Harvard University
Harvard University Biostatistics Working Paper Series

Year 2015 *Paper* 188

Simulation of Semicompeting Risk Survival Data and Estimation Based on Multistate Frailty Model

Fei Jiang* Sebastien Haneuse†

*Harvard University, homebovine@gmail.com
†Harvard University, shaneuse@gmail.com

Simulation of Semicompeting Risk Survival Data and Estimation Based on Multistate Frailty Model

Fei Jiang and Sebastien Haneuse

Abstract

We develop a simulation procedure to simulate the semicompeting risk survival data. In addition, we introduce an EM algorithm and a B–spline based estimation procedure to evaluate and implement Xu et al. (2010)'s nonparametric likelihood es- timation approach. The simulation procedure provides a route to simulate samples from the likelihood introduced in Xu et al. (2010)'s. Further, the EM algorithm and the B–spline methods stabilize the estimation and gives accurate estimation results. We illustrate the simulation and the estimation procedure with simluation examples and real data analysis.

Simulation of Semicompeting Risk Survival Data and Estimation Based on Multistate Frailty Model

Fei Jiang, Sebastien Haneuse

Department of Biostatistics
Harvard University
Boston, MA 02115, U.S.A.

Abstract

We develop a simulation procedure to simulate the semicompeting risk survival data. In addition, we introduce an EM algorithm and a B–spline based estimation procedure to evaluate and implement Xu et al. (2010)'s nonparametric likelihood estimation approach. The simulation procedure provides a route to simulate samples from the likelihood introduced in Xu et al. (2010)'s. Further, the EM algorithm and the B–spline methods stabilize the estimation and gives accurate estimation results. We illustrate the simulation and the estimation procedure with simluation examples and real data analysis.

Keywords: B–spline method; EM algorithm; frailty; nonparametric maximum likelihood.

1 Introduction

In randomized trial studies, a patient could experience multiple failure events. A typical situation is that the patient experiences a disease related event, e.g. the recurrence of the disease, before the failure of death. Such data structure is referred to as semicompeting risk data (Fine et al., 2001). Unlike the competing risk data, the order of two event times are restricted so that the disease related event must not happen before the death. In the other words, the death event terminates the disease related event, but not vice verse. In such setting, the joint distribution of the disease related event, namely the nonterminal event, and the death event, namely the terminal event, are often of interest. Xu et al. (2010) suggested to use the illness-death multistate formulation to model the semicompeting risk data through specifying the transition intensities between three distinct states: the on study state (State 1), the illness state (State 2), and the death state (State 3). The three intensities were assumed to satisfy the proportional hazard assumptions, while the baseline hazards were modelled nonparametrically. In addition, they introduced a multiplicative shared frailty into the model to construct joint distribution of the terminal and non-terminal event times. Xu et al. (2010)'s method has wide applications. For example, in a pancreatic cancer study at the Centers for Medicare and Medicaid Services, two events were considered, a nonterminal event, readmission, and a terminal event, death. Xu et al. (2010)'s multistate formulation can be applied to model the two event times so as to explore the covariate effects on the event times jointly. To facilitate the practical application of Xu et al. (2010)'s method, we develop a simulation strategy to evaluate the method and an estimation procedure to implement the approach.

In the semicompeting risk survival settings, the data generating mechanism is not as straightforward as those in the usual multivariate survival settings because we must consider the nature ordering of the two event times. Let T_1, T_2 be the non-terminal event and terminal event times. The main question is how to generate T_1, T_2 such that T_1, T_2 and T_2 given T_1 follow specific marginal and conditional distributions. In this work, we develop a method to generate semicompeting risk samples. Further, we wrote out the the survival functions in close form expressions under the assumption that the baseline hazard are proportional to each other.

To implement Xu et al. (2010)'s method, one challenge is to estimate the nonparametric baseline hazard functions. Under the univariate survival case, the parameters of interest, say the covariate effects, can be estimated through partial likelihoods without knowing the baseline hazard. However, because of the inclusion of the frailty, the estimation of the baseline hazards is necessary for obtaining the estimates for the covariate effects. Xu et al. (2010) reduces the estimation problem to a parametric maximum likelihood estimation by treating the baseline hazard at the event times as additional finite dimensional parameters. However, in this way, the number of the parameters grows at the same rate as that of the event time so that when the event rate is high in the sample or the sample size is large, the estimation would be unstable (Nielsen et al., 1992). Therefore, we first propose a EM algorithm to stabilize the computation. With the EM algorithm, we do not need to specify the forms of the baseline hazards. Secondly, we use B–spline to approximate the baseline hazards and implement the nonparametric maximimal likelihood estimation procedure. With the B–spline approximation, we reduce the nonparametric estimation problem to a parametric estimation, while the parameter number grows at the rate slower than the growth rate of the event time so that the estimation is numerically more stable than Xu et al. (2010)'s method.

The rest of the paper is organized as follows. We describe the simulation procedure in Section 2. We introduce and evaluate the EM algorithm in Section 3. Then we introduce and evaluate the B–spline based estimation procedure in Section 4. We conclude the paper in Section 5.

2 The simulation procedures

In this section, we develop a simulation mechanism for generating the semicompeting risk data based on the joint densities derived in Xu et al. (2010).

We first describe a general approach to generate the data. Then, we show that under certain assumptions, the data generating functions can have closed forms. Let

$$\begin{aligned}
\lambda_1(t|\Gamma, \mathbf{X}) &= \Gamma\lambda_{01}(t)\alpha_1(\mathbf{X}) \\
\lambda_2(t|\Gamma, \mathbf{X}) &= \Gamma\lambda_{02}(t)\alpha_2(\mathbf{X}) \\
\lambda_3(t|t_1, \Gamma, \mathbf{X}) &= \Gamma\lambda_{03}(t)\alpha_3(\mathbf{X}), t > t_1
\end{aligned}$$

to be the hazard functions for nonterminal event, terminal event without nonterminal event, and terminal following a nonterminal event, where $\alpha_1, \alpha_2, \alpha_3$ are functions on the covariates, $\lambda_{01}(t), \lambda_{02}(t), \lambda_{03}(t)$ are the baselines hazards. Further let $\Lambda_{01}(t), \Lambda_{02}(t), \Lambda_{03}(t)$ to be the corresponding cumulative baseline hazards. Under the assumption that Γ is gamma distributed with mean 1 and variance θ, Xu et al. (2010) show that

$$f_{12}(t_1,t_2) = (\theta+1)\lambda_{01}(t_1)\lambda_{03}(t_2)\left[1+\theta\Lambda_{01}(t_1)+\theta\Lambda_{02}(t_1)+\theta\{\Lambda_{03}(t_2)-\Lambda_{03}(t_1)\}\right]^{-1/\theta-2}$$

and

$$f_{\infty 2}(t_2) = \lambda_{02}(t_2)\left\{1+\theta\Lambda_{01}(t_2)+\theta\Lambda_{02}(t_2)\right\}^{-1/\theta-1}$$

The joint density can be written as

$$\begin{aligned} f(t_1,t_2) &= f_{12}(t_1,t_2)I(t_1\le t_2)+I(t_1=\infty)f_{\infty 2}(t_2) \\ &= I(t_1\le t_2)(\theta+1)\lambda_{01}(t_1)\lambda_{03}(t_2)\left[1+\theta\Lambda_{01}(t_1)+\theta\Lambda_{02}(t_1)+\theta\{\Lambda_{03}(t_2)-\Lambda_{03}(t_1)\}\right]^{-1/\theta-2} \\ &\quad +I(t_1=\infty)\lambda_{02}(t_2)\left\{1+\theta\Lambda_{01}(t_2)+\theta\Lambda_{02}(t_2)\right\}^{-1/\theta-1} \end{aligned}$$

As a result, we can obtain

$$\Pr(T_{i1}=\infty) = \int_0^\infty f_{\infty 2}(t_2)dt_2$$

We further derive the density of T_{i1} when $T_{i1}<\infty$ from f_{12} as,

$$f_{12}(t_1) = \int_{t_1}^\infty f_{12}(t_1,t_2)dt_2.$$

Therefore, combine with the $\Pr(T_{i1}=\infty)$ we have

$$f_1(t_1) = I(t_1<\infty)\int_{t_1}^\infty f_{12}(t_1,t_2)dt_2 + I(t_1=\infty)\int_0^\infty f_{\infty 2}(t_2)dt_2.$$

Then the survival function

$$\begin{aligned} S_1(t_1) &= \Pr(t_1\le T_{i1}<\infty)+\Pr(T_{i1}=\infty) \\ &= \int_{t_1}^\infty f_1(t)dt+\Pr(T_{i1}=\infty) \end{aligned}$$

and

$$S_1(t_1|T_{i1}<\infty) = \int_{t_1}^\infty f_1(t)dt/\{1-\Pr(T_{i1}=\infty)\}$$

In addition, we can also derive joint probability $\Pr\{T_{i2}>t_2, T_{i1}\in[t_1,t_1+dt)\}, t_1<t_2<\infty$, which is

$$\begin{aligned} &\Pr\{T_{i2}>t_2, T_{i1}\in[t_1,t_1+dt)\} \\ =& \int_{t_2}^\infty f_{12}(t_1,t)dt \end{aligned}$$

Therefore, the conditional survival function for T_{i2} given $T_{i1}=t_1<\infty$ is,

$$\begin{aligned} S_{21}(t_2|t_1) &= \Pr(T_{i2}>t_2|T_{i1}=t_1) \\ &= \Pr\{T_{i2}>t_2, T_{i1}\in[t_1,t_1+dt)\}/f_{12}(t_1). \end{aligned}$$

Combining with $\Pr(T_{i1}=\infty)$, we now can use flip coin method to generate the data. With probability $\Pr(T_{i1}=\infty)$, we generate T_{i2} from the survival function $S_{21}(t_2|\infty)$, and with probability $1-\Pr(T_{i1}=\infty)$, we generate T_{i1} from $S_1(t_1|T_{i1}<\infty)$, and then generate T_{i2} from $S_{21}(t_2|t_1), t_1<\infty$ conditioning on the observed value of $T_{i1}=t_1$.

2.1 The simulation procedure under specific hazard functions

We define the hazard functions $\lambda_{01}(t) \equiv \alpha_1(\mathbf{X}_i)\lambda_0(t)$, $\lambda_{02}(t) \equiv \alpha_2(\mathbf{X}_i)\lambda_0(t)$, $\lambda_{03}(t) \equiv \alpha_3(\mathbf{X}_i)\lambda_0(t)$, where $\alpha_j(\mathbf{X}_i), j = 1, 2, 3$ are functions of covariates depending on parameters $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$. And the hazard functions for T_1, T_2 with and without the nonterminal event occurs before as $\lambda_1(t_1|\gamma) = \gamma\lambda_{01}(t_1)$, $\lambda_2(t_2|\gamma, t_1) = \gamma\lambda_{03}(t_2)$, and $\lambda_2(t_2|\gamma) = \gamma\lambda_{02}(t_2)$, respectively. The three hazards are proportional to each other in the sense that $\lambda_{0l}(t)/\lambda_{0k}(t)$ does not depend on t, $l, k = 1, 2, 3, l \neq k$. For illustration, we can use the Weibull hazard functions with the identical shape parameters, κ, i.e., $\lambda_{0j}(t) = \alpha_j(\mathbf{X}_i)\lambda_0(t) = \kappa h_j \exp(\boldsymbol{\beta}_j\mathbf{X}_i)t^{\kappa-1}$, where $h_j, j = 1, 2, 3$ are positive constant. Here $\alpha_j(\mathbf{X}_i) = h_j \exp(\boldsymbol{\beta}_j\mathbf{X}_i)$, and $\lambda_0(t) = \kappa t^{\kappa-1}$. The piecewise constant hazard functions in Xu et al. (2010) is another example which also follows our assumption.

Now we derive the survival functions used to generated data. For the semi-competing risk data, let

$$f_{12}(t_1, t_2) = (\theta+1)\lambda_{01}(t_1)\lambda_{03}(t_2)\left[1 + \theta\Lambda_{01}(t_1) + \theta\Lambda_{02}(t_1) + \theta\{\Lambda_{03}(t_2) - \Lambda_{03}(t_1)\}\right]^{-1/\theta-2}$$

and

$$f_{\infty 2}(t_2) = \lambda_{02}(t_2)\left\{1 + \theta\Lambda_{01}(t_2) + \theta\Lambda_{02}(t_2)\right\}^{-1/\theta-1}$$

the joint density can be written as

$$\begin{aligned}
f(t_1, t_2) &= f_{12}(t_1, t_2)I(t_1 \leq t_2) + I(t_1 = \infty)f_{\infty 2}(t_2) \\
&= I(t_1 \leq t_2)(\theta+1)\lambda_{01}(t_1)\lambda_{03}(t_2)\left[1 + \theta\Lambda_{01}(t_1) + \theta\Lambda_{02}(t_1) + \theta\{\Lambda_{03}(t_2) - \Lambda_{03}(t_1)\}\right]^{-1/\theta-2} \\
&\quad + I(t_1 = \infty)\lambda_{02}(t_2)\left\{1 + \theta\Lambda_{01}(t_2) + \theta\Lambda_{02}(t_2)\right\}^{-1/\theta-1} \\
&= I(t_1 \leq t_2)(\theta+1)\alpha_1(\mathbf{X}_i)\lambda_0(t_1)\alpha_3(\mathbf{X}_i)\lambda_0(t_2)[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) \\
&\quad + \theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t_2) - \Lambda_0(t_1)\}]^{-1/\theta-2} + I(t_1 = \infty)\alpha_2(\mathbf{X}_i)\lambda_0(t_2)\{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_2) \\
&\quad + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_2)\}^{-1/\theta-1}.
\end{aligned}$$

From the above expression, we can derive the

$$\begin{aligned}
\Pr(T_{i1} = \infty) &= \int_0^\infty f_{\infty 2}(t_2)dt_2 \\
&= \int_0^\infty \kappa\alpha_2(\mathbf{X}_i)\lambda_0(t_2)\{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_2) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_2)\}^{-1/\theta-1}dt_2 \\
&= -\theta\frac{\alpha_2(\mathbf{X}_i)}{\theta\alpha_1(\mathbf{X}_i) + \theta\alpha_2(\mathbf{X}_i)}\{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_2) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_2)\}^{-1/\theta}\Big|_0^\infty \\
&= \frac{\alpha_2(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i) + \alpha_2(\mathbf{X}_i)}.
\end{aligned}$$

We further derive the density of T_{i1} when $T_{i1} < \infty$ from f_{12} as,

$$
\begin{aligned}
f_{12}(t_1) &= \int_{t_1}^{\infty} f_{12}(t_1, t_2) dt_2 \\
&= \int_{t_1}^{\infty} (\theta+1)\kappa\alpha_1(\mathbf{X}_i)\lambda_0(t_1)\kappa\alpha_3(\mathbf{X}_i)\lambda_0(t_2)[1+\theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1)+\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) \\
&\quad +\theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t_2)-\Lambda_0(t_1)\}]^{-1/\theta-2} dt_2 \\
&= -\frac{\theta}{\theta+1}(\theta+1)\alpha_1(\mathbf{X}_i)\lambda_0(t_1)\frac{\alpha_3(\mathbf{X}_i)}{\theta\alpha_3(\mathbf{X}_i)}[1+\theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1)+\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) \\
&\quad +\theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t_2)-\Lambda_0(t_1)\}]^{-1/\theta-1}\Big|_{t_1}^{\infty} \\
&= \alpha_1(\mathbf{X}_i)\lambda_0(t_1)[1+\theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1)+\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1/\theta-1}
\end{aligned}
$$

Therefore, combine with the $\Pr(T_{i1} = \infty)$ we have

$$
\begin{aligned}
f_1(t_1) &= I(t_1 < \infty)\alpha_1(\mathbf{X}_i)\lambda_0(t_1)[1+\theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1)+\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1/\theta-1} \\
&\quad +I(t_1 = \infty)\frac{\alpha_2(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i)+\alpha_2(\mathbf{X}_i)}.
\end{aligned}
$$

Then the survival function

$$
\begin{aligned}
S_1(t_1) &= \Pr(t_1 \le T_{i1} < \infty) + \Pr(T_{i1} = \infty) \\
&= \int_{t_1}^{\infty} f_1(t)dt + \Pr(T_{i1} = \infty) \\
&= \frac{\alpha_1(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i)+\alpha_2(\mathbf{X}_i)}[1+\theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1)+\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1/\theta} \\
&\quad +\frac{\alpha_2(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i)+\alpha_2(\mathbf{X}_i)}
\end{aligned}
$$

and

$$
\begin{aligned}
S_1(t_1|T_{i1} < \infty) &= \frac{\alpha_1(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i)+\alpha_2(\mathbf{X}_i)}[1+\theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1)+\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1/\theta}/\Pr(T_{i1} < \infty) \\
&= [1+\theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1)+\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1/\theta}.
\end{aligned}
$$

The hazard function for $t_1 < \infty$ is

$$
\lambda_1(t_1) = I(t_1 < \infty)\{\alpha_1(\mathbf{X}_i)+\alpha_2(\mathbf{X}_i)\}\lambda_0(t_1)[1+\theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1)+\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1}.
$$

We can also derive joint probability $\Pr\{T_{i2} > t_2, T_{i1} \in [t_1, t_1 + dt)\}, t_1 < t_2 < \infty$, which is

$$\begin{aligned}
& \Pr\{T_{i2} > t_2, T_{i1} \in [t_1, t_1 + dt)\} \\
= & \int_{t_2}^{\infty} f_{12}(t_1, t)dt \\
= & \int_{t_2}^{\infty} (\theta + 1)\alpha_1(\mathbf{X}_i)\lambda_0(t_1)\alpha_3(\mathbf{X}_i)\lambda_0(t)[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) \\
& +\theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t) - \Lambda_0(t_1)\}]^{-1/\theta-2}dt \\
= & -\frac{\theta}{1+\theta}(\theta + 1)\alpha_1(\mathbf{X}_i)\lambda_0(t_1)\frac{\alpha_3(\mathbf{X}_i)}{\theta\alpha_3(\mathbf{X}_i)}[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) \\
& +\theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t) - \Lambda_0(t_1)\}]^{-1/\theta-1}\Big|_{t_2}^{\infty} \\
= & \alpha_1(\mathbf{X}_i)\lambda_0(t_1)[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t_2) - \Lambda_0(t_1)\}]^{-1/\theta-1}.
\end{aligned}$$

Therefore, the conditional survival function for T_{i2} given $T_{i1} = t_1 < \infty$ is,

$$\begin{aligned}
S_{21}(t_2|t_1) &= \Pr(T_{i2} > t_2|T_{i1} = t_1) \\
&= \frac{[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t_2) - \Lambda_0(t_1)\}]^{-1/\theta-1}}{[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1/\theta-1}}.
\end{aligned}$$

And the conditional hazard function is

$$\begin{aligned}
\lambda_{21}(t_2|t_1) &= -\partial \log S_{21}(t_2|t_1)/\partial t_2 \\
&= (\theta + 1)\alpha_3(\mathbf{X}_i)\lambda_0(t_2)[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t_2) - \Lambda_0(t_1)\}]^{-1}.
\end{aligned}$$

Further, when $t_1 = \infty$, we have

$$\begin{aligned}
S_{21}(t_2|\infty) &= \Pr(T_{i2} > t_2, T_{i1} = \infty)/\Pr(T_{i1} = \infty) \\
&= \left\{\frac{\alpha_2(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i) + \alpha_2(\mathbf{X}_i)}\right\}^{-1} \int_{t_2}^{\infty} \alpha_2(\mathbf{X}_i)\lambda_0(t)\{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t) \\
&\quad +\theta\alpha_2(\mathbf{X}_i)\Lambda_0(t)\}^{-1/\theta-1}dt \\
&= -\left\{\frac{\alpha_2(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i) + \alpha_2(\mathbf{X}_i)}\right\}^{-1} \frac{\alpha_2(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i) + \alpha_2(\mathbf{X}_i)}\{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t)\}^{-1/\theta}\Big|_{t_2}^{\infty} \\
&= \{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_2) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_2)\}^{-1/\theta}
\end{aligned}$$

which gives the hazard function

$$\begin{aligned}
\lambda_{21}(t_2|\infty) &= -\partial \log S_{21}(t_2|\infty)/\partial t_2 \\
&= \{\alpha_1(\mathbf{X}_i) + \alpha_2(\mathbf{X}_i)\}\lambda_0(t_2)\{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_2) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_2)\}^{-1}
\end{aligned}$$

In conclusion, the above derivations give the three hazard functions are

$$\begin{aligned}
\lambda_1(t_1) &= I(t_1 < \infty)\{\alpha_1(\mathbf{X}_i) + \alpha_2(\mathbf{X}_i)\}\lambda_0(t_1)[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1} \\
\lambda_{21}(t_2|t_1) &= I(t_1 \le t_2)(\theta + 1)\alpha_3(\mathbf{X}_i)\lambda_0(t_2)[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) \\
&\quad +\theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t_2) - \Lambda_0(t_1)\}]^{-1} \\
\lambda_{21}(t_2|\infty) &= I(t_1 = \infty)\{\alpha_1(\mathbf{X}_i) + \alpha_2(\mathbf{X}_i)\}\lambda_0(t_2)\{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_2) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_2)\}^{-1}
\end{aligned}$$

and the corresponding survival functions are

$$\begin{aligned} S_1(t_1|T_{i1} < \infty) &= [1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1/\theta} \\ S_{21}(t_2|t_1) &= Pr(T_{i2} \geq t_2|T_{i1} = t_1) \\ &= \frac{[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_3(\mathbf{X}_i)\{\Lambda_0(t_2) - \Lambda_0(t_1)\}]^{-1/\theta-1}}{[1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_1) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_1)]^{-1/\theta-1}} \\ S_{21}(t_2|\infty) &= Pr(T_{i2} \geq t_2|T_{i1} = \infty) \\ &= \{1 + \theta\alpha_1(\mathbf{X}_i)\Lambda_0(t_2) + \theta\alpha_2(\mathbf{X}_i)\Lambda_0(t_2)\}^{-1/\theta}. \end{aligned}$$

Combining with the fact that $\Pr(T_{i1} = \infty) = \frac{\alpha_2(\mathbf{X}_i)}{\alpha_1(\mathbf{X}_i)+\alpha_2(\mathbf{X}_2)}$, we now can use flip coin method to generate the data. With probability $\Pr(T_{i1} = \infty)$, we generate T_{i2} from the survival function $S_{21}(t_2|\infty)$, and with probability $1 - \Pr(T_{i1} = \infty)$, we generate T_{i1} from $S_1(t_1|T_{i1} < \infty)$, and then generate T_{i2} from $S_{21}(t_2|t_1), t_1 < \infty$ conditioning on the observed value of $T_{i1} = t_1$.

3 An EM Algorithm to Implement Xu et al. (2010)'s Method.

3.1 The estimation without left truncation

Xu et al. (2010) introduced a nonparametric likelihood method for the estimation under the multistate model. The estimation were performed under the assumption that the baseline hazards had piece–wise constant forms. Assuming the frailty is gamma distributed, Xu et al. (2010) derived the score functions for the covariate effects, the parameters associated with the baseline hazards, and the parameter associated with the gamma frailty. Because the baseline hazards are nonparametrically specified and the parameter size grows with the observed event times, the computation is not stable when solving the score functions directly for a large number of parameters. We propose to use the EM algorithm described in Nielsen et al. (1992) to stabilize the estimation procedures. The method imputes the latent frailties by their posterior mean obtained from conditioning on the observed random variables. More specifically, by the property of the gamma distribution, we obtain the posterior mean of Γ as

$$E(\Gamma_i|\mathbf{o_i}) = \frac{\delta_1 + \delta_2 + 1/\theta}{1/\theta + \Lambda_{01}(y_{i1})\exp(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x}_i) + \Lambda_{02}(y_{i1})\exp(\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x}_i) + \Lambda_{03}(y_{i1}, y_{i2})\exp(\boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{x}_i)} \quad (1)$$

where $\Lambda_{03}(y_{i1}, y_{i2}) = \Lambda_{03}(y_{i2}) - \Lambda_{03}(y_{i1})$. Now knowing that, if $E(\Gamma_i|\mathbf{o_i})$ is known, the estimation procedures are the same as the ones used in the Cox regression model. More specifically, we could write the score functions for $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ as follows

$$\begin{aligned}
\mathbf{U}_{\beta_1} &= \sum_{i=1}^{n} \delta_{i1} \left[\mathbf{x}_i - \frac{\sum_{j=1}^{n} I(y_{i1} \le y_{j1}) E\left\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) \mathbf{x}_j | \mathbf{o}_j\right\}}{\sum_{j=1}^{n} I(y_{i1} \le y_{j1}) E\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \right] \\
\mathbf{U}_{\beta_2} &= \sum_{i=1}^{n} (1-\delta_{i1})\delta_{i2} \left[\mathbf{x}_i - \frac{\sum_{j=1}^{n} I(y_{i2} \le y_{j1}) E\left\{\Gamma_j \exp(\boldsymbol{\beta}_2^{\mathrm{T}} \mathbf{x}_j) \mathbf{x}_j | \mathbf{o}_j\right\}}{\sum_{j=1}^{n} I(y_{i2} \le y_{j1}) E\{\Gamma_j \exp(\boldsymbol{\beta}_2^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \right] \\
\mathbf{U}_{\beta_3} &= \sum_{i=1}^{n} \delta_{i1}\delta_{i2} \left[\mathbf{x}_i - \frac{\sum_{j=1}^{n} I(y_{j1} < y_{i2} \le y_{j2}) E\left\{\Gamma_j \exp(\boldsymbol{\beta}_3^{\mathrm{T}} \mathbf{x}_j) \mathbf{x}_j | \mathbf{o}_j\right\}}{\sum_{j=1}^{n} I(y_{j1} < y_{i2} \le y_{j2}) E\{\Gamma_j \exp(\boldsymbol{\beta}_3^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \right].
\end{aligned}$$

After obtaining $\boldsymbol{\beta}_j, j = 1, 2, 3$, we can estimate the baseline hazard by using Nelson-Aalen type estimators, i.e,

$$\begin{aligned}
\widehat{\lambda}_{01}(y_{i1}) &= \frac{\delta_{i1}}{\sum_{j=1}^{n} I(y_{i1} \le y_{j1}) E\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \\
\widehat{\lambda}_{02}(y_{i1}) &= \frac{(1-\delta_{i1})\delta_{i2}}{\sum_{j=1}^{n} I(y_{i2} \le y_{j1}) E\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \\
\widehat{\lambda}_{03}(y_{i2}) &= \frac{\delta_{i1}\delta_{i2}}{\sum_{j=1}^{n} I(y_{j1} < y_{i2} \le y_{j2}) E\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}}. \qquad (2)
\end{aligned}$$

Note that $E(\Gamma_i | \mathbf{o}_i)$ is replaced by (1) in the estimation procedures. In estimating $\boldsymbol{\beta}_j$'s, $E(\Gamma_i | \mathbf{o}_i)$ is treated as a function of $\boldsymbol{\beta}_j$, while when estimating λ_{0j}'s, $E(\Gamma_i | \mathbf{o}_i)$ is assumed to be known in advance. Because the estimating equation $\mathbf{U}_{\beta_j}$'s involve λ_{0j}'s through $E(\Gamma_i | \mathbf{o}_i)$, we obtain the estimators by iteratively using $\mathbf{U}_{\beta_j}$'s and (2).

First, for each fixed θ, the estimation procedures start from estimating $\boldsymbol{\beta}_j$ by using $\mathbf{U}_{\beta_j}$'s and initial λ_{0j}'s, which are defaulted to be the Nelson–Aalen estimators without covariates. Then we plug the resulting $\boldsymbol{\beta}_j$'s to (1) to obtain $E(\Gamma_i | \mathbf{o}_i)$. Next, we plug $\boldsymbol{\beta}_j$ and $E(\Gamma_i | \mathbf{o}_i)$ to (2) for updating λ_{0j}'s. After that, we plug the current λ_{0j}'s to $\mathbf{U}_{\beta_j}$'s again to update $\boldsymbol{\beta}_j$'s. We perform the iterative procedure until it converges.

The above procedure are based on a fixed θ. To find the estimator for θ, we use the criteria that the given θ and the estimated $\boldsymbol{\beta}_j$'s and λ_{0j} together maximize the likelihood function that

$$\begin{aligned}
&\prod_{i=1}^{n} \lambda_{01}(Y_{i1})^{\delta_{i1}} \lambda_{02}(Y_{i2})^{\delta_{i2}(1-\delta_{i1})} \lambda_{03}(Y_{i2})^{\delta_{i2}\delta_{i1}} \exp\{\delta_{i1}\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}_i + \delta_{i2}(1-\delta_{i1})\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X}_i + \delta_{i1}\delta_{i2}\boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}_i\} \\
&\times (1+\theta)^{\delta_{i1}\delta_{i2}} [1 + \theta\{\Lambda_{01}(Y_{i1})\exp(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}_i) + \Lambda_{02}(Y_{i1})\exp(\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X}_i) + \Lambda_{03}(Y_{i1}, Y_{i2})\exp(\boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}_i)\}]^{-1/\theta - \delta_{i1} - \delta_{i2}}.
\end{aligned}$$

3.2 The estimation with left truncation

The EM algorithm can be easily extended to the setting when data are left truncation. Follow Su and Wang (2012), we specify the score functions as

$$
\begin{aligned}
\mathbf{U}_{L\beta_1} &= \sum_{i=1}^{n} \delta_{i1} \left[\mathbf{x}_i - \frac{\sum_{j=1}^{n} I(t_j < y_{i1} \le y_{j1}) E\left\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) \mathbf{x}_j | \mathbf{o}_j \right\}}{\sum_{j=1}^{n} I(t_j < y_{i1} \le y_{j1}) E\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \right] \\
\mathbf{U}_{L\beta_2} &= \sum_{i=1}^{n} (1-\delta_{i1})\delta_{i2} \left[\mathbf{x}_i - \frac{\sum_{j=1}^{n} I(t_j < y_{i2} \le y_{j1}) E\left\{\Gamma_j \exp(\boldsymbol{\beta}_2^{\mathrm{T}} \mathbf{x}_j) \mathbf{x}_j | \mathbf{o}_j \right\}}{\sum_{j=1}^{n} I(t_j < y_{i2} \le y_{j1}) E\{\Gamma_j \exp(\boldsymbol{\beta}_2^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \right] \\
\mathbf{U}_{L\beta_3} &= \sum_{i=1}^{n} \delta_{i1}\delta_{i2} \left[\mathbf{x}_i - \frac{\sum_{j=1}^{n} I(y_{j1} < y_{i2} \le y_{j2}) E\left\{\Gamma_j \exp(\boldsymbol{\beta}_3^{\mathrm{T}} \mathbf{x}_j) \mathbf{x}_j | \mathbf{o}_j \right\}}{\sum_{j=1}^{n} I(y_{j1} < y_{i2} \le y_{j2}) E\{\Gamma_j \exp(\boldsymbol{\beta}_3^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \right].
\end{aligned}
$$

and the baseline hazard estimators as

$$
\begin{aligned}
\widehat{\lambda}_{L01}(y_{i1}) &= \frac{\delta_{i1}}{\sum_{j=1}^{n} I(t_j < y_{i1} \le y_{j1}) E\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \\
\widehat{\lambda}_{L02}(y_{i1}) &= \frac{(1-\delta_{i1})\delta_{i2}}{\sum_{j=1}^{n} I(t_j < y_{i2} \le y_{j1}) E\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}} \\
\widehat{\lambda}_{L03}(y_{i2}) &= \frac{\delta_{i1}\delta_{i2}}{\sum_{j=1}^{n} I(y_{j1} < y_{i2} \le y_{j2}) E\{\Gamma_j \exp(\boldsymbol{\beta}_1^{\mathrm{T}} \mathbf{x}_j) | \mathbf{o}_j\}}.
\end{aligned}
$$

Note that the only changes are made on the indicator functions of the time intervals. Other estimations follow the same procedures as the ones using for the data without left truncation.

3.3 Simulation Examples

We implement the simulation and the EM estimation procedure in R. We generate the censoring time from uniform distribution (c_1, c_2). The simulation function is "simCpRsk", which takes arguments "n" to be the sample size, "p" to be the covariate dimension to be, "theta" to be θ , "lambda1, lambda2, lambda3" to be h_1, h_2, h_3, "kappa" to be κ, "beta1, beta2, beta3" to be $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$, "cen1, cen2" to be c_1, c_2. The estimation function is "FrqID" which takes "survData" to be the input dataset, "startValue" to be the user specified initial value for estimating $\boldsymbol{\beta}_j, j = 1, 2, 3$, "stheta" to be the interval for searching the true θ. The rest of the arguments are the control parameters used for determining the convergence of the estimation procedure.

As an example, we generate $n = 250$ samples with one covariate generated from normal (0, 1) distribution. We specify $\theta = 1$, $h_1 = 2, h_2 = 1, h_3 = 0.5, \kappa = 0.5, \beta_1 = 0.5, \beta_2 = 0.1, \beta_3 = -0.3$ and the censoring distribution is uniform (2.5, 3). Under these specifications, the nonterminal event rate is 64% and the terminal event rate is 87%. We obtain the estimators as $\widehat{\beta}_1 = 0.46, \widehat{\beta}_2 = 0.09, \widehat{\beta}_3 = 0.09, \widehat{\theta} = 1.01$. The results are close to the truth. We also check if θ falls into the interior of the user specified interval by plotting the likelihood vesus the values of θ. Figure 1 shows clearly that the likelihood is concave, and the resulting estimators maximize the likelihood.

In addition, we implement the left truncation procedure proposed in Section 3.2. We simulate additional left truncation times $v_i, i = 1, \ldots, n$ from an uniform distribution on the range of (0, 0.5), and select the samples with $y_{i1} > v_i$. The resulting sample contains 218 nontruncated observations. We obtain the estimators as $\widehat{\beta}_1 = 0.48, \widehat{\beta}_2 = 0.10, \widehat{\beta}_3 = 0.1, \theta = 0.97$, which is close to the truth. We further check if θ falls into the interior of the user specified interval by plotting the likelihood vesus the values of θ. Figure 2 shows clearly that the likelihood is concave, and the resulting estimators maximize the likelihood.

In another example, we generate the data using the same mechanism as described above. The results are $\widehat{\beta}_1 = 0.42, \widehat{\beta}_2 = 0.23, \widehat{\beta}_3 = 0.43, \widehat{\theta} = 0.85$. The results deviate from the true values. It raises the question that if the likelihood is concave, and if the results fall into the interior the selected region. We plot the likelihood vesus the values of θ as shown in Figure 3. Figure 3 shows that the likelihood is monotone decreasing, and the maximum is achieved at the boundary of the chosen interval. Therefore, in practice, in order to ensure obtaining accurate estimators, we suggest to implement likelihood checking as shown in Figure 1, 2 and 3.

4 A B-spline Method to Implement Xu et al. (2010)'s Method.

One alterative to the Xu et al. (2010)'s method for the nonparametric maximum likelihood estimation is to use the B-spline to approximate the unknown baseline hazard functions. More specifically, we can use $\mathbf{B}_{rl}(t)^{\mathrm{T}}\boldsymbol{\alpha}_l$ to approximate $\lambda_{0l}(t)$ where $\mathbf{B}_{rl}$ is the B-spline basis with degree r, and $\boldsymbol{\alpha}_l$ is a r dimensional B-spline coefficient. Theoretically, there is a $\mathbf{B}_{rl}(t)^{\mathrm{T}}\boldsymbol{\alpha}_l$ can well approximate $\lambda_{0l}(t)$ when the selected knots $t_1, \ldots, t_N$, $t_k \in [0, \tau], k = 1, \ldots, N$ goes to infinite, where τ is the maximal of the study duration, and N is the number of knots. In practice, we can select a finite number of knots so that the maximization of the nonparametric likelihood reduces to an finite dimension optimization problem, which can be carried out using existing optimization routings. In addition to λ, we can also approximate $\Lambda(t)$ by $\int_0^t \mathbf{B}_{rl}(s)ds\boldsymbol{\alpha}$. Denoting $\int_0^t \mathbf{B}_{rl}(s)ds$ by $\mathbf{IB}_{rl}(t)$, as shown in De Boor et al. (1978), $\mathbf{IB}_{rl}(t)$ is a higher degree B-spline basis and has a closed form expression which facilitates the computation. After replacing $\lambda(t)$, $\Lambda(t)$ by the B-spline approximates, the approximated log likelihood can be written as

$$\begin{aligned}
&\sum_{i=1}^{n} \delta_{i1}\log\{\mathbf{B}_{r1}(Y_{i1})^{\mathrm{T}}\boldsymbol{\alpha}_1\} + \{(1-\delta_{i1})\delta_{i2}\}\log\{\mathbf{B}_{r2}(Y_{i1})^{\mathrm{T}}\boldsymbol{\alpha}_2\} + \delta_{i1}\delta_{i2}\log\{\mathbf{B}_{r3}(Y_{i2})^{\mathrm{T}}\boldsymbol{\alpha}_3\} \\
&\times\delta_{i1}\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}_i + (1-\delta_{i1})\delta_{i2}\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X}_i + \delta_{i1}\delta_{i2}\boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}_i + \delta_{i1}\delta_{i2}\log(1+\theta) - (1/\theta + \delta_{i1} + \delta_{i2}) \\
&\times\log[1 + \theta\{\mathbf{IB}_{r1}(Y_{i1})^{\mathrm{T}}\boldsymbol{\alpha}_1 \exp(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}_i) + \mathbf{IB}_{r2}(Y_{i1})^{\mathrm{T}}\boldsymbol{\alpha}_2 \exp(\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X}_i) \\
&+\{\mathbf{IB}_{r3}(Y_{i2}) - \mathbf{IB}_{r3}(Y_{i1})\}^{\mathrm{T}}\boldsymbol{\alpha}_3 \exp(\boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}_i)\}].
\end{aligned}$$

After specifying the B-spline bases, our goal is to find $\boldsymbol{\beta}_l, \boldsymbol{\alpha}_l, l = 1, 2, 3$ to maximize the log likelihood.

4.1 Simulation study and real data analysis

We perform 100 simulation study with sample size 1000 by using the proposed simulation procedure introduced in Section 2. We generate three covariates from normal distribution.

We specify $h_1 = 1, h_2 = 0.5, h_3 = 1, \kappa = 1/2$. The estimation standard deviations are obtained by using the sandwich estimator. The parameter estimation results are shown in Table 1. The estimated cumulative hazards are shown in Figure 5 and 4.

Table 1 verifies that the proposed B–spline method provides reasonable estimators that are close to the true values.

Further, we apply the B-spline method to analyze Pancreatic cancer data. To stabilize the computation, we divide the patients' survival time by 90 so that the survival times are between 0, 1. We use total 11 covariates. We report the results for the $\widehat{\boldsymbol{\beta}}$ and $\widehat{\theta}$. We use sandwich estimator to obtain the estimation variance and the confidence intervals. We compare the result with the one obtained by the Bayesian method develop by Lee et al. (2015). Table 2 shows the estimators $\widehat{\boldsymbol{\beta}}$ by using the B-spline method. The frailty parameter estimator is 0.51 with 95% confidence interval (0.32, 0.72). Table 3 shows the estimation results from the Bayesian method proposed by Lee et al. (2015). The corresponding frailty parameter estimator is 0.56 with 95% Bayesian confidence interval (0.30, 0.86). The estimators from the B–spline method and the Bayesian method are close to each other. Further, we compare the estimated cumulative hazards in Figure 5 and 4. The figures show that the estimated cumulative hazard are close to each other by using the two methods. These results suggest that the B–spline method and the Bayesian method produce in similar results without prior knowledge about the true parameters.

5 Conclusion

In this work, we develop the simulation and estimation procedures to implement Xu et al. (2010)'s method. We show that the proposed data generating mechanism is consistent with the likelihood introduced in Xu et al. (2010). We propose an EM algorithm and a B–spline based approach to stabilize the estimation procedures. The simulation and estimation are implemented in R. We use a simulation example to evaluate the proposed estimation methods. The simulation results show that the proposed estimation procedures could obtain estimators which are close to the truth. We also compare the results from using the B–spline method with the one from the Bayesian method developed by Lee et al. (2015). Without prior knowledge about the true parameters, the two methods give the similar estimation results.

References

De Boor, C. et al. (1978), *A practical guide to splines*, New York: Springer.

Fine, J. P., Jiang, H., and Chappell, R. (2001), "On semi-competing risks data," *Biometrika*, 88, 907–919.

Lee, K. H., Haneuse, S., Schrag, D., and Dominici, F. (2015), "Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis," *J. R. Stat. Soc. Ser. C Appl. Stat.*, 64, 253–273.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sørensen, T. I. (1992), "A counting process approach to maximum likelihood estimation in frailty models," *Scand. J. Statist.*, 25–43.

Su, Y.-R. and Wang, J.-L. (2012), "Modeling left-truncated and right-censored survival data with longitudinal covariates," *Ann. Statist.*, 40, 1465–1488.

Xu, J., Kalbfleisch, J. D., and Tai, B. (2010), "Statistical analysis of illness–death processes and semicompeting risks data," *Biometrics*, 66, 716–725.

Table 1: Results based on 100 simulation with 1000 sample sizes. The estimation standard deviations is the empirical standard deviation derived from the 100 simulation runs.

	True parameters	Estimators	Estimation standard deviations
θ	0.5	0.471	0.192
β_{11}	0.2	0.206	0.068
β_{12}	0.3	0.304	0.067
β_{13}	0.2	0.193	0.057
β_{21}	0.2	0.212	0.088
β_{22}	0.2	0.201	0.076
β_{23}	0.4	0.388	0.077
β_{31}	0.3	0.301	0.059
β_{32}	0.2	0.199	0.084
β_{33}	0.4	0.412	0.082

Table 2: Results, estimators and 95% confidence intervals, for the B-spline method.

$\widehat{\boldsymbol{\beta}}_1$			$\widehat{\boldsymbol{\beta}}_2$			$\widehat{\boldsymbol{\beta}}_3$		
$\widehat{\boldsymbol{\beta}_1}$	sd($\widehat{\boldsymbol{\beta}_1}$)	CI	$\widehat{\boldsymbol{\beta}_2}$	sd($\widehat{\boldsymbol{\beta}_2}$)	CI	$\widehat{\boldsymbol{\beta}_3}$	sd($\widehat{\boldsymbol{\beta}_3}$)	CI
-0.88	0.17	(-1.22, -0.56)	-2.31	0.27	(-2.83, -1.79)	-1.42	0.38	(-2.16. -0.68)
0.22	0.24	(-0.25, 0.69)	0.24	0.23	(-0.21, 0.70)	0.43	0.36	(-0.27, 1.13)
0.30	0.28	(-0.25, 0.84)	-0.004	0.34	(-0.67, 0.66)	0.03	0.43	(-0.82, 0.87)
-0.01	0.12	(-0.25,0.23)	0.50	0.13	(-0.25,0.75)	0.26	0.21	(-0.15, 0.67)
-0.10	0.05	(-0.19, 0.01)	0.09	0.05	(-0.007, 0.20)	0.15	0.09	(-0.02, 0.33)
-0.16	0.12	(-0.39, 0.06)	-0.31	0.13	(-0.56, -0.06)	-0.58	0.22	(-1.02, -0.15)
0.10	0.07	(-0.05, 0.25)	0.002	0.08	(-0.15, 0.16)	0.02	0.13	(-0.23, 0.28)
0.02	0.14	(-0.24, 0.29)	0.92	0.16	(0.61, 1.22)	0.33	0.24	(-0.14, 0.81)
-0.85	0.52	(-1.86,0.17)	2.90	0.24	(2.43,3.36)	1.87	0.61	(0.68, 3.07)
-0.54	0.21	(-0.96, -0.12)	1.61	0.18	(1.27, 1.96)	1.47	0.49	(0.51, 2.43)
0.28	0.29	(-0.30, 0.86)	1.43	0.34	(0.78, 2.09)	0.34	0.56	(-0.76, 1.43)

Table 3: Estimators, estimation standard deviations (sd) 95% confidence intervals, for the Bayesian method.

$\widehat{\boldsymbol{\beta}}_1$			$\widehat{\boldsymbol{\beta}}_2$			$\widehat{\boldsymbol{\beta}}_3$		
$\widehat{\boldsymbol{\beta}_1}$	sd($\widehat{\boldsymbol{\beta}_1}$)	CI	$\widehat{\boldsymbol{\beta}_2}$	sd($\widehat{\boldsymbol{\beta}_2}$)	CI	$\widehat{\boldsymbol{\beta}_3}$	sd($\widehat{\boldsymbol{\beta}_3}$)	CI
-0.92	0.19	(-1.29, -0.56)	-2.36	0.30	(-2.95, -1.81)	-1.50	0.35	(-2.27. -0.85)
0.21	0.27	(-0.33, 0.73)	0.25	0.24	(-0.21, 0.71)	0.37	0.40	(-0.49, 1.11)
0.27	0.29	(-0.30, 0.82)	-0.022	0.31	(-0.64, 0.57)	-0.06	0.44	(-0.98, 0.74)
-0.01	0.14	(-0.29,0.26)	0.49	0.14	(-0.24,0.77)	0.27	0.22	(-0.17, 0.71)
-0.10	0.05	(-0.20, -0.003)	0.10	0.05	(-0.004, 0.20)	0.16	0.08	(0.002, 0.31)
-0.17	0.13	(-0.43, 0.08)	-0.31	0.14	(-0.57, -0.04)	-0.62	0.22	(-1.05, -0.32)
0.10	0.08	(-0.05, 0.25)	0.001	0.09	(-0.18, 0.17)	0.03	0.13	(-0.23, 0.28)
0.02	0.15	(-0.28, 0.32)	0.91	0.20	(0.52, 1.29)	0.34	0.25	(-0.15, 0.83)
-1.01	0.55	(-2.19,-0.024)	2.85	0.25	(2.35,3.36)	1.84	0.70	(0.48, 3.17)
-0.56	0.23	(-1.01, -0.12)	1.61	0.21	(1.17, 2.02)	1.48	0.50	(0.70, 2.38)
0.27	0.31	(-0.34, 0.88)	1.42	0.35	(0.75, 2.10)	0.29	0.54	(-0.77, 1.32)

Figure 1: Concave likelihood function

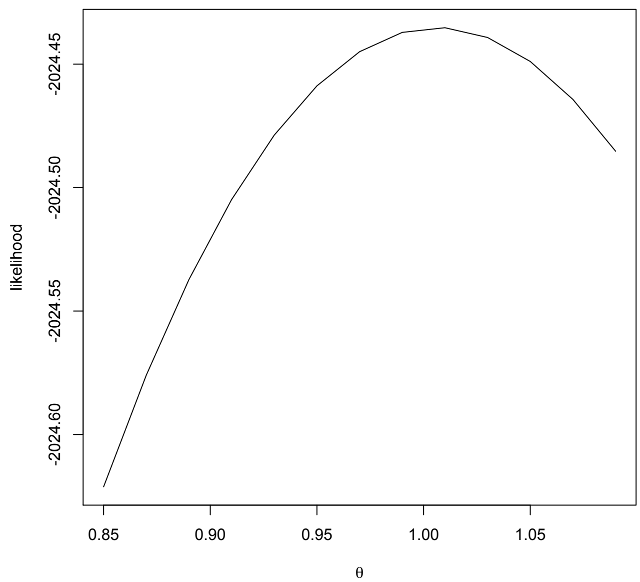


Figure 2: Concave likelihood function: data with left truncation.

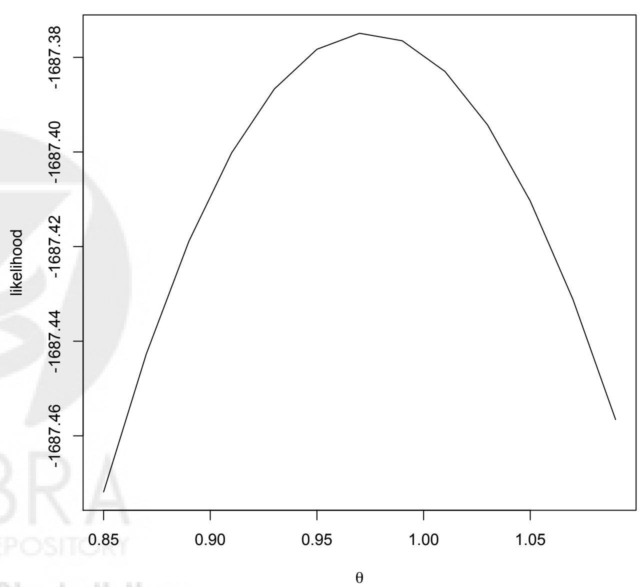
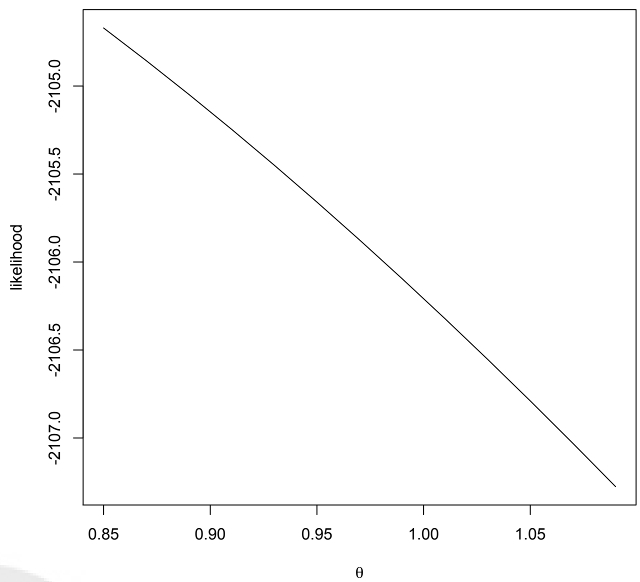


Figure 3: Monotone likelihood function



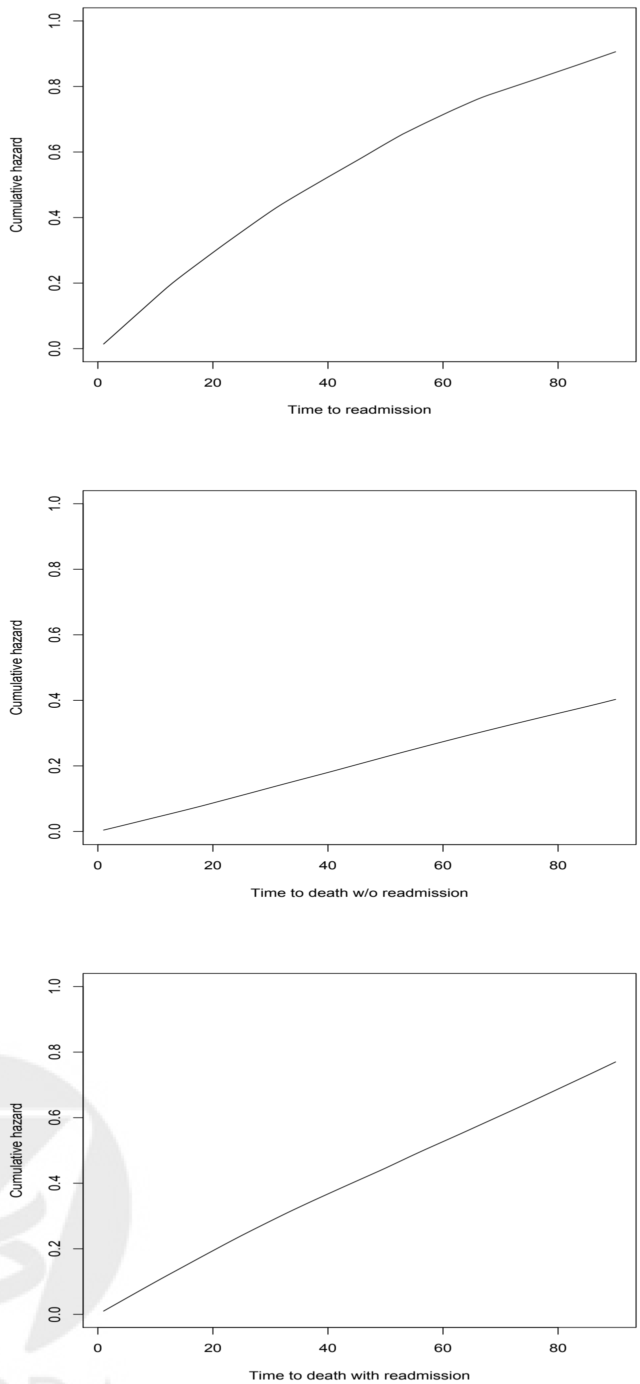


Figure 4: Estimation cumulative baseline hazard using the Bayesian method method.

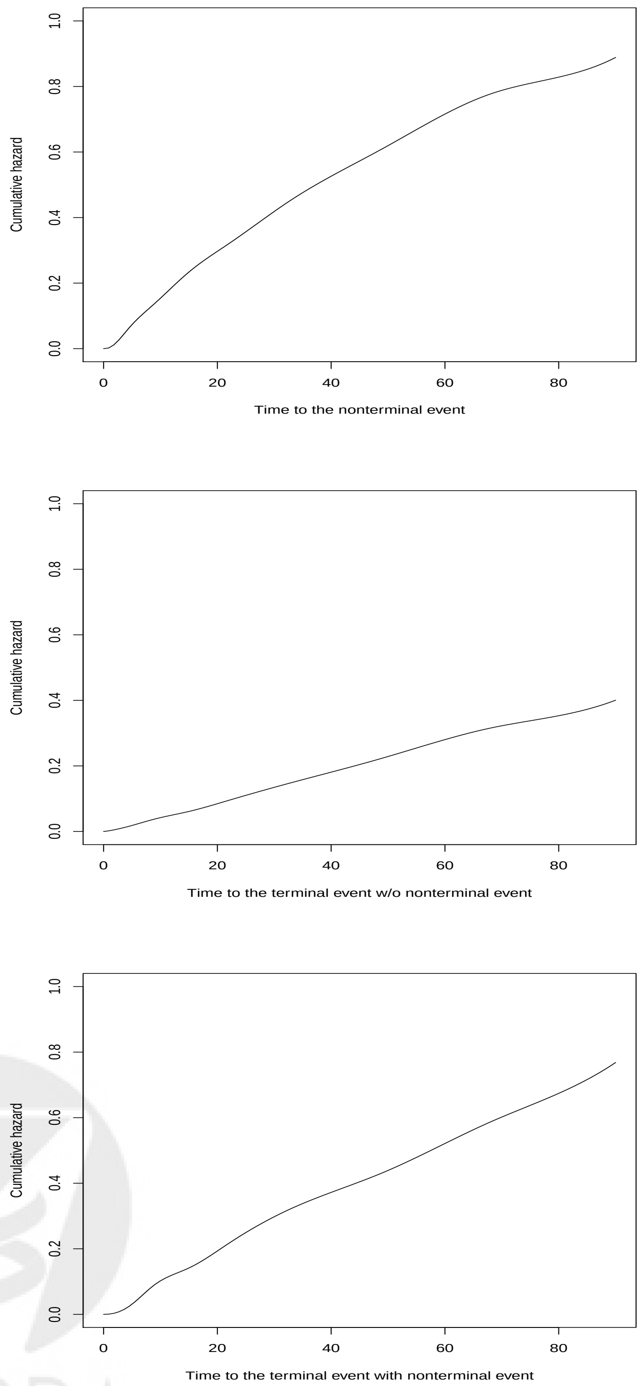


Figure 5: Estimation cumulative baseline hazard using the B–spline method.