_i < s + ds | T_i \geq s, \mathcal{Y}_i(t)), \tag{2}$$

for all s, t such that $0 < s \leq t$ and $ds \to 0$. An equivalent definition is

$$\Pr(\mathcal{Y}_i(t)|\mathcal{Y}_i(s), T_i \geq s) = \Pr(\mathcal{Y}_i(t)|\mathcal{Y}_i(s), T_i = s), s \leq t. \tag{3}$$

On the other hand, endogenous time-varying covariates are the ones that do not satisfy the condition in (2.2). In particular,

$$\Pr(\mathcal{Y}_i(t)|\mathcal{Y}_i(s), T_i \geq s) \neq \Pr(\mathcal{Y}_i(t)|\mathcal{Y}_i(s), T_i = s), s \leq t.$$

In the penalized spline regression models [8, 9], the observed longitudinal covariate is modeled using the truncated power functions with a general power basis of degree p. Moreover, the longitudinal response is also parameterized as a linear mixed-effects model to specify the individual curves and impose the amount of smoothing. As a result, the coefficients of the knots will be constrained to handle smoothing. In particular, the longitudinal submodel for the i^{th} subject at time point t_{ij} is

$$\begin{aligned} y_{ij} &= f(t_{ij}) + g_i(t_{ij}) + \varepsilon(t_{ij}), \varepsilon_i(t_{ij}) \sim N(0, \sigma_\varepsilon^2), \\ f(t_{ij}) &= \beta_0 + \beta_1 t_{ij} + \ldots + \beta_p t_{ij}^p + \sum_{k=1}^{K} u_{pk}(t_{ij} - \mathcal{K}_k)_+^p, \\ g_i(t_{ij}) &= v_{i0} + v_{i1} t_{ij} + v_{i2} t_{ij}^2 + \ldots + v_{ip} t_{ij}^p + \sum_{k=1}^{K} w_{ipk}(t_{ij} - \mathcal{K}_k)_+^p. \end{aligned} \tag{4}$$

Here, the set $\left\{1, t_{ij}, \ldots, t_{ij}^p, (t_{ij} - \mathcal{K}_1)_+^p, \ldots, (t_{ij} - \mathcal{K}_K)_+^p\right\}$ is known as the truncated power basis of degree p. Moreover, $\mathcal{K}_1, \ldots, \mathcal{K}_K$ are fitted K knots, for which K is chosen following Ruppert et al. [9], Chapter 5), Appendix D. The function $f(.)$ is the smooth function which reflects the overall trend of the population. The function $g_i(.)$ is the smooth function which reflects the individual curves. To constrain the coefficient of knots, the vector $(u_{p1}, \ldots, u_{pK})^T$ in the function $f(.)$ is treated as random effects. Therefore, $\beta^T = \left(\beta_0, \ldots, \beta_p\right)$ is a $((p+1) \times 1)$ row vector of fixed effects and $\boldsymbol{b}_i^T = (u_{p1}, \ldots, u_{pK}, v_{i0}, \ldots, v_{ip}, w_{ip1}, \ldots, w_{ipK})$ is a $((p + 2K + 1) \times 1)$ vector of random effects for the i^{th} subject. The assumptions for the random effects for the i^{th} subject are $(v_{i0}, \ldots, v_{ip})^T \sim N(0, \sum)$, $u_{pk} \sim N(0, \sigma_u^2)$, $w_{ipk} \sim N(0, \sigma_w^2)$ and they are independent of one another. We can now rewrite (2.4) as

$$\begin{aligned} y_i(t_{ij}) &= f(t_{ij}) + g_i(t_{ij}) + \varepsilon_i(t_{ij}) \\ &= \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \ldots + \beta_p t_{ij}^p + \sum_{k=1}^{K} (u_{pk} + w_{ipk})(t_{ij} - \mathcal{K}_k)_+^p \\ &\quad + v_{i0} + v_{i1} t_{ij} + v_{i2} t_{ij}^2 + \ldots + v_{ip} t_{ij}^p + \varepsilon_i(t_{ij}). \end{aligned} \tag{5}$$

We let $u_{ipk} = u_{pk} + w_{ipk}$ and note that $u_{ipk} \sim N(0, \sigma_u^2 + \sigma_w^2)$. In order to allow greater flexibility, we assume that $(u_{ip1}, \ldots, u_{ipK})^T \sim N(0, \boldsymbol{D})$, where $\boldsymbol{D} = Diag(D_{11}, \ldots, D_{KK})$. By doing this, the dimension of the vector of random effects, $\boldsymbol{b}_i^T = (v_{i0}, \ldots, v_{ip}, u_{ip1}, \ldots, u_{ipK})$, decreases to $((p + K + 1) \times 1)$. Consequently, the dimension of the multi-integrals in the log-likelihood

function in (3.2) will also decrease. This presentation is crucial for reducing the computational problems while coding. The model in (2.5) now becomes:

$$\begin{aligned} y_i(t_{ij}) &= f(t_{ij}) + g_i(t_{ij}) + \varepsilon_i(t_{ij}) \\ &= \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \ldots + \beta_p t_{ij}^p + \sum_{k=1}^{K} u_{ipk}(t_{ij} - \mathcal{K}_k)_+^p \\ &\quad + v_{i0} + v_{i1} t_{ij} + v_{i2} t_{ij}^2 + \ldots + v_{ip} t_{ij}^p + \varepsilon_i(t_{ij}). \end{aligned} \tag{6}$$

The model in (2.6) can be rewritten in matrix notation as:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{b} + \varepsilon, \tag{7}$$

where

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_n \end{bmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{X}_1 & 0 & \ldots & 0 & \boldsymbol{Z}_1 & 0 & \ldots & 0 \\ 0 & \boldsymbol{X}_2 & \ldots & 0 & 0 & \boldsymbol{Z}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{X}_n & 0 & 0 & \ldots & \boldsymbol{Z}_n \end{bmatrix},$$

$$\boldsymbol{X}_i = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 & \cdots & t_{i1}^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 & \cdots & t_{in_i}^p \end{bmatrix}, \quad \boldsymbol{Z}_i = \begin{bmatrix} (t_{i1} - \mathcal{K}_1)_+^p & \cdots & (t_{i1} - \mathcal{K}_K)_+^p \\ \vdots & \vdots & \vdots \\ (t_{in_i} - \mathcal{K}_1)_+^p & \cdots & (t_{in_i} - \mathcal{K}_K)_+^p \end{bmatrix},$$

$$\boldsymbol{b}^T = (v_{10}, \ldots, v_{1p}, \ldots, v_{n0}, \ldots, v_{np}, u_{1p1}, \ldots, u_{1pK}, \ldots, u_{np1}, \ldots, u_{npK}),$$

$$\beta^T = (\beta_0, \ldots, \beta_p).$$

Here, $\boldsymbol{y}$ is the $\left(\sum_{i=1}^{n} n_i \times 1\right)$ matrix of observed longitudinal data; $\boldsymbol{X}$ is the $\left(\sum_{i=1}^{n} n_i \times (p+1)\right)$ matrix of fixed effect covariates; $\boldsymbol{Z}$ is the $\left(\sum_{i=1}^{n} n_i \times (p+K+1)n\right)$ matrix of random effect covariates and ε is the $\left(\sum_{i=1}^{n} n_i \times 1\right)$ matrix of error.

Postulating a proportion hazard model, the penalized spline joint models for longitudinal and time-to-event data is defined by

$$\begin{aligned} h_i(t|\mathcal{M}_i(t), \boldsymbol{w}_i) &= \lim_{dt \to 0} \Pr\{t \le T_i^* < t + dt | T_i^* \ge t, \mathcal{M}_i(t), \boldsymbol{w}_i\}/dt \\ &= h_0(t) \exp\{\gamma^T \boldsymbol{w}_i + \alpha m_i(t)\}, \end{aligned} \tag{8}$$

where $h_0(t)$ is the hazard at baseline and $\boldsymbol{w}_i$ is a vector of baseline covariates (such as treatment indicator, gender of a patient, etc.). Furthermore, $\mathcal{M}_i(t) = \{m_i(s), \ \ 0 \le s < t\}$ denotes the history of the true unobserved longitudinal process up to time point t.

Using (2.7), the longitudinal submodel for the i^{th} subject is given by

$$\begin{cases} m_i(t) = m_i(t) + \varepsilon_i(t), \varepsilon_i(t) \sim N\left(0, \sigma_\varepsilon^2\right) \\ y_i(t) = \boldsymbol{X}_i^T(t)\beta + \boldsymbol{X}_i^T(t)\boldsymbol{v}_i + \boldsymbol{Z}_i^T(t)\boldsymbol{u}_i + \varepsilon_i(t) \\ \boldsymbol{v}_i \sim N\left(\mathbf{0}, \sum\right), \boldsymbol{u}_i \sim N(\mathbf{0}, \boldsymbol{D}), \end{cases} \tag{9}$$

where the covariance matrix of random effects $\boldsymbol{b}_i^T = \left(v_{i0}, \ldots, v_{ip}, u_{ip1}, \ldots, u_{ipK}\right)$ is given as

$$\boldsymbol{G} = Cov(\boldsymbol{b}_i) = \begin{bmatrix} \sum & 0 \\ 0 & \boldsymbol{D} \end{bmatrix}.$$

To complete the specification of the model in (2.8), we now need to define the form for the baseline risk function $h_0(.)$. Motivated by the fact that in real life, $h_0(.)$ is usually unknown. Therefore, two options are adopted to determine the form of the function $h_0(.)$ in this chapter. First, a standard option is to use a known parametric distribution for the risk function. For this option, the Gompertz distribution is chosen. Second, the piecewise constant model is chosen when $h_0(.)$ is considered completely unspecified.

Therefore, the proposed penalized spline joint models considered in this chapter are as follows:

Model 1: Penalized spline joint model with hazard rate at base line having Gompertz distribution

$$\begin{cases} h_i(t|\mathcal{M}_i(t), \boldsymbol{w}_i) = \lambda_1 \exp(\lambda_2 t) \exp\left\{\gamma^T \boldsymbol{w}_i + \alpha m_i(t)\right\} \\ m_i(t) = \boldsymbol{X}_i^T(t)\beta + \boldsymbol{X}_i^T(t)\boldsymbol{v}_i + \boldsymbol{Z}_i^T(t)\boldsymbol{u}_i. \end{cases} \tag{10}$$

Model 2: Penalized spline joint model with a piecewise-constant baseline risk function

$$\begin{cases} h_i(t|\mathcal{M}_i(t), \boldsymbol{w}_i) = \sum_{q=1}^{Q} \xi_q I\left(\nu_{q-1} < t \leq \nu_q\right) \exp\left\{\boldsymbol{\gamma}^T \boldsymbol{w}_i + \alpha m_i(t)\right\} \\ m_i(t) = \boldsymbol{X}_i^T(t)\beta + \boldsymbol{X}_i^T(t)\mathbf{v}_i + \boldsymbol{Z}_i^T(t)\boldsymbol{u}_i, \end{cases} \tag{11}$$

where $0 = \nu_0 < \nu_1 < \ldots < \nu_Q$ denotes a split of the time scale, with ν_Q being larger than the largest observed time and ξ_q denotes the value of the baseline hazard in the interval $\left[\nu_{q-1}, \nu_q\right)$. In both models, $\boldsymbol{X}_i$, $\boldsymbol{Z}_i$, β, $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$ are given in (2.7).

3. Parameter estimation

After defining the two penalized spline joint models, we now present the joint likelihood and score functions of the parameters in the models. The ECM algorithm is also presented in this section.

http://dx.doi.org/10.5772/intechopen.75975

3.1. Likelihood and score functions

Following Rizopoulos [2], we assume that the vector of time-independent random effects b_i underlies both the longitudinal and survival processes. This means that

$$\begin{aligned} p\left(T_i, \delta_i, \boldsymbol{y_i} | \boldsymbol{b_i}; \boldsymbol{\theta}\right) &= p(T_i, \delta_i | \boldsymbol{b_i}; \boldsymbol{\theta}) p\left(\boldsymbol{y_i} | \boldsymbol{b_i}; \boldsymbol{\theta}\right) \\ p\left(\boldsymbol{y_i} | \boldsymbol{b_i}; \boldsymbol{\theta}\right) &= \prod_j p\left\{\boldsymbol{y_i}\left(t_{ij}\right) | \boldsymbol{b_i}; \boldsymbol{\theta}\right\}, \end{aligned} \tag{12}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_t^T, \boldsymbol{\theta}_y^T, \boldsymbol{\theta}_b^T\right)^T$ denotes the full parameter vector with $\boldsymbol{\theta}_t = \left(\gamma^T, \alpha, \theta_{h_0}^T\right)^T$ denoting the parameter vector for the survival outcomes. Furthermore, $\boldsymbol{\theta}_y = \left(\beta^T, \sigma_\varepsilon^2\right)^T$ is the parameter vector for longitudinal outcomes and $\boldsymbol{\theta}_b = vech(\boldsymbol{G})$ is the vector-half of the variance matrix of random effects. In addition, we assume that the hazard rate at time t conditional on the covariate path depends on the current value of longitudinal outcomes and the censoring mechanism is independent of the true event times and future longitudinal measurements. Under these assumptions, the log-likelihood formulation of the penalized spline joint models can be written as

$$\begin{aligned} l(\boldsymbol{\theta}) &= l\left(\boldsymbol{\theta} | T_i, \delta_i, \mathbf{y}_i\right) \\ &= \sum_i \log \int_{\mathbf{b}_i} p\left(T_i, \delta_i | \boldsymbol{b}_i; \boldsymbol{\theta}_t, \beta\right) p\left(\boldsymbol{y}_i | \boldsymbol{b}_i; \boldsymbol{\theta}_y\right) p\left(\boldsymbol{b}_i; \boldsymbol{\theta}_b\right) d\boldsymbol{b}_i, \end{aligned} \tag{13}$$

where the conditional density for survival part has the form of

$$\begin{aligned} p\left(T_i, \delta_i | \boldsymbol{b}_i; \boldsymbol{\theta}_t, \beta\right) &= h\left(T_i | \mathcal{M}_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t, \beta\right)^{\delta_i} S\left(T_i | \mathcal{M}_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t, \beta\right) \\ &= \left[h_0(t) \exp\left\{\gamma^T \boldsymbol{w}_i + \alpha m_i(t)\right\}\right]^{\delta_i} \exp\left[-\int_0^{T_i} h_0(s) \exp\left\{\gamma^T \boldsymbol{w}_i + \alpha m_i(s) ds\right\}\right]. \end{aligned} \tag{14}$$

Here, $S(t)$ is the survival function at time t.

Moreover, the density for the longitudinal part with the random effects is given by

$$\begin{aligned} p\left(\boldsymbol{y_i} | \boldsymbol{b}_i; \boldsymbol{\theta}_y\right) p\left(\boldsymbol{b_i}; \boldsymbol{\theta}_b\right) &= \prod_j p\left\{y_i\left(t_{ij}\right) | \boldsymbol{b_i}; \boldsymbol{\theta}_y\right\} p\left(\boldsymbol{b_i}; \boldsymbol{\theta}_b\right) \\ &= \frac{1}{\left(2\pi\sigma_\varepsilon^2\right)^{\frac{n_i}{2}}} \exp\left\{-\frac{\left\|y_i\left(t_{ij}\right) - \boldsymbol{X}_i^T\left(t_{ij}\right)\beta - \boldsymbol{X}_i^T\left(t_{ij}\right)\boldsymbol{v}_i - \boldsymbol{Z}_i^T\left(t_{ij}\right)\boldsymbol{u}_i\right\|^2}{2\sigma_\varepsilon^2}\right\} \\ &\times (2\pi)^{-\frac{q_b}{2}} \det(\boldsymbol{G})^{-1/2} \exp\left(-\boldsymbol{b}_i^T \boldsymbol{G}^{-1} \boldsymbol{b_i}/2\right), \end{aligned} \tag{15}$$

where q_b denotes the dimensionality of the random effects vector.

We consider the log likelihood of the (T_i, δ_i, $\boldsymbol{y}_i$, $\boldsymbol{b}_i$) over the unknown $\boldsymbol{\theta}_t$, β and $\boldsymbol{b_i}$

$$1 \log l\left(\theta|T_i, \delta_i, \mathbf{y}_i, \mathbf{b_i}\right) = \log p\left(T_i, \delta_i|\mathbf{b}_i; \theta_t, \beta\right) + \log p\left(\mathbf{y}_i|\mathbf{b}_i; \beta\right) + \log p(\mathbf{b}_i; \mathbf{G}).$$

The function for maximizing the log likelihood involves the density function of survival time and least squares with a penalty term, which is

$$\log p\left(T_i, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta}_t, \beta\right) - \frac{\left(\boldsymbol{y}_i - \boldsymbol{X}_i\beta - \boldsymbol{X}_i\boldsymbol{v}_i - \boldsymbol{Z}_i\boldsymbol{u}_i\right)^T\left(\boldsymbol{y}_i - \boldsymbol{X}_i\beta - \boldsymbol{X}_i\boldsymbol{v}_i - \boldsymbol{Z}_i\boldsymbol{u}_i\right)}{\sigma_\varepsilon^2} - \boldsymbol{b}_i^T\boldsymbol{G}^{-1}\boldsymbol{b}_\mathbf{i}. \quad (16)$$

According to Ruppert et al. [9], the term $\sigma_\varepsilon^2\boldsymbol{b}_i^T\boldsymbol{G}^{-1}\boldsymbol{b}_\mathbf{i}$ is called a roughness penalty and the variance components matrix defined as $\boldsymbol{F} = \sigma_\varepsilon^2\boldsymbol{G}^{-1}$. Using a Lagrange multiplier argument, the variance components matrix is the condition to constrain the coefficients of the knots $\boldsymbol{u}_i$. These will restrict the influence of the variables $(t - K_k)_+^p$ and will lead to smoother spline functions.

Using (3.2), the score vector for the penalized spline joint models can be expressed as:

$$\begin{aligned} S(\boldsymbol{\theta}) &= \sum_i \frac{\partial}{\partial\boldsymbol{\theta}^T} \log \int p\left(T_i, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta}_t, \beta\right) p\left(\boldsymbol{y}_i|\boldsymbol{b}_i; \boldsymbol{\theta}_y\right) p(\boldsymbol{b}_i; \boldsymbol{\theta}_b) d\boldsymbol{b}_i \\ &= \sum_i \int \frac{\partial}{\partial\boldsymbol{\theta}^T} \log \left\{p\left(T_i, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta}_t, \beta\right) p(\boldsymbol{y}_i|\boldsymbol{b}_i; \boldsymbol{\theta}_y) p(\boldsymbol{b}_i; \boldsymbol{\theta}_b)\right\}. p\left(\boldsymbol{b}_i|T_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}\right) d\boldsymbol{b}_i. \end{aligned} \quad (17)$$

The requirement for numerical integration with respect to the random effects is one of the main difficulties in the joint models [2]. The maximum likelihood estimation (MLE) is typically obtained using standard maximization algorithms such as expectation maximization algorithm or Newton-Raphson algorithm.

3.2. The ECM algorithm

The EM algorithm has been widely used in the joint models, such as for the standard joint model of Rizopoulos [2] and for the generalized linear mixed joint model [10]. The ECM algorithm is a natural extension of EM algorithm for which the maximization process on the M-step is conditional on some functions of the parameters under estimation. It also can reduce computer time. The ECM algorithm will be used to obtain the maximum likelihood estimates of the penalized spline joint models following McLachlan and Krishnan [11] in this chapter.

In these models, the random effects $\boldsymbol{b}_i$ are considered as missing data. Hence, it is difficult to estimate directly the parameter vector $\boldsymbol{\theta}$ that maximizes the observed data log likelihood $l(\boldsymbol{\theta})$ in (3.2). Alternatively, we can estimate the parameter vector $\boldsymbol{\theta}$ that maximizes the expected value of the complete data log-likelihood which is $E\Big\{\log p\left(T_i, \delta_i, \boldsymbol{y}_\mathbf{i}, \boldsymbol{b}_i; \boldsymbol{\theta}\right)|T_i, \delta_i, \boldsymbol{y}_\mathbf{i}; \boldsymbol{\theta}^{(it)}\Big\}$, where $\theta^{(it)}$ is the parameter vector given at the i^{th} iteration.

The following are the steps of this algorithm.

Step 1: *Initialization*

We first initialize the parameters. We assume that there are m parameters in the models and the starting value of the parameter vector is $\boldsymbol{\theta}^{(0)} = \left(\theta_1^{(0)}, \ldots, \theta_m^{(0)}\right)$. Based on these initial values, we calculate the log-likelihood using (3.2).

Step 2: *The E-step for the penalized joint models*

We fill in the missing data and replace the log-likelihood function of the observed data with the expected function of the complete data log-likelihood as follows:

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(it)}\right) &= \sum_i \int \log\left\{p\left(T_i, \delta_i, \boldsymbol{y_i}, \boldsymbol{b}_i; \boldsymbol{\theta}\right)\right\}.p\left(\boldsymbol{b}_i|T_i, \delta_i, \boldsymbol{y_i}; \boldsymbol{\theta}^{(it)}\right)d\boldsymbol{b}_i \\ &= \sum_i \int \left(\log p(T_i, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta}) + \log p(\boldsymbol{y_i}|\boldsymbol{b_i}; \boldsymbol{\theta}) + \log p(\boldsymbol{b}_i; \boldsymbol{\theta})\right).p\left(\boldsymbol{b}_i|T_i, \delta_i, \boldsymbol{y_i}; \boldsymbol{\theta}^{(it)}\right)d\boldsymbol{b}_i. \end{aligned} \tag{18}$$

Step 3: *The conditional M-step for the penalized joint models.*

This step will be implemented in four stages as follows:

3.1 Given the current value of the parameter vector at the i^{th} iteration $\boldsymbol{\theta}^{(it)} = \left(\theta_1^{(it)}, \theta_2^{(it)}, \ldots, \theta_m^{(it)}\right)$, we calculate the log likelihood at $l\left(\boldsymbol{\theta}^{(it)}\right) = \sum_i \log \int_{b_i} p\left(T_i, \delta_i, \boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\theta}^{(it)}\right)d\boldsymbol{b}_i$.

3.2 Propose the new value for the first parameter $\theta_1^{(prop)}$ which maximizes $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(it)}\right)$. Then, we calculate the log likelihood at $l\left(\boldsymbol{\theta}^{(prop)}\right)$ where $\boldsymbol{\theta}^{(prop)} = \left(\theta_1^{(prop)}, \theta_2^{(it)}, \ldots, \theta_m^{(it)}\right)$.

3.3 Set $\boldsymbol{\theta}_1^{(it)} = \boldsymbol{\theta}^{(prop)}$ if $l\left(\boldsymbol{\theta}^{(prop)}\right) \geq l\left(\boldsymbol{\theta}^{(it)}\right)$, otherwise set $\boldsymbol{\theta}_1^{(it)} = \boldsymbol{\theta}^{(it)}$.

3.4 Similarly, based on the value of the parameter vector $\boldsymbol{\theta}_1^{(it)}$, we update the new value for the second parameter and continue updating for the last parameter, $\boldsymbol{\theta}_m^{(it)}$ and set $\boldsymbol{\theta}^{(it+1)} = \boldsymbol{\theta}_m^{(it)}$.

Step 4: *Iterate among steps 2–3 until the algorithm numerically converges.*

To update the new values for parameters in the conditional M-step, we have the closed-form estimates for the measurement of error variance σ^2 and the covariance matrix of the random effects respectively by maximizing the expected function $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(it)}\right)$. Unfortunately, we cannot obtain closed-form expressions for the remaining of the parameters. We thus employ the one-step Newton-Raphson approach to get the updates for $\beta^{(it+1)}$, $\gamma^{(it+1)}$, $\alpha^{(it+1)}$ and $\boldsymbol{\theta}_{h_0}^{(it+1)}$ respectively as detailed in Appendix B.

Following Louis [12], standard errors for the parameter estimates can be calculated by using the estimated observed information matrix

$$\widehat{var}\left(\widehat{\boldsymbol{\theta}}\right) = \left\{\mathcal{I}\left(\widehat{\boldsymbol{\theta}}\right)\right\}^{-1},$$

where

$$\mathcal{I}\left(\widehat{\boldsymbol{\theta}}\right) = -\sum_{i=1}^{n} \frac{\partial S_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}.$$

4. Empirical results

This section presents two simulation studies for Model 1, whereas Model 2 will be applied for a case study only. In Section 4.1, we simulate data from Model 1 with three internal knots in the longitudinal submodel and Gompertz distribution for the baseline risk function. In Section 4.2, we simulate data from Model 1 having Gompertz distribution for the baseline risk function and non-linear logarithm subject-specific trajectories. The ECM algorithm, written in R code, is applied to estimate the true values of parameters in both cases.

4.1. Simulation study 1

4.1.1. Data description

Recall the penalized spline joint Model 1 of (2.10) with three internal knots in longitudinal submodel and Gompertz distribution for the baseline risk function in the form of

$$h_i(t) = h_0(t)\exp\left(\gamma x_i + \alpha(m_i(t))\right) = \lambda_1 \exp\left(\lambda_2 t\right)\exp\left\{\gamma x_i + \alpha m_i(t)\right\}, \tag{19}$$

where $h_0(t)$ is the hazard function at baseline having Gompertz distribution, x_i is baseline covariate and $m_i(t)$ denotes the true and unobserved value of the longitudinal at time t. The form of $m_i(t)$ is given by

$$m_i(t) = \beta_0 + \beta_1 t + u_{i1}(t-\mathcal{K}_1)_+ + u_{i2}(t-\mathcal{K}_2)_+ + u_{i3}(t-\mathcal{K}_3)_+ + v_{i0}, \tag{20}$$

where $\boldsymbol{b}_i = (u_{11}, u_{12}, u_{13}, v_{i0})^T$ is the vector of random effects and is assumed to have a normal distribution with mean zero and the diagonal covariance matrix $\boldsymbol{D} = Diag(D_{11}, D_{22}, D_{33}, D_{44})$. $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3$ denote the three internal knots put into the model. The observed longitudinal value at time point t for the i^{th} subject is of the form

$$y_i(t) = m_i(t) + \varepsilon_i(t), \tag{21}$$

where the error variable $\varepsilon_i(t)$ is assumed to come from a normal distribution with mean zero and variance σ^2.

From this model, the vector of all parameters $\boldsymbol{\theta}$ of the models in (4.1) and (4.2) is $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_t^T, \boldsymbol{\theta}_y^T, \boldsymbol{\theta}_b^T\right)^T$, where $\boldsymbol{\theta}_t = (\gamma, \alpha, \lambda_1, \lambda_2)^T$ denotes the parameter vector for the survival

outcomes. Furthermore, $\boldsymbol{\theta}_y = \left(\beta_0, \beta_1, \sigma_\varepsilon^2\right)^T$ is the parameter vector for longitudinal outcomes and $\boldsymbol{\theta}_b = \boldsymbol{D}$ is the variance matrix of random effects.

To simulate the observed survival time T_i of the joint model in (4.1), we applied the methods adapted by Bender et al. [13], Austin [14] and Crowther and Lambert [15] to generate the true survival time. We further assumed that the censoring mechanism is exponentially distributed with parameter λ. The observed survival time was the minimum of the censoring time and the true survival time. We generated the survival time T_i for $n = 500$ subjects with the parameters: $\beta_0 = 5,\ \beta_1 = 2,\ \lambda_1 = 0.1,\ \lambda_2 = 0.5,\ \gamma = 0.5, \alpha = 0.05, \delta = 2$ and $D = Diag(2, 2, 2, 4)$. Then we generated the longitudinal responses $m_i(t)$ using (4.2). The simulated model is therefore

$$\begin{cases} h_i(t) = 0.1 \exp(0.5t) \exp\{0.5x_i + 0.05m_i(t)\} \\ m_i(t) = 5 + 2t + u_{i1}(t-1)_+ + u_{i2}(t-2)_+ + u_{i3}(t-3)_+ + v_{i0}. \end{cases} \tag{22}$$

The sample of simulated data is presented in Appendix A. The curve of Kaplan-Meier estimate for the survival function of simulated data (left panel) and the longitudinal trajectories for the whole simulated sample (right panel) are presented in **Figure 1**. The dashed lines in the left panel correspond to 95% pointwise confidence intervals. It is clear from the plot of Kaplan-Meier estimator that the survival probability starts from 1 and decreases gradually until at the 5^{th} month of the study. After this, it is nearly zero after 6 months or so. The right panel is the longitudinal trajectories for the first 100 subjects reflecting the form as in (4.2).

4.1.2. Parameter estimation

The ECM algorithm, as described in Section 3.2, is now implemented to estimate all parameters in (4.4). The initial values of the parameters were set at $\beta_0 = 1,\ \beta_1 = 1,\ \lambda_1 = 0.05, \lambda_2 = 0.1,$ $\gamma = 0.1,\ \alpha = 0.01,\ \sigma = 1,\ D_{11} = 3,\ D_{22} = 3,\ D_{33} = 3,\ D_{44} = 3$, respectively. However, these initial values can also be set randomly. The traces of each of these parameters are presented in **Figures 2** and **3**, respectively. The traces of estimates show the way how the algorithm updates

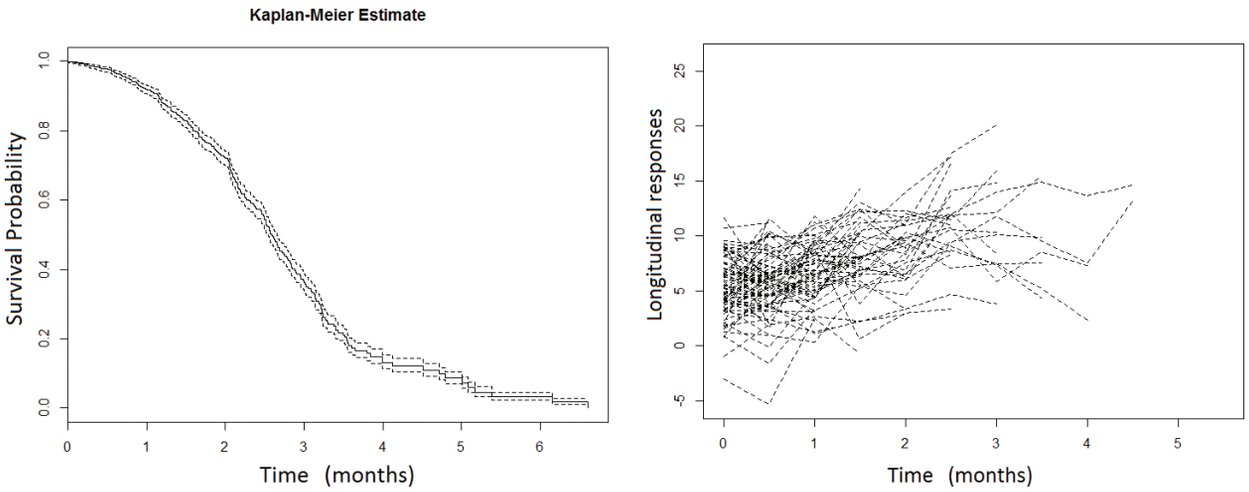

Figure 1. Kaplan-Meier estimate of the survival function of the simulated data of (4.4) (left panel). Longitudinal trajectories of the first 100 subjects from the simulated sample of (4.4) (right panel).

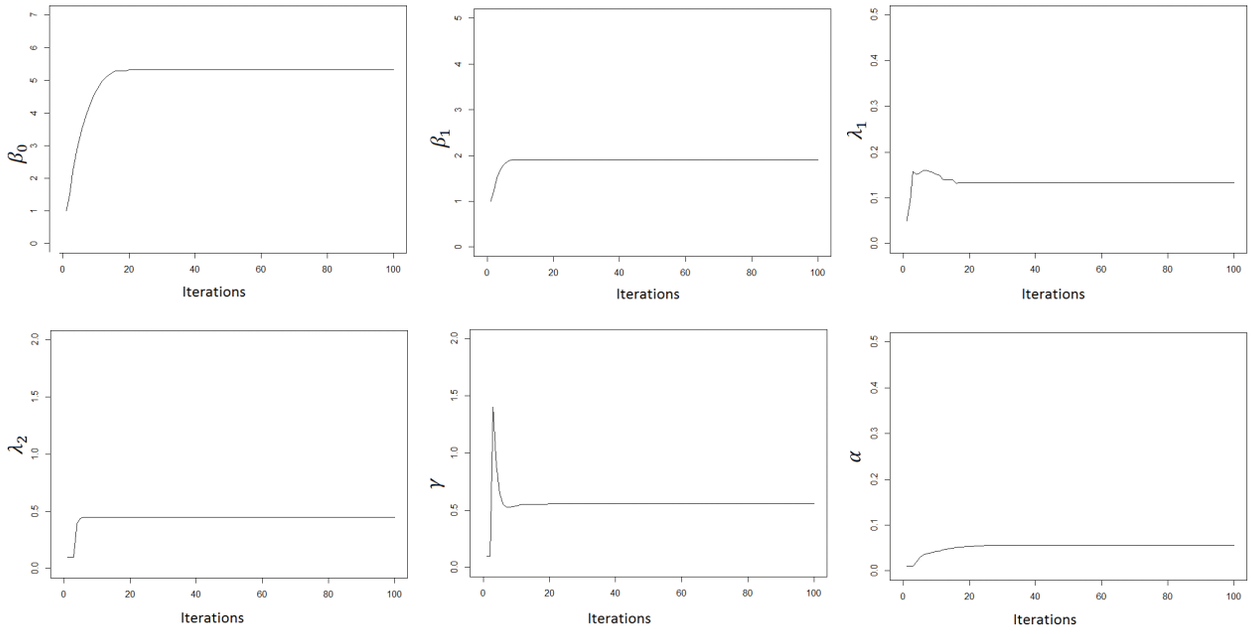

Figure 2. The traces of parameters β_0, β_1, λ_1, λ_2, γ, α for 100 iterations.

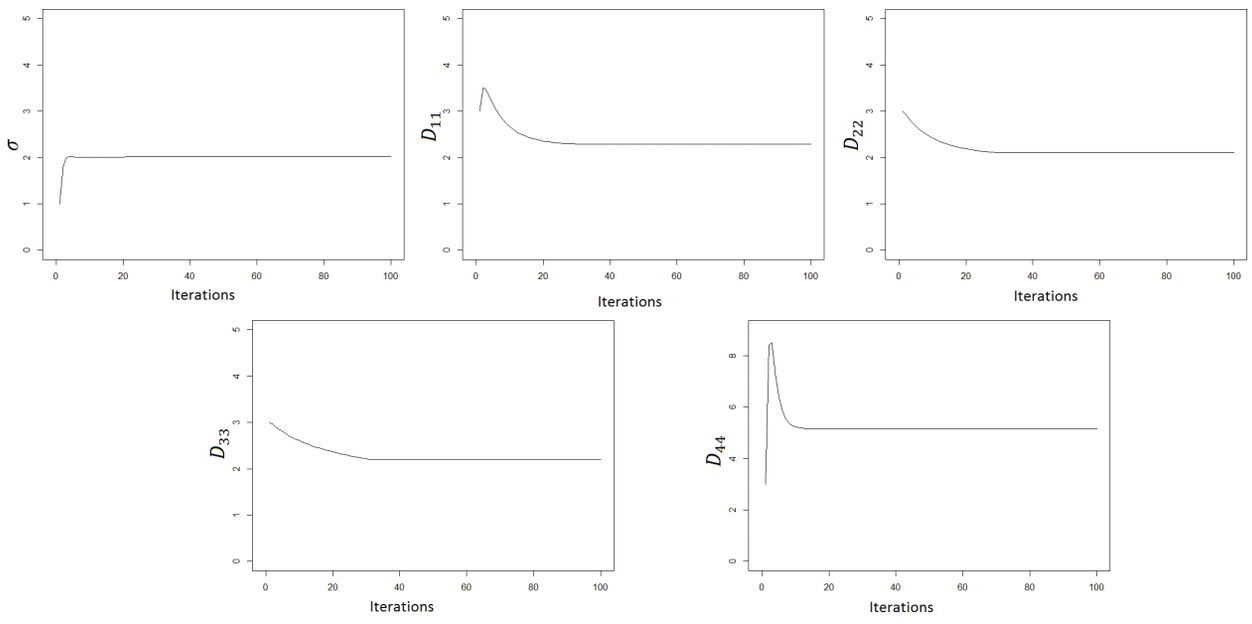

Figure 3. The traces of parameters σ, D_{11}, D_{22}, D_{33}, D_{44} for 100 iterations.

new values of the parameters. In addition, they also demonstrate the convergence of the algorithm after 10–20 iterations. In particular, the parameters β_0, β_1, λ_2, α, σ, D_{22} and D_{33} converge linearly to the true values while the parameters λ_1, γ, D_{11}, and D_{44} oscillate before converging to the true values.

We now run the simulation for 30 independent samples with different sample sizes ($n = 200$, 300 and 500). Then, we calculate the means, standard deviations (SD) and mean square error (MSE) of parameters as presented in **Table 1**. The point estimates of each parameter are reasonably close to the true values when the sample sizes are 300 and 500. This is also supported by the values of SD and MSE which decrease gradually when the sample size increases. In addition to this, we also compare the parameter estimates with different censoring rates (20% and 40%) for a sample size of 500 in 5, Appendix E. The result shows that when the sample size is large the censoring rate has little influence on the estimates.

4.2. Simulation study 2

4.2.1. Data description

We now perform a simulation study on proportional hazard model having Gompertz distribution at baseline and non-linear subject-specific trajectory. In particular, the model is in the form of

$$h_i(t) = h_0(t)\exp\left(\gamma x_i + \alpha(m_i(t))\right) = \lambda_1 \exp\left(\lambda_2 t\right)\exp\left\{\gamma x_i + \alpha m_i(t)\right\}, \tag{23}$$

where $h_0(t)$ is the hazard function at baseline having Gompertz distribution, x_i is baseline covariate and $m_i(t)$ denotes the true and unobserved value of the longitudinal at time t. The observed longitudinal value at time point t for the i^{th} subject has the non-linear form

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) \\ &= 5\log(1+t) + b_{i1}t + b_{i0} + \varepsilon_i(t), \end{aligned} \tag{24}$$

Parameter	True value	$n = 200$			$n = 300$			$n = 500$		
		Estimate	SD	MSE	Estimate	SD	MSE	Estimate	SD	MSE
β_0	5	4.21	0.72	0.76	4.68	0.50	0.32	5.10	0.30	0.27
β_1	2	1.69	0.75	0.57	1.86	0.75	0.28	2.10	0.57	0.18
λ_1	0.1	0.12	0.13	0.00	0.12	0.12	0.00	0.11	0.10	0.00
λ_2	0.5	0.50	0.15	0.02	0.57	0.14	0.01	0.48	0.14	0.02
γ	0.5	0.50	0.17	0.03	0.49	0.12	0.04	0.51	0.09	0.01
α	0.05	0.03	0.04	0.00	0.04	0.05	0.00	0.04	0.04	0.00
σ	2	2.06	0.13	0.01	2.02	0.06	0.00	2.02	0.06	0.00
D_{11}	2	2.87	0.92	0.62	2.59	0.73	0.53	2.27	0.80	0.22
D_{22}	2	2.03	0.42	0.16	2.21	0.46	0.23	2.10	0.43	0.05
D_{33}	2	2.10	0.37	0.17	0.34	0.50	0.34	2.22	0.59	0.10
D_{44}	4	5.24	1.82	0.76	4.32	0.74	0.60	4.24	0.63	0.18

Table 1. Summary statistics for parameter estimation of the simulated data of the model in (4.4) for different sample sizes.

where $\varepsilon_i(t) \sim N(0, \sigma^2)$. In the model of (4.6), the mean longitudinal response of the population is assumed to have a non-linear logarithm curve. Different subjects are assumed to have different intercepts and slopes. In particular, we assume that $b_i = (b_{i0}, b_{i1})^T$ having a bivariate normal distribution with mean $\mu = (3, 2)$ and covariance matrix $D = Diag(1, 1)$. The true values of the other parameters we put in the model were $\lambda_1 = 0.01$, $\lambda_2 = 0.1, \gamma = 0.5$, $\alpha = 0.2$, $\sigma = 2$, respectively. In addition, the censoring mechanism is assumed exponentially distributed with a parameter of $\lambda = 0.25$.

Based on the model in (4.5) and the simulation study 1, we simulated survival times T_i for 500 subjects with 35% censoring rate. In particular, the ending time for the study was 5 months and all subjects alive by the end of the study (i.e. time 5) were assumed to be censored. This design was also reflected of many clinical studies in real life. In this sample, there were 329 uncensored subjects and 1387 observations for 500 subjects. For each subject, 1–5 longitudinal measurements were recorded. On average, there were three longitudinal measurements per subject. In **Figure 4**, the Kaplan-Meier estimate for survival curve is presented for the simulated data of (4.5) with 95% pointwise confidence intervals in the left panel. Moreover, the subject-specific longitudinal profiles for six randomly selected subjects is drawn in the right panel. It can be seen that some of the subjects in this dataset showed non-linear evolutions in their longitudinal values. Each subject has its own trajectory.

4.2.2. Parameter estimation

We will be using Model 1 in (4.1) and in (4.2) to handle the non-linear longitudinal trajectory in the simulated data in (4.5). In this model, we put three internal knots at 25, 50 and 75%, respectively, of the follow up time. Then, the ECM algorithm, as explained in Section 3, is implemented once again to estimate all parameters in the model.

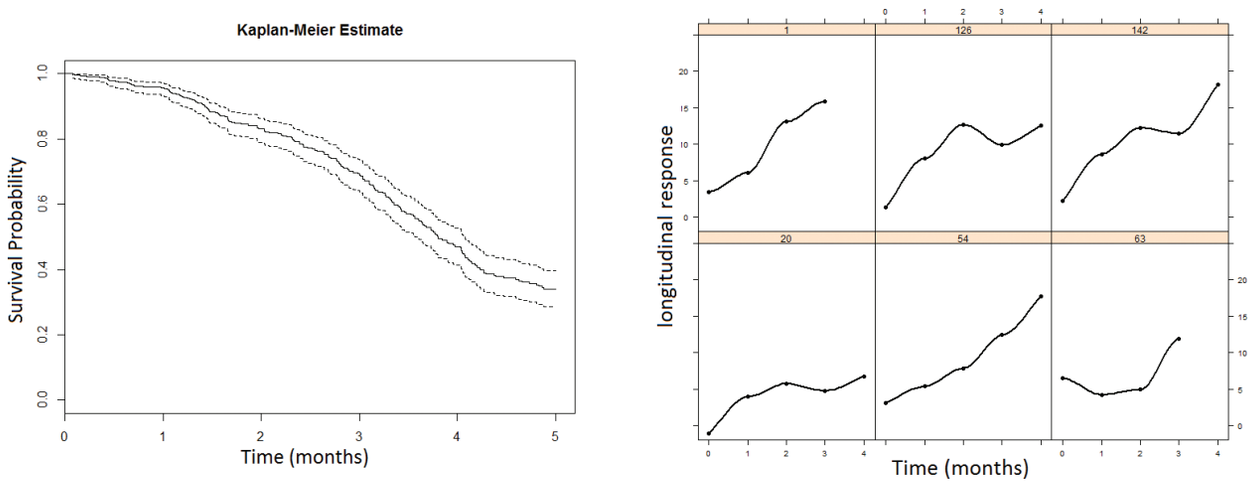

Figure 4. Kaplan-Meier estimate of the survival function of the simulated data of (4.5) (left panel). Longitudinal trajectories for the six randomly selected subjects of (4.6) (right panel).

Parameter	True value	Estimate	SD	95% CI
β_0	—	3.399	0.673	[3.158;3.640]
β_1	—	4.330	0.142	[4.280;4.380]
λ_1	0.01	0.013	0.021	[0.007;0.021]
λ_2	0.1	0.083	0.184	[0.017;0.148]
γ	0.5	0.640	0.386	[0.486;0.778]
α	0.2	0.186	0.142	[0.136;0.237]
σ	2	1.993	0.061	[1.971;2.015]
D_{11}	—	0.977	0.190	[0.909;1.044]
D_{22}	—	1.365	0.183	[1.300;1.430]
D_{33}	—	1.976	0.154	[1.921;2.031]
D_{44}	—	1.393	0.196	[1.322;1.463]

Table 2. Summary statistics for parameter estimation of the simulated data of the model in (4.5) and (4.6).

The results for parameter estimation are presented in **Table 2**. The means, standard deviations and 95% confidence intervals of parameter estimates are calculated for 30 independent samples. The point estimates for $\lambda_1,\ \lambda_2,\ \gamma,\ \alpha,\ \sigma^2$ are reasonably close to the true values. Similarly, the 95% CIs include the true values of $\lambda_1, \lambda_2, \gamma, \alpha, \sigma^2$.

Based on the estimated values of parameters, we generate back the estimated survival time by approximating values of random effects from linear mixed-effects function. The detail of the generation is explained in Appendix C. Then, we use the Kaplan-Meier estimate to compare between the survival function of the simulated dataset (the black solid line) and the estimated survival function (the dashed line) which are presented in the left panel of **Figure 5**.

Moreover, we also draw the smooth and predicted longitudinal profiles for 12 patients chosen randomly in the right panel of **Figure 5**. The dot points are the true observed longitudinal values from simulated data. The solid lines are the smooth longitudinal profiles of the true observed longitudinal values using the loess smoother and the dashed lines are the predicted profiles of 12 randomly selected individuals. It is clear that the Kaplan-Meier estimates from simulated data overlaps the Kaplan-Meier estimates based on the predicted value in the left panel of **Figure 4**. The penalized spline regression model in (2.10) was a good fit for subject-specific curves in the right panel of **Figure 5**.

In summary, simulation studies have shown the stability of the algorithm and the goodness of fit of the penalized spline models. From the simulation study 1, it is shown that the updating process through the ECM algorithm converges quickly to the true values of the parameters. In addition, the simulation study 2 shows that the model can well predict the survival function and individual trajectories respectively.

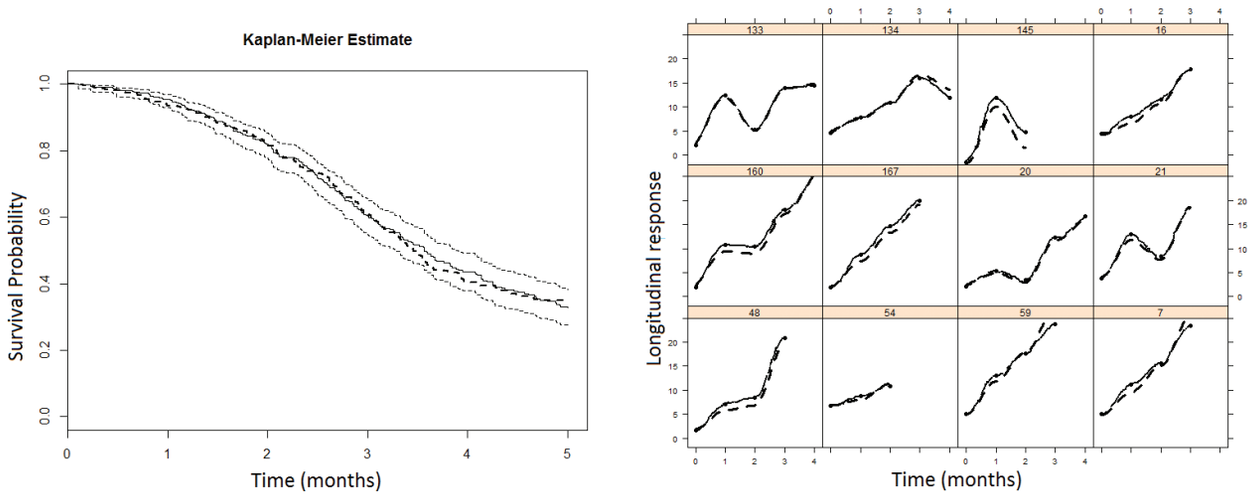

Figure 5. Kaplan-Meier estimate of the survival function from simulated failure times (the solid line) with 95% confidence intervals (dot lines), from Model 1 (4.5) (the dashed line) (left panel). Observed longitudinal trajectories (the solid line) and predicted longitudinal trajectories (the dashed line) for the twelve randomly selected subjects (right panel).

4.3. The AIDS data set

In the AIDS dataset, there were 467 patients with advanced human immunodeficiency virus infection during antiretroviral treatment who had failed or were intolerant to zidovudine therapy. Patients in the study were randomly assigned to receive either *didanosine* drug (*ddI*) or *zalcitabine* drug (*ddC*). CD4 cells are a type of white blood cells made in the spleen, lymph nodes and thymus gland and are part of the infection-fighting system. CD4 cell counts were recorded at the time of study entry as well as at 2, 6, 12 and 18 months thereafter. The detail regarding the design of this study can be found in Abrams et al. [16]. By the end of the study, there were 188 patients died, resulting in about 59.7% censoring. There were 1405 longitudinal responses recorded.

Previously, Rizopoulos [2] used his standard joint model for the AIDS data which consider the variability between subjects mostly depend on the intercept. However, the model could not predict observed longitudinal data accurately. When the time unit is changed from month to year in the data, the variability between subjects depends not only on the intercept but also on the *obstime* variable. In addition, the longitudinal trajectories plot also shows many non-linear curves as depicted in the right panel of **Figure 6**.

Given the non-linearity, it is appropriate to apply our models, Model 1 and Model 2, for the AIDS data. In particular, we use the two joint models in (2.10) and (2.11) with the four internal knots are placed at 20, 40, 60, 80%, respectively of the observed failure times for hazard rate at baseline. Then, the ECM algorithm is implemented to estimate all parameters in the two models. A summary of statistics for parameter estimation using Model 1 and Model 2 is presented in **Table 3**.

Following Rizopoulos [2], in Model 1 and Model 2, the univariate Wald tests are applied for the fixed effects $\beta = \left(\beta_0, \beta_1\right)^T$ in the longitudinal submodel, the regression coefficient γ and the association parameter α respectively. The results from **Table 3** show that the point estimates of β_0, β_1, γ, α are all statistically significant for both models at a significance level of 5%.

http://dx.doi.org/10.5772/intechopen.75975

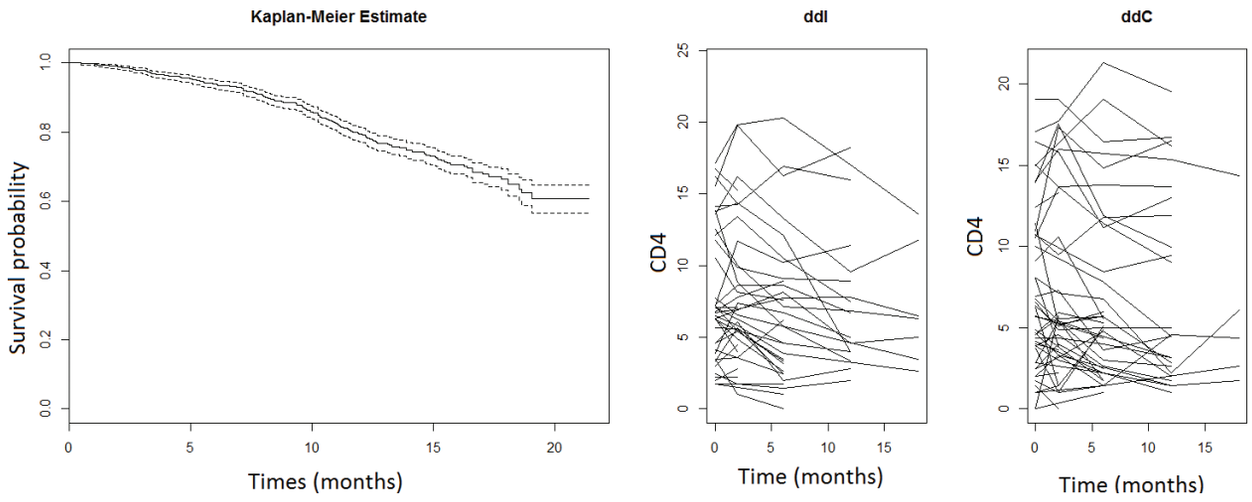

Figure 6. Kaplan-Meier estimate of the survival function of the AIDS data (left panel). Longitudinal trajectories for CD4 cell count of the first 100 patients for two groups (right panel).

Model 1					**Model 2**				
Parameter	**Estimate**	**Std. error**	**z-value**	**p-value**	**Parameter**	**Estimate**	**Std. error**	**z-value**	**p-value**
β_0	7.87	0.06	127.07	<0.001	β_0	7.81	0.07	114.34	<0.001
β_1	−1.69	0.11	−14.77	<0.001	β_1	−1.62	0.12	−12.99	<0.001
γ	0.22	0.11	2.06	0.039	γ	0.31	0.10	3.03	0.002
α	−0.20	0.01	−15.84	<0.001	α	−0.24	0.01	−18.15	<0.001
λ_1	1.68	0.07			λ_1	1.04	0.11		
λ_2	0.33	0.00			λ_2	1.79	0.23		
σ	2.36	0.36			λ_3	1.38	0.38		
D_{11}	2.18	0.14			λ_4	1.67	0.42		
D_{22}	1.04	0.07			λ_5	2.48	0.66		
D_{33}	0.85	0.06			σ	2.62	0.45		
D_{44}	11.87	0.78			D_{11}	1.02	0.07		
					D_{22}	0.97	0.06		
					D_{33}	0.99	0.07		
					D_{44}	11.40	0.75		

Table 3. Summary statistics for parameter estimation of the AIDS data of Model 1 and Model 2 respectively.

We conduct the Kaplan-Meier estimate of the survival function from the observed survival time (the light solid line) and the dot lines correspond to 95% pointwise confidence intervals in **Figure 6** (left panel). The predicted survival function from Model 1 is the dashed line and the predicted survival function from Model 2 is the bold solid line. The plots show that Model 2 works very well in this case as shown in **Figure 7**. Moreover, Model 2 is also preferred in

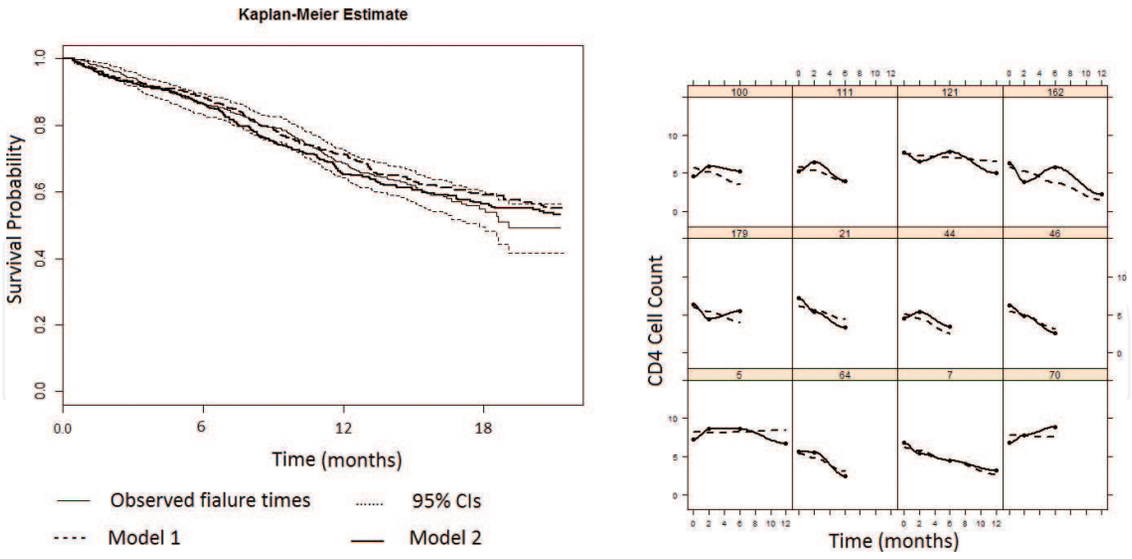

Figure 7. Kaplan-Meier estimates of the survival function from observed failure times, from model 1 and from model 2 (left panel). Observed longitudinal trajectories (the solid line) and predicted longitudinal trajectories (the dashed line) for the 12 randomly selected patients (right panel).

practice because $h_0(.)$ usually is considered as unspecified in order to avoid the impact of misspecifying the distribution of survival times.

Based on the model of longitudinal regression in (4.2), we also draw the smooth and predicted longitudinal profiles for nine patients from the AIDS dataset as depicted in **Figure 7** (right panel). The dot points are the true observed longitudinal values. The solid lines are the smooth longitudinal profiles using the loess smoother and the dashed lines are the predicted profiles of nine randomly selected individuals. Most of the predicted profiles are quite close to the observed ones.

5. Discussion

In this chapter, two joint models using a penalized spline with a truncated polynomial basis have been proposed to model a non-linear longitudinal outcome and a time-to-event data. The use of a truncated polynomial basis gives us an intuitive and obvious way to model non-linear longitudinal outcome. By adding some penalties for the coefficients of the knots and using linear mixed-effects models, the smoothing is controlled and the individual curves are specified.

We have conducted a sensitivity analysis on the assumption of normality for either random effects or errors. The t-distribution with the degree of freedom 5 is applied for each of them. The results show that the estimates of parameters are sensitive when both of terms are not normally distributed.

The main findings we may derive from this chapter are, at least, threefold: (1) the ECM algorithm provides a reasonable quick convergence algorithm for the proposed models; (2) the fitted joint models are able to measure the association between the internal time-dependent

covariates and the risk for an event and (3) the two models provide a good prediction for both the longitudinal and survival functions, as presented in empirical results.

The limitations of this study are, at least, threefold: (1) the number of internal knots is limited to three due to computational costs; (2) the polynomial power functions can form an ill-conditioned basis for the models and (3) the estimation results are sensitive when both random effects and error are not normally distributed.

Based on the limitations, our future work will focus on using new methods for approximating the integrals to reduce the computational problems or relaxing the normality assumption. Furthermore, we will apply a different basis for joint models, that is the penalized B-spline. In terms of parameter estimation, we are considering a different approach to estimate the parameters in the models using a Bayesian approach, via Markov chain Monte Carlo (MCMC) algorithms.

A. Appendix A

One sample of simulated data of the penalized spline joint model in (4.4) is presented in **Table 4** for the first three patients. The subjects are measured bimonthly and the entry time is 0 for all

Id	Obstime	Time	x	y	Death	Z_1	Z_2	Z_3	Z_4
1	0.0	4.97	0	1.41	1	0.0	0.0	0.0	1
1	0.5	4.97	0	6.45	1	0.0	0.0	0.0	1
1	1.0	4.97	0	4.10	1	0.0	0.0	0.0	1
1	1.5	4.97	0	1.50	1	0.5	0.0	0.0	1
1	2.0	4.97	0	4.07	1	1.0	0.0	0.0	1
1	2.5	4.97	0	6.16	1	1.5	0.5	0.0	1
1	3.0	4.97	0	3.60	1	2.0	1.0	0.0	1
1	3.5	4.97	0	8.32	1	2.5	1.5	0.5	1
1	4.0	4.97	0	6.32	1	3.0	2.0	1.0	1
2	0.0	2.79	0	6.81	1	0.0	0.0	0.0	1
2	0.5	2.79	0	7.77	1	0.0	0.0	0.0	1
2	1.0	2.79	0	9.75	1	0.0	0.0	0.0	1
2	1.5	2.79	0	11.04	1	0.5	0.0	0.0	1
2	2.0	2.79	0	7.20	1	1.0	0.0	0.0	1
3	0.0	1.90	0	−1.84	0	0.0	0.0	0.0	1
3	0.5	1.90	0	1.12	0	0.0	0.0	0.0	1
3	1.0	1.90	0	0.78	0	0.0	0.0	0.0	1

Table 4. A snapshot of simulated data for penalized spline joint model in (4.4).

subjects. *Obstime* variable includes the time points at which these measurements are recorded. *Time* variable includes the observed survival times when subject meets an event. x is a time-constant binary random variable with parameter $p = 0.5$. Column y contains the longitudinal responses. *Death* variable is the event status indicator. This variable receives value 1 when the true survival time is less than or equal to the censoring time and 0 otherwise. We define the four random effects variables which are $Z_1 = (obstime - \mathcal{K}_1)_+$, $Z_2 = (obstime - \mathcal{K}_2)_+$, $Z_3 = (obstime - \mathcal{K}_3)_+$ and $Z_4 = \mathbf{1}$. For the longitudinal process, there are 1902 of observations for 500 subjects. For each subject, 1-7 longitudinal measurements are recorded. On average, there are four longitudinal measurements per subject. For the event process, there are 297 subjects who meet for an event which is equivalent to 59.4% of the whole sample.

B. Appendix B

The integrals with respect to the random effects in (3.7) do not have closed-form solutions. Therefore, in this chapter, we implement the Gaussian-Hermite quadrature rule as in Rizopoulos [5] to approximate the integrals. In our simulation study and R coding, we use the Gaussian-Hermite quadrature rule with 10 quadrature points.

The updating formulas of the parameters in Step 3 have different forms for each parameter following Rizopoulos [2]. We have the closed-form estimates for the measurement error variance σ_ε^2 in the longitudinal model and the covariance matrix of the random effects as follows:

$$\widehat{G}^{(it+1)} = \frac{1}{n}\sum_i \int \boldsymbol{b}_i^T \boldsymbol{b}_i p\left(\boldsymbol{b}_i | T_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}^{(it)}\right) d\boldsymbol{b}_i = \frac{1}{N}\sum_i v\tilde{\boldsymbol{b}}_i^{(it)} + \tilde{\boldsymbol{b}}_i^{(it)}\tilde{\boldsymbol{b}}_i^{(it)T}, \tag{25}$$

where $\tilde{\boldsymbol{b}}_i = E\left(\boldsymbol{b}_i | T_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}\right) = \int \boldsymbol{b}_i p\left(\boldsymbol{b}_i | T_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}\right) d\boldsymbol{b}_i$ and $\tilde{v}\boldsymbol{b}_i = \mathrm{var}\left(\boldsymbol{b}_i | T_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}\right) = \int (\boldsymbol{b}_i - \tilde{\boldsymbol{b}}_i) p\left(\boldsymbol{b}_i | T_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}\right) d\boldsymbol{b}_i$. The updating formula for σ_ε^2 is

$$\widehat{\sigma}_{\varepsilon(it+1)}^2 = \frac{1}{n}\sum_i \int \boldsymbol{W}^T \boldsymbol{W} p\left(\boldsymbol{b}_i | T_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}^{(it)}\right) d\boldsymbol{b}_i, \tag{26}$$

where $\boldsymbol{W} = \boldsymbol{y}_i - \boldsymbol{X}_i\beta - \boldsymbol{X}_i\boldsymbol{u}_i - \boldsymbol{Z}_i\boldsymbol{v}_i$.

Unfortunately, we cannot obtain closed-form expressions for the fixed effects β and the parameters of the survival submodel γ, α, *and* $\boldsymbol{\theta}_{h_0}$. We thus employ the one-step Newton-Raphson approach to obtain the updated $\beta^{(it+1)}$, $\gamma^{(it+1)}$,$\alpha^{(it+1)}$ and $\boldsymbol{\theta}_{h_0}^{(it+1)}$. In particular, we have

$$\begin{aligned} S(\boldsymbol{\theta}) &= \frac{\partial Q\left(\boldsymbol{\theta} | \boldsymbol{\theta}^{(it)}\right)}{\partial \boldsymbol{\theta}} \\ \widehat{\boldsymbol{\theta}}^{(it+1)} &= \widehat{\boldsymbol{\theta}}^{(it)} - \left[\frac{\partial S\left(\widehat{\boldsymbol{\theta}}^{(it)}\right)}{\partial \boldsymbol{\theta}}\right]^{-1} S\left(\widehat{\boldsymbol{\theta}}^{(it)}\right), \end{aligned} \tag{27}$$

where $S(\boldsymbol{\theta})$ is the score vector corresponding to parameter $\boldsymbol{\theta}$ and the score vector has the form of

$$S(\boldsymbol{\theta}) = \frac{\partial Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(it)}\right)}{\partial \boldsymbol{\theta}} = \sum_i \int \frac{\partial}{\partial \boldsymbol{\theta}^T} \log\left\{p\left(T_i, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta}^{(it)}\right) p(\boldsymbol{y}_i|\boldsymbol{b}_i; \boldsymbol{\theta}^{(it)}) p(\boldsymbol{b}_i; \boldsymbol{\theta}^{(it)})\right\} . p\left(\boldsymbol{b}_i|T_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}^{(it)}\right) d\boldsymbol{b}_i.$$

C. Appendix C

There are four cases for simulating survival time T_i of the model (4.1) as follows.

When the survival time $t < \mathcal{K}_1$, we calculate the cumulative hazard function $H_i(t) = \int_0^t h_i(s)ds$. Based on the relation between the survival function $S_i(t)$, cumulative hazard function $H_i(t)$ and cumulative distribution $F_i(t)$, we have

$$S_i(t) = \exp\left(-H_i(t)\right) = 1 - F_i(t). \tag{28}$$

Following (5.4), we set

$$u = 1 - F_i(T_i), \tag{29}$$

where u is a random variable with $u \sim Uni[0, 1]$. The survival time t is the solution of the equation

$$U = \exp\left(-H_i(t)\right) = \exp\left(-\int_0^t h_i(s)ds\right).$$

The condition $t < \mathcal{K}_1$ is equal to

$$-\log(U) < \int_0^{\mathcal{K}_1} h(s)ds.$$

When $\mathcal{K}_1 \leq t < \mathcal{K}_2$, we calculate the cumulative hazard function $H_i(t) = \int_0^{\mathcal{K}_1} h_i(s)ds + \int_{\mathcal{K}_1}^t h_i(s)ds$. The survival time t is the solution of the equation

$$U = \exp\left[-\left\{\int_0^{\mathcal{K}_1} h_i(s)ds + \int_{\mathcal{K}_1}^t h_i(s)ds\right\}\right],$$

where U is a value of $u \sim Uni[0, 1]$. The condition $\mathcal{K}_1 \leq t < \mathcal{K}_2$ is equal to

$$-\log(U) < \int_0^{\mathcal{K}_1} h_i(s)ds + \int_{\mathcal{K}_1}^{\mathcal{K}_2} h_i(s)ds.$$

When $\mathcal{K}_2 \le t < \mathcal{K}_3$, we calculate the cumulative hazard function $H_i(t) = \int_0^{\mathcal{K}_1} h_i(s)ds + \int_{\mathcal{K}_1}^{\mathcal{K}_2} h_i(s)ds + \int_{\mathcal{K}_2}^{t} h_i(s)ds$. The survival time t is the solution of the equation

$$U = \exp\left[-\left\{\int_0^{\mathcal{K}_1} h_i(s)ds + \int_{\mathcal{K}_1}^{\mathcal{K}_2} h_i(s)ds + \int_{\mathcal{K}_2}^{t} h_i(s)ds\right\}\right],$$

where U is a value of $u \sim Uni[0,1]$. The condition $\mathcal{K}_2 \le t < \mathcal{K}_3$ is equal to

$$-\log(U) < \int_0^{\mathcal{K}_1} h_i(s)ds + \int_{\mathcal{K}_1}^{\mathcal{K}_2} h_i(s)ds + \int_{\mathcal{K}_2}^{\mathcal{K}_3} h_i(s)ds.$$

When $\mathcal{K}_3 \le t$, the cumulative hazard function has the form $H_i(t) = \int_0^{\mathcal{K}_1} h_i(s)ds + \int_{\mathcal{K}_1}^{\mathcal{K}_2} h_i(s)ds + \int_{\mathcal{K}_2}^{\mathcal{K}_3} h_i(s)ds + \int_{\mathcal{K}_3}^{t} h_i(s)ds$. Survival time t is the solution of the equation

$$U = \exp\left[-\left\{\int_0^{\mathcal{K}_1} h_i(s)ds + \int_{\mathcal{K}_1}^{\mathcal{K}_2} h_i(s)ds + \int_{\mathcal{K}_2}^{\mathcal{K}_3} h_i(s)ds + \int_{\mathcal{K}_3}^{t} h_i(s)ds\right\}\right].$$

D. Appendix D

In particular, Ruppert et al. [9] introduced a default choices for knot location and number of knots. The idea is to choose sufficient knots to resolve the essential structure in the underlying regression function. But for more complicated penalized spline models, there are computational advantages to keeping the number of knots relatively low. A reasonable default is to choose the knots to ensure that there are a fixed number of unique observations, say 4–5, between each knot. For large data sets, this can lead to an excessive numbers of knots; therefore, a maximum number of allowable knots (say, 20–40 total) are recommended.

According to Ruppert et al. [9], the choice for knot position is

$\mathcal{K}_k = \left(\frac{k+1}{K+2}\right) th$ sample quantile of the unique x_i for $k = 1, \ldots, K$.

The simple choice of K is

$K = min\left(\frac{1}{4} \times number\ of\ unique\ x_i, 35\right)$.

E. Appendix E

See **Table 5**.

Parameter	True value	Censored (20%)			Censored (40%)		
		Estimate	SD	MSE	Estimate	SD	MSE
β_0	5	4.85	0.30	0.25	5.10	0.30	0.27
β_1	2	1.86	0.45	0.20	2.10	0.57	0.18
λ_1	0.1	0.13	0.12	0.00	0.11	0.10	0.00
λ_2	0.5	0.52	0.07	0.00	0.49	0.14	0.02
γ	0.5	0.48	0.10	0.00	0.51	0.09	0.00
α	0.05	0.05	0.02	0.00	0.04	0.04	0.00
σ	2	2.02	0.05	0.00	2.02	0.06	0.00
D_{11}	2	2.21	0.67	0.17	2.27	0.80	0.22
D_{22}	2	2.16	0.27	0.09	2.10	0.43	0.05
D_{33}	2	2.26	0.27	0.01	2.22	0.60	0.10
D_{44}	4	4.20	0.53	0.20	4.24	0.63	0.18

Table 5. Summary statistics for parameter estimation of the simulated data of the model in (22) for different censoring rates.

Author details

Huong Thi Thu Pham[1]* and Hoa Pham[2]

*Address all correspondence to: pham0092@flinders.edu.au

1 School of Computer Science, Engineering and Mathematics, Flinders University, Adelaide, South Australia, Australia

2 Mathematical Department, An Giang University, Long Xuyen, Viet Nam

References

[1] Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological). 1972;**34**(2):187-220

[2] Rizopoulos D. Joint Models for Longitudinal and Time-To-Event Data with Applications in R. Chapman & Hall/CRC Biostatistics series; 2012

[3] Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. Biometrics. 2005;**61**(1):64-73

[4] Ding J, Wang J-L. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. Biometrics. 2008;**64**(2):546-556

[5] Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. Computational Statistics and Data Analysis. 2011;**56**:491-501

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

[7] Kalbfleisch J, Prentice R. The Statistical Analysis of Failure Time Data. 2nd ed. New York: Wiley; 2002

[8] Durban M, Harezlak J, Wand M, Carroll R. Simple fitting of subject-specific curves for longitudinal data. Statistics in Medicine. 2005;**24**(8):1153-1167

[9] Ruppert D, Wand M, Carroll R. Semiparametric Regression. Cambridge: Cambridge University Press; 2003

[10] Viviani S, Alfo M, Rizopoulos D. Generalized linear mixed joint model for longitudinal and survival outcomes. Statistics and Computing. 2014;**24**(3):417-427

[11] McLachlan G, Krishnan T. The EM Algorithm and Extensions. Vol. 382. John Wiley & Sons; 2007

[12] Louis TA. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological). 1982:226-233

[13] Bender, Augustin Blettner, Bender, Bender R, Augustin T, Blettner, M. 2005. Generating survival times to simulate cox proportional hazards models, Statistics in Medicine. **24**(11): 1713-1723

[14] Austin PC. Generating survival times to simulate cox proportional hazards models with time-varying covariates. Statistics in Medicine. 2012;**31**(29):3946-3958

[15] Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data, Statistics in Medicine. 2013;**32**(23):4118-4134

[16] Abrams D, Goldman A, Launer C, Korvick J, Neaton J, Crane L, Grodesky M, Wakefield S, Muth K, Kornegay S. Comparative trial of didanosine and zalcitabine in patients with human immunodeficiency virus infection who are intolerant of or have failed zidovudine therapy. New England Journal of Medicine. 1994;**330**:657-662