



Published in final edited form as:

Lifetime Data Anal. 2014 January ; 20(1): . doi:10.1007/s10985-013-9254-8.

Bayesian Gamma Frailty Models for Survival Data with Semi-Competing Risks and Treatment Switching

Yuanye Zhang,

Novartis Institutes for BioMedical Research, Inc., 220 Massachusetts Avenue, Cambridge, MA 02139

Ming-Hui Chen,

Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269

Joseph G. Ibrahim,

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599

Donglin Zeng,

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599

Qingxia Chen,

Department of Biostatistics, Vanderbilt University, Nashville, TN 37232

Zhiying Pan, and

Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320

Xiaodong Xue

Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320

Ming-Hui Chen: ming-hui.chen@uconn.edu

Abstract

Motivated from a colorectal cancer study, we propose a class of frailty semi-competing risks survival models to account for the dependence between disease progression time, survival time, and treatment switching. Properties of the proposed models are examined and an efficient Gibbs sampling algorithm using the collapsed Gibbs technique is developed. A Bayesian procedure for assessing the treatment effect is also proposed. The Deviance Information Criterion (DIC) with an appropriate deviance function and Logarithm of the Pseudomarginal Likelihood (LPML) are constructed for model comparison. A simulation study is conducted to examine the empirical performance of DIC and LPML and as well as the posterior estimates. The proposed method is further applied to analyze data from a colorectal cancer study.

Keywords

Competing risks; Panitumumab; Partial treatment switching; Posterior propriety; Semi-Markov model

1 Introduction

In chronic disease or cancer studies and clinical trials, it is very common to have both terminating events and nonterminating events in the data. This type of situation is referred as semi-competing risks, in which an event time can be censored by another event time but not vice versa. A terminating event potentially censors a nonterminating event, but the nonterminating event does not prevent subsequent observation of the terminating event. An

example of this is the colorectal cancer clinical trial that we examine here, called the panitumumab 408 study, which was conducted by Amgen Inc. (see Section 5). In this study, disease progression is a nonterminating event, death is a terminating event, and disease progression can be censored by death but not vice-versa. In addition to semi-competing risks, treatment switching may also occur in clinical trials. In such trials, patients in the control arm who experience an intermediate event, such as disease progression, may begin taking the experimental treatment. In the panitumumab 408 study, there were a substantial proportion of patients in the control arm who switched treatment after disease progression (see Section 5). As discussed in Marcus and Gibbons (2001), an intent-to-treat (ITT) analysis leads to attenuated treatment effect estimates, and thus one must properly model the data accommodating this switching effect and then appropriately estimate the treatment effect.

In semi-competing risks data, there are two major issues: dependent censoring and identifiability. In order to deal with these issues, several modeling and inference approaches have been developed. One major approach is to model the joint distribution of T_D and T_E , where T_D denotes the time to terminating event and T_E denotes the time to nonterminating event. Day and Bryant (1997) used frailty models for the joint survival function using a relevant censoring process. Later, Fine et al. (2001) adopted this model and proposed a novel estimator for the marginal distribution of T_E based on a bivariate location-shift model with a completely unspecified underlying distribution for T_D and T_E . Although this method is appropriate for modeling one recurrent event taking into account dependent censoring, it cannot be applied to more than two recurrent events. Furthermore, various types of copula models have been applied for modeling the joint distribution of (T_E, T_D) (Wang, 2003; Ghosh, 2006; Peng and Fine, 2007). Another approach is to model the gap time T_G between T_D and T_E (Mandel, 2010). Nonparametric estimation of the gap time distribution and regression methods for gap time hazard functions have been developed. A third approach is similar to the above gap time model. In addition to modeling T_E and T_G , another event time T_D^* is introduced, which denotes the terminating event that happens without the nonterminating event T_E . Shen and Thall (1998) used such a model for obtaining the marginal distributions of T_E , T_G and T_D^* . They assumed that the distributions of T_E and T_D^* are mutually independent. For the bivariate distribution of T_E and T_G , they used a bivariate generalized von Morgenstern distribution, which characterizes the positive or negative association between these two times using a single parameter. A conditional model is also developed (Zeng et al., 2012). Instead of modeling the joint distribution of T_E and T_G , a conditional model of T_G given T_E is used. Multistate modeling is another approach for survival data with semi-competing risks, in which no event, nonterminating event, and terminating event can be viewed as the three states in a multistate process. The focus of multistate modeling is mainly on the transition probabilities between different states. Aalen-Johansen estimators (Aalen et al., 1978; Andersen et al., 1993) can be used to estimate these transition probabilities. However, this approach does not provide much information on the dependence structure between the time to nonterminating event and the time to terminating event. Except for Zeng et al. (2012), most of the aforementioned articles do not directly deal with both semi-competing risks and treatment switching.

In this paper, we introduce a Bayesian frailty model for survival data with semi-competing risks in the presence of partial treatment switching (i.e., not every subject in the control arm switched to active treatment). In the frequentist inference, the Monte Carlo EM (MCEM) algorithm is often used to obtain the maximum likelihood estimates in the presence of the unobserved frailty variables. However, the MCEM algorithm may fail to converge when fitting a semi-competing risks frailty model with unknown parameters in the frailty distribution since the estimates of these unknown parameters are unstable. To overcome this

challenging computational issue, we develop an efficient Gibbs sampling algorithm via the introduction of latent variables, reparameterization, and the collapsed Gibbs sampler. The Bayesian framework also allows us to characterize the conditions for model identifiability by examining posterior propriety. In addition, to appropriately estimate the treatment effect, we extend the method of Zeng et al. (2012) to derive the predictive survival function with partial treatment switching under the semi-competing risks frailty model and carry out Bayesian inference on this quantity without resorting to asymptotics.

The rest of the paper is organized as follows. Section 2 presents a detailed development of the semi-competing risks model via a gamma frailty including explicit expressions for the likelihood function based on the observed data. In Section 3, we characterize posterior propriety conditions under this complex model, provide the Bayesian formulation of the predictive survival function with partial treatment switching, develop an efficient Gibbs sampling algorithm, and introduce two Bayesian model comparison criteria. A simulation study is carried out to examine the empirical performance of the posterior estimates and Bayesian model criteria in Section 4, and a detailed analysis of a subset of the data from the panitumumab 408 study is presented in Section 5. We conclude the paper with a brief discussion in Section 6. The proofs of all theorems and detailed derivations of the computational development are given in the Appendices.

2 The Semi-Competing Risks Frailty Models

2.1 Models

To introduce the proposed model, we use the following notation. As motivated from the panitumumab 408 study, we consider disease progression as a nonterminating event. However, the proposed model can be applied to any other type of nonterminating event. Let E be a dichotomous variable to denote the disease progression status of subjects, where $E = 1$ if the subject is in the disease progression population, which include subjects who eventually develop disease progression before death, and $E = 0$ if otherwise. Also let T_D denote the time from study entry to death for subjects with $E = 0$. For the disease progression population ($E = 1$), we further let T_E denote the time from study entry to disease progression and let T_G denote the time from disease progression to death. A graphical illustration of these variables is shown in Fig. 1.

The proposed statistical model consists of the following three components. The first component is to model the disease progression status E given the baseline covariates \mathbf{x} and the treatment indicator A ($A = 1$ if the subject is on the treatment arm and $A = 0$ if the subject is on the placebo or control arm). To this end, we assume

$$\text{logit}(P(E=1|A, \mathbf{x}, \alpha)) = \log \left\{ \frac{P(E=1|A, \mathbf{x}, \alpha)}{1 - P(E=1|A, \mathbf{x}, \alpha)} \right\} = \alpha_0 + A\alpha_1 + \mathbf{x}'\alpha_2, \quad (2.1)$$

where α_0 , α_1 , and α_2 are unknown coefficients and $\alpha = (\alpha_0, \alpha_1, \alpha_2)'$. The second component models the survival distribution of the non-progression population given \mathbf{x} and A , which is defined by

$$h_D(t|A, \mathbf{x}, E=0) = h_0(t) \exp\{A\beta_0 + \mathbf{x}'\gamma_0\}, \quad (2.2)$$

where $h_D(t|A, \mathbf{x}, E=0)$ is the conditional hazard function of T_D given the covariates, $h_0(t)$ is an unknown baseline hazard function, and (β_0, γ_0) are unknown regression coefficients.

As shown in Fig. 1, T_E and T_G are potentially dependent. To capture this dependence, we assume the frailty model

$$h_E(t|A, \mathbf{x}, E=1, \omega) = h_1(t) \exp\{A\beta_1 + \mathbf{x}'\gamma_1\} \omega \text{ and } h_G(t|A, \mathbf{z}, V, E=1, \omega) = h_2(t) \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} \omega, \quad (2.3)$$

where $h_E(t|A, \mathbf{x}, E = 1, \omega)$ is the conditional hazard function for T_E , $h_G(t|A, \mathbf{z}, E = 1, \omega)$ is the conditional hazard function for T_G , both $h_1(t)$ and $h_2(t)$ are unknown baseline hazard functions, and the β 's and γ 's are regression coefficients. Here, V is the treatment switching indicator (1 = switching; 0 = no switching) and \mathbf{z} reflects the covariates collected at baseline or at disease progression, which could be prognostic factors for the treatment switching decision. In (2.3), ω is a latent gamma-frailty, which is assumed to follow a Gamma distribution, $\text{Gamma}(1/\tau, 1/\tau)$, with mean one, variance τ ($\tau > 0$), and density given by

$$f(\omega|\tau) = \frac{(1/\tau)^{1/\tau}}{\Gamma(1/\tau)} \omega^{1/\tau-1} \exp(-\omega/\tau).$$

Given ω , T_E and T_G are conditionally independent.

Unconditionally, T_E and T_G are dependent and, moreover, the local measure of dependence (Oakes, 1989) between T_E and T_G is $\varphi_{FM} = 1 + \tau$, indicating a positive association between T_E and T_G . When $\tau \rightarrow 0$, $\varphi_{FM} \rightarrow 1$ and T_E and T_G become independent. The model defined by (2.1) – (2.3) is thus called the semi-competing risks frailty model abbreviated by FM. As an alternative to (2.3), we may consider the following models for T_E and T_G :

$$h_E(t|A, \mathbf{x}, E=1) = h_1(t) \exp\{A\beta_1 + \mathbf{x}'\gamma_1\} \text{ and } h_G(t|A, \mathbf{z}, V, T_E, E=1) = h_2(t) \exp\{A\beta_{21} + V(1-A)\beta_{22} + T_E\gamma_{21} + \mathbf{z}'\gamma_{22}\}, \quad (2.4)$$

where γ_{21} is the regression coefficient corresponding to T_E and γ_{22} is the corresponding vector of regression coefficients associated with \mathbf{z} . The model defined by (2.1), (2.2), and (2.4) is called the conditional semi-competing risks model, denoted by CM. After some algebra, we can show that the local measure of dependence between T_E and T_G under CM is given by $\varphi_{CM} = [\int_{t_E} \{\exp[-H_2(t_G) \exp\{A\beta_{21} + V(1-A)\beta_{22} + \gamma_{21}u + \mathbf{z}'\gamma_{22}\}] h_1(u) \exp(A\beta_1 + \mathbf{x}'\gamma_1) \times \exp\{-H_1(u) \exp(A\beta_1 + \mathbf{x}'\gamma_1)\} du \exp(t_E \gamma_{21})] \times [\int_{t_E} \{\exp[-H_2(t_G) \exp\{A\beta_{21} + V(1-A)\beta_{22} + \gamma_{21}u + \mathbf{z}'\gamma_{22}\}] h_1(u) \exp(A\beta_1 + \mathbf{x}'\gamma_1) \times \exp\{-H_1(u) \exp(A\beta_1 + \mathbf{x}'\gamma_1)\} \exp(u\gamma_{21})\} du]^{-1}$ for $t_E > 0$ and $t_G > 0$, where $H_j(t) = \int_0^t h_j(u) du$ for $j = 1, 2$. Unlike φ_{FM} , φ_{CM} depends on (t_E, t_G) . It is easy to see that $\varphi_{CM} > 1$ when $\gamma_{21} < 0$, $\varphi_{CM} = 1$ when $\gamma_{21} = 0$, and $\varphi_{CM} < 1$ when $\gamma_{21} > 0$. This result implies that CM allows a positive or negative association between T_E and T_G . As discussed in Zhao (2009), FM is a homogeneous Markov model when $h_2(t)$ is constant while CM is a homogeneous semi-Markov model since the hazard function for T_G in (2.4) depends on the progression time T_E . On the other hand, the marginal distributions of T_E and T_G after integrating out the gamma frailty belong to the class of generalized odds-rate hazards (GORH) models (see Banerjee et al, 2007). As the GORH model is a non-proportional hazards model, FM is more robust to the proportional hazards assumption than CM.

We further assume piecewise exponential models for the baseline hazard functions $h_0(t)$, $h_1(t)$, and $h_2(t)$. For $k = 0, 1, 2$, let $0 < s_{k1} < s_{k2} < \dots < s_{kJ_k}$ be a finite partition of the time axis. Thus, we have the J_k intervals: $(0, s_{k1}]$, $(s_{k1}, s_{k2}]$, \dots , $(s_{kJ_{k-1}}, s_{kJ_k}]$, where $s_{kJ_k} = \infty$. In the j^{th} interval, we assume a constant baseline hazard, $h_k(y|\lambda_k) = \lambda_{kj}$ for $y \in (s_{k,j-1}, s_{kj}]$. Letting $\lambda_k = (\lambda_{k1}, \lambda_{k2}, \dots, \lambda_{kJ_k})'$, the cumulative baseline hazard function corresponding to $h_k(t)$ is given by

$$H_k(y|\lambda_k) = \lambda_{kj}(y - s_{k,j-1}) + \sum_{g=1}^{j-1} \lambda_{kg}(s_{kg} - s_{k,g-1}) \text{ when } s_{k,j-1} < y \leq s_{kj} \quad (2.6)$$

for $k = 0, 1, 2$.

2.2 Likelihood Function

Suppose we have n subjects. Let y_i denote the observed death time or censoring time, \mathbf{x}_i is the vector of baseline covariates, A_i is the treatment indicator, y_{Ei} is the observed disease progression time, \mathbf{z}_i is the vector of covariates collected at baseline or at disease progression, and V_i is the indicator for treatment switching for the i^{th} subject for $i = 1, \dots, n$. Also let v_i be the censoring variable such that $v_i = 1$ if y_i is a death time and $v_i = 0$ if y_i is a right censoring time, and let d_i be the indicator variable such that $d_i = 1$ if y_{Ei} is a disease progression time and 0 if there is no disease progression for the i^{th} individual. When $d_i = 0$, y_{Ei} is assumed to be equal to infinity. Finally, we use E_i to denote the disease progression indicator such that $E_i = 1$ if subject i is in the disease progression population and 0 otherwise. Let

$$P(E_i=0|A_i, \mathbf{x}_i, \alpha)$$

$$\begin{aligned} &= [1 + \exp(\alpha_0 + A_i\alpha_1 + \mathbf{x}'_i\alpha_2)]^{-1}, S_0(y_i|A_i, \mathbf{x}_i, \beta_0, \gamma_0, \lambda_0) = \exp\{-H_0(y_i|\lambda_0)\exp(A_i\beta_0 + \mathbf{x}'_i\gamma_0)\}, f_0(y_i|A_i, \mathbf{x}_i, \beta_0, \\ &+ V_i \times (1 - A_i)\beta_{22} \\ &+ \mathbf{z}'_i\gamma_2\} \text{ and } f_2(y_{Gi}|A_i, V_i, \mathbf{z}_i, \omega_i, \beta_2, \gamma_2, \lambda_2) \\ &= \omega_i h_2(y_{Gi}|\lambda_2) \exp\{A_i\beta_{21} \\ &+ V_i(1 \\ &- A_i)\beta_{22} \\ &+ \mathbf{z}'_i\gamma_2\} S_2(y_{Gi}|A_i, V_i, \mathbf{z}_i, \omega_i, \beta_2, \gamma_2, \lambda_2) \end{aligned}$$

, where $\beta_2 = (\beta_{21}, \beta_{22})'$ and ω_i is a latent frailty for the i^{th} subject. Based on the nature of the semi-competing risks, the observations in the observed data can be classified into four different cases. Under FM, the likelihoods for these four cases are derived as follows.

Case 1—Subject died at time y_i and no disease progression was observed. Then we have $E_i = 0$, $d_i = 0$ and $v_i = 1$ and the observation is $D_i = (E_i = 0, y_i, d_i = 0, v_i = 1, \mathbf{x}_i, A_i)$. The likelihood function is given as follows:

$$L_{1i}(\alpha, \beta_0, \gamma_0, \lambda_0 | D_i) = P(E_i=0|A_i, \mathbf{x}_i, \alpha) f_0(y_i|A_i, \mathbf{x}_i, \beta_0, \gamma_0, \lambda_0). \quad (2.5)$$

Case 2—Subject was observed to have disease progression at y_{Ei} and died at y_i . Then we have $E_i = 1$, $d_i = 1$, and $v_i = 1$, and the observation is $D_i = (E_i = 1, y_{Ei}, y_{Gi} = y_i - y_{Ei}, d_i = 1, v_i = 1, \mathbf{x}_i, A_i, V_i(1 - A_i), \mathbf{z}_i)$ with the likelihood function given by:

$$L_{2i}(\alpha, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2 | D_i, \omega_i) = P(E_i=1|A_i, \mathbf{x}_i, \alpha) f_1(y_{Ei}|A_i, \mathbf{x}_i, \omega_i, \beta_1, \gamma_1, \lambda_1) \times f_2(y_{Gi}|A_i, V_i, \mathbf{z}_i, \omega_i, \beta_2, \gamma_2, \lambda_2), \quad (2.6)$$

where $P(E_i = 1|A_i, \mathbf{x}_i, \alpha) = 1 - P(E_i = 0|A_i, \mathbf{x}_i, \alpha)$.

Case 3—Subject was observed to have disease progression at y_{Ei} and right censored at y_i . Then we have $E_i = 1$, $d_i = 1$, and $v_i = 0$, and the observation is $D_i = (E_i = 1, y_{Ei}, y_{Gi} = y_i - y_{Ei}, d_i = 1, v_i = 0, \mathbf{x}_i, A_i, V_i(1 - A_i), \mathbf{z}_i)$ with the likelihood function given by

$$L_{3i}(\alpha, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2 | D_i, \omega_i) = P(E_i=1|A_i, \mathbf{x}_i, \alpha) f_1(y_{Ei}|A_i, \mathbf{x}_i, \omega_i, \beta_1, \gamma_1, \lambda_1) \times S_2(y_{Gi}|A_i, V_i, \mathbf{z}_i, \omega_i, \beta_2, \gamma_2, \lambda_2). \quad (2.7)$$

Case 4—Subject was only observed to be right censored at y_i and no disease progression occurred before y_i . Then we have $d_i = 0$ and $v_i = 0$ and the observation is $D_i = (y_i, d_i = 0, v_i = 0, \mathbf{x}_i, A_i)$ and for such a subject, it is possible that $E_i = 1$ or $E_i = 0$. The likelihood function is given by

$$\begin{aligned}
 &L_{4i}(\alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1 | D_i, \omega_i) \\
 &= P(E_i \\
 &= 1 | A_i, \mathbf{x}_i, \alpha) S_1(y_i | A_i, \mathbf{x}_i, \omega_i, \beta_1, \gamma_1, \lambda_1) \quad (2.8) \\
 &+ P(E_i \\
 &= 0 | A_i, \mathbf{x}_i, \alpha) S_0(y_i | A_i, \mathbf{x}_i, \beta_0, \gamma_0, \lambda_0).
 \end{aligned}$$

Let $D_{obs} = (D_i, i = 1, \dots, n)$ denote the observed data, where D_i is defined by (2.5) – (2.8). Then, the observed-data likelihood function under FM is given by

$$\begin{aligned}
 &L(\alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2, \tau | D_{obs}) \\
 &= \prod_{i=1}^n \left\{ [L_{1i}(\alpha, \beta_0, \gamma_0, \lambda_0 | D_i)]^{1\{d_i=0, \nu_i=1\}} [L_{2i}(\alpha, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2 | D_i)]^{1\{d_i=1, \nu_i=1\}} \right. \\
 &\quad \times [L_{3i}(\alpha, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2 | D_i)]^{1\{d_i=1, \nu_i=0\}} [L_{4i}(\alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1 | D_i)]^{1\{d_i=0, \nu_i=0\}}, \quad (2.9)
 \end{aligned}$$

where $1\{B\}$ denotes the indicator function such that $1\{B\} = 1$ if B is true and 0 otherwise, $L_{1i}(\alpha, \beta_0, \gamma_0, \lambda_0 | D_i)$ is defined by (2.5), $L_{2i}(\alpha, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2 | D_i) = P(E_i = 1 | A_i, \mathbf{x}_i, \boldsymbol{\omega}) [1 + \tau] h_1(y_{Ei} | \lambda_1)$

$$\begin{aligned}
 &\exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) h_2(y_{Gi} | \lambda_2) \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\} \\
 &\quad \times \{1 + \tau [H_1(y_{Ei} | \lambda_1) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + H_2(y_{Gi} | \lambda_2) \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\}]\}^{-(2 + \frac{1}{\tau})}, L_{3i}(\alpha, \beta_1, \gamma_1, \lambda_1) \\
 &= P(E_i \\
 &= 1 | A_i, \mathbf{x}_i, \alpha) h_1(y_{Ei} | \lambda_1) \exp(A_i \beta_1 \\
 &\quad + \mathbf{x}'_i \gamma_1) \\
 &\quad \times \{1 + \tau [H_1(y_{Ei} | \lambda_1) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + H_2(y_{Gi} | \lambda_2) \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\}]\}^{(1 - \frac{1}{\tau})}, \text{ and } L_{4i}(\alpha, \beta_0, \gamma_0, \lambda_0) \\
 &= P(E_i \\
 &= 1 | A_i, \mathbf{x}_i, \alpha) [1 + \tau H_1(y_i | \lambda_1) \times \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1)]^{-\frac{1}{\tau}} \\
 &+ P(E_i \\
 &= 0 | A_i, \mathbf{x}_i, \alpha) S_0(y_i | A_i, \mathbf{x}_i, \beta_0, \gamma_0, \lambda_0).
 \end{aligned}$$

The likelihood under CM can be derived in a similar way. Specifically, under CM, $P(E_i = 0 | A_i, \mathbf{x}_i, \boldsymbol{\omega})$, $S_0(y_i | A_i, \mathbf{x}_i, \beta_0, \gamma_0, \lambda_0)$, $f_0(y_i | A_i, \mathbf{x}_i, \beta_0, \gamma_0, \lambda_0)$, and (2.5) remain the same while (2.6) – (2.8) are obtained using

$$\begin{aligned}
 &S_1(y_{Ei} | A_i, \mathbf{x}_i, \beta_1, \gamma_1, \lambda_1) \\
 &= \exp\{-H_1(y_{Ei} | \lambda_1) \\
 &= \exp(A_i \beta_1 \\
 &\quad + \mathbf{x}'_i \gamma_1)\}, f_1(y_{Ei} | A_i, \mathbf{x}_i, \beta_1, \gamma_1, \lambda_1) = h_1(y_{Ei} | \lambda_1) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) S_1(y_{Ei} | A_i, \mathbf{x}_i, \beta_1, \gamma_1, \lambda_1), S_2(y_{Gi} | A_i, V_i, \mathbf{z}_i, \gamma_2) \\
 &+ V_i(1 \\
 &\quad - A_i) \beta_{22} \\
 &\quad + y_{Ei} \gamma_{21} + \mathbf{z}'_i \gamma_2\} S_2(y_{Gi} | A_i, V_i, \mathbf{z}_i, y_{Ei}, \beta_2, \gamma_2, \lambda_2) \text{ where } \gamma_2 \\
 &= (\gamma_{21}, \gamma'_{22})'
 \end{aligned}$$

3 Posterior Inference and Computation

3.1 Prior and Posterior Distributions

Let $\theta = (\alpha', \beta_0, \gamma_0', \lambda_0', \beta_1, \gamma_1', \lambda_1', \beta_2', \gamma_2', \lambda_2', \tau)$ denote the vector of all the model parameters. To carry out a Bayesian analysis, we need to specify a prior distribution for θ . We assume that $\alpha, (\beta_0, \gamma_0), (\beta_1, \gamma_1), (\beta_2, \gamma_2), \lambda_0, \lambda_1, \lambda_2$ and τ are independent, *a priori*, and the following priors are specified for these parameters:

$\alpha \sim N_{p_a}(0, \Sigma_a), (\beta_0, \gamma_0)' \sim N_{p_0}(0, \Sigma_0), (\beta_1, \gamma_1)' \sim N_{p_1}(0, \Sigma_1), (\beta_2, \gamma_2)' \sim N_{p_2}(0, \Sigma_2)$, and $\tau \sim \text{IG}(a_\tau, b_\tau)$, which is an inverse Gamma distribution with mean $b_\tau/(a_\tau - 1)$ and variance $(b_\tau)^2/[(a_\tau - 1)^2(a_\tau - 2)]$, where p_a, p_0, p_1 , and p_2 are the dimensions corresponding to the respective vectors of the model parameters, and $\Sigma_a, \Sigma_0, \Sigma_1, \Sigma_2$, and (a_τ, b_τ) are pre-specified hyperparameters. Independently, we assume $\lambda_{kj} \sim \text{Gamma}(a_{kj}, b_{kj})$ for $j = 1, \dots, J_k$ and $k = 0, 1, 2$. Let $\pi_a(\alpha), \pi_0(\beta_0, \gamma_0), \pi_1(\beta_1, \gamma_1), \pi_2(\beta_2, \gamma_2), \pi(\tau|a_\tau, b_\tau), \pi_0\lambda(\lambda_0), \pi_1\lambda(\lambda_1)$, and $\pi_2\lambda(\lambda_2)$ denote the above prior distributions, respectively. Then, the joint prior of θ is given by $\pi(\theta) \propto \pi_a(\alpha)\pi_0(\beta_0, \gamma_0)\pi_1(\beta_1, \gamma_1)\pi_2(\beta_2, \gamma_2)\pi(\tau|a_\tau, b_\tau)\pi_0\lambda(\lambda_0)\pi_1\lambda(\lambda_1)\pi_2\lambda(\lambda_2)$. In the simulation study in Section 4 and the analysis of the real data from a colorectal cancer study in Section 5, these hyperparameters were specified as $\Sigma_a = 1000I_{p_a}, \Sigma_0 = 1000I_{p_0}, \Sigma_1 = 1000I_{p_1}, \Sigma_2 = 1000I_{p_2}, a_\tau = b_\tau = 0.01$, and $a_{kj} = b_{kj} = 0$ for $j = 1, \dots, J_k$ and $k = 0, 1, 2$, where $I_{p_a}, I_{p_0}, I_{p_1}$, and I_{p_2} are the identity matrices. Using (2.9) and $\pi(\theta)$, the posterior distribution of θ given the observed data D_{obs} under FM is of the form

$$\pi(\theta|D_{obs}) \propto L(\theta|D_{obs})\pi(\theta). \quad (3.1)$$

When $\pi(\theta)$ is proper, the posterior distribution $\pi(\theta|D_{obs})$ is also proper. However, even when $\pi(\theta)$ is improper, the posterior distribution can still be proper under certain mild conditions. To formally establish posterior propriety in this case, let \mathcal{N}_j denote the set which consists of subjects who were in Case j and $n_j = |\mathcal{N}_j|$, which is the total number of subjects in Case j for $j = 1, \dots, 4$, respectively. Write $X_a = ((1 - 2E_i)\mathbf{x}_{E_i}, i \in \cup_{j=1}^3 \mathcal{N}_j)'$, which is an $(n_1 + n_2 + n_3) \times p_a$ matrix with rows $(1 - 2E_i)\mathbf{x}_{E_i}'$, where $\mathbf{x}_{E_i} = (1, A_i, \mathbf{x}_i)'$. Let $\delta_{i0j} = 1$ if $y_i \in (s_{0,j-1}, s_{0j}]$ and 0 otherwise for $j = 1, 2, \dots, J_0$ for $i \in \mathbf{y}_1$; $\delta_{i1j} = 1$ if $y_{Ei} \in (s_{1,j-1}, s_{1j}]$ and 0 otherwise for $j = 1, 2, \dots, J_1$ and $i \in \mathcal{N}_2 \cup \mathcal{N}_3$; and $\delta_{i2j} = 1$ if $y_i - y_{Ei} \in (s_{2,j-1}, s_{2j}]$ and 0 otherwise for $j = 1, 2, \dots, J_2$ and $i \in \mathcal{N}_2 \cup \mathcal{N}_3$. Define X_0 to be an $n_1 \times (J_0 + p_0)$ matrix with rows $(\delta_{i01}, \dots, \delta_{i0J_0}, A_i, \mathbf{x}_i')$ for $i \in \mathcal{N}_1, X_1$ an $(n_2 + n_3) \times (J_1 + p_1)$ matrix with rows $(\delta_{i11}, \dots, \delta_{i1J_1}, A_i, \mathbf{x}_i')$ for $i \in \mathcal{N}_2, \cup \mathcal{N}_3, X_2$ an $n_2 \times (J_2 + p_2)$ matrix with rows $(\delta_{i21}, \dots, \delta_{i2J_2}, A_i, V_i(1 - A_i), \mathbf{z}_i')$ for $i \in \mathcal{N}_2$. We are led to the following theorem.

Theorem 1—Assume $\pi_a(\alpha) \propto 1, \pi_0(\beta_0, \gamma_0) \propto 1, \pi_1(\beta_1, \gamma_1) \propto 1, \pi_2(\beta_2, \gamma_2) \propto 1$, and $a_{kj} = b_{kj} = 0$ for $j = 1, \dots, J_k$ and $k = 0, 1, 2$. If the following conditions are satisfied: (i) X_a is of full rank; (ii) there exists a positive vector $\mathbf{c} = (c_1, \dots, c_{n^*})' \in \mathbb{R}^{n_1+n_2+n_3}$, i.e., each component $c_i > 0$, such that $X_a'\mathbf{c} = 0$; (iii) X_0, X_1 , and X_2 are of full rank; and (iv) $a_\tau > 0$ and $b_\tau > 0$, then the joint posterior $\pi(\theta|D_{obs})$ in (3.1) is proper, i.e., $\int L(\theta|D_{obs})\pi(\theta)d\theta < \infty$.

The proof of Theorem 1 is given in Appendix A. When $a_{kj} = b_{kj} = 0$ for $j = 1, \dots, J_k$ and $k = 0, 1, 2$, we specify improper (Jeffreys's) priors for all the λ_{jk} 's, namely, $\pi_{k\lambda}(\lambda_{kj}) \propto \frac{1}{\lambda_{kj}}$ for $j = 1, \dots, J_k$ and $k = 0, 1, 2$. Conditions (i) and (ii) ensure posterior propriety for α , Condition (iii) leads to the posterior propriety of $(\lambda_0, \beta_0, \gamma_0)$ and Conditions (iii) and (iv) are required for the posterior propriety of $(\lambda_1, \beta_1, \gamma_1, \lambda_2, \beta_2, \gamma_2, \tau)$. Condition (iii) is quite mild and

essentially requires that at least one event (death or disease progression) occurs in each interval $(s_{k,j-1}, s_{kj}]$, and the corresponding covariate matrix is of full rank. These conditions are easily satisfied in most applications and are quite easy-to-check.

3.2 The Predictive Survival Function with Partial Treatment Switching

An inferential research goal in this research is to compare the survival function of the death time in the setting when no subjects have switched treatment. Let $T_D^*(a)$ denote a potential survival time when a subject receives treatment a at the time of randomization and stays on the same treatment over the entire study duration. Let $S_a(t|\theta) = P(T_D^*(a) > t|\theta)$. Following Zeng et al. (2012), we state the following two assumptions: (i) Treatment A is completely randomized and $T_D^*(a) = T_D(a)$ if a subject never switches treatment; and (ii) Given $(A = 0, z, T_E = u)$ or $(A = 1, z, T_E = u)$, V is independent of the potential outcomes $\{T_D^*(0), T_D^*(1)\}$. We note that these two assumptions are only used to compute $S_a(t|\theta)$. Similar to Zeng et al. (2012), under Assumptions (i) and (ii), we have

$$\begin{aligned}
 S_a(t|\theta) &= \int_{\mathbf{x}} P(T_D > t | A \\
 &= a, \mathbf{x}, E \\
 &= 0, \beta_0, \gamma_0, \lambda_0) P(E \\
 &= 0 | A \\
 &= a, \mathbf{x}, \alpha) f_{\mathbf{x}}(\mathbf{x} | A \\
 &= a) d\mathbf{x} \\
 &+ \int_{\mathbf{x}, \mathbf{z}, \omega, u} P(T_G > t \\
 &- u | A \\
 &= a, \mathbf{z}, V \\
 &= 0, E \\
 &= 1, \beta_2, \gamma_2, \lambda_2, \omega) \times f_1(u | A \\
 &= a, \mathbf{x}, \omega, \beta_1, \gamma_1, \lambda_1, \omega) du f(\omega | \tau) d\omega \\
 &\times f_{\mathbf{z}}(\mathbf{z} | A \\
 &= a, \mathbf{x}, E \\
 &= 1) P(E \\
 &= 1 | A \\
 &= a, \mathbf{x}, \alpha) f_{\mathbf{x}}(\mathbf{x} | A = a) dz d\mathbf{x},
 \end{aligned} \tag{3.2}$$

where $f_{\mathbf{x}}(\mathbf{x} | A = a)$ is the conditional density of \mathbf{X} given $A = a$, and $f_{\mathbf{z}}(\mathbf{z} | A = a, \mathbf{x}, E = 1)$ is the conditional density of \mathbf{Z} given $A = a, \mathbf{x}$, and $E = 1$. When $J_0 = J_1 = J_2 = 1$, after some algebra, we obtain

$$\begin{aligned}
 S_a(t|\theta) &= \int_{\mathbf{x}} \exp\{-t\lambda_0 \exp(A\beta_0 + \mathbf{x}'\gamma_0)\} P(E \\
 &= 0|A \\
 &= a, \mathbf{x}, \alpha) f_X(\mathbf{x}|A \\
 &= a) d\mathbf{x} \\
 &+ \int_{\mathbf{x}, z} \left(\left[\frac{1}{1 + \lambda_1 \exp(A\beta_1 + \mathbf{x}'\gamma_1)t\tau} \right]^{\frac{1}{\tau}} \right. \\
 &+ \frac{\lambda_1 \exp(A\beta_1 + \mathbf{x}'\gamma_1)}{\lambda_2 \exp\{A\beta_{21} + V(1 - A)\beta_{22} + z'\gamma_2\} - \lambda_1 \exp(A\beta_1 + \mathbf{x}'\gamma_1)} \\
 &\times \left. \left\{ \left[\frac{1}{1 + \lambda_1 \exp(A\beta_1 + \mathbf{x}'\gamma_1)t\tau} \right]^{\frac{1}{\tau}} - \left[\frac{1}{1 + \lambda_2 \exp\{A\beta_{21} + V(1 - A)\beta_{22} + z'\gamma_2\}t\tau} \right]^{\frac{1}{\tau}} \right\} \times f_Z(z|A \right. \\
 &= a, \mathbf{x}, E \\
 &= 1) P(E \\
 &= 1|A \\
 &= a, \mathbf{x}, \alpha) f_X(\mathbf{x}|A=a) dz d\mathbf{x}.
 \end{aligned} \tag{3.3}$$

A detailed derivation of (3.3) is given in Appendix B. We assume nonparametric distributions for $f_X(\mathbf{x} | A = a)$ and $f_Z(z | X, A = a, E = 1)$ as follows:

$$\begin{aligned}
 f_X(\mathbf{x}|A=a) &= \sum_{j=1}^n I(X_j = \mathbf{x}, A_j = a) / \sum_{j=1}^n I(A_j = a) \text{ and} \\
 f_Z(z|X, A=a, E=1) &= \frac{\sum_{j \in (\text{Cases}_{2,3})} \delta(Z_j = z) K_{a_n}(\|X_j - X\|) I(A_j = a)}{\sum_{j \in (\text{Cases}_{2,3})} K_{a_n}(\|X_j - X\|) I(A_j = a)}.
 \end{aligned}$$

Since $S_a(t|\theta)$ is a function of θ , the posterior estimates of $S_a(t|\theta)$ can be easily obtained using the MCMC samples from the posterior distribution of θ .

3.3 Posterior Computation

Due to the complexity of the likelihood structure for the proposed frailty model, an analytical evaluation of the posterior distribution is not possible. In order to carry out posterior inference, we develop an efficient Gibbs sampling algorithm to sample θ from the posterior distribution in (3.1). We first consider the transformation $\lambda_k^* = \tau \lambda_k$. The Jacobian of

this transformation is $\left| \frac{\partial \lambda_k}{\partial \lambda_k^*} \right| = \tau^{-J_k}$ for $k = 1, 2$. Write $\theta^* = (\alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \tau)$. After the transformation, the posterior distribution of θ^* is given by

$$\pi(\theta^* | D_{obs}) \propto L(\alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1^*/\tau, \beta_2, \gamma_2, \lambda_2^*/\tau, \tau | D_{obs}) \pi_a(\alpha) \pi_0(\beta_0, \gamma_0) \pi_1(\beta_1, \gamma_1) \times \pi_2(\beta_2, \gamma_2) \pi(\tau | a_\tau, b_\tau) \pi_{0\lambda}(\lambda_0) \pi_{1\lambda}(\lambda_1^*/\tau) \pi$$

where $L(\alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \tau | D_{obs})$ is defined in (2.9).

To facilitate the posterior computation, we introduce two sets of latent variables $\mathbf{E}^* = (E_i^*, i \in \mathcal{N}_4)$ and $\mathbf{w} = (w_1, w_2, \dots, w_n)$ so that the augmented posterior distribution of $(\theta^*, \mathbf{E}^*, \mathbf{w})$ is given by

$$\begin{aligned} \pi(\theta^*, \mathbf{w}, \mathbf{E}^* | D_{obs}) &\propto \prod_{i=1}^n \left\{ [L_{1i}(\alpha, \beta_0, \gamma_0, \lambda_0 | D_i)]^{1\{\nu_i=1, d_i=0\}} [L_{2i}^*(\alpha, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2, w_i | D_i)]^{1\{\nu_i=1, d_i=1\}} \right. \\ &\times [L_{3i}^*(\alpha, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2, w_i | D_i)]^{1\{\nu_i=0, d_i=1\}} \\ &\times [L_{4i}^*(\alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1^*, w_i, E_i^* | D_i)]^{1\{\nu_i=0, d_i=0\}} \left. \right\} \times \pi_a(\alpha) \pi_0(\beta_0, \gamma_0) \\ &\times \pi_1(\beta_1, \gamma_1) \pi_2(\beta_2, \gamma_2) \pi(\tau | a_\tau, b_\tau) \pi_{0\lambda}(\lambda_0) \left[\prod_{k=1}^2 \pi_{k\lambda}(\lambda_k^* / \tau) \right] \tau^{-(J_1+J_2)}, \end{aligned} \tag{3.5}$$

Where $L_{1i}(\alpha, \beta_0, \gamma_0, \lambda_0 | d_i)$ is defined by (2.5),

$$\begin{aligned} L_{2i}^*(\alpha, \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, w_i | D_i) &= P(E_i \\ &= 1 | A_i, \mathbf{x}_i, \alpha) \\ &\times \left[(1 \right. \\ &+ \tau) / \tau^2 \Big] h_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 \\ &+ \mathbf{x}_i' \gamma_1) h_2(y_{G_i} | \lambda_2^*) \exp\{A_i \beta_{21} \\ &+ V_i(1 \\ &- A_i) \beta_{22} \\ &+ z_i' \gamma_2 \Big] \left[\frac{w_i^{1+\frac{1}{\tau}}}{\Gamma(2+\frac{1}{\tau})} \right] \exp(\\ &- w_i [1 \\ &+ H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 \\ &+ \mathbf{x}_i' \gamma_1) \\ &+ H_2(y_{G_i} | \lambda_2^*) \exp\{A_i \beta_{21} \\ &+ V_i(1 \\ &- A_i) \beta_{22} \\ &+ z_i' \gamma_2 \Big] \Big], L_{3i}^*(\alpha, \beta_1, \gamma_1, \lambda_1, \beta_2, \gamma_2, \lambda_2, w_i | D_i) = P(E_i = 1 | A_i, \mathbf{x}_i, \alpha) (1/\tau) h_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}_i' \gamma_1) \left[\frac{w_i^{\frac{1}{\tau}}}{\Gamma(1+\frac{1}{\tau})} \right] \\ &+ V_i(1 \\ &- A_i) \beta_{22} \\ &+ z_i' \gamma_2 \Big] \Big], \text{ and } L_{4i}^*(\alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1^*, w_i, E_i^* | D_i) \\ &= \{ P(E_i = 1 | A_i, \mathbf{x}_i, \alpha) \}^{-E_i^*} \left[\frac{w_i^{\frac{1}{\tau}-1} E_i^*}{\Gamma(\frac{1}{\tau})} \right] \exp\{ \\ &- w_i E_i^* [1 \\ &+ H_1(y_i | \lambda_1^*) \exp(A_i \beta_1 \\ &+ \mathbf{x}_i' \gamma_1) \Big] \} \{ P(E_i = 0 | A_i, \mathbf{x}_i, \alpha) S_0(y_i | A_i, \mathbf{x}_i, \beta_0, \gamma_0, \lambda_0) \}^{1-E_i^*} \end{aligned}$$

. It can be shown that $\int_{E^*} \pi(\theta^*, \mathbf{w}, \mathbf{E}^* | D_{obs}) d\mathbf{w} = \pi(\theta^* | D_{obs})$, which is given by (3.4). We note that the latent variables (the w_i 's) in (3.5) are different than those ω_i 's in (2.6) – (2.8).

Let $A|B$ denote the conditional distribution of A given B . To run the Gibbs sampling algorithm, we sample from the following conditional distributions in turn: (i)

$[\lambda_0, \lambda_1^*, \lambda_2^* | \beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2, \mathbf{w}, \mathbf{E}^*, D_{obs}]$; (ii)

$[\beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2, \tau, \mathbf{w}, \mathbf{E}^* | \alpha, \lambda_0, \lambda_1^*, \lambda_2^*, D_{obs}]$; and (iii) $[\alpha | \mathbf{E}^*, D_{obs}]$. For (ii), we use the modified collapsed Gibbs technique (Liu, 1994; Chen et al., 2000). It is easy to show that

$$\begin{aligned} & [\beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2, \tau, \mathbf{w}, \mathbf{E}^* | \alpha, \lambda_0, \lambda_1^*, \lambda_2^*, D_{obs}] \\ &= [\beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2, \tau, \mathbf{E}^* | \alpha, \lambda_0, \lambda_1^*, \lambda_2^*, D_{obs}] [\mathbf{w} | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \tau, \mathbf{E}^*, D_{obs}]. \end{aligned} \quad (3.6)$$

For (ii), following Chen et al. (2000) and using (3.6), we run a sub-Gibbs sampling algorithm to draw from the following conditional distributions: (iia)

$[\beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2, | \alpha, \lambda_0, \lambda_1^*, \lambda_2^*, \tau, \mathbf{E}^*, D_{obs}]$; (iib) $[\mathbf{E}^* | \alpha, \beta_0, \gamma_0, \lambda_0, \beta_1, \gamma_1, \lambda_1^*, \tau, D_{obs}]$;

(iic) $[\tau | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs}]$; and (iid) $[\mathbf{w} | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs}]$. Next, we will only discuss the properties of the conditional distribution

$[\tau | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs}]$ and how to sample τ from this conditional distribution. All other conditional distributions are discussed in detail in Appendix B. We first consider the transformation $\tau^* = 1/\tau$. Then, the conditional posterior density of τ^* is given by

$$\pi(\tau^* | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs}) \propto \prod_{i=1}^n \left\{ \left(\frac{1}{\tau^*} + 1 \right) (\tau^*)^2 \times [1 + H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + H_2(y_{G_i} | \lambda_2^*) \exp\{A_i \beta_2 + V_i(1 - A_i) \beta_2\}] \right\}$$

We are led to the following theorem.

Theorem 2—Assume that

$\sum_{i=1}^n [1\{d_i=1, \nu_i=1\} + 1\{d_i=1, \nu_i=0\}] + a_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_K} a_{kj} - 3 > 0$. Then we have (i) the conditional density of τ^* given by (3.7) is log-concave; and (ii) the mode of (3.7) is analytically available and given by

$$\hat{\tau}^* = \frac{-(B_1 + B_2 + B_3) - \sqrt{(B_1 + B_2 + B_3)^2 - 4B_1B_2}}{2B_1} \quad (3.8)$$

$$\begin{aligned}
 B_1 &= \sum_{i=1}^n \{ -[1\{d_i \\
 &= 1, \nu_i \\
 &= 1\} + 1\{d_i \\
 &= 1, \nu_i \\
 &= 0\}] \log[1 \\
 &+ H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 \\
 &+ \mathbf{x}'_i \gamma_1) \\
 &+ H_2(y_{G_i} | \lambda_2^*) \exp\{A_i \beta_{21} \\
 &+ V_i(1 - A_i) \beta_{22} \\
 &+ \mathbf{z}'_i \gamma_2\}] \\
 &- E_i^* 1\{d_i = 0, \nu_i = 0\} \log[1 \\
 &+ H_1(y_i | \lambda_1^*) \exp(A_i \beta_1 \\
 &+ \mathbf{x}'_i \gamma_1)] \} - [b_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_k} (b_{kj} \lambda_{kj}^*)], \beta_2 \\
 &= \sum_{i=1}^n [1\{d_i \\
 &= 1, \nu_i \\
 &= 1\} + 1\{\nu_i \\
 &= 0, d_i \\
 &= 1\}] \\
 &+ a_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_k} a_{kj} \\
 &- 3, \text{ and } B_3 \\
 &= \sum_{i=1}^n [1\{d_i \\
 &= 1, \nu_i = 1\}].
 \end{aligned}$$

where

The proof of Theorem 2 is given in Appendix A. The assumption

$\sum_{i=1}^n [1\{d_i = 1, \nu_i = 1\} + 1\{d_i = 1, \nu_i = 0\}] + \alpha_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_k} a_{kj} - 3 > 0$ ensures the log-concavity and the existence of the mode. This assumption is quite mild. As long as there are more than three patients with disease progression, this assumption still holds even when the improper priors with $a_\tau = 0$ and $a_{kj} = 0$ for all k and j are specified for τ and the λ_{jk} 's. With the log-concavity property, τ^* can be exactly drawn from the conditional distribution in (3.7) using the adaptive rejection algorithm of Gilks and Wild (1992). After τ^* is generated, we let $\tau = 1/\tau^*$ and then the value of τ is a sample from the conditional distribution $[\tau | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs}]$ in (ii). With the analytical form of the mode, the performance of the rejection algorithm can be improved substantially as the algorithm does not need to search for the mode.

3.4 Model Comparison

To carry out Bayesian model comparison, we consider the deviance information criterion (DIC) and the Logarithm of the PseudoMarginal Likelihood (LPML). We define the deviance $\text{Dev}(\boldsymbol{\theta}) = -2 \log L(\boldsymbol{\theta} | D_{obs})$, where $L(\boldsymbol{\theta} | D_{obs})$ is the observed-data likelihood defined in (2.9). Let $\bar{\boldsymbol{\theta}}$ and $\overline{\text{Dev}} = E[\text{Dev}(\boldsymbol{\theta}) | D_{obs}]$ denote the posterior mean of $\boldsymbol{\theta}$ and $\text{Dev}(\boldsymbol{\theta})$, respectively. According to Spiegelhalter et al. (2002), the DIC measure is defined as $\text{DIC} = \text{Dev}(\bar{\boldsymbol{\theta}}) + 2p_D$, where $p_D = \overline{\text{Dev}} - D(\bar{\boldsymbol{\theta}})$ is the effective number of model parameters. The smaller the DIC value, the better the model fits the data. LPML is another useful Bayesian

measure of goodness-of-fit statistic, which is defined based on the Conditional Predictive Ordinate (CPO). For the i^{th} observation, we define CPO as $\text{CPO}_i = \int L(\theta|D_i) \pi(\theta|D^{(-i)}) d\theta$, where D_i is the observed data defined in Section 2.2, $L(\theta|D_i)$ is the observed likelihood for the i^{th} subject, which is the term inside the product in (2.9), $D^{(-i)}$ is the data with D_i deleted, and $\pi(\theta|D^{(-i)})$ is the posterior density of θ based on the data $D^{(-i)}$. According to Ibrahim et al. (2001), $\text{LPML} = \sum_{i=1}^n \log(\text{CPO}_i)$. The larger the LPML value, the better the model fits the data.

4 A Simulation Study

To examine the empirical performance of the posterior estimates and DIC and LPML, we carry out a simulation study. Five hundred simulated data sets with $n = 500$ as well as $n = 1,000$ were generated. In the simulation study, the baseline treatment A was generated from a Bernoulli(0.5), corresponding to a randomized trial with a 1:1 sample size allocation; two baseline covariates X_1 and X_2 were independently generated from a $U(-1, 1)$ and a Bernoulli(0.6), respectively. Given A and (X_1, X_2) , E was generated from model (2.1) with the coefficients (including an intercept) being 1.6, -1.8 , 1, and 0.1, respectively. When $E = 0$, we simulated T_D from model (2.2) with $H_0(t) = t$, $\beta_0 = -1$ and $(\gamma_{01}, \gamma_{02}) = (1, 0.2)$. For $E = 1$, we first generated ω from a Gamma($1/\tau, 1/\tau$) with $\tau = 1$. Then, T_E was generated from model (2.3) with $H_1(t) = 5t$, $\beta_1 = -0.5$ and $(\gamma_{11}, \gamma_{12}) = (1, 0)$ and an additional prognostic factor Z at disease progression was generated from a $U(0, 10)$ while the selection into treatment switching (V) for a subject in the control arm ($A = 0$) was from a Bernoulli(p), where $p = \exp(-0.5 + 0.3T_E + 0.2X_1 + 0.5Z) / [1 + \exp(-0.5 + 0.3T_E + 0.2X_1 + 0.5Z)]$. Moreover, T_G was generated from the model in (2.3) with $H_2(t) = t$, $\beta_{21} = -0.3$, $\beta_{22} = -0.5$, and $\gamma_{21} = -0.5$, $\gamma_{22} = 0.5$, $\gamma_{23} = -0.4$. Finally, the censoring time was generated from a $U(1, 7)$ and the study duration was $\tau^* = 3$. The latter yielded the average proportions of Cases 1 to 4 as 23%, 39%, 19%, and 18%.

For each simulated dataset, we fit the proposed FM with various values of (J_0, J_1, J_2) and computed DIC and LPML. The mean values of the DICs and LPMLs over the 500 simulated datasets were 2986.22 and -1493.24 for $(J_0, J_1, J_2) = (1, 1, 1)$; 2998.05 and -1499.39 for $(J_0, J_1, J_2) = (5, 5, 5)$; and 3013.21 and -1507.22 for $(J_0, J_1, J_2) = (10, 10, 10)$. We note that the true value of (J_0, J_1, J_2) is $(1, 1, 1)$. Thus, both DIC and LPML correctly identified the true model. Under the best combination of (J_0, J_1, J_2) , namely, $(1, 1, 1)$, the average of the posterior means (EST), and the average of the posterior standard deviations (SD), the simulation standard error (SE), the root of the mean squared error (RMSE), and the coverage probability (CP) of the 95% highest posterior density (HPD) intervals for each parameter as well as $S_d(t|\theta)$ were computed. The results are given in Table 1. Table 1 shows excellent empirical performance of the posterior estimates for all the parameters as well as the survival probabilities for both $n = 500$ and $n = 1000$. In particular, the ESTs are nearly identical to the true values, the SDs are very close to the SEs, and the CPs are very close to 95%. For each simulated dataset, we also fit CM as discussed in Section 2.1 and computed the corresponding DIC and LPML for $(J_0, J_1, J_2) = (1, 1, 1)$. The box plots of the DIC and LPML differences between CM and FM are shown in Fig. 2. From this figure, we see that all of the DIC differences are above 0 and all LPML differences are below 0, indicating that the frailty model fits the data better than the conditional model for all 500 simulated data sets, which is expected since the data were generated from the frailty model. These results further empirically confirm that FM is indeed quite different from CM, and DIC and LPML are two effective Bayesian model comparison measures for identifying the true models.

5 Analysis of the Panitumumab Study

We carry out here a detailed analysis of a subset of the data from the panitumumab study (PMAB408) conducted by Amgen Inc. (van Cutsem et al., 2007 and Amado et al., 2008). PMAB408 was an open label, randomized, phase III multicenter study designed to compare the efficacy and safety of panitumumab plus best supportive care (P+BSC) versus BSC alone in subjects with EGFr-expressing metastatic colorectal cancer who had documented disease progression during or after prior standard treatment with fluoropyrimidine, irinotecan, and oxaliplatin chemotherapy. Subjects were randomly assigned to receive P+BSC (treatment) or BSC (control). The baseline covariates include initial treatment (P+BSC versus BSC), age in years at screening, baseline Eastern Cooperative Oncology Group (ECOG) score (score 0 or 1 versus ≥ 2 (bECOG01)), primary tumor diagnosis type (rectal versus colon (Rectal)), gender, and region (western Europe (WesternEU), eastern and central Europe (CenEstEU), and rest of the world). In the subset of the data, there were 223 and 231 patients in the control and treatment arms, respectively. There were 424 subjects who died (208 and 207 in the control and treatment arms, respectively), 387 subjects (201 and 186 in the control and treatment arms, respectively) who developed disease progression, and 59 subjects (18 and 41 in the control and treatment arms, respectively) who died without disease progression. The median age was 62.5 years with interquartile range (55, 69) years. There were 388 patients with ECOG score 0 or 1, 287 were males, 151 had rectal cancer, 352 were from Western Europe, 39 were from Eastern and Central Europe, and 63 were from the rest of the world. The median follow-up time was 189.5 days and the interquartile range of the follow-up time was (93, 334) days. Among those 387 patients who developed disease progression, the median disease progression time is 53 days and the interquartile range is (45, 84) days. Of these 201 patients who developed disease progression in the control arm, 167 patients were switched to the treatment arm at the time of disease progression.

The model for the time in months to disease progression includes all the baseline covariates. Among the 387 patients who developed disease progression, the median age at the time of disease progression was 62.1 years with interquartile range (55.0, 69.1), the numbers of patients who had partial response, stable disease, and progressive disease were 19, 86, and 282, respectively. There were 348 patients with baseline ECOG score 0 or 1, 286 patients had a last ECOG score on or prior to disease progression 0 or 1, and 180 patients had grade 2 or above adverse events. The covariates for the time in months from disease progression to death include treatment, bECOG01, age at disease progression, best tumor response with partial response (BTR PR) or stable disease (BTR SD) versus progressive disease according to investigator assessment, last ECOG score on or prior to disease progression (score 0 or 1 versus ≥ 2 (LECOG01)), and adverse events (AE).

We fit both FM and CM with different values of J_0 , J_1 and J_2 to the panitumumab data. The DIC and LPML values are given in Table 2. We see from Table 2 that $(J_0, J_1, J_2) = (1, 30, 5)$ achieves the smallest DIC value and the largest LPML value among the 7 combinations of (J_0, J_1, J_2) considered here under both FM and CM and the best DIC and LPML values were 3475.27 and -1741.32 under FM and 3482.62 and -1746.76 under CM, respectively. We also observe that for each of these seven combinations of (J_0, J_1, J_2) , FM consistently has a smaller DIC value and a larger LPML value than CM, implying that FM fits the panitumumab data better than CM.

Table 3 shows the posterior estimates of the model parameters under FM with $(J_0, J_1, J_2) = (1, 30, 5)$. The 95% HPD intervals for treatment were $(-1.753, -0.484)$ under the E model, $(-1.145, 0.175)$ under the T_D model, $(-1.733, -1.148)$ under the T_E model, and $(-1.479, -0.441)$ under the T_G model.

These results imply that treatment is associated with E , T_E , T_G but not with T_D . The other important prognostic factors include bCOG01 under the E , T_D , and T_E model, LECOG01 under the T_G model, age under the T_E model, and AE under the T_G model as their corresponding 95% HPD intervals do not contain 0. The treatment switching variable, V , is also associated with T_G . The posterior mean and 95% HPD interval for τ were 0.322 and (0.163, 0.490), which implies that there is a moderate dependence between T_E and T_G . We also fit the best CM with $(J_0, J_1, J_2) = (1, 30, 5)$ to the panitumumab data and the posterior mean and 95% HPD interval for γ_{21} in (2.4) were -0.083 and $(-0.161, -0.008)$, which implies that there is a positive association between T_E and T_G . Panel (a) in Figure 3 shows the estimated differences of the survival probabilities and their pointwise 95% confidence intervals (CIs) between the two treatment groups of P+BSC and BSC using the intent-to-treat (ITT) Kaplan-Meier approach and Panel (b) plots the posterior estimates, $E[S_1(t|\theta) - S_0(t|\theta)|D_{obs}]$, where $S_0(t|\theta)$ and $S_1(t|\theta)$ are given in (3.2), and the corresponding pointwise 95% HPD intervals of $S_1(t|\theta) - S_0(t|\theta)$ between these two treatment groups. From Panel (a) of Figure 3, we see that the ITT approach yields no difference between two treatment groups as all 95% CIs contain 0. In contrast, as shown in Panel (b) of Figure 3, all the posterior estimates of $S_1(t|\theta) - S_0(t|\theta)$ are above 0 and the corresponding 95% HPD intervals are above 0 after 2.25 months. We note that the maximum estimated difference $E[S_1(t|\theta) - S_0(t|\theta)|D_{obs}]$ was attained at approximately 9 months and the corresponding posterior mean and 95% HPD interval were 0.165 and (0.110, 0.227). These posterior estimates indicate that P+BSC does yield a higher survival probability than BSC.

In all of the Bayesian computations, we used 20,000 Gibbs samples after a burn-in of 1000 for each model to compute all the posterior estimates, including posterior means, posterior standard deviations, and 95% HPD intervals. Codes were written for the FORTRAN 95 compiler using IMSL subroutines with double precision accuracy. The convergence of the Gibbs sampler was checked using several diagnostic procedures discussed in Chen et al. (2000). The autocorrelations for all model parameters disappeared before lag 10.

6 Discussion

In this paper, we have proposed a novel semi-competing risks Bayesian frailty model that accommodates treatment switching and dependence between the progression time and survival time. This type of scenario arises often in clinical trials in which, once a patient experiences an event, such as progression, they immediately switch to the experimental treatment. As a result of the switch, the model attempts to capture the treatment effect when no subjects would have switched treatment. The innovation in the model lies in the fact that the observed data likelihood is modeled and is based on four possible scenarios, and the model itself has three components. This type of model is quite different from what has been proposed in the literature. Another innovation here lies in the Bayesian approach to fit the model. Efficient MCMC methods based on the collapsed Gibbs sampler facilitate a flexible Bayesian model that is computationally feasible and identifiable. Such a model does not appear computationally feasible from a frequentist perspective. As shown in the simulation studies and real data analysis, our proposed model has several advantages over the conditional model (CM) proposed by others in the literature. It appears to have better performance under certain scenarios and produces a better model fit according to DIC and LPML. The proposed model is useful for practitioners encountering treatment switching studies in the presence of semi-competing risks where one is interested in assessing the treatment effect on overall survival.

Acknowledgments

The authors wish to thank the Editor-in-Chief, the Associate Editor, and the referee 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.

References

- Aalen OO, Johansen S, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*. 1978; 5:141–150.
- Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD, Chang DD. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *Journal of Clinical Oncology*. 2008; 28:1626–1634. [PubMed: 18316791]
- Andersen, PK.; Borgan, O.; Gill, RD.; Keiding, N. *Statistical Models Based on Counting Processes*. New York: Springer; 1993.
- Banerjee T, Chen MH, Dey DK, Kim S. Bayesian analysis of generalized odds-rate hazards models for survival data. *Lifetime Data Analysis*. 2007; 13:241–260. [PubMed: 17401683]
- Chen MH, Shao QM. Propriety of posterior distribution for dichotomous quantal response models with general link functions. *Proceedings of the American Mathematical Society*. 2001; 129:293–302.
- Chen, MH.; Shao, QM.; Ibrahim, JG. *Monte Carlo Methods in Bayesian Computation*. New York: Springer; 2000.
- Day R, Bryant J. Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators. *Biometrika*. 1997; 84:45–56.
- Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika*. 2001; 88:907–919.
- Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Appl Stat*. 1992; 41:337–348.
- Ghosh D. Semiparametric inferences for association with semi-competing risks data. *Statistics in Medicine*. 2006; 25:2059–2070. [PubMed: 16196081]
- Ibrahim, JG.; Chen, MH.; Sinha, D. *Bayesian Survival Analysis*. New York: Springer; 2001.
- Liu JS. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*. 1994; 89:958–966.
- Mandel M. The competing risks illness-death model under cross-sectional sampling. *Biostatistics*. 2010; 11:290–303. [PubMed: 19933879]
- Marcus SM, Gibbons RD. Estimating the efficacy of receiving treatment in randomized clinical trials with noncompliance. *Health Services & Outcomes Research Methodology*. 2001; 2:247–258.
- Oakes D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*. 1989; 84:487–493.
- Peng L, Fine JP. Regression Modeling of Semicompeting Risks Data. *Biometrics*. 2007; 63:96–108. [PubMed: 17447934]
- Shen Y, Thall PF. Parametric likelihoods for multiple non-fatal competing risks and death. *Statistics in Medicine*. 1998; 17:999–1015. [PubMed: 9612887]
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with Discussion). *Journal of the Royal Statistical Society, Series B*. 2002; 64:583–639.
- Van Cutsem E, Peeters M, Siena S, Humblet Y, Hendlisz A, Neyn B, Canon JL, Van Laethem JL, Maurel J, Richardson G, Wolf M, Amado RG. Open-label Phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *Journal of Clinical Oncology*. 2007; 25:1658–1664. [PubMed: 17470858]
- Wang W. Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society, Series B*. 2003; 65:257–273.
- Zeng D, Chen Q, Chen MH, Ibrahim JG. Amgen research group. Estimating treatment effects with treatment switching via semi-competing risks models: An application to a colorectal cancer study. *Biometrika*. 2012; 99:167–184. [PubMed: 23049136]

Zhao, L. Unpublished Ph.D dissertation. Simon Fraser University; 2009. Multi-state Processes with Duration-dependent Transition Intensities: Statistical Methods and Applications.

Appendix A: Proofs of Theorems

Proof of Theorem 1

Using (3.1) with the prior distributions assumed in the theorem, we have

$$\pi(\theta|D_{obs}) \propto L(\theta|D_{obs}) \left[\prod_{k=0}^2 \prod_{j=1}^{J_K} \frac{1}{\lambda_{kj}} \right] \left[\frac{1}{\tau^{a\tau+1}} \exp\left(-\frac{b\tau}{\tau}\right) \right].$$

Using (2.9), it is easy to show that

$$L(\theta|D_{obs}) \leq L_a(\alpha|D_{obs}) L_1(\beta_0, \gamma_0, \lambda_0|D_{obs}) L_2(\beta_1, \gamma_1, \lambda_1, \beta_{21}, \beta_{22}, \gamma_2, \lambda_2, \tau|D_{obs}) L_3(\beta_1, \gamma_1, \lambda_1, \tau|D_{obs}).$$

$$\text{where } L_a(\alpha|D_{obs}) = \prod_{i \in \cup_{\ell=1}^3 \mathcal{N}_\ell} \frac{\exp\{E_i(\alpha_0 + A_i \alpha_1 + \mathbf{x}'_i \alpha_2)\}}{1 + \exp(\alpha_0 + A_i \alpha_1 + \mathbf{x}'_i \alpha_2)},$$

$$L_1(\beta_0, \gamma_0, \lambda_0|D_{obs}) = \prod_{i \in \mathcal{N}_1} h_0(y_i|\lambda_0) \exp(A_i \beta_0 + \mathbf{x}'_i \gamma_0) \exp\{-H_0(y_i|\lambda_0) \exp(A_i \beta_0 + \mathbf{x}'_i \gamma_0)\}, \quad (\text{A. } 1)$$

$$L_2(\beta_1, \gamma_1, \lambda_1, \beta_{21}, \beta_{22}, \gamma_2, \lambda_2, \tau|D_{obs})$$

$$= \prod_{i \in \mathcal{N}_2} \left[(1+\tau) h_1(y_{E_i}|\lambda_1) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) \times h_2(y_{G_i}|\lambda_2) \exp\{A_i \beta_{21} + V_i(1-A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\} \times \left\{ 1 + \tau \left(H_1(y_{E_i}|\lambda_1) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) \right) \right\} \right] \quad (\text{A. } 2)$$

and

$$L_3(\beta_1, \gamma_1, \lambda_1, \tau|D_{obs}) = \prod_{i \in \mathcal{N}_3} h_1(y_{E_i}|\lambda_1) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) \left\{ 1 + \tau H_1(y_{E_i}|\lambda_1) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) \right\}^{-(1+\frac{1}{\tau})}. \quad (\text{A. } 3)$$

To prove propriety of the posterior, it is sufficient to show (a) $\int L_a(\mathbf{a}|D_{obs}) d\mathbf{a} < \infty$; (b)

$$\int L_1(\beta_0, \gamma_0, \lambda_0|D_{obs}) \left(\prod_{j=1}^{J_0} \frac{1}{\lambda_{0j}} \right) d\beta_0 d\gamma_0 d\lambda_0 < \infty \text{ and (c)}$$

$$\int L_2(\beta_1, \gamma_1, \lambda_1, \beta_{21}, \beta_{22}, \gamma_2, \lambda_2, \tau|D_{obs}) L_3(\beta_1, \gamma_1, \lambda_1, \tau|D_{obs}) \left[\prod_{k=1}^2 \prod_{j=1}^{J_K} \frac{1}{\lambda_{kj}} \right] \left[\frac{1}{\tau^{a\tau+1}} \exp\left(-\frac{b\tau}{\tau}\right) \right] d\beta_1 d\gamma_1 d\lambda_1 d\beta_{21} d\beta_{22} d\gamma_2 d\lambda_2 d\tau < \infty.$$

Under Conditions (i) and (ii), Theorem 2.1 of Chen and Shao (2001) directly leads to (a).

Let j_i be an index such that $s_{0,j_i-1} < y_i \leq s_{0,j_i}$. Then we have $\delta_{i0j} = 1$ for $j = j_i$ and $\delta_{i0j} = 0$ for $j \neq j_i$ and

$$\begin{aligned}
 & h_0(y_i|\lambda_0)\exp(A_i\beta_0 \\
 & \quad + \mathbf{x}'_i\gamma_0)\exp\{ \\
 & \quad - H_0(y_i|\mathbf{x}_i)\exp(A_i\beta_0 + \mathbf{x}'_i\gamma_0)\} \leq \lambda_{0j_i}\exp(A_i\beta_0 + \mathbf{x}'_i\gamma_0)\exp\{ \\
 & \quad - \lambda_{0j_i}(y_i \\
 & \quad - s_{0,j_i-1})\exp(A_i\beta_0 \\
 & \quad + \mathbf{x}'_i\gamma_0)\} \leq M_{11},
 \end{aligned} \tag{A.4}$$

where $M_1 > 0$ is a constant. Consider the transformation $\xi_{0j} = \log(\lambda_{0j})$, where $d\xi_{0j} = \frac{d\lambda_{0j}}{\lambda_{0j}}$ for $j = 1, \dots, J_0$. Under condition (iii), there exist $J_0 + p_0$ distinct $i_1, \dots, i_{J_0+p_0} \in \mathcal{N}_1$ such that the $(J_0 + p_0) \times (J_0+p_0)$ matrix X_0^* , which has rows $(\delta_{i_\ell 01}, \dots, \delta_{i_\ell 0J_0}, A_{i_\ell}, \mathbf{x}'_{i_\ell})$ for $\ell = 1, \dots, J_0+p_0$, is of full rank. Using (A.4), we have

$$\int L_1(\beta_0, \gamma_0, \lambda_0|D_{obs}) \left(\prod_{j=1}^{J_0} \frac{1}{\lambda_{0j}} \right) d\beta_0 d\gamma_0 d\lambda_0 \leq M_{12} \int \left[\prod_{\ell=1}^{J_0+p_0} \exp(\xi_{0j_{i_\ell}} + A_{i_\ell}\beta_0 + \mathbf{x}'_{i_\ell}\gamma_0) \times \exp\{-y_{i_\ell} - s_{0,j_{i_\ell}-1}\} \exp(\xi_{0j_{i_\ell}} + A_{i_\ell}\beta_0 + \mathbf{x}'_{i_\ell}\gamma_0) \right]$$

where $M_{12} > 0$ is a constant and $\xi_0 = (\xi_{01}, \dots, \xi_{0J_0})'$. Now, we take a one-to-one transformation $\phi_0 = (\phi_{01}, \dots, \phi_{0,J_0+p_0})' = X_0^*(\xi_0', \beta_0, \gamma_0)'$. Using (A.5), we have

$$\begin{aligned}
 & \int L_1(\beta_0, \gamma_0, \lambda_0|D_{obs}) \left(\prod_{j=1}^{J_0} \frac{1}{\lambda_{0j}} \right) d\beta_0 d\gamma_0 d\lambda_0 \leq M_{13} \prod_{\ell=1}^{J_0+p_0} \int_{-\infty}^{\infty} \exp(\phi_{0\ell}) \exp\{ \\
 & \quad - (y_{i_\ell} - s_{0,j_{i_\ell}-1}) \exp(\phi_{0\ell}) d\phi_{0\ell} \\
 & \quad = M_{13} \prod_{\ell=1}^{J_0+p_0} (y_{i_\ell} - s_{0,j_{i_\ell}-1})^{-1} > \infty,
 \end{aligned} \tag{A.6}$$

where $M_{13} > 0$ is a constant, which completes the proof of (b).

For (c), we first rewrite (A.2) as follows:

$$L_2(\beta_1, \gamma_1, \lambda_1, \beta_{21}, \beta_{22}, \gamma_2, \lambda_2, \tau|D_{obs}) = L_{2a}(\beta_1, \gamma_1, \lambda_1, \tau|D_{obs}) L_{2b}(\beta_{21}, \beta_{22}, \gamma_2, \lambda_2|D_{obs}, \beta_1, \gamma_1, \lambda_1, \tau),$$

where

$$L_{2a}(\beta_1, \gamma_1, \lambda_1, \tau|D_{obs}) = \prod_{i \in \mathcal{N}_2} h_1(y_{E_i}|\lambda_1)\exp(A_i\beta_1 + \mathbf{x}'_i\gamma_1) \left\{ 1 + \tau H_1(y_{E_i}|\lambda_1)\exp(A_i\beta_1 + \mathbf{x}'_i\gamma_1) \right\}^{-(2+\frac{1}{\tau})}, \tag{A.7}$$

$$L_{2b}(\beta_{21}, \beta_{22}, \gamma_2, \lambda_2|D_{obs}, \beta_1, \gamma_1, \lambda_1, \tau)$$

$$= \prod_{i \in \mathcal{N}_2} \left[(1+\tau)h_2(y_{G_i}|\lambda_2)\exp(A_i\beta_{21} + V_i(1 - A_i)\beta_{22} - \mathbf{z}'_i\gamma_2) \right] \times \left\{ 1 + \tau b_i H_2(y_{G_i}|\lambda_2)\exp\{A_i\beta_{21} + V_i(1 - A_i)\beta_{22} + \mathbf{z}'_i\gamma_2\} \right\}^{-1} \tag{A.8}$$

and $b_i = \left\{ 1 + \tau H_1(y_{E_i}|\lambda_1)\exp(A_i\beta_1 + \mathbf{x}'_i\gamma_1) \right\}^{-1}$. Let j_i denote the index such that $s_{2,j_i-1} < y_{G_i} \leq s_{2,j_i}$. Then, we have

$$L_{2b}(\beta_{21}, \beta_{22}, \gamma_2, \lambda_2 | D_{obs}, \beta_1, \gamma_1, \lambda_1, \tau) \leq \prod_{i \in \mathcal{N}_2} [(1+\tau)\lambda_{2j_i} \exp(A_i\beta_{21} + V_i(1-A_i)\beta_{22} + z'_i\gamma_2) \times \{1 + \tau b_i \lambda_{2j_i} (y_{G_i} - s_{2j_i-1}) \exp\{A_i\beta_{21} + V_i(1-A_i)\beta_{22} + z'_i\gamma_2\}^{-1}]$$

Observing that $(1+rv)^{1+\frac{1}{r}} \geq 1+v$ for all $r > 0$ and $v > 0$, we obtain

$$(1+\tau)\lambda_{2j_i} \exp\{A_i\beta_{21} + V_i(1-A_i)\beta_{22} + z'_i\gamma_2\} \times \{1 + \tau b_i \lambda_{2j_i} (y_{G_i} - s_{2j_i-1}) \exp\{A_i\beta_{21} + V_i(1-A_i)\beta_{22} + z'_i\gamma_2\}^{-1}\}^{-\left(2+\frac{1}{\tau}\right)} \leq \frac{(1+\tau)\lambda_{2j_i} \exp\{A_i\beta_{21} + V_i(1-A_i)\beta_{22} + z'_i\gamma_2\}}{1 + (1+\tau)b_i \lambda_{2j_i} (y_{G_i} - s_{2j_i-1}) \exp\{A_i\beta_{21} + V_i(1-A_i)\beta_{22} + z'_i\gamma_2\}^{-1}} \tag{A.10}$$

Under condition (iii), X_2 is of full rank. Therefore, there exist $J_2 + p_2$ distinct $i_1, \dots, i_{J_2+p_2} \in \mathcal{N}_2$ such that the $(J_2 + p_2) \times (J_2 + p_2)$ matrix X_2^* , which has rows $(\delta_{i_\ell 21}, \dots, \delta_{i_\ell 2J_2}, A_{i_\ell}, z'_{i_\ell})$ for $\ell = 1, \dots, J_2 + p_2$, is of full rank. Let $\mathcal{N}_{21} = \{i_1, \dots, i_{J_2+p_2}\}$ and $\mathcal{N}_{22} = \mathcal{N}_2 - \mathcal{N}_{21}$. Using (A.10), we have

$$L_{2b}(\beta_{21}, \beta_{22}, \gamma_2, \lambda_2 | D_{obs}, \beta_1, \gamma_1, \lambda_1, \tau) \leq M_{21} \left[\prod_{i \in \mathcal{N}_{22}} \frac{1}{b_i} \right] \times \left(\prod_{i \in \mathcal{N}_{21}} [(1+\tau)\lambda_{2j_i} \exp\{A_i\beta_{21} + V_i(1-A_i)\beta_{22} - z'_i\gamma_2\} \times \{1 + \tau b_i \lambda_{2j_i} (y_{G_i} - s_{2j_i-1}) \exp\{A_i\beta_{21} + V_i(1-A_i)\beta_{22} + z'_i\gamma_2\}^{-1}]^{-\left(2+\frac{1}{\tau}\right)} \right) \tag{A.11}$$

where $M_{21} > 0$ is a constant. Similar to (A.5) and (A.6), using (A.11), we can show that

$$\begin{aligned} \int L_{2b}(\beta_{21}, \beta_{22}, \gamma_2, \lambda_2 | D_{obs}, \beta_1, \gamma_1, \lambda_1, \tau) \left[\prod_{j=1}^{J_2} \frac{1}{\lambda_{2j}} \right] d\beta_{21} d\beta_{22} d\gamma_2 d\lambda_2 &\leq M_{22} \left[\prod_{i \in \mathcal{N}_{22}} \frac{1}{b_i} \right] \times \left(\prod_{i \in \mathcal{N}_{22}} [(1+\tau) \int_{-\infty}^{\infty} \exp(\phi_{2i}) \{1 + \tau b_i (y_{G_i} - s_{2j_i-1}) \exp(\phi_{2i})\}^{-\left(2+\frac{1}{\tau}\right)} d\phi_{2i}] \right) \\ &= M_{22} \left[\prod_{i \in \mathcal{N}_{22}} \frac{1}{b_i} \right] \tag{A.12} \\ &\times \left[\prod_{i \in \mathcal{N}_{21}} \frac{(1+\tau)}{\left(1+\frac{1}{\tau}\right)\tau b_i (y_{G_i} - s_{2j_i-1})} \right] \\ &= M_{23} \prod_{i \in \mathcal{N}_{22}} (1+\tau H_1(y_{E_i} | \lambda_1) \exp(A_i\beta_1 + \mathbf{x}'_i\gamma_1)) \end{aligned}$$

where M_{22} and M_{23} are two positive constants. Now, using (A.2), (A.3), (A.7), and (A.12), we obtain

$$\begin{aligned} \int \left\{ L_2(\beta_1, \gamma_1, \lambda_1, \beta_{21}, \beta_{22}, \gamma_2, \lambda_2, \tau | D_{obs}) L_3(\beta_1, \gamma_1, \lambda_1, \tau | D_{obs}) \times \left[\prod_{k=1}^2 \prod_{j=1}^{J_K} \frac{1}{\lambda_{kj}} \right] \times \left[\frac{\exp\left(-\frac{b\tau}{\tau}\right)}{\tau^{a\tau+1}} \right] \right\} d\beta_{21} d\beta_{22} d\gamma_2 d\lambda_2 &\leq M_{23} \left[\prod_{i \in \mathcal{N}_{22}} \right. \\ &\times \left. \left[\prod_{j=1}^{J_1} \frac{1}{\lambda_{1j}} \right] \left[\frac{\exp\left(-\frac{b\tau}{\tau}\right)}{\tau^{a\tau+1}} \right] \right]. \tag{A.13} \end{aligned}$$

The right hand side of (A.13) is precisely the kernel of the posterior distribution of $(\beta_1, \gamma_1, \lambda_1, \tau)$ under the GORH model of Banerjee et al. (2007). Thus, under conditions (iii) and (iv), following the proof of Theorem 3.1 of Banerjee et al. (2007), we can show that the integration of the right hand side of (A.13) over $(\beta_1, \gamma_1, \lambda_1, \tau)$ is finite, which completes the proof of Theorem 1.

Proof of Theorem 2

For the posterior conditional distribution, the first derivative of the log-likelihood function is given by

$$\frac{\partial \log \pi(\tau^* | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs})}{\partial \tau^*} \propto \sum_{i=1}^n \left\{ \left(\frac{1}{\tau^*} + \frac{1}{\tau^* + 1} - \log[1 - H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + H_2(y_{G_i} | \lambda_2^*) \exp(A_i \beta_2 + \mathbf{x}'_i \gamma_2)] \right) \right. \\ \left. + \frac{(a_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_K} a_{kj} - 3)}{\tau^*} - [b_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_K} (b_{kj} \lambda_{kj}^*)] \right\}$$

. The second derivative is given by

$$\frac{\partial^2 \log \pi(\tau^* | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs})}{\partial (\tau^*)^2} \propto \sum_{i=1}^n \left\{ \left[-\frac{1}{(\tau^*)^2} - \frac{1}{(\tau^* + 1)^2} \right]^{1\{d_i=1, \nu_i=1\}} \right. \\ \left. + \left[-\frac{1}{(\tau^*)^2} \right]^{1\{d_i=1, \nu_i=0\}} \right\} - \frac{(a_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_K} a_{kj} - 3)}{(\tau^*)^2}$$

Assuming that $\sum_{i=1}^n [1\{\nu_i=1, d_i=1\} + 1\{\nu_i=0, d_i=1\}] + a_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_K} a_{kj} - 3 > 0$, we

will always have $\frac{\partial^2 \log \pi(\tau^* | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs})}{\partial (\tau^*)^2} < 0$. Therefore, the conditional density of τ^* given by (3.7) is log-concave. Letting

$\frac{\partial \log \pi(\tau^* | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs})}{\partial \tau^*} = 0$, then we have $B_1 + \frac{B_2}{\tau^*} + \frac{B_3}{1 + \tau^*} = 0$, where B_1 , B_2 and B_3 are defined in Theorem 2. After some algebra, the solution is given by

$$\hat{\tau}^* = \frac{-(B_1 + B_2 + B_3) - \sqrt{(B_1 + B_2 + B_3)^2 - 4B_1B_2}}{2B_1}$$

. The reasons are as follows: with $b_\tau > 0$, $b_{kj} > 0$ and $\lambda_{kj}^* > 0$, it is obvious that $B_1 < 0$; and with the previous assumption that

$\sum_{i=1}^n [1\{\nu_i=1, d_i=1\} + 1\{\nu_i=0, d_i=1\}] + a_\tau + \sum_{k=1}^2 \sum_{j=1}^{J_K} a_{kj} - 3 > 0$, then $B_2 > 0$, so we have $(B_1 + B_2 + B_3)^2 - 4B_1B_2 > 0$. Therefore, the equation has two roots. Since $\tau^* > 0$, then we only keep the positive solution $\hat{\tau}^*$ since the other root is negative. Since we just showed that the conditional density of τ^* given by (3.7) is log-concave, it follows that the mode of (3.7) is analytically available and given by $\hat{\tau}^*$.

Appendix B: Computational Development

B.1. Derivation of the Potential Survival Function

After some algebra, we obtain

$$\begin{aligned}
 S_a(t|\theta) &= \int_{\mathbf{x}} P(T_D > t | A \\
 &= a, \mathbf{x}, E \\
 &= 0, \beta_0, \gamma_0, \lambda_0) P(E \\
 &= 0 | A \\
 &= a, \mathbf{x}, \alpha) f_X(\mathbf{x} | A \\
 &= a) d\mathbf{x} \\
 &+ \int_{\mathbf{x}, \mathbf{z}, \omega} \{ P(T_E > t | A = a, \mathbf{x}, E = 1, \beta_1, \gamma_1, \lambda_1, \omega) + \int_0^t P(T_G > t - u | A = a, \mathbf{z}, V = 0, E = 1, \beta_2, \gamma_2, \lambda_2, \omega) \times f_1(u | A = a, \mathbf{x}, \omega, \beta_1, \gamma_1) \\
 &\times f_Z(\mathbf{z} | A \\
 &= a, \mathbf{x}, E \\
 &= 1) P(E \\
 &= 1 | A \\
 &= a, \mathbf{x}, \alpha) f_X(\mathbf{x} | A = a) dz d\mathbf{x}.
 \end{aligned}$$

When $J_0 = J_1 = J_2 = 1$, we obtain $P(T_D > t | A = a, \mathbf{x}, E = 0, \beta_0, \gamma_0, \lambda_0) = \exp\{-t\lambda_0 \exp(A\beta_0 + \mathbf{x}'\gamma_0)\}$, $\int_{\omega} P(T_E > t | A = a, \mathbf{x}, E = 1, \beta_1, \gamma_1, \lambda_1, \omega) f(\omega | \tau) d\omega = \int_{\omega} \exp[-\omega t \lambda_1 \exp(A\beta_1 + \mathbf{x}'\gamma_1)] f(\omega | \tau) d\omega = \left[\frac{1}{1 + \lambda_1 \exp(A\beta_1 + \mathbf{x}'\gamma_1) t \tau} \right]^{\frac{1}{\tau}}$, and

$$\begin{aligned}
 \int_{\omega} \{ \int_0^t P(T_G > t \\
 - u | A \\
 = a, \mathbf{z}, V \\
 = 0, E \\
 = 1, \beta_2, \gamma_2, \lambda_2, \omega) f_1(u | A = a, \mathbf{x}, \omega, \beta_1, \gamma_1, \lambda_1) du \} f(\omega | \tau) d\omega &= \frac{\lambda_1 \exp(A\beta_1 + \mathbf{x}'\gamma_1)}{\lambda_2 \exp\{A\beta_{21} + V(1 - A)\beta_{22} + \mathbf{z}'\gamma_2\} - \lambda_1 \exp\{A\beta_1 + \mathbf{x}'\gamma_1\}} \\
 - \left[\frac{1}{1 + \lambda_2 \exp\{A\beta_{21} + V(1 - A)\beta_{22} + \mathbf{z}'\gamma_2\} t \tau} \right]^{\frac{1}{\tau}} \}
 \end{aligned}$$

For the more general case, where the values of J_0, J_1 and J_2 are not specified, we have

$$\begin{aligned}
 &P(T_D > t | A = a, \mathbf{x}, E \\
 &= 0, \beta_0, \gamma_0, \lambda_0) \\
 &= \exp \left\{ - \sum_{j=1}^{J_0} [1_{\{s_{0,j-1} < t \leq s_{0j}\}} \{ \lambda_{0j}(t \right. \\
 &- s_{0,j-1}) + \sum_{g=1}^{j-1} \lambda_{0g}(s_{0g} \\
 &- s_{0,g-1})] \exp(A\beta_0 + \mathbf{x}'\gamma_0), \int_{\omega} P(T_E > t | A = a, \mathbf{x}, E = 1, \beta_1, \gamma_1, \lambda_1, \omega) f(\omega | \tau) d\omega = \int_{\omega} \exp \left\{ - \omega \sum_{j=1}^{J_1} [1_{\{s_{1,j-1} < t \leq s_{1j}\}} \{ \lambda_{1j}(t \right. \\
 &- s_{1,j-1}) + \sum_{g=1}^{j-1} \lambda_{1g}(s_{1g} \\
 &- s_{1,g-1})] \exp(A\beta_1 \\
 &+ \mathbf{x}'\gamma_1) \} f(\omega | \tau) d\omega \\
 &= \left[\frac{1}{1 + \sum_{j=1}^{J_1} [1_{\{s_{1,j-1} < t \leq s_{1j}\}} \{ \lambda_{1j}(t - s_{1,j-1}) + \sum_{g=1}^{j-1} \lambda_{1g}(s_{1g} - s_{1,g-1}) \} \exp[A\beta_1 + \mathbf{x}'\gamma_1] \tau]} \right]^{\frac{1}{\tau}}.
 \end{aligned}$$

Before the next derivation, we need to align the partitions of the time axis for h_1 and h_2 . Let $0 < s_{3,1} < s_{3,2} < \dots < s_{3,J_3}$ be the ordered distinct values of s_{k,J_k} , where $k = 1, 2$. For a given time point t , there exists j_t such that $t \in (s_{3,j_t-1}, s_{3,j_t})$. In order to facilitate the computation, let $s_{3,J_3+1} = s_{3,J_3}$, $s_{3,J_3} = s_{3,J_3-1}, \dots, s_{3,j_t} = t$. Then the corresponding constant hazards for each interval are as follows: $\lambda_{3kj} = \lambda_{kl}$ if $s_{k,l-1} < s_{3,j-1} < s_{3,j} < s_{k,l}$ for $j \in (1, J_3 + 1)$ and $l \in (1, J_k)$, where $k = 1, 2$. Next let

$$\begin{aligned}
 Q &= \int_0^t P(T_G > t \\
 &- u | A \\
 &= a, \mathbf{z}, V \\
 &= 0, E \\
 &= 1, \beta_2, \gamma_2, \lambda_2, \omega) f_1(u | A \\
 &= a, \mathbf{x}, \omega, \beta_1, \gamma_1, \lambda_1) du \\
 &= \sum_{j=1}^{j_t} \left\{ \omega \exp(A\beta_1 \right. \\
 &+ \mathbf{x}'\gamma_1) \int_{s_{3,j-1}}^{s_{3,j}} [P(T_G > t - u | A = a, \mathbf{z}, V = 0, E = 1, \beta_2, \gamma_2, \lambda_2, \omega) \\
 &\times \lambda_{31j} \exp \left\{ - \left[\sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} - s_{3,l-1}) + \lambda_{31j}(u - s_{3,j-1}) \right] \omega \exp(A\beta_1 + \mathbf{x}'\gamma_1) \right\} du \left. \right\},
 \end{aligned}$$

and

$$\begin{aligned}
 Q_j &= \omega \exp(A\beta_1 + \mathbf{x}'\gamma_1) \int_{s_{3,j-1}}^{s_{3,j}} [P(T_G > t - u | A = a, \mathbf{z}, V = 0, E = 1, \beta_2, \gamma_2, \lambda_2, \omega) \\
 &\quad \times \lambda_{31j} \exp\{-[\sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} - s_{3,l-1}) + \lambda_{31j}(u - s_{3,j-1})] \omega \exp(A\beta_1 + \mathbf{x}'\gamma_1)\}] du.
 \end{aligned}$$

While $u \in (s_{3,j-1}, s_{3,j}]$, we have $(t - u) \in (t - s_{3,j}, t - s_{3,j-1}]$, then there exist j_l and j_u such that $(t - s_{3,j}) \in (s_{3,j_l-1}, s_{3,j_l}]$ and $(t - s_{3,j-1}) \in (s_{3,j_u-1}, s_{3,j_u}]$. Let $r = j_u - j_l$. Then the range of r is from 0 to J_3 . Therefore, when $r = 0$, we have

$$\begin{aligned}
 Q_j &= \frac{\lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)}{\lambda_{32j_u} \exp\{A\beta_{21} + V(1 - A)\beta_{22} + \mathbf{z}'\gamma_2\} - \lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)} \\
 &\quad \times \left\{ \exp\left[-\left\{\sum_{l=1}^{j_u-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) + \lambda_{32j_u}(t - s_{3,j} - s_{3,j_u-1})\right\}\omega\right] \right. \\
 &\quad \times \exp\{A\beta_{21} + V(1 - A)\beta_{22} + \mathbf{z}'\gamma_2\} \\
 &\quad \left. - \left\{\sum_{l=1}^j \lambda_{31l}(s_{3,l} - s_{3,l-1})\right\}\omega \exp(A\beta_1 + \mathbf{x}'\gamma_1) - \exp\left[-\left\{\sum_{l=1}^{j_u-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) + \lambda_{32j_u}(t - s_{3,j-1} - s_{3,j_u-1})\right\}\omega\right] \right. \\
 &\quad \times \exp\{A\beta_{21} + V(1 - A)\beta_{22} + \mathbf{z}'\gamma_2\} \\
 &\quad \left. - \sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} - s_{3,l-1})\omega \exp(A\beta_1 + \mathbf{x}'\gamma_1)\right\}.
 \end{aligned}$$

Then

$$\int_{\omega} Q_j f(\omega|\tau) d\omega = \frac{\lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)}{\lambda_{32j_u} \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} - \lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)} \times \left[1 + \tau \left(\left\{ \sum_{l=1}^{j_u-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) + \lambda_{32j_u}(t + s_{3,j} - s_{3,j_u-1}) \right\} \times \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} + \left\{ \sum_{l=1}^j \lambda_{31l}(s_{3,l} - s_{3,l-1}) \right\} \right) \right]$$

When $r = 1$, we have

$$Q_j = \frac{\lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)}{\lambda_{32j_u} \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} - \lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)} \times \left\{ \exp \left[- \left\{ \sum_{l=1}^{j_u-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) \right\} \omega \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} - \left\{ \sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} - s_{3,l-1}) + \lambda_{31j}(t - s_{3,j-1} - s_{3,j_u-1}) \right\} \omega \exp(A\beta_1 + \mathbf{x}'\gamma_1) \right] - \exp \left[- \left\{ \sum_{l=1}^{j_u-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) + \lambda_{32j_u}(t - s_{3,j_u-1} - s_{3,j-1}) \right\} \omega \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} - \left\{ \sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} - s_{3,l-1}) \right\} \omega \exp(A\beta_1 + \mathbf{x}'\gamma_1) \right] \right\}$$

$$\begin{aligned}
& + \sum_{j^*=1}^{r-1} \frac{\lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)}{\lambda_{32(j_u-j^*)} \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} - \lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)} \\
& \quad \times \left\{ \exp\left[\begin{aligned} & - \sum_{l=1}^{j_u-j^*-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) \omega \exp\{A\beta_{21} \\ & + V(1-A)\beta_{22} \\ & + \mathbf{z}'\gamma_2\}, \\ & - \sum_{l=1}^{j-1} \{\lambda_{31l}(s_{3,l} - s_{3,l-1}) + \lambda_{31j}(t - s_{3,j_u-(j^*+1)} - s_{3,j-1})\} \omega \exp(A\beta_1 \\ & + \mathbf{x}'\gamma_1) \right] - \exp\left[- \left\{ \sum_{l=1}^{j_u-j^*} \lambda_{32l}(s_{3,l} - s_{3,l-1}) \right\} \omega \exp\{A\beta_{21} \right. \\ & \left. + V(1-A)\beta_{22} \right. \\ & \left. + \mathbf{z}'\gamma_2\} \right. \\ & \left. - \sum_{l=1}^{j-1} \{\lambda_{31l}(s_{3,l} - s_{3,l-1}) + \lambda_{31j}(t - s_{3,j_u-j^*} - s_{3,j-1})\} \omega \exp(A\beta_1 + \mathbf{x}'\gamma_1) \right] \right\}
\end{aligned}$$

$$\begin{aligned}
& + \frac{\lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)}{\lambda_{32(j_u-r)} \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} - \lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)} \\
& \quad \times \{ \exp[\\
& \quad - \{ \sum_{l=1}^{j_u-r-1} \lambda_{32l}(s_{3,l} \\
& \quad - s_{3,l-1}) + \lambda_{32(j_u-r)}(t \\
& \quad - s_{3,j_u-r-1} \\
& \quad - s_{3,j}) \} \omega \\
& \quad \times \exp\{A\beta_{21} \\
& \quad + V(1 \\
& \quad - A)\beta_{22} \\
& \quad + \mathbf{z}'\gamma_2\} \\
& \quad - \sum_{l=1}^j \{ \lambda_{31l}(s_{3,l} \\
& \quad - s_{3,l-1}) \} \omega \exp(A\beta_1 + \mathbf{x}'\gamma_1)] - \exp[-\{ \sum_{l=1}^{j_u-r} \lambda_{32l}(s_{3,l} \\
& \quad - s_{3,l-1}) \} \omega \exp\{A\beta_{21} \\
& \quad + V(1 \\
& \quad - A)\beta_{22} \\
& \quad + \mathbf{z}'\gamma_2\} \\
& \quad - \{ \sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} \\
& \quad - s_{3,l-1}) + \lambda_{31j}(t \\
& \quad - s_{3,j_u-r} - s_{3,j-1}) \} \omega \exp(A\beta_1 + \mathbf{x}'\gamma_1)] \},
\end{aligned}$$

for which

$$\begin{aligned}
 & \int_{\omega} Q_j f(\omega | \tau) d\omega \\
 &= \frac{\lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)}{\lambda_{32j_u} \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} - \lambda_{31j} \exp(A\beta_1 + \mathbf{x}'\gamma_1)} \\
 & \times \left\{ \left[1 + \tau \left[\sum_{l=1}^{j_u-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) \right] \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} + \left\{ \sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} - s_{3,l-1}) + \lambda_{31j}(t - s_{3,j-1} - s_{3,j_u-1}) \right\} \epsilon \right. \right. \\
 & \quad \left. \left. - \left[1 + \tau \left[\sum_{l=1}^{j_u-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) + \lambda_{32j_u}(t - s_{3,j_u-1} - s_{3,j-1}) \right] \times \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} + \sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} - s_{3,l-1}) \right] \right\} \\
 & \times \left\{ \left[1 + \tau \left[\sum_{l=1}^{j_u-j^*-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} + \sum_{l=1}^{j-1} \left\{ \lambda_{31l}(s_{3,l} - s_{3,l-1}) + \lambda_{31j}(t - s_{3,j_u-(j^*+1)} - s_{3,j} \right. \right. \right. \right. \\
 & \quad \left. \left. \left. \times \left[1 + \tau \left[\sum_{l=1}^{j_u-r-1} \lambda_{32l}(s_{3,l} - s_{3,l-1}) + \lambda_{32(j_u-r)}(t - s_{3,j_u-r-1} - s_{3,j}) \right] \times \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} + \sum_{l=1}^j \left\{ \lambda_{31l}(s_{3,l} - \right. \right. \right. \right. \\
 & \quad \left. \left. \left. - \left[1 + \tau \left[\sum_{l=1}^{j_u-r} \lambda_{32l}(s_{3,l} - s_{3,l-1}) \right] \exp\{A\beta_{21} + V(1-A)\beta_{22} + \mathbf{z}'\gamma_2\} + \left\{ \sum_{l=1}^{j-1} \lambda_{31l}(s_{3,l} - s_{3,l-1}) + \lambda_{31j}(t - s_{3,j_u-r} - s_{3,j-1}) \right\} \right] \right\} \right. \\
 & \quad \left. \left. \left. \right. \right. \right.
 \end{aligned}$$

B.2. Sampling from the Conditional Posterior Distributions

In the posterior computation section, a series of conditional posterior distributions are listed. Now we will show how to sample from these distributions.

- (i) $[\lambda_0, \lambda_1^*, \lambda_2^* | \beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2, \mathbf{w}, \mathbf{E}^*, D_{obs}]$. It is easy to see that conditional on $(\beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2, \mathbf{w}, \mathbf{E}^*, D_{obs})$, $(\lambda_0, \lambda_1^*, \lambda_2^*)$ are independent. Therefore, the conditional distributions can be sampled separately. Let $\delta_{ikj} = 1$ if the i^{th} subject failed or was censored in the j^{th} interval for $j = 1, 2, \dots, J_k$ and 0 otherwise. It can be shown that (ia) $[\lambda_{0j} | \beta_0, \gamma_0, \mathbf{E}^*, D_{obs}] \sim \text{Gamma}(a_{0j}^\pi, b_{0j}^\pi)$, where

$$\begin{aligned}
 a_{0j}^\pi &= \sum_{i=1}^n [\delta_{i0j} 1\{d_i=0, \nu_i=1\}] + a_{0j}, \text{ and } b_{0j}^\pi = \sum_{s_{0,j-1} < y_i \leq s_{0j}} (y_i - S_{0,j-1}) \exp \\
 & (A_i \beta_0 + \mathbf{x}'_i \gamma_0) [1\{d_i=0, \nu_i=1\} + (1 - E_i^*) 1\{d_i=0, \nu_i=0\}] + \sum_{y_i > s_{0j}} (s_{0,j} - s_{0,j-1}) \\
 & \exp(A_i \beta_0 + \mathbf{x}'_i \gamma_0) [1\{d_i=0, \nu_i=1\} + (1 - E_i^*) 1\{d_i=0, \nu_i=0\}] + b_{0j}; \text{ (ib)}
 \end{aligned}$$

$$[\lambda_{1j}^* | \beta_1, \gamma_1, \mathbf{w}, \mathbf{E}^*, D_{obs}] \sim \text{Gamma}(a_{1j}^\pi, b_{1j}^\pi), \text{ where } a_{1j}^\pi = \sum_{i=1}^n \delta_{i1j} [1\{d_i=1, \nu_i=1\} + 1\{d_i=1, \nu_i=0\}] + a_{1j}$$

; and (ic) $[\lambda_{2j}^* | \beta_2, \gamma_2, \mathbf{w}, \mathbf{E}^*, D_{obs}] \sim \text{Gamma}(a_{2j}^\pi, b_{2j}^\pi)$, where

$$\begin{aligned}
 a_{2j}^\pi &= \sum_{i=1}^n [\delta_{i2j} 1\{d_i=1, \nu_i=1\}] + a_{2j} \text{ and } b_{2j}^\pi = \sum_{s_{2,j-1} < y_i < s_{2j}} \mathbf{w}_i (y_i - S_{2,j-1}) \\
 & \exp \\
 & \{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\} [1\{d_i=1, \nu_i=1\} + 1\{d_i=1, \nu_i=0\}] + \sum_{y_i > s_{2j}} \mathbf{w}_i (s_{2,j} - s_{2,j-1}) \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\}
 \end{aligned}$$

- (ii) $[[\beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2, \tau, \mathbf{w}, \mathbf{E}^* | \alpha, \lambda_0, \lambda_1^*, \lambda_2^*, D_{obs}]]$.

- (iii) $[[\beta_0, \gamma_0, \beta_1, \gamma_1, \beta_2, \gamma_2 | \alpha, \lambda_0, \lambda_1^*, \lambda_2^*, \tau, \mathbf{E}^*, D_{obs}]]$. From the joint posterior distribution, it is obvious that conditional on $[[\alpha, \lambda_0, \lambda_1^*, \lambda_2^*, \tau, \mathbf{E}^*, D_{obs}]]$, the

parameters $[\beta_0, \gamma_0], [\beta_1, \gamma_1], [\beta_2, \gamma_2]$ are independent. Therefore, we can sample the following conditional posterior distributions separately.

(iiaa) $[\beta_0, \gamma_0 | \lambda_0, \mathbf{E}^*, D_{obs}]$. This density is proportional to

$$\prod_{i=1}^n \left\{ \left[\exp(A_i \beta_0 + \mathbf{x}'_i \gamma_0) \exp\{-H_0(y_i | \lambda_0) \exp(A_i \beta_0 + \mathbf{x}'_i \gamma_0)\} \right]^{1\{d_i=0, \nu_i=1\}} \times \left[\exp\{-H_0(y_i | \lambda_0) \exp(A_i \beta_0 + \mathbf{x}'_i \gamma_0)\} \right]^{(1-E_i^*)1\{d_i=0, \nu_i=0\}} \exp\{- (\beta_0, \gamma_0)' \Sigma_0^{-1} (\beta_0, \gamma_0)\} \right\}.$$

It is easy to show that this conditional distribution is log-concave in each component of β_0 and γ_0 .

(iiab) $[\beta_1, \gamma_1 | \alpha, \lambda_0, \lambda_1^*, \lambda_2^*, \tau, \mathbf{E}^*, D_{obs}]$. This density is proportional to

$$\prod_{i=1}^n \left\{ \left([1 + H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + H_2(y_{G_i} | \lambda_2^*) \times \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\}]^{-\frac{1}{\tau}-2} \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) \right)^{1\{d_i=1, \nu_i=1\}} \times \left(\exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) [1 + H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + H_2(y_{G_i} | \lambda_2^*) \times \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\}]^{-\frac{1}{\tau}-1} \right)^{1\{d_i=1, \nu_i=0\}} \times \left\{ [1 + H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1)]^{-\frac{1}{\tau}} \right\}^{E_i^* 1\{d_i=0, \nu_i=0\}} \exp\{- (\beta_1, \gamma_1)' \Sigma_1^{-1} (\beta_1, \gamma_1)\} \right\}.$$

It can also be shown that this conditional distribution is log-concave in each component of β_1 and γ_1 .

(iiac) $[\beta_2, \gamma_2 | \alpha, \lambda_0, \lambda_1^*, \lambda_2^*, \tau, \mathbf{E}^*, D_{obs}]$. This density is proportional to

$$\prod_{i=1}^n \left\{ \left([1 + H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + H_2(y_{G_i} | \lambda_2^*) \times \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\}]^{-\frac{1}{\tau}-2} \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\} \right)^{1\{d_i=1, \nu_i=1\}} \times \left([1 + H_1(y_{E_i} | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + H_2(y_{G_i} | \lambda_2^*) \times \exp\{A_i \beta_{21} + V_i(1 - A_i) \beta_{22} + \mathbf{z}'_i \gamma_2\}]^{-\frac{1}{\tau}-1} \right)^{1\{d_i=1, \nu_i=0\}} \exp\{- (\beta_2, \gamma_2)' \Sigma_2^{-1} (\beta_2, \gamma_2)\} \right\}.$$

Similarly, this conditional distribution can be shown to be log-concave in each component of β_2 and γ_2 .

(iib) $[\mathbf{E}^* | \alpha, \beta_0, \gamma_0, \beta_1, \gamma_1, \lambda_0, \lambda_1^*, \tau, D_{obs}]$. This conditional posterior distribution is given by

$$[\mathbf{E}^* | \alpha, \beta_0, \gamma_0, \beta_1, \gamma_1, \lambda_0, \lambda_1^*, \tau, D_{obs}] \propto \prod_{i=1}^n \left\{ \left([H_1(y_i | \lambda_1^*) \exp(A_i \beta_1 + \mathbf{x}'_i \gamma_1) + 1]^{-\frac{1}{\tau}} [1 + \exp\{- (\alpha_0 + A_i \alpha_1 + \mathbf{x}'_i \alpha_2)\}]^{-1} \right)^{E_i^*} \times \left(\exp[- (\alpha_0 + A_i \alpha_1 + \mathbf{x}'_i \alpha_2)] \right)^{1 - E_i^*} \right\}.$$

(iic) $[\tau | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \mathbf{E}^*, D_{obs}]$. This one is shown in Theorem 2.

(iid) $[\omega | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, \tau, \mathbf{E}^*, D_{obs}]$. To sample from this conditional distribution, we can sample $[\omega | \beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2 \text{ and } \lambda_2^*, 1/\tau^*, \mathbf{E}^*, D_{obs}]$ instead. It

can be shown that $[\omega|\beta_1, \gamma_1, \lambda_1^*, \beta_2, \gamma_2, \lambda_2^*, 1/\tau^*, \mathbf{E}^*, D_{obs}] \sim \text{Gamma}(a_w, b_w)$,

$$\begin{aligned}
 a_w &= (\tau^* + 2)1\{d_i \\
 &= 1, \nu_i \\
 &= 1\} + (\tau^* + 1)1\{d_i \\
 &= 1, \nu_i \\
 &= 0\} + \tau^* E_i^* 1\{d_i \\
 &= 0, \nu_i \\
 &= 0\} \text{ and } b_w \\
 &= [H_1(y_{E_i}|\lambda_1^*)\exp(A_i\beta_1 \\
 &+ \mathbf{x}'_i\gamma_1) + H_2(y_{G_i}|\lambda_2^*)\exp(A_i\beta_{21} \\
 &+ V_i(1 \\
 &\quad - A_i)\beta_{22} \\
 &\quad + \mathbf{z}'_i\gamma_2) + 1][1\{d_i=1, \nu_i=1\} + 1\{d_i=1, \nu_i=0\}] \\
 &+ [H_1(y_i|\lambda_1^*)\exp(A_i\beta_1 \\
 &\quad + \mathbf{x}'_i\gamma_1) \\
 &\quad + 1]E_i^* 1\{d_i=0, \nu_i=0\}
 \end{aligned}$$

(iii) $[\alpha|\mathbf{E}^*, D_{obs}]$. This conditional posterior distribution is given by

$$\begin{aligned}
 [\alpha|\mathbf{E}^*, D_{obs}] &\propto \prod_{i=1}^n \left\{ \left([1 + \exp(\alpha_0 + A_i\alpha_1 + \mathbf{x}'_i\alpha_2)]^{-1} \right)^{1\{d_i=0, \nu_i=1\} + (1-E_i^*)1\{d_i=0, \nu_i=0\}} \right. \\
 &\quad \left. \times \left([1 + \exp\{-(\alpha_0 + A_i\alpha_1 + \mathbf{x}'_i\alpha_2)\}]^{-1} \right)^{1\{d_i=1, \nu_i=1\} + 1\{d_i=1, \nu_i=0\} + E_i^* 1\{d_i=0, \nu_i=0\}} \right\} \times \exp\{-(\alpha_0, \alpha_1, \alpha_2)' \Sigma_a^{-1}(\alpha_0, \alpha_1, \alpha_2)\}
 \end{aligned}$$

It is easy to show that this density is log-concave in each component of α . Therefore, we apply the adaptive rejection algorithm of Gilks and Wild (1992) to draw α .

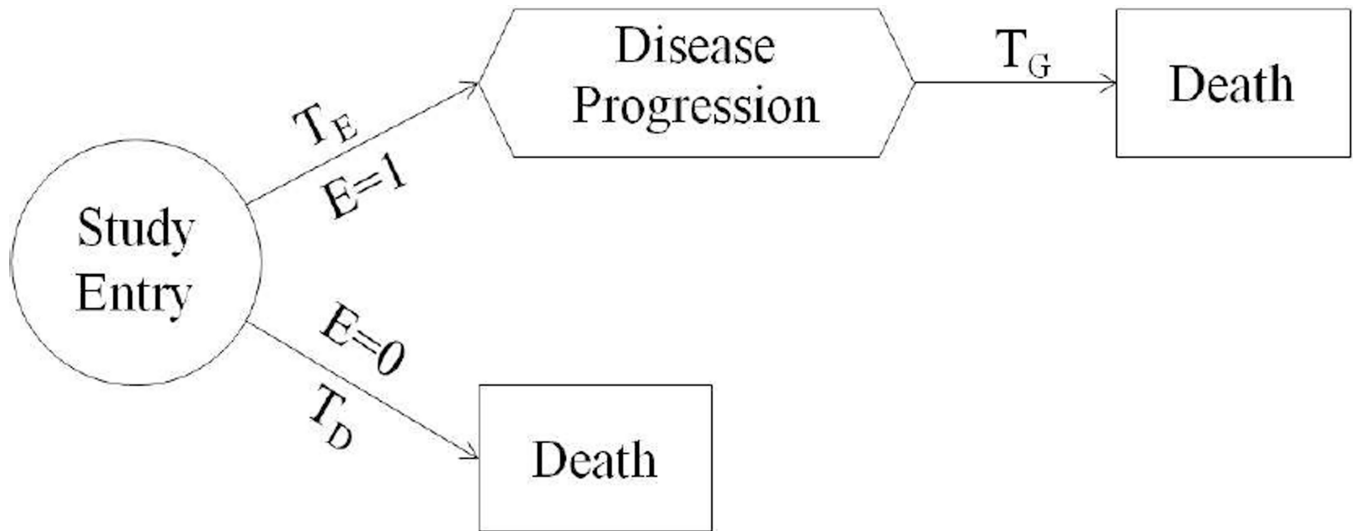


Fig. 1. A graphical illustration of key random variables in the semi-competing risks model.

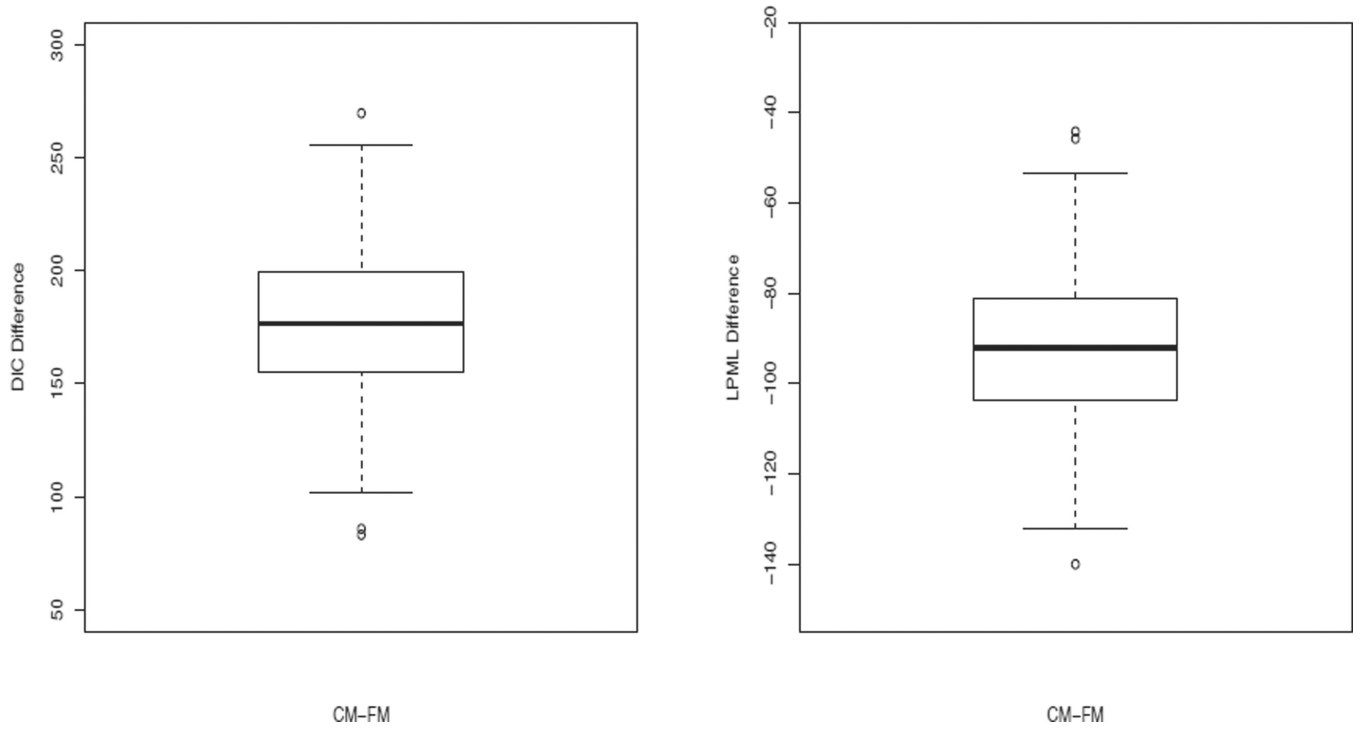
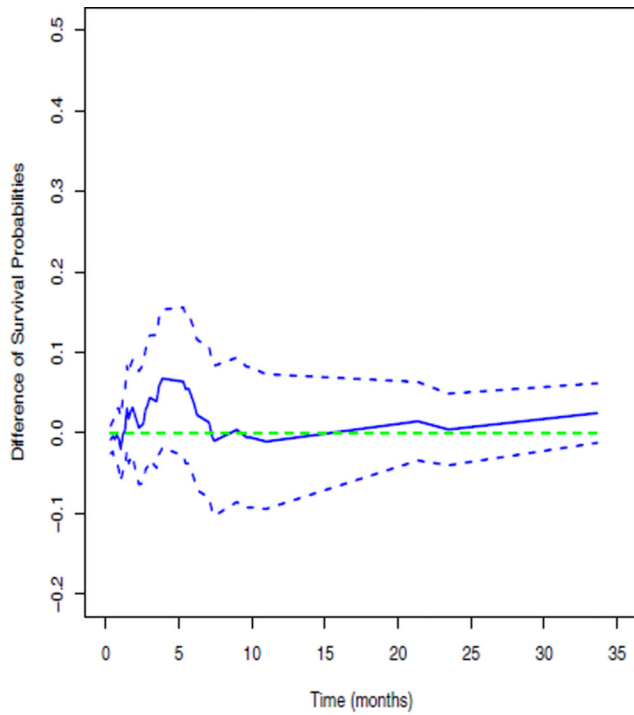
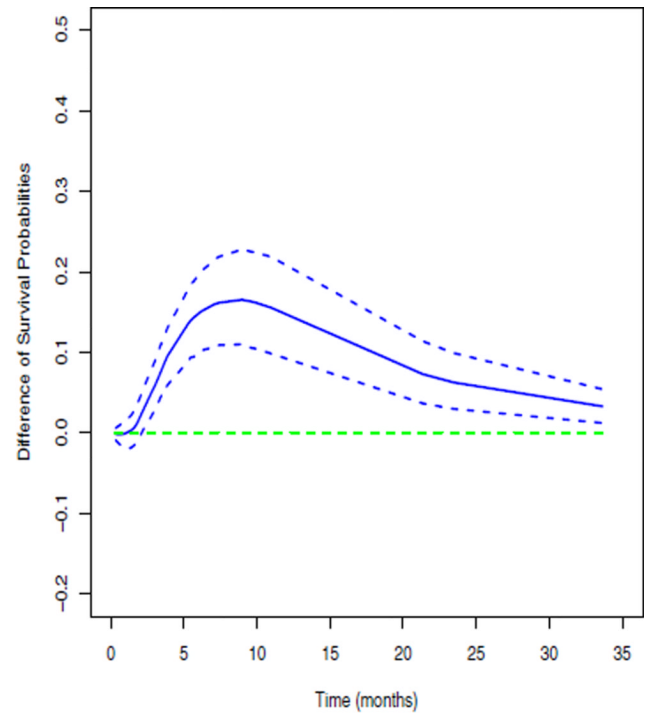


Fig. 2. Box Plots of the DIC and LPML Differences between CM and FM.



(a) ITT



(b) Bayesian

Fig. 3. The estimated differences with 95% intervals of the survival curves between the treatment and control arms.

Table 1

Posterior estimates under FM in the simulation study

Parameter	True	$n = 500$				$n = 1,000$					
		EST	SD	SE	RMSE	CP%	EST	SD	SE	RMSE	CP%
<u>T_D model</u>											
β_0	-1.0	-1.00	0.25	0.24	0.24	0.24	0.18	0.18	0.18	0.18	0.95
γ_{01}	1.0	0.99	0.20	0.19	0.19	0.19	0.14	0.14	0.14	0.14	0.96
γ_{02}	0.2	0.21	0.23	0.22	0.22	0.22	0.16	0.16	0.15	0.15	0.96
<u>T_E model</u>											
β_1	-0.5	-0.53	0.22	0.22	0.22	0.22	0.16	0.16	0.15	0.15	0.95
γ_{11}	1.0	1.02	0.19	0.20	0.20	0.20	0.13	0.14	0.14	0.14	0.95
γ_{12}	0.0	0.01	0.21	0.20	0.20	0.20	0.15	0.15	0.15	0.15	0.95
<u>T_G model</u>											
β_{21}	-0.3	-0.31	0.24	0.23	0.23	0.23	0.17	0.17	0.17	0.17	0.95
β_{22}	-0.5	-0.49	0.23	0.23	0.23	0.23	0.16	0.16	0.16	0.16	0.93
τ	1.0	1.05	0.15	0.14	0.14	0.14	0.10	0.10	0.10	0.10	0.95
γ_{22}	-0.5	-0.49	0.18	0.19	0.19	0.19	0.13	0.13	0.13	0.13	0.94
γ_{23}	0.5	0.51	0.21	0.21	0.21	0.21	0.15	0.16	0.16	0.16	0.92
γ_{24}	-0.4	-0.40	0.32	0.32	0.32	0.32	0.22	0.22	0.23	0.23	0.95
<u>E model</u>											
α_0	1.6	1.63	0.24	0.24	0.24	0.24	0.17	0.16	0.16	0.16	0.97
α_1	-1.8	-1.81	0.25	0.26	0.26	0.26	0.18	0.18	0.18	0.18	0.95
α_{21}	1.0	1.01	0.22	0.24	0.24	0.24	0.15	0.16	0.16	0.16	0.95
α_{22}	0.1	0.11	0.24	0.25	0.25	0.25	0.17	0.17	0.17	0.17	0.94
<u>Estimated Survival function of control arm</u>											
$S_0(\tau^{*/2})$	0.44	0.44	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.96
$S_0(\tau^*)$	0.26	0.28	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.95
<u>Estimated Survival function of treatment arm</u>											
$S_1(\tau^{*/2})$	0.57	0.57	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.93

Parameter	True	$n = 500$						$n = 1,000$					
		EST	SD	SE	RMSE	CP%	EST	SD	SE	RMSE	CP%		
$S_{1(\tau^*)}$	0.36	0.37	0.03	0.03	0.03	0.93	0.37	0.02	0.02	0.02	0.91		

Table 2

DIC and LPML Values for the Panitumumab Data

Parameter	FM			CM				
	J_0	J_1	J_2	DIC	LPML	p_D		
1	30	5	3475.27	67.13	-1741.32	3482.64	67.32	-1746.76
3	30	5	3479.02	69.20	-1742.90	3486.14	69.48	-1748.76
5	30	5	3480.09	71.24	-1743.92	3486.92	71.44	-1748.94
1	25	5	3493.38	61.83	-1749.55	3500.00	62.20	-1755.11
1	35	5	3477.89	72.31	-1743.16	3484.61	72.54	-1748.62
1	30	3	3484.99	65.15	-1746.02	3488.20	65.37	-1749.05
1	30	10	3481.34	72.23	-1744.74	3490.95	72.46	-1751.73

Table 3

Posterior Estimates for the Panitumumab Data under FM with $(J_0, J_1, J_2) = (1, 30, 5)$

Parameter	E Model		T_D Model		
	EST	SD	EST	SD	
Intercept	1.512	0.989			
Treatment	-1.115	0.324	-0.481	0.337	
Age	-0.010	0.014	0.023	0.015	
bECOG01	1.987	0.346	-0.617	0.296	
Rectal	0.303	0.337	-0.060	0.319	
Male	-0.317	0.330	-0.326	0.306	
CentEastEU	0.001	0.628	-0.308	0.638	
WesternEU	0.334	0.422	0.231	0.395	
		T_E Model		T_G Model	
Treatment	-1.443	0.150	-0.975	0.265	-1.479, -0.441
Age	-0.019	0.006	-1.475	0.256	-1.968, -0.962
bECOG01	-0.869	0.220	-0.007	0.007	-0.020, 0.006
Rectal	-0.112	0.133	-0.254	0.347	-0.942, 0.413
Male	-0.129	0.132	-0.088	0.192	-0.460, 0.296
CentEastEU	0.190	0.274	-0.445	0.242	-0.908, 0.034
WesternEU	-0.074	0.188	-1.186	0.177	-1.522, -0.829
			0.348	0.141	(0.082, 0.631)
			0.322	0.083	(0.163, 0.490)