

Bayesian inference of the fully specified subdistribution model for survival data with competing risks

Miaomiao Ge · Ming-Hui Chen

Received: 28 September 2011 / Accepted: 19 March 2012 / Published online: 8 April 2012
© Springer Science+Business Media, LLC 2012

Abstract Competing risks data are routinely encountered in various medical applications due to the fact that patients may die from different causes. Recently, several models have been proposed for fitting such survival data. In this paper, we develop a fully specified subdistribution model for survival data in the presence of competing risks via a subdistribution model for the primary cause of death and conditional distributions for other causes of death. Various properties of this fully specified subdistribution model have been examined. An efficient Gibbs sampling algorithm via latent variables is developed to carry out posterior computations. Deviance information criterion (DIC) and logarithm of the pseudomarginal likelihood (LPML) are used for model comparison. An extensive simulation study is carried out to examine the performance of DIC and LPML in comparing the cause-specific hazards model, the mixture model, and the fully specified subdistribution model. The proposed methodology is applied to analyze a real dataset from a prostate cancer study in detail.

Keywords Latent variables · Markov chain Monte Carlo · Partial likelihood · Proportional hazards

1 Introduction

Competing risks data are frequently encountered in various medical applications due to the fact that patients may die from different causes. Studies on this topic have been

M. Ge
Clinical Bio Statistics, Boehringer Ingelheim Pharmaceuticals, Inc.,
900 Ridgefield Road, Ridgefield, CT 06877, USA

M.-H. Chen (✉)
Department of Statistics, University of Connecticut, 215 Glenbrook Road,
U-4120, Storrs, CT 06269, USA
e-mail: ming-hui.chen@uconn.edu

active and productive. [Gail \(1975\)](#) proposed a multivariate model of failure times due to different causes. [Tsiatis \(1975\)](#) showed that for any joint distribution of n failure times there exists a joint distribution of n independent failure times such that the marginal cause-specific cumulative incident functions from the two joint distributions coincide, which implies that the correlations between the failure times are not identifiable in the multivariate failure time model. [Prentice et al. \(1978\)](#) introduced a cause-specific hazards model. [Larson and Dinse \(1985\)](#) established a mixture model with hazards function conditional on failure from a specific cause. [Fine and Gray \(1999\)](#) discussed the subdistribution model with proportional hazards assumption to assess the covariates effect on the cumulative incidence function of the cause of interest. Recently, [Fan \(2008\)](#) introduced a Bayesian nonparametric methodology based on the full likelihood for the proportional subdistribution hazards model. [Elashoff et al. \(2007, 2008\)](#) jointly modeled the longitudinal measurements and survival data with competing risks, where they extended respectively the cause-specific hazards model and the mixture model for survival data, and used latent random variables to link together the sub-models for longitudinal measurements and survival data.

The Bayesian literature on competing risks analysis is still sparse. [Fan \(2008\)](#) developed Bayesian methods by extending the subdistribution model of [Fine and Gray \(1999\)](#) for each cause-specific risk. More recently, [Hu et al. \(2009\)](#) and [Huang et al. \(2011\)](#) developed the Bayesian methods for a joint analysis of longitudinal measurements and survival data with competing risks, in which cause-specific hazards sub-models were considered for modeling survival times. As pointed out in [Fine and Gray \(1999\)](#), one of the nice properties of the subdistribution model is that the effect of a covariate on the marginal probability function can be directly assessed. However, the subdistribution model proposed by [Fine and Gray \(1999\)](#) cannot be compared to two other established models as the competing risks for other causes are not specified in their model. Due to this reason, we develop a fully specified subdistribution model with subdistribution hazard for the primary cause of death and conditional hazards for other causes of death. Under this fully specified subdistribution model, we are able to establish a theoretical connection between the partial likelihood of [Fine and Gray \(1999\)](#) and the one under the fully specified subdistribution model for the cause of primary interest when all failure times are observed. We notice that this connection may not be established under the models discussed in [Fan \(2008\)](#). With this new development, formal model comparisons between the fully specified subdistribution model and two other established models, namely, the cause-specific hazards model ([Prentice et al. 1978](#)) and the mixture model ([Larson and Dinse 1985](#)), can be carried out via Bayesian deviance information criterion (DIC) and logarithm of the pseudo-marginal likelihood (LPML). Furthermore, the fully specified subdistribution model also facilitates an efficient implementation of the Gibbs sampling algorithm.

The rest of the article is organized as follows. In Sect. 2, we present a detailed development of the fully specified subdistribution model and examine various properties of it. The prior and posterior are discussed and an efficient Gibbs sampling algorithm via a set of latent variables is developed in Sect. 3. In Sect. 4, we briefly review the cause-specific hazards model ([Prentice et al. 1978](#)) and the mixture model ([Larson and Dinse 1985](#)), and provide necessary mathematical formulations for DIC and LPML under these two models and the fully specified subdistribution model. In Sect. 5, we

present the design of a simulation study and the simulation algorithms for generating the data under the three competing risk models. We notice that these three competing risk models have never been formally compared based on our best knowledge. In Sect. 6, we analyze a real data from a prostate cancer study in detail. We conclude the paper with brief discussion and some extensions of the proposed model in Sect. 7.

2 Subdistribution based models for competing risks

2.1 Preliminary

We consider two competing risks throughout the paper and the extension to more than two competing risks will be discussed in Sect. 7. Let T_j be the time to failure due to cause j for $j = 1, 2$ and δ be the index of cause of death. Also let $T = \min\{T_1, T_2\}$. Assume cause 1 is the cause of primary interest. The subdistribution hazard for cause 1 defined in Gray (1988) is given as follows:

$$\begin{aligned} h_1(t) &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{\Pr(t \leq T \leq t + \Delta t, \delta = 1 | T \geq t \cup (T \leq t \cap \delta \neq 1))}{\Delta t} \right\} \\ &= \frac{\partial F_1(t)/\partial t}{1 - F_1(t)}, \end{aligned} \quad (2.1)$$

where $F_1(t) = \Pr(T \leq t, \delta = 1)$. As discussed in Fine and Gray (1999), to develop the regression model of (2.1) with the proportional hazards assumption, $h_1(t|\mathbf{x}) = h_{10}(t) \exp(\mathbf{x}'\boldsymbol{\beta}_1)$ and $F_1(t|\mathbf{x}) = 1 - \exp\left\{-\int_0^t h_{10}(u) \exp(\mathbf{x}'\boldsymbol{\beta}_1) du\right\}$, where \mathbf{x} is a vector of covariates and $\boldsymbol{\beta}_1$ is a vector of the corresponding regression coefficients. As pointed out in Fine and Gray (1999), the covariate effects can be directly assessed on the cumulative incidence function for primary cause under the subdistribution model. However, the distributions for failure times due to other causes are never specified in Fine and Gray (1999).

2.2 A fully specified subdistribution model for two competing risks

Let $T_j^* = T_j \times I(\delta = j) + \infty \times I(\delta \neq j)$, $j = 1, 2$, where we define $\infty \times 0 = 0$. Write $T^* = \min\{T_1^*, T_2^*\}$ as the observed time to failure. We propose the cause-specific cumulative incidence functions for both causes as follows:

$$\begin{aligned} F_1(t) &= \Pr(T^* \leq t, \delta = 1) = \Pr(T_1 \leq t, \delta = 1), \\ F_2(t) &= \Pr(T^* \leq t, \delta = 2) = M_2(t) \Pr(\delta = 2), \end{aligned} \quad (2.2)$$

where $M_2(t)$ is the cumulative incidence function conditional on cause 2. The fact that $\lim_{t \rightarrow \infty} F_j(t) = \Pr(\delta = j) < 1$ implies that $F_j(t)$ is improper. Yet $M_2(t)$ is proper due to $\lim_{t \rightarrow \infty} M_2(t) = \lim_{t \rightarrow \infty} [F_2(t)/\Pr(\delta = 2)] = 1$. Note that in (2.2), we do not directly model the correlation between T_1 and T_2 , which is not identifiable

as shown in Tsiatis (1975). Instead, $F_1(t)$ and $F_2(t)$ are related to each other via $Pr(\delta = 2) = 1 - Pr(\delta = 1) = 1 - F_1(\infty)$.

We apply the definition of subdistribution hazard in Fine and Gray (1999) for cause 1 by $h_1(t) = \frac{\partial F_1(t)}{\partial t} \times \frac{1}{1 - F_1(t)}$. Then $H_1(t)$ is improper because $\lim_{t \rightarrow \infty} H_1(t) = -\log [1 - \lim_{t \rightarrow \infty} F_1(t)] < \infty$. We specify a proportional hazards model with an improper baseline hazard function for $F_1(t|\mathbf{x})$ as

$$F_1(t|\mathbf{x}) = 1 - \exp \left\{ - H_{10}(t) \exp(\mathbf{x}'\boldsymbol{\beta}_1) \right\}. \tag{2.3}$$

For cause 2, we propose a proportional hazards model for $M_2(t|\mathbf{x})$ as

$$M_2(t|\mathbf{x}) = 1 - \exp \left\{ - H_{20}(t) \exp(\mathbf{x}'\boldsymbol{\beta}_2) \right\}. \tag{2.4}$$

The model defined by (2.3) and (2.4) is thus called the fully specified subdistribution (FS) model. Under the FS model, $Pr(\delta = 2|\mathbf{x}) = 1 - Pr(\delta = 1|\mathbf{x}) = \exp\{-H_{10}(\infty) \exp(\mathbf{x}'\boldsymbol{\beta}_1)\}$.

Assume there are n observations with the vector of observed time $\mathbf{t} = (t_1, t_2, \dots, t_n)'$, the matrix of covariates $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$, and the vector of cause indicator $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)'$, where δ_i takes possible values of 0, 1, and 2, corresponding to ‘‘censored’’, ‘‘died due to cause 1’’, and ‘‘died due to cause 2’’ for the i th subject, respectively. Under the model defined in (2.3) and (2.4), the likelihood function is given by

$$\begin{aligned} &L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_{10}, h_{20}|\mathbf{t}, X, \boldsymbol{\delta}) \\ &= \prod_{i=1}^n \left[h_{10}(t_i) \exp(\mathbf{x}'_i\boldsymbol{\beta}_1) \exp \left\{ - H_{10}(t_i) \exp(\mathbf{x}'_i\boldsymbol{\beta}_1) \right\} \right]^{I(\delta_i=1)} \\ &\quad \times \left[h_{20}(t_i) \exp(\mathbf{x}'_i\boldsymbol{\beta}_2) \exp \left\{ - H_{20}(t_i) \exp(\mathbf{x}'_i\boldsymbol{\beta}_2) - H_{10}(\infty) \exp(\mathbf{x}'_i\boldsymbol{\beta}_1) \right\} \right]^{I(\delta_i=2)} \\ &\quad \times \left[\exp \left\{ - H_{10}(t_i) \exp(\mathbf{x}'_i\boldsymbol{\beta}_1) \right\} \right. \\ &\quad \left. - \left(1 - \exp \left\{ - H_{20}(t_i) \exp(\mathbf{x}'_i\boldsymbol{\beta}_2) \right\} \right) \times \exp \left\{ - H_{10}(\infty) \exp(\mathbf{x}'_i\boldsymbol{\beta}_1) \right\} \right]^{I(\delta_i=0)}. \end{aligned} \tag{2.5}$$

2.3 Justification of Fine and Gray’s partial likelihood

The FS model is not only a natural expansion of the subdistribution model of Fine and Gray (1999) but also provides novel justifications of Fine and Gray’s partial likelihood under certain conditions. Assume there are n complete observations with the vector of observed time $\mathbf{t} = (t_1, t_2, \dots, t_n)'$, the matrix of covariates $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$, and the vector of cause indicator $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)'$. The partial likelihood of $\boldsymbol{\beta}_1$ for cause 1 given in Fine and Gray (1999) is of the form

$$L_p(\beta_1 | t, X, \delta) = \prod_{i=1}^n \left[\frac{\exp(\mathbf{x}'_i \beta_1)}{\sum_{j \in R_i^*} \exp(\mathbf{x}'_j \beta_1)} \right]^{I(\delta_i=1)}, \quad (2.6)$$

where R_i^* is defined as a special risk set at failure time t_i given by

$$R_i^* = \{j : (t_j \geq t_i) \cup (t_j \leq t_i \cap \delta_j \neq 1)\}. \quad (2.7)$$

Note that the risk set R_i^* is quite different than the risk set in Cox's partial likelihood (Cox 1972, 1975) as the patients who died from cause 2 before t_i are also included in R_i^* .

Three theorems are established below to show that the partial likelihood (2.6) can be obtained under the FS regression model via three different approaches with detailed proofs given in Appendix A. Denote D_1 as the number of deaths due to cause 1. Let $y_i = t_i$ when $\delta_i = 1$, and $y_i = \infty$ when $\delta_i \neq 1$. Write $\mathbf{y} = (y_{(1)}, y_{(2)}, \dots, y_{(n)})'$, where $0 = y_{(0)} < y_{(1)} < \dots < y_{(D_1)} < y_{(D_1+1)} = \dots = y_{(n)} = \infty$. Since all observations are failure times, the likelihood function of β_1 for cause 1 given the n complete observations is the part of the likelihood function in (2.5) involving β_1 :

$$L(\beta_1, h_{10} | \mathbf{y}, X, \delta) = \prod_{i=1}^n \left[h_{10}(y_{(i)}) \exp(\mathbf{x}'_i \beta_1) \right] \times \exp \left\{ - H_{10}(y_{(i)}) \exp(\mathbf{x}'_i \beta_1) \right\}.$$

Theorem 1 *With n complete observations, assume that in the FS model the baseline hazard rate h_{10} is zero after the last failure time due to cause 1. The partial likelihood function (2.6) can be attained by the profile likelihood approach, which is to plug in the profile maximum likelihood estimator of h_{10} in the likelihood function $L(\beta_1, h_{10} | \mathbf{y}, X, \delta)$.*

Theorem 2 *With n complete observations, assume that in the FS model the baseline hazard rate $h_{10}(t)$ is zero after the last failure time due to cause 1 and the prior of $h_{10}(t)$ is degenerate at 0 everywhere except at y_i 's when $\delta_i = 1$. Let $h_{10}(y_i) = \lambda_i$ when $\delta_i = 1$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{D_1})'$. We further assume independent Jeffreys-type priors for the λ_i 's, i.e., $\pi(\boldsymbol{\lambda}) \propto \prod_{i=1}^{D_1} 1/\lambda_i$. Then, the partial likelihood function (2.6) is obtained by*

$$L_p(\beta_1 | \mathbf{y}, X, \delta) \propto \int L(\beta_1, h_{10} | \mathbf{y}, X, \delta) \pi(\boldsymbol{\lambda}) d\lambda_1 \dots d\lambda_{D_1},$$

where $L(\beta_1, h_{10} | \mathbf{y}, X, \delta)$ is defined in (2.8).

Theorem 3 *With n complete observations, assume that in the FS model the baseline hazard rate $h_{10}(t)$ is zero after the last failure time due to cause 1 and that $H_{10}(t)$ has a Gamma process prior, i.e., $h_{1i} \sim \text{Gamma}(c_0 h_{0i}, c_0)$, where $c_0 > 0$, $h_{1i} = H_{10}(y_{(i)}) - H_{10}(y_{(i-1)})$ for $i = 1, 2, \dots, D_1$, $h_{1, D_1+1} = \dots = h_{1n} = 0$, $h_{0i} = H_0(y_{(i)}) - H_0(y_{(i-1)})$, $H_0(y)$ is increasing and differentiable at y_1, \dots, y_{D_1}*

with $H_0(0) = 0$, and the h_{1i} are independent of each other. Then the partial likelihood function (2.6) can be approximated by

$$L_p(\beta_1|y, X, \delta) \approx \lim_{c_0 \downarrow 0} g(c_0) E^{h_{10}} [L(\beta_1, h_{10}|y, X, \delta)],$$

where $g(c_0)$ is a function of c_0 , which is free from β_1 .

Fine and Gray (1999) also showed that the partial likelihood arises from complete data using a certain reduced data structure, without any assumptions on the models for the subdistribution for other causes. The results established in the above theorems give insight into Fine and Gray’s partial likelihood.

3 Prior, posterior, and computational development

3.1 Prior and posterior

For the sake of simpler calculation, a special case of the gamma process prior for the cumulative baseline hazard function assumed in Theorem 3 is considered here for the FS model. Assume the baseline hazard functions respectively have piecewise constant forms, which are, with $K_j + 1$ partitions of the time axis, $0 = s_{j0} < s_{j1} < s_{j2} < \dots < s_{jK_j} < \infty$,

$$\begin{cases} h_{10}(t) = \lambda_{1k}, & s_{1,k-1} < t \leq s_{1k}, & k = 1, 2, \dots, K_1; \\ h_{10}(t) = \lambda_{1,K_1+1} \exp\{-(t - s_{1K_1})\}, & t > s_{1K_1}; \\ h_{20}(t) = \lambda_{2k}, & s_{2,k-1} < t \leq s_{2k}, & k = 1, 2, \dots, K_2; \\ h_{20}(t) = \lambda_{2K_2}, & t > s_{1K_2}. \end{cases} \tag{3.1}$$

To construct posterior distributions for the unknown parameters, we assume β_j follows an improper uniform prior, λ_{jk} follows a Jeffreys-type prior, and λ_{1,K_1+1} follows a gamma prior. We further assume that β_j , λ_{jk} , and λ_{1,K_1+1} are independent for $k = 1, \dots, K_j$ and $j = 1, 2$. Let $\lambda_1 = (\lambda_{11}, \lambda_{12}, \dots, \lambda_{1,K_1+1})'$ and $\lambda_2 = (\lambda_{21}, \lambda_{22}, \dots, \lambda_{2K_2})'$, then the joint prior of $(\beta_1, \beta_2, \lambda_1, \lambda_2)$ is specified as follows

$$\pi(\beta_1, \beta_2, \lambda_1, \lambda_2) = \pi(\beta_1)\pi(\beta_2)\pi(\lambda_1)\pi(\lambda_2) \propto \left[\prod_{j=1}^2 \prod_{k=1}^{K_j} \frac{1}{\lambda_{jk}} \right] \pi(\lambda_{1,K_1+1}), \tag{3.2}$$

where $\pi(\lambda_{1,K_1+1}) \propto \lambda_{1,K_1+1}^{a-1} \exp(-b\lambda_{1,K_1+1})$ with $a > 0$ and $b > 0$, which are prespecified hyperparameters. The joint posterior distribution is given by

$$\begin{aligned} \pi(\beta_1, \beta_2, \lambda_1, \lambda_2|t, X, \delta) &\propto \pi^*(\beta_1, \beta_2, \lambda_1, \lambda_2|t, X, \delta) \\ &\equiv L(\beta_1, \beta_2, \lambda_1, \lambda_2|t, X, \delta)\pi(\beta_1, \beta_2, \lambda_1, \lambda_2), \end{aligned} \tag{3.3}$$

where $L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 | \mathbf{t}, X, \boldsymbol{\delta})$ is given by (2.5) with $h_{10}(t)$ and $h_{20}(t)$ defined in (3.1). Recently, Wang et al. (2012) established a theoretical connection between the gamma process prior specified for the cumulative baseline function and the independent gamma priors for the λ_{jk} for the interval-censored survival data. Also, the independent gamma priors assumed for the baseline hazard function approximate the gamma process prior when $c_0 \rightarrow 0+$. Thus, the priors in (3.2) for the λ_{jk} can be considered as a special case of the gamma process priors specified for the cumulative baseline function in this sense.

Let $v_{jik} = 1$ if the i th subject failed or was censored in the k th interval $(s_{j,k-1}, s_{jk}]$, and 0 otherwise for $k = 1, 2, \dots, K_j + 1$, and $i = 1, 2, \dots, n$, where $s_{j,K_j+1} = \infty$, for $j = 1, 2$. Also let X_j be a matrix with its i th row equal to $I(\delta_i = j)(v_{ji1}, \dots, v_{jiK_j}, \mathbf{x}'_i)$ for $j = 1, 2$. Then, we are led to the following theorem regarding the propriety of the posterior distribution of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ with an improper prior given by (3.2).

Theorem 4 Assume that (i) when $\delta_i > 0$, $t_i > 0$ and for $k = 1, 2, \dots, K_j$ for $j = 1, 2$, and (ii) X_1 and X_2 are of full rank. Then, the posterior distribution $\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 | \mathbf{t}, X, \boldsymbol{\delta})$ in (3.3) with the prior specified in (3.2) is proper.

A proof of Theorem 4 is given in Appendix A. Theorem 4 gives very mild conditions for ensuring propriety of the joint posterior distribution of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ under the fully specified subdistribution model. The conditions (i) and (ii) essentially require that all event times are strictly positive, at least one event occurs in each chosen interval $(s_{j,k-1}, s_{jk}]$, and the corresponding covariate matrix is of full rank. Notice that we do not require any events for the last interval (s_{1K_1}, ∞) for the primary cause as we specify a proper prior for λ_{1,K_1+1} . These conditions are easily satisfied in most applications and are quite easy-to-check. Under certain additional conditions, we can also show that when $\pi(\lambda_{1,K_1+1}) \propto 1$ in (3.2), the resulting posterior of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ is still proper. We also notice that following Chen et al. (2006), the posterior propriety by assuming full gamma process priors on H_{10} and H_{20} for $c_0 > 0$ can be established and, however, stronger propriety conditions are required in this case.

3.2 Computational development

Due to the complexity of the likelihood structure of the FS model, an analytical evaluation of the posterior distribution does not appear to be possible. In order to carry out posterior inference, we adopt Markov chain Monte Carlo (MCMC) methods and develop a computationally efficient Gibbs sampling algorithm to sample from the posterior distribution in (3.3).

In order to avoid the complicated form in the censored part of the likelihood function, we introduce a latent variable η_i to indicate whether or not subject i would eventually fail from cause 1 or 2 and another latent variable u_i to be the failure time such that $u_i \geq t_i$ when subject i was censored at t_i and $\eta_i = 1$. Then the complete likelihood is constructed as

$$\begin{aligned}
 &L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_{10}, h_{20}|t, X, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{u}) \\
 &= \prod_{i=1}^n \left[h_{10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \exp \left\{ - H_{10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} \right]^{I(\delta_i=1)} \\
 &\quad \times \left[h_{20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \exp \left\{ - H_{20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) - H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} \right]^{I(\delta_i=2)} \\
 &\quad \times \left[h_{10}(u_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \exp \left\{ - H_{10}(u_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} \right]^{I(\delta_i=0, \eta_i=1)} \\
 &\quad \times \left[\exp \left\{ - H_{20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) - H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} \right]^{I(\delta_i=0, \eta_i=2)}, \tag{3.4}
 \end{aligned}$$

where $\boldsymbol{\eta} = (\eta_i : \delta_i = 0, 1 \leq i \leq n)$ and $\mathbf{u} = (u_i : \delta_i = 0, \eta_i = 1, 1 \leq i \leq n)$. Based on the complete data likelihood, the augmented posterior of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u})$ is given by

$$\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}|t, X, \boldsymbol{\delta}) \propto L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_{10}, h_{20}|t, X, \boldsymbol{\delta}, \boldsymbol{\eta}, \mathbf{u})\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2), \tag{3.5}$$

where $h_{10}(t)$ and $h_{20}(t)$ are defined in (3.1). It is easy to show that

$$\sum_{\boldsymbol{\eta}} \int \pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}|t, X, \boldsymbol{\delta}) d\mathbf{u} = \pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2|t, X, \boldsymbol{\delta}),$$

where $\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2|t, X, \boldsymbol{\delta})$ is the posterior given in (3.3). This result ensures that whenever

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}) \sim \pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}|t, X, \boldsymbol{\delta}), \text{ then } (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \sim \pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2|t, X, \boldsymbol{\delta}).$$

The introduction of the latent variables $\boldsymbol{\eta}$ and \mathbf{u} greatly facilitates a convenient implementation of the Gibbs sampling algorithm. To develop an efficient Gibbs sampling algorithm, we use the collapsed Gibbs method of Liu (1994). First, we group $(\boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u})$ together. Then, the Gibbs sampling algorithm requires to sample from the following conditional posterior distributions in turn: (i) $[\boldsymbol{\beta}_1|\boldsymbol{\lambda}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}, t, X, \boldsymbol{\delta}]$; (ii) $[\boldsymbol{\lambda}_1|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}, t, X, \boldsymbol{\delta}]$; and (iii) $[\boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}|\boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, t, X, \boldsymbol{\delta}]$. For (i), the conditional posterior density of $\boldsymbol{\beta}_1$ given $(\boldsymbol{\lambda}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}, t, X, \boldsymbol{\delta})$ is log-concave in each component of $\boldsymbol{\beta}_1$. Thus, we can use the adaptive rejection algorithm of Gilks and Wild (1992) to sample $\boldsymbol{\beta}_1$. For (ii), it can be shown that given $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}, t, X, \boldsymbol{\delta})$, the λ_{1k} 's are conditionally independent and each of them follows a gamma distribution. Thus, sampling $\boldsymbol{\lambda}_1$ is straightforward. For (iii), it is easy to see that

$$[\boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, \mathbf{u}|\boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, t, X, \boldsymbol{\delta}] = [\boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}|\boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, t, X, \boldsymbol{\delta}][\mathbf{u}|\boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\eta}, t, X, \boldsymbol{\delta}]. \tag{3.6}$$

In (3.6), we collapse out \mathbf{u} in the conditional distribution $[\boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta}|\boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, t, X, \boldsymbol{\delta}]$. However, jointly sampling $(\boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \boldsymbol{\eta})$ from their conditional distribution is not

possible. Thus, we need to run a sub-Gibbs sampling algorithm to sample $(\beta_2, \lambda_2, \eta)$ from this conditional posterior distribution. The approach is called the modified collapsed Gibbs sampling algorithm. As shown in [Chen et al. \(2000\)](#), the modified collapsed Gibbs sampling algorithm yields the target posterior as its stationary distribution. The sub-Gibbs sampling algorithm requires to sample from the following three additional conditional posterior distributions in turns: (iiia) $[\beta_2 | \beta_1, \lambda_1, \lambda_2, \eta, t, X, \delta]$; (iiib) $[\lambda_2 | \beta_1, \lambda_1, \beta_2, \eta, t, X, \delta]$; and (iiic) $[\eta | \beta_1, \lambda_1, \beta_2, \lambda_2, t, X, \delta]$. For (iiia) the conditional posterior density of β_2 is log-concave in each component of β_2 and we again use adaptive rejection algorithm of [Gilks and Wild \(1992\)](#) to sample β_2 . For (iiib), the λ_{1k} 's are conditionally independent and each of them follows a gamma distribution. Finally, for (iiic), η_i 's are conditionally independent and each η_i follows a Bernoulli distribution. The technical detail of sampling η from $[\eta | \beta_1, \lambda_1, \beta_2, \lambda_2, t, X, \delta]$ and sampling \mathbf{u} from $[\mathbf{u} | \beta_1, \lambda_1, \lambda_2, \eta, t, X, \delta]$ is given in Appendix B.

4 Model comparison

The cause-specific hazards model and the mixture model are two well-established models for competing risks survival data. We discuss details of these two models and compare them with the FS model both theoretically, in simulation, and in an analysis of a real dataset.

4.1 Other models for comparison

Cause-specific Hazards Model As discussed in [Gaynor et al. \(1993\)](#), the cause-specific hazard function is denoted by

$$h_{Cj}(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{\Pr(t \leq T < t + \Delta t, \delta = j | T \geq t)}{\Delta t} \right\}, \quad j = 1, 2.$$

Under the proportional hazards assumption, $h_{Cj}(t|\mathbf{x}) = h_{Cj0}(t) \exp(\mathbf{x}'\beta_j)$ and the cumulative incidence function of cause j is given by

$$F_j(t|\mathbf{x}) = \Pr(T \leq t, \delta = j) = \int_0^t h_{Cj0}(u) \exp(\mathbf{x}'\beta_j) \\ \times \exp \left\{ - \sum_{j=1}^2 H_{Cj0}(u) \exp(\mathbf{x}'\beta_j) \right\} du.$$

The covariate effects can be directly assessed on the cause-specific hazard functions. But they cannot be directly estimated by β_j alone on the cause-specific cumulative incidence function of cause j , as the cause-specific cumulative incidence function of cause 1 also depends on regression coefficients β_2 for cause 2.

Mixture Model Larson and Dinse (1985) discussed the mixture model. Assume the types of cause-specific failures follow a multinomial distribution. Define the probability of failing from cause j as $p_j = Pr(\delta = j)$ for $j = 1, 2$, where $p_1 + p_2 = 1$. Define $h_{Mj}(t)$ as the hazard function conditional on failure from cause j ,

$$h_{Mj}(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{Pr(t \leq T < t + \Delta t | \delta = j, T \geq t)}{\Delta t} \right\}.$$

Under the proportional hazards assumption, $h_{Mj}(t|\mathbf{x}) = h_{Mj0}(t) \exp(\mathbf{x}'\boldsymbol{\beta}_j)$ and the cause-specific cumulative incidence function is given by

$$F_j(t|\mathbf{x}) = Pr(T \leq t, \delta = j|\mathbf{x}) = p_j \left(1 - \exp \left\{ - H_{Mj0}(t) \exp(\mathbf{x}'\boldsymbol{\beta}_j) \right\} \right).$$

It is observed that both cause-specific hazard functions and cause-specific cumulative incidence functions depend on regression coefficients of the corresponding cause as well as the probability of failing from that cause.

Notice that the definition of the subdistribution hazard is different than the definition of the cause-specific hazard function or the definition of the conditional hazard function in the mixture model. Thus, if the proportional structure on the hazard function of one model is true, the hazard functions of other two models could never achieve the Cox proportional hazards assumption.

4.2 Model comparison measures

Deviance information criterion (DIC) (Spiegelhalter et al. 2002) and logarithm of the pseudomarginal likelihood (LPML) (Ibrahim et al. 2001) are used here to compare the cause-specific hazards model, the mixture model, and the FS model. Let $\boldsymbol{\theta}$ denote a collection of model parameters. DIC is defined as $DIC = D(\hat{\boldsymbol{\theta}}) + 2p_D$, where $D(\boldsymbol{\theta})$ is a deviance function, $p_D = \bar{D} - D(\hat{\boldsymbol{\theta}})$, \bar{D} and $\hat{\boldsymbol{\theta}}$ are the posterior means of $D(\boldsymbol{\theta})$ and $\boldsymbol{\theta}$. The formula of LPML is given by $LPML = \sum_{i=1}^n \log(CPO_i)$, where the Conditional Predictive Ordinate (CPO), $CPO_i = f(t_i|\mathbf{x}_i, D^{(i)}) = \int f(t_i|\boldsymbol{\theta}, \mathbf{x}_i)\pi(\boldsymbol{\theta}|D^{(i)})$, $D^{(i)}$ is the data with the i th observation deleted, and $\pi(\boldsymbol{\theta}|D^{(i)})$ is the posterior distribution based on the data $D^{(i)}$. According to Gelfand and Dey (1994), LPML implicitly includes a similar dimensional penalty as AIC asymptotically.

For the proposed FS model, $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2)$, and the deviance function $D(\boldsymbol{\theta})$ is given by $D(\boldsymbol{\theta}) = -2 \log L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_{10}, h_{20}|\mathbf{t}, X, \boldsymbol{\delta})$, where $L(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_{10}, h_{20}|\mathbf{t}, X, \boldsymbol{\delta})$ is given by (2.5) and $h_{10}(t)$ and $h_{20}(t)$ are defined in (3.1).

For the cause-specific hazards model, suppose the piecewise exponential models for $h_{Cj0}(t)$, $j = 1, 2$. The deviance function is defined by

$$D(\boldsymbol{\theta}) = -2 \log L_C(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_{C10}, h_{C20}|\mathbf{t}, X, \boldsymbol{\delta}),$$

where

$$\begin{aligned}
 L_C(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_{C10}, h_{C20} | \mathbf{t}, X, \boldsymbol{\delta}) &= \prod_{i=1}^n \left\{ h_{C10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\}^{I(\delta_i=1)} \\
 &\quad \times \left\{ h_{C20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \right\}^{I(\delta_i=2)} \\
 &\quad \times \exp \left\{ -H_{C10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right. \\
 &\quad \left. - H_{C20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \right\}.
 \end{aligned}$$

From the above likelihood function, it is easy to see that under the cause-specific hazards model, $\text{DIC} = \text{DIC}_1 + \text{DIC}_2$, where DIC_j is the DIC of the survival model with single cause j by treating other causes of death as censored.

For the mixture model, let p_{1i} denote the probability of death due to cause 1 for the i th subject, $\mathbf{p}_1 = (p_{11}, p_{12}, \dots, p_{1n})'$. The likelihood function is

$$\begin{aligned}
 L_M(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, h_{M10}, h_{M20}, \mathbf{p}_1 | \mathbf{t}, X, \boldsymbol{\delta}) \\
 &= \prod_{i=1}^n \left[p_{1i} h_{M10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \exp \left\{ -H_{M10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} \right]^{I(\delta_i=1)} \\
 &\quad \times \left[(1 - p_{1i}) h_{M20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \exp \left\{ -H_{M20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \right\} \right]^{I(\delta_i=2)} \\
 &\quad \times \left[p_{1i} \exp \left\{ -H_{M10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} \right. \\
 &\quad \left. + (1 - p_{1i}) \exp \left\{ -H_{M20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \right\} \right]^{I(\delta_i=0)}.
 \end{aligned}$$

Assume $p_{1i} = \frac{\exp(\mathbf{z}'_i \boldsymbol{\phi})}{1 + \exp(\mathbf{z}'_i \boldsymbol{\phi})}$, where \mathbf{z}_i is a vector of covariates, which may be a subset of \mathbf{x}_i .

For the cause-specific hazards model and mixture model, the form of the baseline hazard functions h_{Cj0} and h_{Mj0} and the prior of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ are assumed in the same way as those in the FS model excluding the last piece $h_{10}(t)$ when $t \geq s_{1, K_1}$. Similar to the FS model, the propriety of the joint posterior distribution of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ can also be established under an improper joint prior $\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$, which is similar to (3.2).

5 A simulation study

In this section, we carry out a simulation study to compare the cause-specific hazards model, mixture model, and fully specified subdistribution model via DIC and LPML. To generate data from the FS model, we assume there are two causes with cause 1 to be the cause of interest, and there are two covariates with true parameters $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})'$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22})'$, which are chosen such that $Pr(\delta = 1)$ is around 1/3. Covariates x_{i1} are generated from $N(0, 1)$ and x_{i2} given x_{i1} are generated

from Bernoulli($p(x_{i1})$), where $p(x_{i1}) = \frac{\exp\{0.5+0.3x_{i1}\}}{1+\exp\{0.5+0.3x_{i1}\}}$. Assume the failure times of two causes follow distinct piecewise exponential distributions, where for cause 1 the time is partitioned as $s_{10} = 0, s_{11} = 8, s_{12} = 12, s_{13} = 15, s_{14} = 16,$ and $s_{15} = 17$ with corresponding $\lambda_1 = (0.001, 0.01, 0.03, 0.02, 0.3)'$, and for cause 2 the time is partitioned as $s_{20} = 0, s_{21} = 3, s_{22} = 5, s_{23} = 8, s_{24} = 10, s_{25} = 11, s_{26} = 12, s_{27} = 13, s_{28} = 15, s_{29} = 17,$ and $s_{210} = 18$ with corresponding $\lambda_2 = (0.001, 0.005, 0.01, 0.02, 0.04, 0.07, 0.1, 0.15, 0.2, 1.0)'$. Generate r_i from $U(0, 1)$. If $r_i < Pr(\delta_i = 1|\mathbf{x}_i)$, then generate t_i^* of cause 1 from a piecewise exponential distribution

$$f_1(t_i^*|\mathbf{x}_i) = \frac{h_{10}(t_i^*) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \exp \left\{ - H_{10}(t_i^*) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\}}{1 - \exp \left\{ - H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\}}.$$

If $r_i \geq Pr(\delta_i = 1|\mathbf{x}_i)$, then generate t_i^* of cause 2 from a piecewise exponential distribution

$$f_2(t_i^*|\mathbf{x}_i) = \frac{h_{20}(t_i^*) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \exp \left\{ - H_{20}(t_i^*) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) - H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\}}{\exp \left\{ - H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} \left[1 - \exp \left\{ - H_{20}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \right\} \right]}.$$

The censoring time c_i is generated from a uniform distribution, $U(a_c, b_c)$, where $0 < a_c < b_c$ are chosen so that the proportion of death is around 2/5, and then t_i is taken to be $t_i = \min\{t_i^*, c_i\}$. Note that under the FS model, $Pr(\delta_i = 1|\mathbf{x}_i) = F_1(\infty|\mathbf{x}_i) = 1 - \left\{ - H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\}$.

For the mixture model, the settings of model parameters and covariates are similar to those for the FS model, while $Pr(\delta = 1)$ is calculated by $p_1 = \frac{\exp(\mathbf{z}'\boldsymbol{\phi})}{1+\exp(\mathbf{z}'\boldsymbol{\phi})}$, where $\boldsymbol{\phi} = (\phi_1, \phi_2)'$ is chosen such that p_1 is around 1/3. If r_i from $U(0, 1)$ falls in $(0, Pr(\delta = 1))$, then t_i^* of cause 1 is generated from a piecewise exponential distribution

$$f_1(t_i^*|\mathbf{x}_i) = h_{10}(t_i^*) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \exp \left\{ - H_{10}(t_i^*) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\}.$$

Otherwise, t_i^* of cause 2 is generated from a piecewise exponential distribution

$$f_2(t_i^*|\mathbf{x}_i) = h_{20}(t_i^*) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \exp \left\{ - H_{20}(t_i^*) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) \right\}.$$

For the cause-specific hazards model, the parameter and covariates settings are also similar as above. According to [Lu and Tsiatis \(2001\)](#), t_i is generated by $t_i = \min\{t_{1i}, t_{2i}, c_i\}$, where t_{1i}, t_{2i} , and c_i are generated independently, respectively, from a piecewise exponential distribution

$$f_1(t_{1i}|\mathbf{x}_i) = h_{10}(t_{1i}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \exp \left\{ - H_{10}(t_{1i}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\},$$

Table 1 Frequency of ranking each model as best under different scenarios

(K_1, K_2)	DIC			LPML		
	FS best	C best	M best	FS best	C best	M best
Data simulated from FS model						
(5, 10)	0.626	0.002	0.372	0.608	0.000	0.392
(10, 20)	0.764	0.004	0.232	0.778	0.008	0.214
(15, 30)	0.892	0.006	0.102	0.918	0.006	0.076
Data simulated from C model						
(5, 10)	0.194	0.524	0.282	0.190	0.482	0.328
(10, 20)	0.130	0.722	0.148	0.132	0.716	0.152
(15, 30)	0.156	0.732	0.112	0.158	0.726	0.116
Data simulated from M model						
(5, 10)	0.096	0.044	0.860	0.100	0.072	0.828
(10, 20)	0.104	0.094	0.802	0.110	0.142	0.748
(15, 30)	0.098	0.154	0.748	0.120	0.192	0.688

a piecewise exponential distribution

$$f_2(t_{2i} | \mathbf{x}_i) = h_{20}(t_{2i}) \exp(\mathbf{x}_i' \boldsymbol{\beta}_2) \exp \left\{ - H_{20}(t_{2i}) \exp(\mathbf{x}_i' \boldsymbol{\beta}_2) \right\},$$

and a uniform distribution, $U(a_c, b_c)$, such that the proportion of death is around 2/5.

500 data sets with $n = 500$ observations in each dataset were generated from each of the three models, respectively, as described above. Each simulated dataset was fitted by all three models and the corresponding DICs and LPMLs were calculated. From Table 1, we see that the best model chosen by DIC and LPML is always consistent with the true model where the data were simulated. The mean DIC and mean LPML under each model and scenario are shown in Table 2. The true value, estimate, standard deviation, mean square error, and coverage probability of each covariate coefficient when estimating from the true model for all of the three models are given in Table 3. It is observed that the standard deviations and mean square errors are moderate and stable, and that the coverage probabilities are always around 0.95 under all three scenarios of (K_1, K_2) .

To further compare the models, the median and interquartile range (IQR) of the pairwise differences of DIC and LPML for two models under each scenario are calculated, and the corresponding boxplots are shown in Fig. 1. When the data were simulated from another model, the mixture model and fully specified subdistribution model fit better than the cause-specific hazards model.

6 Analysis of the prostate cancer data

A subset of data from the prostate cancer studies published in Choueiri et al. (2010) is analyzed using the three models. The response variable was the time from prostate-specific antigen (PSA) failure to death or to the last follow-up, whichever came

Table 2 Mean of DIC and LPML of the three models

(K_1, K_2)	DIC			LPML		
	FS model	C model	M model	FS model	C model	M model
Data simulated from FS model						
(5, 10)	2142.0	2167.1	2144.2	-1071.1	-1083.7	-1072.0
(10, 20)	2119.2	2145.6	2123.2	-1060.6	-1073.9	-1062.6
(15, 30)	2126.5	2153.9	2131.8	-1065.8	-1079.6	-1068.4
Data simulated from C model						
(5, 10)	2064.8	2055.8	2056.6	-1032.5	-1028.0	-1028.2
(10, 20)	2039.0	2030.7	2033.4	-1020.5	-1016.4	-1017.8
(15, 30)	2044.6	2037.5	2040.8	-1024.9	-1021.4	-1023.0
Data simulated from M model						
(5, 10)	2286.8	2283.7	2274.7	-1144.0	-1142.3	-1137.7
(10, 20)	2285.9	2282.3	2276.1	-1144.4	-1142.4	-1139.5
(15, 30)	2295.9	2291.8	2287.0	-1150.9	-1148.7	-1146.5

Bold values are the best values of DIC or LPML corresponding to the best model

Table 3 Posterior estimates of β under the three models in simulation studies

$(K_1, K_2) =$		FS model			C model			M model		
		(5, 10)	(10, 20)	(15, 30)	(5, 10)	(10, 20)	(15, 30)	(5, 10)	(10, 20)	(15, 30)
β_{11}	True	0.2			0.2			0.2		
	Est	0.2	0.2	0.2	0.18	0.19	0.19	0.19	0.20	0.20
	SD	0.09	0.09	0.09	0.11	0.11	0.11	0.13	0.13	0.13
	MSE	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02
	CP	0.94	0.94	0.94	0.95	0.95	0.95	0.94	0.94	0.94
β_{12}	True	0.8			1.0			1.5		
	Est	0.78	0.78	0.78	1.00	1.02	1.02	1.47	1.48	1.48
	SD	0.23	0.23	0.23	0.24	0.24	0.24	0.35	0.36	0.37
	MSE	0.05	0.05	0.06	0.06	0.06	0.06	0.14	0.15	0.16
	CP	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.94
β_{21}	True	0.3			0.3			0.3		
	Est	0.29	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
	SD	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	MSE	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	CP	0.97	0.96	0.96	0.95	0.95	0.95	0.95	0.95	0.95
β_{22}	True	1.0			0.2			0.5		
	Est	0.98	1.00	1.00	0.19	0.20	0.20	0.53	0.53	0.54
	SD	0.17	0.17	0.17	0.15	0.15	0.15	0.18	0.18	0.18
	MSE	0.03	0.03	0.03	0.02	0.02	0.02	0.03	0.03	0.03
	CP	0.94	0.95	0.95	0.96	0.95	0.96	0.96	0.96	0.96

Note that Est, SD, MSE, and CP denote the average of the posterior means, the average of the posterior standard deviations, and the mean square error, and the coverage probability of the 95 % HPD intervals over 500 simulations

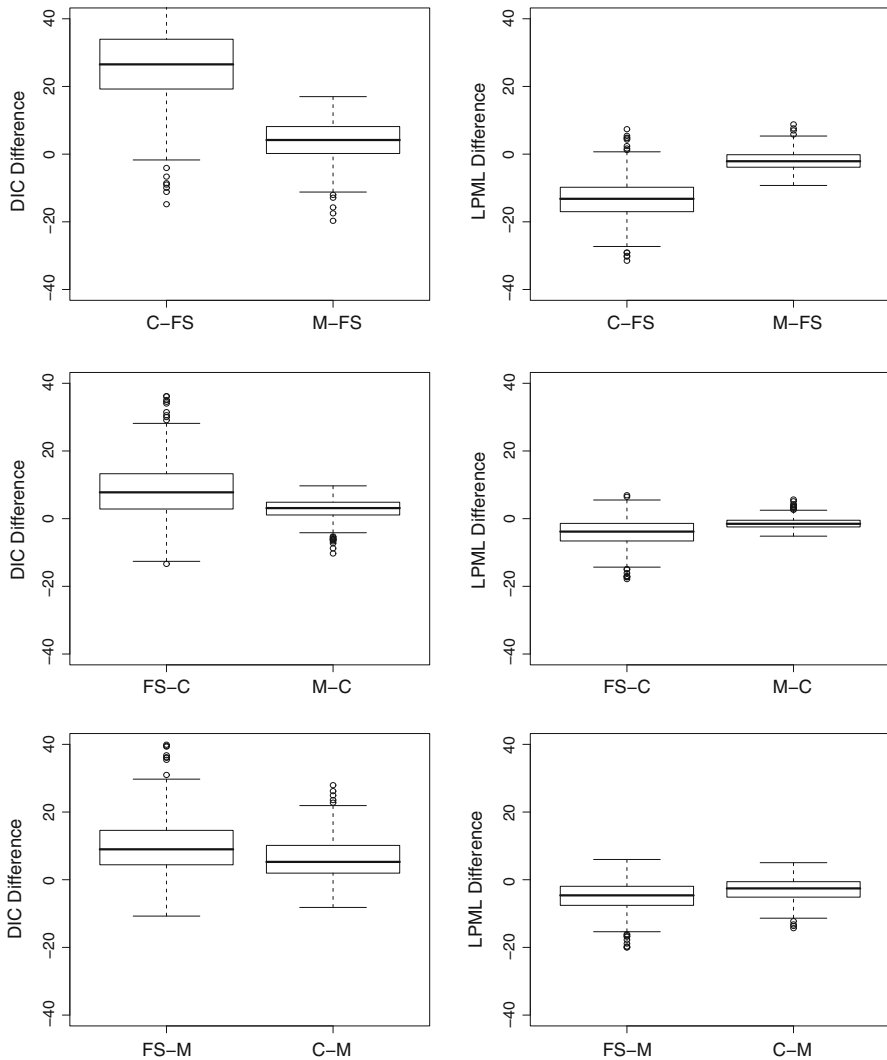


Fig. 1 Box plots of DIC and LPML differences between models

first. The median follow-up time after PSA failure was 11.2 years with IQR=(5.8, 16.0). The sample size was 546, with 54 prostate cancer deaths and 151 other causes of deaths. Seven covariates were considered in the analysis, including patient's age at the date of PSA failure, the natural logarithm of PSA ($\log\text{psa}$), prostatectomy Gleason score (7 ($\text{GS7} = 1$) or otherwise ($\text{GS7} = 0$)), prostatectomy Gleason score (8 to 10 ($\text{GS8H} = 1$) or otherwise ($\text{GS8H} = 0$)), prostatectomy T classification (T3 and higher ($\text{T3} = 1$) or otherwise ($\text{T3} = 0$)), surgical margin status (positive ($\text{margin} = 1$) or otherwise ($\text{margin} = 0$)), and PSA doubling time (DT) (less than 6 months ($\text{DT6} = 1$) or otherwise ($\text{DT6} = 0$)). In this analysis, the prostate cancer cause of death is the

Table 4 DIC, dimension penalty, and LPML of three models

K_1	K_2	C model			M model			FS model		
		DIC	p_D	LPML	DIC	p_D	LPML	DIC	p_D	LPML
10	10	1586.5	34.3	-795.3	1580.0	42.5	-791.9	1565.3	34.4	-784.2
	20	1578.6	44.6	-792.5	1574.9	52.7	-791.2	1560.1	44.7	-782.9
	30	1600.8	55.0	-805.6	1596.0	63.0	-803.1	1581.5	55.1	-795.0
15	10	1584.2	39.9	-795.7	1576.4	48.3	-791.9	1564.4	39.8	-785.0
	20	1575.8	49.9	-792.7	1571.1	58.4	-790.3	1558.9	49.9	-783.5
	30	1598.0	60.3	-805.2	1592.1	68.8	-802.6	1580.1	60.5	-795.4
20	10	1599.8	45.1	-805.4	1592.6	53.6	-801.4	1579.4	45.2	-794.6
	20	1592.0	55.5	-802.6	1587.9	64.1	-800.3	1574.7	55.8	-793.4
	30	1614.1	65.9	-815.2	1609.6	74.7	-813.0	1595.4	65.9	-805.2

Bold values are the best values of DIC or LPML corresponding to the best model

Table 5 Estimates of β_1 under the subdistribution model and the fully specified subdistribution model

Variable	Subdistribution model			Fully Specified Subdistribution model		
	Estimate	SE	95 % CI	Estimate	SD	95 % HPD interval
age	0.017	0.022	(-0.026, 0.059)	0.020	0.020	(-0.017, 0.061)
logpsa	-0.057	0.143	(-0.337, 0.223)	0.036	0.139	(-0.225, 0.316)
GS7	-0.173	0.431	(-1.018, 0.672)	-0.123	0.411	(-0.935, 0.676)
GS8H	0.298	0.400	(-0.486, 1.082)	0.153	0.402	(-0.615, 0.951)
T3	0.453	0.409	(-0.348, 1.255)	0.598	0.417	(-0.195, 1.414)
margin	0.483	0.305	(-0.115, 1.080)	0.417	0.295	(-0.124, 1.035)
DT6	0.990	0.278	(0.445, 1.535)	0.919	0.271	(0.407, 1.467)

primary cause and the other cause of death is any causes of death other than prostate cancer. We let $\beta_1 = (\beta_{11}, \beta_{12}, \dots, \beta_{17})$ denote the vector of the corresponding regression coefficients for the prostate cancer cause of death.

Different values of K_1 and K_2 were tried out for optimizing the model fitting. In the mixture model, z is the vector of all covariates without preselecting for a fair comparison between the models. The values of DIC, p_D , and LPML for the 3×3 combinations of K_1 and K_2 under all three models are reported in Table 4. We see that $(K_1, K_2) = (15, 20)$ is the optimum combination of (K_1, K_2) for almost all models, and that the fully specified subdistribution model always outperforms the other two models by achieving the smallest DIC and the largest LPML.

The subdistribution model of Fine and Gray (1999) was also fit the data. For the prostate cancer death, the estimates, standard errors (SEs) and 95 % confidence intervals (CIs) of β_1 under the subdistribution model of Fine and Gray (1999) and the posterior means (estimates), posterior standard deviations (SDs), and 95 % highest posterior density (HPD) intervals of β_1 under the FS model for the scenario of $(K_1, K_2) = (15, 20)$ are shown in Table 5. We see, from Table 5, that all estimates were quite close and

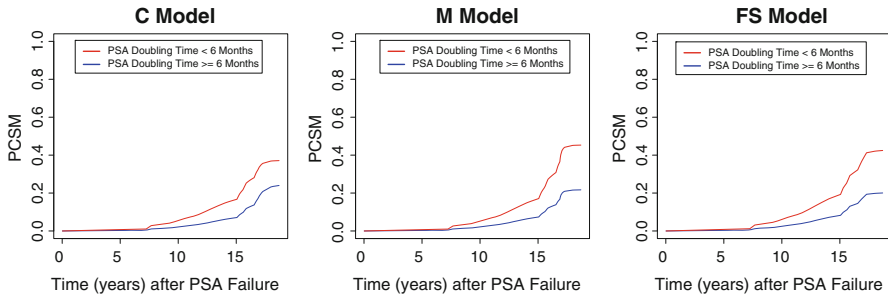


Fig. 2 Plots of the PCSMs under the three models

the two models gave consistent conclusions in terms of significance of covariates at a significance level of 0.05. Note that the estimates of β_1 under the subdistribution model of [Fine and Gray \(1999\)](#) were computed using the R-package `cmprsk`.

Let the prostate cancer specific mortality (PCSM) be the cumulative incident function corresponding to the primary cause of death due to prostate cancer. The covariate effect was further investigated by comparing the posterior means of the PCSM at different times stratified by PSA doubling time ($DT6 = 1$ versus $DT6 = 0$) under each of the three models, where the other covariates were fixed at age = mean age (66.5), $\logpsa = \text{mean } \logpsa$ (2.5), $GS7 = 1$, $GS8H = 0$, $T3 = 1$, and $\text{margin} = 1$. The PCSM plots are shown in [Fig. 2](#). The shapes of the PCSM curves under the three models were similar except at the tail part and the difference between the two curves was slightly smaller under the cause-specific hazards model. For example, at the 10th and 15th year after PSA failure, the posterior means of PCSM under the cause-specific hazards model were 0.055 and 0.166 for patients with PSA doubling time less than 6 months and 0.021 and 0.071 for patients with PSA doubling time greater than or equal to 6 months; under the mixture model the posterior means of PCSM were 0.052 and 0.17 for patients with PSA doubling time less than 6 months and 0.022 and 0.073 for patients with PSA doubling time greater than or equal to 6 months; under the FS model the posterior means of PCSM were 0.061 and 0.192 for patients with PSA doubling time less than 6 months and 0.025 and 0.082 for patients with PSA doubling time greater than or equal to 6 months. Those PCSM plots indicate that the patients with PSA doubling time less than 6 months had worse PCSMs than those with PSA doubling time greater than or equal to 6 months. This covariate effect can directly be seen from [Table 5](#) under the FS model as $DT6$ was significant at a significance level of 0.05. In addition, the proportional hazards structure of the FS model also allows us to compute the adjusted hazard ratio (AHR), which is defined as $\exp(\beta_{17})$, of $DT6$ for the PCSM. Specifically, the posterior mean and 95% HPD interval of the AHR of $DT6$ were 2.601 and (1.369, 4.061), respectively. However, this covariate effect could not be directly assessed under the other two models. For example, under the mixture model with $K_1 = 15$, $K_2 = 20$, for the hazard regression sub-model corresponding to the prostate cancer death, the posterior mean, SD, and 95% HPD interval of β_{17} for $DT6$ were 0.122, 0.384, and (-0.627, 0.865) while the posterior mean, SD, and 95% HPD interval of ϕ_7 for $DT6$ were 0.519, 0.168, and (0.187, 0.839) in the logistic regression sub-model for p_1 , indicating that $DT6$ was significant.

7 Discussion

In this paper, we have developed a fully specified subdistribution model of Fine and Gray (1999) and provided a justification of Fine and Gray's partial likelihood via the profile likelihood approach and the Bayesian approach. Our Bayesian justification is the first such development in the context of competing risk models after the Bayesian justification of Cox's partial likelihood (Kalbfleisch 1978; Sinha et al. 2003) as the risk set at time t in Fine and Gray's partial likelihood includes all patients who are still alive prior to t as well as the patients who were died from other causes of death up to t , which is quite different than the usual risk set in Cox's partial likelihood (Cox 1972, 1975). To fit the proposed FS model, a piecewise exponential model with Jeffreys-type priors, which is a special case of the gamma process prior when $c_0 \rightarrow 0+$, is assumed for the baseline hazard function. Compared to the full gamma process priors, the gamma priors based on the piecewise exponential model relax the conditions for the posterior propriety and facilitate the development of an efficient Gibbs sampling algorithm for carrying out the posterior computation.

In Sect. 5, we conducted an extensive simulation study in examining the performance of DIC and LPML in identifying the model from which the data were generated. Our simulation results empirically showed that when the data are from one (the true model) of the three models (cause-specific hazards model, mixture model, and fully specified subdistribution model), it is unlikely that the other two models would have smaller DICs and larger LPMLs than the true model. This may be due to different proportional hazard functions assumed under these three models. For the prostate cancer data, the FS model had much smaller DIC and larger LPML than those under the cause-specific model and mixture model, implying that the FS model was much more appropriate for fitting this dataset than the other two models.

The fully specified subdistribution model can be further extended to the cases with more than 2 competing risks. Assume there are J competing risks with cause 1 as the cause of interest. Denote $T_j^* = T_j \times I(\delta = j) + \infty \times I(\delta \neq j)$, $j = 1, 2, \dots, J$, and $T^* = \min\{T_1^*, T_2^*, \dots, T_J^*\}$. The cause-specific cumulative incidence functions can be constructed as follows: $F_1(t) = Pr(T^* \leq t, \delta = 1) = Pr(T_1 \leq t, \delta = 1)$, and $F_j(t) = Pr(T^* \leq t, \delta = j) = M_j(t)Pr(\delta \neq j - 1 | \delta \neq 1, \dots, \delta \neq j - 2) \dots Pr(\delta \neq 2 | \delta \neq 1)Pr(\delta \neq 1)$, $j = 2, \dots, J$, where $M_j(t)$ is the probability of failure from cause j by time t conditional on not failing from causes 1, 2, ..., $j - 1$.

In this paper, we only considered fixed covariates. Including time-dependent covariates in the FS model, as well as jointly modeling longitudinal measurements (e.g., a series of PSA measures over-time) and survival endpoints of cause-specific death times are important future research topics, which are under investigation currently. Models with frailty terms (Clayton 1978; Vaupel et al. 1979) are commonly used for correlated survival data with multivariate risk factors. Dixon et al. (2011) introduced a multivariate subdistribution hazard model including frailty to induce correlations among clustered survival times. The FS model can be extended for correlated survival data in the presence of competing risks in the manner to include frailties. This is another interesting topic for future research.

In all the Bayesian computations, we used 10,000 Gibbs samples after a burn-in of 1000 iterations for each model to compute all the posterior estimates, including posterior means, posterior standard deviations, 95% HPD intervals, DICs and LPMLs. We also generated 50,000 Gibbs samples after a burn-in of 1000 to re-compute those posterior quantities and the results were very similar. The HPD intervals were computed via the Monte Carlo method developed by [Chen and Shao \(1999\)](#). Codes were written for FORTRAN 95 compiler, and we used IMSL subroutines with double precision accuracy. The fortran codes for the FS model are available upon request.

Acknowledgments The authors wish to thank the Editor-in-Chief, the Associate Editor, and the two referees for their helpful comments and suggestions, which have led to an improved version of this article. This research was partially supported by NIH grants #GM 70335 and #CA 74015.

Appendix A: Proofs of Theorems

Proof of Theorem 1 With the assumption that h_{10} is zero after the last observation of failure due to cause 1, the likelihood function is now

$$L(\boldsymbol{\beta}_1, h_{10} | \mathbf{y}, X, \boldsymbol{\delta}) = \prod_{i=1}^{D_1} \left[h_{10}(y_{(i)}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right] \times \exp \left\{ - \sum_{i=1}^n H_{10}(y_{(i)}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\}.$$

The profile likelihood approach assumes that h_{10} is zero except for the failure times due to cause 1. Then

$$\begin{aligned} L(\boldsymbol{\beta}_1, h_{10} | \mathbf{y}, X, \boldsymbol{\delta}) &= \left[\prod_{i=1}^{D_1} h_{10}(y_{(i)}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right] \\ &\quad \times \exp \left\{ - \sum_{i=1}^{D_1} \sum_{k=1}^i h_{10}(y_{(k)}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right. \\ &\quad \left. - \sum_{i=D_1+1}^n \sum_{l=1}^{D_1} h_{10}(y_{(l)}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} \\ &= \left[\prod_{i=1}^{D_1} h_{10}(y_{(i)}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right] \\ &\quad \times \exp \left\{ - \sum_{i=1}^{D_1} h_{10}(y_{(i)}) \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1) \right\}. \end{aligned}$$

Therefore, the profile maximum likelihood estimator of h_{10} is given by

$$\hat{h}_{10}(y_{(i)}) = \frac{1}{\sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)}.$$

Plugging $\hat{h}_{10}(y_{(i)})$ in $L(\beta_1, h_{10}|y, X, \delta)$ results in the profile likelihood function given by

$$L_p^*(\beta_1|y, X, \delta) \propto \prod_{i=1}^{D_1} \frac{\exp(x'_i \beta_1)}{\sum_{j \in R_i^*} \exp\{x'_j \beta_1\}},$$

which is (2.6). □

Proof of Theorem 2 Assume the prior of h_{10} only has value λ_i at times $y_{(i)}$ such that $\delta_i = 1$. Then $\lambda_i = H_{10}(y_{(i)}) - H_{10}(y_{(i-1)})$, and $\lambda_{D_1+1} = \dots = \lambda_n = 0$, $i = 1, \dots, D_1$. The survival function at time t is

$$\begin{aligned} Pr(T_1^* > t) &= \prod_{i=1}^n [1 - F_1(y_{(i)})] \\ &= \exp \left\{ - \sum_{i=1}^{D_1} H_{10}(y_{(i)}) \exp(x'_i \beta_1) - \sum_{i=D_1+1}^n H_{10}(\infty) \exp(x'_i \beta_1) \right\} \\ &= \exp \left\{ - \sum_{i=1}^{D_1} \sum_{k=1}^i \lambda_k \exp(x'_i \beta_1) - \sum_{i=D_1+1}^n \sum_{l=1}^{D_1} \lambda_l \exp(x'_i \beta_1) \right\} \\ &= \exp \left\{ - \sum_{i=1}^{D_1} \lambda_i \sum_{j \in R_i^*} \exp(x'_j \beta_1) \right\}. \end{aligned} \tag{A.1}$$

Then the likelihood function in (2.8) reduces to

$$\begin{aligned} L(\beta_1, \lambda|y, X, \delta) &= \prod_{i=1}^{D_1} [\lambda_i \exp(x'_i \beta_1)] \times \exp \left\{ - \lambda_i \sum_{j \in R_i^*} \exp(x'_j \beta_1) \right\} \\ &= \prod_{i=1}^{D_1} \left[\frac{\exp(x'_i \beta_1)}{\sum_{j \in R_i^*} \exp(x'_j \beta_1)} \right] \times \left[\lambda_i \sum_{j \in R_i^*} \exp(x'_j \beta_1) \right] \\ &\quad \times \exp \left\{ - \lambda_i \sum_{j \in R_i^*} \exp(x'_j \beta_1) \right\}. \end{aligned}$$

Since $\pi(\lambda) \propto \prod_{i=1}^{D_1} 1/\lambda_i$, we have

$$\begin{aligned} &\int L(\beta_1, \lambda|y, X, \delta) \pi(\lambda) d\lambda \\ &= \int \prod_{i=1}^{D_1} \frac{1}{\lambda_i} \times \left[\frac{\exp(x'_i \beta_1)}{\sum_{j \in R_i^*} \exp(x'_j \beta_1)} \right] \times \left[\lambda_i \sum_{j \in R_i^*} \exp(x'_j \beta_1) \right] \end{aligned}$$

$$\begin{aligned} & \times \exp \left\{ -\lambda_i \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1) \right\} d\lambda \\ & = \prod_{i=1}^{D_1} \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_1)}{\sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)}. \end{aligned}$$

□

Proof of Theorem 3 Assume H_{10} follows a gamma process prior. Let $h_{1i} = H_{10}(y_{(i)}) - H_{10}(y_{(i-1)})$, $h_{1i} \sim G(c_0 h_{0i}, c_0)$, $i = 1, \dots, D_1$. h_{1i} 's are independent of each other, and $h_{1, D_1+1} = \dots = h_{1n} = 0$. Similar to (A.1), we can show that the survival function at time t is given by

$$Pr(\mathbf{T}_1^* > t) = \prod_{i=1}^n \left[1 - F_1(y_{(i)}) \right] = \exp \left\{ -\sum_{i=1}^{D_1} h_{1i} \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1) \right\}.$$

Taking expectation with respect to the gamma process prior gives

$$\begin{aligned} E^{h_{10}} \left[Pr(\mathbf{T}_1^* > t) \right] & = \prod_{i=1}^{D_1} \left[\frac{c_0}{c_0 + \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)} \right]^{c_0 h_{0i}} \\ & = \prod_{i=1}^{D_1} \exp \left\{ c_0 \sum_{k=1}^i h_{0k} \log \left(1 - \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_1)}{c_0 + \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)} \right) \right\}. \end{aligned}$$

Now the expectation of the likelihood function in (2.8) with respect to the gamma process prior reduces to

$$\begin{aligned} E^{h_{10}} \left[L(\boldsymbol{\beta}_1, h_{10} | y, X, \delta) \right] & = \prod_{i=1}^{D_1} \exp \left\{ c_0 \sum_{k=1}^i h_{0k} \log \left(1 - \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_1)}{c_0 + \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)} \right) \right\} \\ & \quad \times \left[-c_0 \sum_{k=1}^i \frac{dh_{0k}}{dy_{(i)}} \log \left(1 - \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_1)}{c_0 + \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)} \right) \right]. \end{aligned}$$

Since $\lim_{c_0 \downarrow 0} \exp \left\{ c_0 \sum_{k=1}^i h_{0k} \log \left(1 - \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_1)}{c_0 + \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)} \right) \right\} = 1$ and

$$\lim_{c_0 \downarrow 0} \log \left(1 - \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_1)}{c_0 + \sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)} \right) \approx -\frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_1)}{\sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)},$$

we have

$$\lim_{c_0 \downarrow 0} \frac{E^{h_{10}} \left[L(\boldsymbol{\beta}_1, h_{10} | y, X, \delta) \right]}{\prod_{i=1}^{D_1} \left[c_0 \sum_{k=0}^i \frac{dh_{0k}}{dy_{(i)}} \right]} \approx \prod_{i=1}^{D_1} \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}_1)}{\sum_{j \in R_i^*} \exp(\mathbf{x}'_j \boldsymbol{\beta}_1)}.$$

□

Proof of Theorem 4 To show that (3.3) is proper, it is needed to show that

$$\int \pi^*(\beta_1, \beta_2, \lambda_1, \lambda_2 | t, X, \delta) d\beta_1 d\beta_2 d\lambda_1 d\lambda_2 < \infty,$$

where $\pi^*(\beta_1, \beta_2, \lambda_1, \lambda_2 | t, X, \delta)$ is the unnormalized joint posterior density defined in (3.3). After some algebra, we can show that

$$\begin{aligned} & \pi^*(\beta_1, \beta_2, \lambda_1, \lambda_2 | t, X, \delta) \\ & \leq \prod_{i=1}^n \prod_{k=1}^{K_1} \left[\lambda_{1k} \exp(x'_i \beta_1) \exp \left\{ - \exp(x'_i \beta_1) \lambda_{1k} (t_i - s_{1, k-1}) \right\} \right]^{v_{1ik} I(\delta_i=1)} \\ & \times \lambda_{1k}^{-1} \lambda_{1, K_1+1}^{a-1} \exp(-b\lambda_{1, K_1+1}) \\ & \times \prod_{i=1}^n \prod_{k=1}^{K_2} \left[\lambda_{2k} \exp(x'_i \beta_2) \exp \left\{ - \exp(x'_i \beta_2) \lambda_{2k} (t_i - s_{2, k-1}) \right\} \right]^{v_{2ik} I(\delta_i=2)} \times \lambda_{2k}^{-1} \\ & \equiv \pi^{**}(\beta_1, \lambda_1 | t, X, \delta) \times \pi^{**}(\beta_2, \lambda_2 | t, X, \delta). \end{aligned}$$

It suffices to show that

$$\int \pi^{**}(\beta_1, \lambda_1 | t, X, \delta) d\beta_1 d\lambda_1 < \infty,$$

since $\int \pi^{**}(\beta_2, \lambda_2 | t, X, \delta) d\beta_2 d\lambda_2 < \infty$ can be proved in a similar way.

Consider the transformation $u_j = \log(\lambda_{1j})$, $j = 1, 2, \dots, K_1$, and let $\mathbf{u} = (u_1, u_2, \dots, u_{K_1})'$. Then, we have

$$\begin{aligned} & \int \pi^{**}(\beta_1, \lambda_1 | t, X, \delta) d\beta_1 d\lambda_1 \\ & = \int \pi^{**}(\beta_1, \mathbf{u} | t, X, \delta) d\beta_1 d\mathbf{u} \times \int \lambda_{1, K_1+1}^{a-1} \exp(-b\lambda_{1, K_1+1}) d\lambda_{1, K_1+1} \\ & \propto \int \prod_{i=1}^n \prod_{k=1}^{K_1} \left[\exp(u_k + x'_i \beta_1) \right. \\ & \quad \left. \times \exp \left\{ - \exp(u_k + x'_i \beta_1) (t_i - s_{1, k-1}) \right\} \right]^{v_{1ik} I(\delta_i=1)} d\beta_1 d\mathbf{u}. \end{aligned} \tag{A.2}$$

It is easy to show that $\exp(u_k + x'_i \beta_1) \times \exp \left\{ - \exp(u_k + x'_i \beta_1) (t_i - s_{1, k-1}) \right\} \leq M_1$, where $M_1 > 0$ is a constant. Under condition (ii), X_1 is of full rank. Thus, there exist distinct $i_1, i_2, \dots, i_{K_1+p_1}$, where $p_1 = \dim(\beta_1)$, such that the $(K_1 + p_1) \times (K_1 + p_1)$ matrix X_1^* , which has rows $(\mathbf{x}'_{1\ell})' = (v_{ji_\ell 1}, \dots, v_{ji_\ell K_1}, \mathbf{x}'_{i_\ell})$ for $\ell = 1, \dots, K_1 + p_1$, is of full rank. Let k_{i_ℓ} be an integer such that $t_{i_\ell} \in (s_{1, k_{i_\ell}-1}, s_{1, k_{i_\ell}}]$ for $\ell = 1, \dots, K_1 + p_1$. We take a one-to-one transformation $\xi_1 = X_1^*(\mathbf{u}', \beta_1)'$. Using (A.2), we have

$$\begin{aligned}
& \int \pi^{**}(\boldsymbol{\beta}_1, \boldsymbol{\lambda}_1 | t, X, \boldsymbol{\delta}) d\boldsymbol{\beta}_1 d\boldsymbol{\lambda}_1 \\
& \leq M_2 \int \prod_{\ell=1}^{K_1+p_1} \left(\exp\{(\mathbf{x}'_{1\ell})'(\mathbf{u}', \boldsymbol{\beta}')'\} \right. \\
& \quad \left. \times \exp\left[-\exp\{(\mathbf{x}'_{1\ell})'(\mathbf{u}', \boldsymbol{\beta}')'(t_i - s_{1, k_{i\ell}-1})\}\right] \right) dud\boldsymbol{\beta} \\
& = M_3 \prod_{\ell=1}^{K_1+p_1} \int_{-\infty}^{\infty} \exp(\xi_{1\ell}) \exp\left[-\exp(\xi_{1\ell})(t_i - s_{1, k_{i\ell}-1})\right] d\xi_{1\ell} \\
& = M_3 \prod_{\ell=1}^{K_1+p_1} (t_i - s_{1, k_{i\ell}-1})^{-1} < \infty,
\end{aligned}$$

where M_2 and M_3 are two positive constants. This completes the proof. \square

Appendix B: Generating η from $[\eta | \boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, t, X, \boldsymbol{\delta}]$ and u from $[u | \boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \eta, t, X, \boldsymbol{\delta}]$

Generating η from $[\eta | \boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, t, X, \boldsymbol{\delta}]$ When $\delta_i = 0$, we generate η_i by $I(\eta_i = 1) \sim \text{Bin}(1, p_{i1})$ and $I(\eta_i = 2) = 1 - I(\eta_i = 1)$, where $p_{i1} = \frac{a_i}{a_i + b_i}$, $a_i = \exp\{-H_{10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\} - \exp\{-H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\}$, and $b_i = \exp\{-H_{20}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_2) - H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\}$.

Generating u from $[u | \boldsymbol{\beta}_1, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \eta, t, X, \boldsymbol{\delta}]$ When $\delta_i = 0$ and $\eta_i = 1$, we generate u_i from a truncated piecewise exponential distribution $f(u_i)$, where

$$f(u_i) \propto h_{10}(u_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \exp\{-H_{10}(u_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\}, \quad u_i \geq t_i.$$

Denote δ_{ji} as the index such that $s_j, \delta_{ji-1} \leq t_i < s_j, \delta_{ji}$. Let

$$\begin{aligned}
a_{1i} &= \exp\{-H_{10}(t_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\} - \exp\{-H_{10}(s_{1, \delta_{1i}}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\}, \\
a_{ki} &= \exp\{-H_{10}(s_{1, \delta_{1i+k-2}}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\} \\
& \quad - \exp\{-H_{10}(s_{1, \delta_{1i+k-1}}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\}, \\
& \quad k = 2, \dots, K_1 - \delta_{1i} + 1. \\
a_{K_1 - \delta_{1i} + 2, i} &= \exp\{-H_{10}(s_{1, K_1}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\} - \exp\{-H_{10}(\infty) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1)\}.
\end{aligned}$$

Generate v_i from a $U(0, 1)$ distribution. If v_i falls into the k_{v_i} th interval such that

$$\frac{a_{1,i} + \dots + a_{k_{v_i}-1,i}}{a_{1,i} + \dots + a_{K_1 - \delta_{1i} + 2,i}} < v_i \leq \frac{a_{1,i} + \dots + a_{k_{v_i},i}}{a_{1,i} + \dots + a_{K_1 - \delta_{1i} + 2,i}},$$

the inverse distribution function method is used to calculate u_i as

$$v_i = \frac{a_{1,i} + \cdots + a_{k_v-1,i} + \exp \left\{ -H_{10}(s_{1,k_v-1}) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\} - \exp \left\{ -H_{10}(u_i) \exp(\mathbf{x}'_i \boldsymbol{\beta}_1) \right\}}{a_{1,i} + \cdots + a_{K_1-\delta_{1i}+2,i}}.$$

References

- Chen MH, Ibrahim JG, Shao QM (2006) Posterior propriety and computation for the Cox regression model with applications to missing covariates. *Biometrika* 93:791–807
- Chen MH, Shao QM (1999) Monte Carlo estimation of Bayesian credible and HPD intervals. *J Comput Graph Stat* 8:69–92
- Chen MH, Shao QM, Ibrahim JG (2000) Monte Carlo methods in Bayesian computation. Springer, New York
- Choueiri TK, Chen MH, D'Amico AV, Sun L, Nguyen PL, Hayes JH, Robertson CN, Walther PJ, Polascik TJ, Albala DM, Moul JW (2010) Impact of postoperative prostate-specific antigen disease recurrence and the use of salvage therapy on the risk of death. *Cancer* 116:1887–1892
- Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 65:141–151
- Cox DR (1972) Regression models and life tables (with discussion). *J R Stat Soc B* 34:187–220
- Cox DR (1975) Partial likelihood. *Biometrika* 62:269–276
- Dixon SN, Darlington GA, Desmond AF (2011) A competing risks model for correlated data based on the subdistribution hazard. *Lifetime Data Anal* 17:473–495
- Elashoff RM, Li G, Li N (2007) An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Stat Med* 26:2813–2835
- Elashoff RM, Li G, Li N (2008) A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* 64:762–771
- Fan X (2008) Bayesian nonparametric inference for competing risks data. Unpublished Ph.D. Dissertation, Division of Biostatistics, Medical College of Wisconsin
- Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 94:496–509
- Gail M (1975) A review and critique on some models used in competing risk analysis. *Biometrics* 31:209–222
- Gaynor JJ, Feuer EJ, Tan CC, Wu DH, Little CR, Strauss DJ, Clarkson BD, Brennan MF (1993) On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data. *J Am Stat Assoc* 88:400–409
- Gelfand AE, Dey DK (1994) Bayesian model choice: asymptotics and exact calculations. *J R Stat Soc B* 56:501–514
- Gilks WR, Wild P (1992) Adaptive rejection sampling for Gibbs sampling. *Appl Stat* 41:337–348
- Gray RJ (1988) A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat* 16:1141–1154
- Hu W, Li G, Li N (2009) A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Stat Med* 28:1601–1619
- Huang X, Li G, Elashoff RM, Pan J (2011) A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Anal* 17:80–100
- Ibrahim JG, Chen MH, Sinha D (2001) Bayesian survival analysis. Springer, New York
- Kalbfleisch JD (1978) Non-parametric Bayesian analysis of survival time data. *J R Stat Soc B* 40:214–221
- Larson MG, Dinse GE (1985) A mixture model for the regression analysis of competing risks data. *Appl Stat* 34:201–211
- Liu JS (1994) The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *J Am Stat Assoc* 89:958–966
- Lu K, Tsiatis AA (2001) Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. *Biometrics* 57:1191–1197
- Prentice RL, Kalbfleisch JD, Peterson A, Flournoy N, Farewell V, Breslow N (1978) The analysis of failure times in the presence of competing risks. *Biometrics* 34:541–554

- Sinha D, Ibrahim JG, Chen MH (2003) A Bayesian justification of Cox's partial likelihood. *Biometrika* 90:629–641
- Spiegelhalter DJ, Best NG, Carlin BP, Linde A (2002) Bayesian measures of model complexity and fit (with Discussion). *J R Stat Soc B* 64:583–639
- van der Tsiatis A (1975) A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci USA* 72:20–22
- Vaupel JW, Manton KG, Stallard E (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16:439–454
- Wang X, Sinha A, Yang J, Chen MH (2012) Bayesian inference of interval-censored survival data. In: Chen DG, Sun J, Peace KE (eds) *Interval-censored time-to-event data: methods and applications*. Chapman & Hall, Boca Raton, FL (in press)