Biometrics 000, 1–25 DOI: 000
000 0000

A Bayesian Multi-Risks Survival (MRS) Model in the Presence of Double Censorings

Mário de Castro[1], Ming-Hui Chen[2], Yuanye Zhang[3], and Anthony V. D'Amico[4]

[1]Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos, SP, Brazil

[2]Department of Statistics, University of Connecticut, Storrs, CT, U.S.A.

[3]Agios Pharmaceuticals, Cambridge, MA, U.S.A.

[4]Department of Medical Oncology and Radiation Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Boston, MA, U.S.A.

email: mcastro@icmc.usp.br

Summary: Semi-competing risks data include the time to a nonterminating event and the time to a terminating event, while competing risks data include the time to more than one terminating event. Our work is motivated by a prostate cancer study, which has one nonterminating event and two terminating events with both semi-competing risks and competing risks present as well as two censoring times. In this paper, we propose a new multi-risks survival (MRS) model for this type of data. In addition, the proposed MRS model can accommodate non-informative right-censoring times for nonterminating and terminating events. Properties of the proposed MRS model are examined in detail. Theoretical and empirical results show that the estimates of the cumulative incidence function (CIF) for a nonterminating event may be biased if the information on a terminating event is ignored. A Markov chain Monte Carlo sampling algorithm is also developed. Our methodology is further assessed using simulations and also an analysis of the real data from a prostate cancer study. As a result, a prostate-specific antigen (PSA) velocity greater than 2.0 ng/ml per year and higher biopsy Gleason scores are positively associated with a shorter time to death due to prostate cancer.

Key words: Bayesian model selection criteria; MCMC methods; multistate model; prostate cancer data; semi-competing risks; time-to-event data.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/biom.13228

1. Introduction

In many medical studies, both terminating events and nonterminating events are often observed in the data. This type of situation is termed as semi-competing risks, in which an event time can be censored by another event time but not conversely. A terminating event potentially censors a nonterminating event, whereas a nonterminating event does not preclude subsequent occurrence of the terminating event. The analysis of this kind of data poses some challenges. Both semi-competing risks and competing risks are present. Therefore, we develop a new multi-risks survival (MRS) model to fit data with one nonterminating event and two terminating events (death due to the disease and death due to other causes). Moreover, (i) disease progression (a nonterminating event) and death can be censored at the same time and (ii) disease progression can be censored at a certain time and death can be censored later on (non-informative right censoring). Due to double censorings (non-informative right-censoring times for terminating and non-terminating events), the proposed MRS model is suitable to fit data with this structure.

Our proposed MRS model is related to the model in Conlon *et al.* (2014) (see also Conlon *et al.*, 2015). Under their model, "cured" and "not cured" (with respect to disease recurrence) are considered as latent states. We emphasize that in the models in Conlon *et al.* (2014) and Lee *et al.* (2015), the cause of death is not considered. However, in several studies, the probability of death due to the disease is of scientific interest (see, e.g., D'Amico *et al.*, 2005). More differences between our model and the models in Conlon *et al.* (2014) and Lee *et al.* (2015) are highlighted in Section 7.

The contributions in this paper are both theoretical and empirical. We propose a novel hybrid model comprising a cure rate model, a cause-specific hazards model, and a frailty variable in an integrated way so that the proposed MRS model can accommodate different features. The cure rate model takes care of those subjects who are non-susceptible to disease

progression. The cause-specific model plays a part because we have two terminating events. The frailty variable is aimed to capture dependence among the three observable times.

Our formulation can deal with competing and semi-competing risks in a setup involving informative and non-informative censoring times. We call attention to the fact that the estimates of the cumulative incidence function (CIF) for a nonterminating event (e.g., disease progression) may be biased if the information on a terminating event (e.g., death) is ignored. We give a proof of this fact. This issue is also examined in our simulation study.

Our paper unfolds as follows. The data set is described in Section 2. Section 3 is dedicated to the model formulation. Section 4 covers Bayesian inference, including the prior and posterior distributions, a Gibbs sampling scheme for drawing samples from the posterior distribution, computation of CIF's, and model comparison. The performance of the MRS model is assessed through a simulation study carried out in Section 5. In Section 6, the methodology is applied for analyzing the prostate cancer data set presented in Section 2. Discussions and concluding remarks are given in Section 7. Additional contents for this article are available online. Specifically, Tables S1–S5 and Figures S1–S6 referred to the text are in Web Appendix E.

2. Motivating Data

Our work is motivated by prostate cancer (PC) data collected in a study analyzed in D'Amico *et al.* (2005). The study cohort is formed by 358 men treated definitively with radiation therapy (RT) for localized PC between 1989 and 2002. Follow-up started on the last day of RT and went until March 1, 2005, or the date of death, whichever happened first. Patients had a serum prostate-specific antigen (PSA) measurement at a median of every 6 months and yearly digital rectal examinations. After PSA recurrence, defined as per the American Society for Therapeutic Radiology and Oncology being a PSA level measured using a blood test that is 2 points above the lowest observed PSA value following the completion of treatment for prostate cancer, PSA levels were gauged at a median of 4 months.

The PSA recurrence time is the earliest follow-up time for which the PSA measurement is 2 points above the lowest observed PSA value following the completion of treatment for prostate cancer. Therefore, PSA measurement is as exact as the sampling permits and it is not an interval-censored time. PSA measurements following prostate cancer therapy are recommended to be obtained every 3 months for 2 years and then every 6 months for an additional 3 years and annually thereafter as noted in D'Amico *et al.* (2008a). Therefore, PSA recurrence is always modeled in the medical literature as a continuous time to the event as noted in D'Amico *et al.* (2005) and D'Amico *et al.* (2008b).

In this study, PSA recurrence is a nonterminating event, whereas death is a terminating event. The primary endpoints are disease progression (represented by the first PSA recurrence), death due to PC, and death due to other causes. In our data, PSA recurrence can be censored by death but not vice versa. A graphical representation of the data set is displayed in Figure 1, where n denotes the number of subjects. The cases in Figure 1 are described in Section 3.1.

[Figure 1 about here.]

Existing methods can be applied for analyzing the data in Figure 1 after ignoring or collapsing parts of the data, which may represent a simplification of a complex process. Some of the existing approaches are described in the sequel. When the cause of death is not available, all causes can be combined in a sole terminating event and the data can be analyzed either ignoring the time to PSA recurrence or not. Either way is unsatisfactory for our motivating study because the estimation of the probability of death due to PC is of scientific interest (see, e.g., D'Amico *et al.*, 2005). The time to PSA recurrence can be modeled alone ignoring the time to death in different ways using standard parametric and semiparametric models as well as cure rate models. In Sections 4.3 and 5, we provide evidence that a cure rate model (Chen *et al.*, 1999) may lead to misleading results.

The structure of the data in Figure 1 might be collapsed so that only terminating events are analyzed under a competing risks setup with the time to PSA recurrence entering as a covariate or not. However, the dependence structure linking nonterminating and terminating events can be more intricate. If we do not separate the causes of death and include PSA recurrence as a state, the data can be modeled via a semi-competing risks approach, as in Zeng *et al.* (2012) and Zhang *et al.* (2014). It is assumed that PSA recurrence would never occur for the 38 patients with time to PSA recurrence censored in Figure 1. This assumption is unrealistic for our data because PSA recurrence certainly occurs among those patients in Case C7.

Multistate modeling (see, e.g., Putter *et al.*, 2007; Beyersmann *et al.*, 2012; de Castro *et al.*, 2015) is an alternative for survival data with semi-competing risks. In principle, a four-state model comprising (1) study entry, (2) PSA recurrence, (3) death due to PC, and (4) death due to other causes with possible transitions $1 \to 2$, $1 \to 4$, $2 \to 3$, and $2 \to 4$ might be proposed. Transition $1 \to 4$ implies that (i) a patient will never experience PSA recurrence ("cured") or (ii) a patient is not cured from PC but dies from other causes before disease progression (semi-competing risks). In case (i), the patient will die only due to other causes, whereas in case (ii) the patient would have PSA recurrence if he does not die from other causes. Due to two censoring times, both cure and semi-competing risks are not directly observable in our data so that transition $1 \to 4$ is not directly observed in our data. In order to fit a multistate model, patients from Cases C2 ($n = 37$) and C7 ($n = 1$) in Figure 1 are assigned to transitions $1 \to 4$ and $2 \to 4$, respectively. In other words, PSA recurrence censoring is ignored and this may lead to biased estimates of the CIF's for disease progression and death due to other causes, as can be seen from the results in Section 5.

Table S1 presents a summary of the data set. The pretreatment covariates we consider include (i) age in years, (ii) PSA level in ng/ml, (iii) GS7, with value 1 if biopsy Gleason score

is equal to 7 and 0 otherwise, (iv) GS8H, with value 1 if biopsy Gleason score $\in \{8, 9, 10\}$ and 0 otherwise, (v) Vel2, pretreatment PSA velocity, taking value 1 if the change in PSA level during the year prior to diagnosis is greater than 2.0 ng/ml per year and 0 otherwise, and (vi) T2, the 1992 American Joint Commission on Cancer clinical tumor category (Cstage), with value 1 if the clinical tumor category is T2 or higher and 0 if the clinical tumor category is T1. The categorization of the Gleason score is already established in the clinical literature (see, e.g., D'Amico *et al.*, 2005). For all 37 patients with disease progression censored and death due to other causes (Case C2 in Figure 1), GS8H is equal to 0. The median age is 72.1 years with interquartile range (IQR) = (68.2, 75.7) years, while the median PSA level is 8.0 ng/ml with IQR = (5.7, 11.8) ng/ml. For the last follow-up time, the median is 4.5 years with IQR = (2.5, 6.8) years.

Death can occur due to PC or to other causes because men of PC bearing age often have other illness (e.g., heart disease) that they can die from after PSA recurrence and before some more indolent forms of PC have time to progress and take their life. Since in the PC data men were treated with radiotherapy, death due to PC only can happen after disease progression. Depending upon an individual's baseline comorbidity, a man with PC who had been treated and has not yet been observed to sustain PSA recurrence, can have an intercurrent event related to a comorbid illness. Such events include myocardial infarction, stroke or Alzheimer's disease, or a second cancer, which now takes precedence and can lead to less regular PSA monitoring. Hence, the PSA level may not be routinely checked unless he is on a clinical trial. In this case, PSA recurrence is censored at the last PSA check time and this censoring time is different from the death censoring time. We emphasize that the data set shown in Figure 1 is very rich.

3. Proposed Methods

3.1 *MRS Model*

First, we lay down some elements underlying the MRS model. Let random variables T_P, T_G, and T_{D2} represent the time to disease progression, the gap time between disease progression and death due to the disease, and the time to death due to other causes, respectively. Then, the time to death due to the disease T_{D1} is given by $T_{D1} = T_P + T_G$. In our motivating data, patients enrolled in the study were treated and they can die from PC (the disease) only after PSA recurrence (the progression of the disease). Thus, a patient will never die from the disease if there is no disease progression.

The diagram in Figure 2 illustrates the relationships among the variables T_P, T_G, and T_{D2}. If $T_P < T_{D2}$, then T_P is observable, whereas if $T_P \geqslant T_{D2}$, T_P is unobservable and death due to other causes potentially censors disease progression, but not vice versa, so that these two events act as semi-competing risks. If $T_P < T_{D2}$, either death due to the disease or death due to other causes may be observed depending on T_G. If T_G is short, death due to the disease may happen and if T_G is long, death due to other causes may happen. Either way, these two terminating events are competing each other. Therefore, $T_{D1} = T_P + T_G$ may censor T_{D2} (informative censoring) but not T_P (recall death due to the disease only can happen after disease progression), while T_{D2} may censor T_{D1} (informative censoring) and T_{D2} may or may not censor T_P. Two paths in Figure 2 head toward the same endpoint due to semi-competing risks. Moreover, disease progression may be censored and death may be censored at the same time or later on (non-informative censoring). Three paths in Figure 2 lead to death due to other causes (T_{D2}) and a multistate model does not cope with these paths.

[Figure 2 about here.]

The MRS model encompasses the variables T_P, T_G, and T_{D2}, which are potentially dependent. This dependence is captured through a subject-specific shared frailty variable w (see

also Lee *et al.*, 2015). For the frailty variable, we assume a gamma ($\mathcal{G}$) distribution, that is, $w \sim \mathcal{G}(1/\tau, 1/\tau)$ with probability density function (PDF) $f(w|\tau) = w^{1/\tau-1}\exp(-w/\tau)/\{\tau^{1/\tau}\Gamma(1/\tau)\}$, where $\Gamma(\cdot)$ denotes the gamma function, so that w has unit mean and variance equal to τ.

We begin with the model for disease progression. We adopt the promotion time cure model in Chen *et al.* (1999). The number of latent causes competing for the progression of the disease is denoted by N, where $N = 0$ for the "cured" subpopulation (no disease progression) and $N \geqslant 1$ for the "non-cured" subpopulation, noticing that $N = 0$ is not directly observable. In the MRS model, the conditional expected value of the number of latent causes N is $\theta(\boldsymbol{x}, w) = w\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)$, where $\kappa > 0$, $\boldsymbol{x} = (x_1, \ldots, x_p)'$ is a vector of covariates not including the intercept, and $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0p})'$ is a vector of coefficients, that is, $\log(\kappa)$ is the intercept. By taking out the intercept from the linear predictor in the expression of $\theta(\boldsymbol{x}, w)$, we gain in simplicity in the computations in Section 4.2.

As in Chen *et al.* (1999), N follows a Poisson distribution with expected value θ, the latent times to disease progression Z_j, $j = 1, 2, \ldots$, are i.i.d. with cumulative distribution function (CDF) $F_0(t)$ and PDF $f_0(t)$, and N and the Z_j's are independent. Notice that $T_P = \min(Z_1, \ldots, Z_N)$, if $N \geqslant 1$. Conditional on w, the improper survival function for the population at time y_P is given by $S_P(y_P|\boldsymbol{x}, w) = \exp\{-wF_0(y_P)\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\}$, and then the improper hazard function is

$$h_P(y_P|\boldsymbol{x}, w) = wf_0(y_P)\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0), \tag{1}$$

where $F_0(y_P) = 1 - S_0(y_P) = 1 - \exp\{-H_0(y_P)\}$, with $H_0(y_P) = \int_0^{y_P} h_0(u)du$ and $h_0(\cdot)$ denotes the baseline hazard function. In Conlon *et al.* (2014), the mixture cure model (Berkson and Gage, 1952) is adopted for the time to recurrence of the disease. Informative priors are assigned in order to alleviate estimation problems with some parameters. In our model, the prior distribution is less informative. Furthermore, the promotion time cure model is a much

more convenient way to accommodate cure and frailty as a multiplicative term. Drawbacks of the mixture cure model and advantages of the promotion time cure model are discussed in Chen *et al.* (1999) and Tsodikov *et al.* (2003), among others.

For the competing risks (death due to the disease and death due to other causes), we adopt the cause-specific hazards model (Prentice *et al.*, 1978; Gaynor *et al.*, 1993). Conditional on w, the hazard functions of T_G and T_{D2}, denoted by $h_1(\cdot)$ and $h_2(\cdot)$, are assumed to have proportional hazards structures

$$h_1(y_G|\boldsymbol{x}, T_P, w) = wh_{10}(y_G)\exp(\boldsymbol{z}'\boldsymbol{\beta}_1) \quad \text{and} \tag{2}$$

$$h_2(y_{D2}|\boldsymbol{x}, w) = wh_{20}(y_{D2})\exp(\boldsymbol{x}'\boldsymbol{\beta}_2), \tag{3}$$

where $h_{10}(\cdot)$ and $h_{20}(\cdot)$ are baseline hazard functions and $\boldsymbol{z} = (\boldsymbol{x}', T_P)'$. For notational simplicity, the vector of covariates $\boldsymbol{x}$ in (1)–(3) is the same. Since the hazard function $h_1(\cdot)$ in (2) depends on T_P, additional dependence (besides the frailty variable) between T_G and T_P is imposed, as in Zeng *et al.* (2012). In this way, the MRS model accommodates different dependence structures using only a univariate frailty variable. Given w, we assume $(T_P, T_G)'$ is independent of T_{D2} and N is independent of T_{D2}. Putting together the relationships among $(T_P, T_G, T_{D2})'$, disease progression/no disease progression (cure model), and the frailty variable (w), we derive the likelihood function in Section 3.2 and the CIF's in Section 4.3.

The baseline hazard function $h_0(\cdot)$ in (1) is represented by a piecewise constant function. First, we create a finite partition of the time axis with J_0 intervals and cut-points $0 = c_{00} < c_{01} < \cdots < c_{0J_0}$, where $c_{0J_0} = \infty$ so that the intervals are $(0, c_{01}]$, $(c_{01}, c_{02}]$, $\ldots$, (c_{0,J_0-1}, ∞). In the j-th interval, we have a constant hazard λ_{0j}, $j = 1, \ldots, J_0$. Then, $h_0(t|\boldsymbol{\lambda}_0) = \lambda_{0j}$, when $c_{0,j-1} < t \leqslant c_{0j}$, and $H_0(t|\boldsymbol{\lambda}_0) = \lambda_{0j}(t - c_{0,j-1}) + \sum_{m=1}^{j-1}\lambda_{0m}(c_{0m} - c_{0,m-1})$, with $\boldsymbol{\lambda}_0 = (\lambda_{01}, \ldots, \lambda_{0J_0})'$. The functions $h_{10}(\cdot)$ in (2) and $h_{20}(\cdot)$ in (3) as well as the cumulative hazard functions $H_{10}(\cdot)$ and $H_{20}(\cdot)$ are defined in a similar way based on J_1 and J_2 intervals, respectively, with parameters $\boldsymbol{\lambda}_m = (\lambda_{m1}, \ldots, \lambda_{mJ_m})'$ and intervals $(0, c_{m1}]$, $(c_{m1}, c_{m2}]$, $\ldots$,

(c_{m,J_m-1}, ∞), for $m = 1, 2$. In the results reported in Sections 5 and 6, for $j = 1, \ldots, J_m$, we chose the limits of the intervals based on the percentiles of the non-censored observed times, for $m = 0, 1, 2$.

From our formulation, the patterns of the observed data are classified in cases displayed in Figure 1 and described as C1: disease progression and death censored at the same time, C2: disease progression censored and death due to other causes, C3: disease progression and death censored, C4: disease progression and death due to other causes, C5: disease progression and death due to the disease, C6: disease progression censored and death censored afterwards, and C7: disease progression censored and death due to the disease. From top to bottom, the endpoints in Figure 2 may lead, respectively, to Cases (C1, C2, C6), (C1, C3, C5, C6, C7), and (C1, C2, C3, C4, C6) in Figure 1.

3.2 *Likelihood Function*

With respect to the observable times in Figure 1, we have y_P, the observed or right-censored disease progression time, and y, the observed or right-censored death time. Let $\boldsymbol{D}_{\text{o}} = (y_{Pi}, y_i, \boldsymbol{x}_i', \mathcal{C}_i : i = 1, \ldots, n)'$ and $\boldsymbol{\gamma} = (\kappa, \boldsymbol{\beta}_0', \boldsymbol{\lambda}_0', \boldsymbol{\beta}_1', \boldsymbol{\lambda}_1', \boldsymbol{\beta}_2', \boldsymbol{\lambda}_2', \tau)'$ be the observed data and the parameters in the model, respectively, where $\mathcal{C}_i \in \{1, \ldots, 7\}$ denotes the case indicator. The likelihood function $L(\boldsymbol{\gamma}|\boldsymbol{D}_{\text{o}}, \boldsymbol{w})$ for the observed data $\boldsymbol{D}_{\text{o}}$ augmented by the frailty vector $\boldsymbol{w}$ can be found in Web Appendix C. After integrating out $\boldsymbol{w}$ from $L(\boldsymbol{\gamma}|\boldsymbol{D}_{\text{o}}, \boldsymbol{w})$, we obtain the likelihood function for observed data $L(\boldsymbol{\gamma}|\boldsymbol{D}_{\text{o}})$. For each case, expressions (4)–(10) give the contributions of the i-th subject to the likelihood function.

C1. Disease progression censored and death censored at the same time ($n = 160$)

$$L_{1i}(\kappa, \boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \tau|\boldsymbol{D}_{\text{o}}) = [1 + \tau\{F_0(y_i|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{20}(y_i|\boldsymbol{\lambda}_2) \times \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)\}]^{-1/\tau}. \quad (4)$$

C2. Disease progression censored and death due to other causes ($n = 37$)

$$L_{2i}(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o}}) = h_{20}(y_i|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)\Big\{\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0)(1+\tau)\int_{y_{P_i}}^{y_i} f_0(u_i|\boldsymbol{\lambda}_0)\Big(1 + \tau[F_0(u_i|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{10}(y_i-u_i|\boldsymbol{\lambda}_1)\exp\{(\boldsymbol{x}_i',u_i)\boldsymbol{\beta}_1\} + H_{20}(y_i|\boldsymbol{\lambda}_2) \times \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)]\Big)^{-1/\tau-2}du_i + [1+\tau\{F_0(y_i|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{20}(y_i|\boldsymbol{\lambda}_2) \times \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)\}]^{-1/\tau-1}\Big\}. \tag{5}$$

C3. Disease progression and death censored ($n = 119$)

$$L_{3i}(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o}}) = f_0(y_{Pi}|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0)[1+\tau\{F_0(y_{Pi}|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{10}(y_i-y_{Pi}|\boldsymbol{\lambda}_1)\exp(\boldsymbol{z}_i'\boldsymbol{\beta}_1) + H_{20}(y_i|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)\}]^{-1/\tau-1}. \tag{6}$$

C4. Disease progression and death due to other causes ($n = 12$)

$$L_{4i}(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o}}) = f_0(y_{Pi}|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0)h_{20}(y_i|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)(1+\tau)[1+\tau \times \{F_0(y_{Pi}|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{10}(y_i-y_{Pi}|\boldsymbol{\lambda}_1)\exp(\boldsymbol{z}_i'\boldsymbol{\beta}_1) + H_{20}(y_i|\boldsymbol{\lambda}_2) \times \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)\}]^{-1/\tau-2}. \tag{7}$$

C5. Disease progression and death due to the disease ($n = 29$)

$$L_{5i}(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o}}) = f_0(y_{Pi}|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0)h_{10}(y_i-y_{Pi}|\boldsymbol{\lambda}_1)\exp(\boldsymbol{z}_i'\boldsymbol{\beta}_1)(1+\tau) \times [1+\tau\{F_0(y_{Pi}|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{10}(y_i-y_{Pi}|\boldsymbol{\lambda}_1)\exp(\boldsymbol{z}_i'\boldsymbol{\beta}_1) + H_{20}(y_i|\boldsymbol{\lambda}_2) \times \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)\}]^{-1/\tau-2}. \tag{8}$$

C6. Disease progression censored and death censored afterwards ($n = 0$)

$$L_{6i}(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o}}) = \Big\{\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0)\int_{y_{P_i}}^{y_i} f_0(u_i|\boldsymbol{\lambda}_0)\Big(1+\tau[F_0(u_i|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{10}(y_i-u_i|\boldsymbol{\lambda}_1)\exp\{(\boldsymbol{x}_i',u_i)\boldsymbol{\beta}_1\} + H_{20}(y_i|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)]\Big)^{-1/\tau-1}du_i + [1+\tau\{F_0(y_i|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{20}(y_i|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)\}]^{-1/\tau}\Big\}. \tag{9}$$

C7. Disease progression censored and death due to the disease ($n = 1$)

$$L_{7i}(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o}}) = \kappa \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0)(1+\tau) \int_{y_{Pi}}^{y_i} \exp\{(\boldsymbol{x}_i', u_i)\boldsymbol{\beta}_1\} f_0(u_i|\boldsymbol{\lambda}_0) h_{10}(y_i - u_i|\boldsymbol{\lambda}_1) \Big(1 + \tau[F_0(u_i|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + H_{10}(y_i - u_i|\boldsymbol{\lambda}_1)\exp\{(\boldsymbol{x}_i', u_i)\boldsymbol{\beta}_1\} + H_{20}(y_i|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_2)]\Big)^{-1/\tau-2} du_i. \tag{10}$$

By combining (4)–(10), we arrive at the likelihood function

$$L(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o}}) = \prod_{i:\,\mathcal{C}_i=1} L_{1i}(\kappa, \boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \tau|\boldsymbol{D}_{\mathrm{o}}) \prod_{j=2}^{7} \prod_{i:\,\mathcal{C}_i=j} L_{ji}(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o}}). \tag{11}$$

If it was possible to follow the patients to the end of their lives, we would have only three cases in Figure 1, namely (i) no disease progression and death due to other causes, (ii) disease progression and death due to other causes (corresponding to Case C4), and (iii) disease progression and death due to the disease (corresponding to Case C5). In Figure 1, disease progression and death were censored for 198 and 279 patients, respectively. Because of double censorings, the likelihood function in (11) comprises seven cases. Cases C1 and C6 arise from non-informative censoring.

4. Bayesian Inference

4.1 *Prior and Posterior*

We specify a prior distribution with independent components such that

$$\pi(\boldsymbol{\gamma}) = \pi(\kappa)\pi(\boldsymbol{\beta}_0)\pi(\boldsymbol{\lambda}_0)\pi(\boldsymbol{\beta}_1)\pi(\boldsymbol{\lambda}_1)\pi(\boldsymbol{\beta}_2)\pi(\boldsymbol{\lambda}_2)\pi(\tau), \tag{12}$$

where $\kappa \sim \mathcal{G}(a_\kappa, b_\kappa)$, $\boldsymbol{\beta}_0 \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{B}_0)$, $\lambda_{0j} \overset{\text{ind.}}{\sim} \mathcal{G}(a_{0j}, b_{0j})$, for $j = 1, \ldots, J_0$, $\boldsymbol{\beta}_1 \sim \mathcal{N}_{p+1}(\boldsymbol{0}, \boldsymbol{B}_1)$, $\lambda_{1j} \overset{\text{ind.}}{\sim} \mathcal{G}(a_{1j}, b_{1j})$, for $j = 1, \ldots, J_1$, $\boldsymbol{\beta}_2 \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{B}_2)$, $\lambda_{2j} \overset{\text{ind.}}{\sim} \mathcal{G}(a_{2j}, b_{2j})$, for $j = 1, \ldots, J_2$, and $\tau \sim \mathcal{IG}(a_\tau, b_\tau)$, with a_κ, b_κ, $\boldsymbol{B}_0$, $\boldsymbol{a}_0 = (a_{01}, \ldots, a_{0J_0})'$, $\boldsymbol{b}_0$, $\boldsymbol{B}_1$, $\boldsymbol{a}_1$, $\boldsymbol{b}_1$, $\boldsymbol{B}_2$, $\boldsymbol{a}_2$, $\boldsymbol{b}_2$, a_τ, and b_τ denoting known hyperparameters, while $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{B})$ and $\mathcal{IG}(a_\tau, b_\tau)$ stand for the p-dimensional normal distribution with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{B}$ and the inverse gamma distribution with $\pi(\tau) = b_\tau^{a_\tau} \tau^{-(a_\tau+1)} \exp(-b_\tau/\tau)/\Gamma(a_\tau)$, respectively. By combining

the likelihood function in (11) and the prior distribution in (12), the posterior distribution is

$$\pi(\boldsymbol{\gamma}|\boldsymbol{D}_{\rm o}) = L(\boldsymbol{\gamma}|\boldsymbol{D}_{\rm o})\pi(\boldsymbol{\gamma}). \tag{13}$$

We specify a weakly informative prior for κ ($a_\kappa = b_\kappa = 1$) and a relatively non-informative prior for τ ($a_\tau = b_\tau = 0.1$). For the regression coefficients, we take $\boldsymbol{B}_0 = \boldsymbol{B}_2 = 10^3\boldsymbol{I}_p$ and $\boldsymbol{B}_1 = 10^3\boldsymbol{I}_{p+1}$, with $\boldsymbol{I}_p$ denoting the $p \times p$ unit matrix. Finally, for the parameters in the baseline hazard functions, the hyperparameters are $a_{mj} = b_{mj} = 0.01$, for $j = 1, \ldots, J_m$ and $m = 0, 1, 2$.

4.2 *Computational Development*

The analytical form of the posterior distribution in (13) is intractable. Therefore, we develop a Gibbs sampling scheme (Robert and Casella, 2004) to draw samples from the posterior distribution. To this end, we consider reparameterizations and introduce latent variables. The details of our computational development are given in Web Appendix D.

In order to facilitate the computations, we make the transformations $\kappa^* = \tau\kappa$, $\lambda^*_{1j} = \tau\lambda_{1j}$, for $j = 1, \ldots, J_1$, and $\lambda^*_{2j} = \tau\lambda_{2j}$, for $j = 1, \ldots, J_2$, with Jacobian given by $\tau^{-(1+J_1+J_2)}$. Let $\boldsymbol{\gamma}^* = (\kappa^*, \boldsymbol{\beta}_0', \boldsymbol{\lambda}_0', \boldsymbol{\beta}_1', \boldsymbol{\lambda}_1^{*\prime}, \boldsymbol{\beta}_2', \boldsymbol{\lambda}_2^{*\prime}, \tau)'$ denote the transformed parameter vector. From (12) and (13), the posterior distribution of $\boldsymbol{\gamma}^*$ is given by

$$\begin{aligned}\pi(\boldsymbol{\gamma}^*|\boldsymbol{D}_{\rm o}) = {} & L(\boldsymbol{\gamma}^{**}|\boldsymbol{D}_{\rm o})\pi(\kappa^*/\tau)\pi(\boldsymbol{\beta}_0)\pi(\boldsymbol{\lambda}_0)\pi(\boldsymbol{\beta}_1)\pi(\boldsymbol{\lambda}_1^*/\tau)\pi(\boldsymbol{\beta}_2)\pi(\boldsymbol{\lambda}_2^*/\tau)\pi(\tau) \\ & \times \tau^{-(1+J_1+J_2)},\end{aligned} \tag{14}$$

where $L(\cdot|\boldsymbol{D}_{\rm o})$ is given in (11) and $\boldsymbol{\gamma}^{**} = (\kappa^*/\tau, \boldsymbol{\beta}_0', \boldsymbol{\lambda}_0', \boldsymbol{\beta}_1', \boldsymbol{\lambda}_1^{*\prime}/\tau, \boldsymbol{\beta}_2', \boldsymbol{\lambda}_2^{*\prime}/\tau, \tau)'$.

Latent indicator variables $\boldsymbol{\mathcal{I}} = (\mathcal{I}_i \in \{0, 1\} : \mathcal{C}_i \in \{2, 6\})'$ augment the data, so that each term of the expressions in (5) and (9) becomes multiplicative. The data are also augmented by the time to disease progression $\boldsymbol{u} = (u_i : y_{Pi} \leqslant u_i \leqslant y_i, \mathcal{C}_i \in \{2, 6, 7\})'$, which is sampled via the griddy-Gibbs sampler (Ritter and Tanner, 1992). After collapsing $\boldsymbol{u}$ (Liu, 1994),

the integrals involved in the sampling of $\boldsymbol{\mathcal{I}}$ are approximated by Gaussian quadrature. The sampling scheme for the components of $\boldsymbol{\gamma}^*$ is described below.

The MRS model is attractive. In particular, the formulation $\theta(\boldsymbol{x}, w) = w\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)$ in Section 3.1 leads to an efficient implementation of the Gibbs sampling algorithm to sample the frailty parameter τ from its full conditional posterior distribution, which is log-concave in $1/\tau$. Moreover, the full conditional posterior distributions of (i) the components of $(\boldsymbol{\beta}_0', \boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ and (ii) of κ^* and the components of $(\boldsymbol{\lambda}_1^{*\prime}, \boldsymbol{\lambda}_2^{*\prime})'$, after log transformations, are also log-concave. Therefore, we can sample from these distributions using the adaptive rejection sampling algorithm (Gilks and Wild, 1992). For the components of $\boldsymbol{\lambda}_0$, we consider the adaptive localized Metropolis algorithm (Chen *et al.*, 2000, Ch. 2).

4.3 *Cumulative Incidence Functions*

Next we present the expressions of the CIF's for disease progression, death due to the disease, and death due to other causes. The detailed derivation of these expressions is provided in Web Appendix A. We define $T_P^* = \infty \times 1\{N = 0\} + T_P \times 1\{N \geqslant 1 \cap\ T_P < T_{D2}\} + \infty \times 1\{N \geqslant 1 \cap T_P \geqslant T_{D2}\}$, $T_{D1}^* = \infty \times 1\{N = 0\} + T_{D1} \times 1\{N \geqslant 1 \cap T_{D1} < T_{D2}\} + \infty \times 1\{N \geqslant 1 \cap T_{D1} \geqslant T_{D2}\}$, and $T_{D2}^* = T_{D2} \times 1\{N = 0\} + T_{D2} \times 1\{N \geqslant 1 \cap\ T_{D2} < T_{D1}\} + \infty \times 1\{N \geqslant 1 \cap\ T_{D2} \geqslant T_{D1}\}$, where $1\{\text{event}\}$ is equal to 1 if the event happens and 0 otherwise, and we define $\infty \times 0 = 0$. These definitions are consistent with the patterns in Figure 2.

Under the MRS model, the CIF's for disease progression, death due to the disease, and death due to other causes are given, respectively, by

$$P(T_P^* \leqslant t | \boldsymbol{x}, \kappa, \boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \tau) = \kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\Big(\int_0^t f_0(u|\boldsymbol{\lambda}_0)[1 + \tau\{F_0(u|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0) + H_{20}(u|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}'\boldsymbol{\beta}_2)\}]^{-1/\tau - 1} du \Big), \quad (15)$$

$$P(T^*_{D1} \leqslant t|\boldsymbol{x},\boldsymbol{\gamma}) = \kappa(1+\tau)\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\Big(\int_0^t f_0(u|\boldsymbol{\lambda}_0)\exp(\boldsymbol{z}'\boldsymbol{\beta}_1)\int_0^{t-u} h_{10}(v|\boldsymbol{\lambda}_1)[1+\tau$$
$$\times\{F_0(u|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)+H_{10}(v|\boldsymbol{\lambda}_1)\exp(\boldsymbol{z}'\boldsymbol{\beta}_1)+H_{20}(v+u|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}'\boldsymbol{\beta}_2)\}]^{-1/\tau-2}dvdu\Big) \tag{16}$$

$$\text{and}\quad P(T^*_{D2} \leqslant t|\boldsymbol{x},\boldsymbol{\gamma}) = \exp(\boldsymbol{x}'\boldsymbol{\beta}_2)\Big\{\int_0^t h_{20}(v|\boldsymbol{\lambda}_2)[1+\tau\{F_0(v|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)+H_{20}(v|\boldsymbol{\lambda}_2)$$
$$\times\exp(\boldsymbol{x}'\boldsymbol{\beta}_2)\}]^{-1/\tau-1}dv\Big\}+\kappa(1+\tau)\exp\{\boldsymbol{x}'(\boldsymbol{\beta}_0+\boldsymbol{\beta}_2)\}\Big(\int_0^t f_0(u|\boldsymbol{\lambda}_0)\int_0^{t-u} h_{20}(v+u|\boldsymbol{\lambda}_2)$$
$$\times[1+\tau\{F_0(u|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)+H_{10}(v|\boldsymbol{\lambda}_1)\exp(\boldsymbol{z}'\boldsymbol{\beta}_1)+H_{20}(v+u|\boldsymbol{\lambda}_2)$$
$$\times\exp(\boldsymbol{x}'\boldsymbol{\beta}_2)\}]^{-1/\tau-2}dvdu\Big), \tag{17}$$

where $\boldsymbol{z}'\boldsymbol{\beta}_1 = (\boldsymbol{x}', u)\boldsymbol{\beta}_1$.

The integrals over $(0,t)$ in (15)–(17) are approximated by Gaussian quadrature. The detailed computation of the integrals over $(0, t-u)$ in (16) and (17) is given in Web Appendix B.

The time to disease progression can be modeled alone ignoring the time to death in different ways using standard parametric and semiparametric models as well as cure rate models. In the following proposition, we characterize the CIF for disease progression by comparing it with the respective CDF from the promotion time cure model in Chen *et al.* (1999).

PROPOSITION 1: The CIF for disease progression in (15) is bounded above by the CDF of the time to disease progression from the promotion time cure model ignoring death, that is, $P(T^*_P \leqslant t|\boldsymbol{x},\kappa,\boldsymbol{\beta}_0,\boldsymbol{\lambda}_0,\boldsymbol{\beta}_2,\boldsymbol{\lambda}_2,\tau) \leqslant [1-\exp\{-F_0(t|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\}]$.

Proof. In (15), first we notice that

$$[1+\tau\{F_0(u|\boldsymbol{\lambda}_0)\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)+H_{20}(u|\boldsymbol{\lambda}_2)\exp(\boldsymbol{x}'\boldsymbol{\beta}_2)\}]^{-1/\tau-1} \leqslant \{1+\tau F_0(u|\boldsymbol{\lambda}_0)\,\kappa\exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\}^{-1/\tau-1}.$$

Performing the change of variable $v = 1 + \tau F_0(u|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)$, we compute

$$\int_0^t \kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0) f_0(u|\boldsymbol{\lambda}_0)\{1 + \tau F_0(u|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\}^{-1/\tau-1} du$$
$$= \frac{1}{\tau}\int_1^{1+\tau F_0(t|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)} v^{-1/\tau-1} dv = 1 - \{1 + \tau F_0(t|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\}^{-1/\tau}.$$

Since $\log(1+u) \leqslant u$ for $u \geqslant 0$, we can show that $1 - \{1 + \tau F_0(t|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\}^{-1/\tau} \leqslant 1 - \exp\{-F_0(t|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\}$. Hence, $P(T_P^* \leqslant t|\boldsymbol{x}, \kappa, \boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \boldsymbol{\beta}_2, \boldsymbol{\lambda}_2, \tau) \leqslant [1 - \exp\{-F_0(t|\boldsymbol{\lambda}_0)\kappa \exp(\boldsymbol{x}'\boldsymbol{\beta}_0)\}]$. The right-hand side in this inequality corresponds to the improper CDF of the promotion time cure model in Chen *et al.* (1999) ignoring death. □

The result in Proposition 1 holds true even if $\tau \to 0$, that is, if the frailty term is constant ($w = 1$). Proposition 1 implies the CIF for disease progression may be upward biased if we consider only the time to disease progression through the promotion time cure model. This issue is further investigated in Section 5. Using the Gibbs samples from the posterior distribution of the parameters in Section 4.2, the posterior estimates of the CIF's in (16)–(17) are obtained.

4.4 *Model Comparison*

To carry out Bayesian model comparison, we use the Deviance Information Criterion (DIC) and the Logarithm of the Pseudo-Marginal Likelihood (LPML). We define the deviance $\text{Dev}(\boldsymbol{\gamma}) = -2\log L(\boldsymbol{\gamma}|\boldsymbol{\tau}, \boldsymbol{D}_\text{o})$, where $L(\boldsymbol{\gamma}|\boldsymbol{D}_\text{o})$ is given in (11). Let $\overline{\boldsymbol{\gamma}}$ and $\overline{\text{Dev}} = E\{\text{Dev}(\boldsymbol{\gamma}|\boldsymbol{D}_\text{o})\}$ denote the posterior means of $\boldsymbol{\gamma}$ and $\text{Dev}(\boldsymbol{\gamma})$, respectively. According to Spiegelhalter *et al.* (2002), the DIC measure is defined as $\text{DIC} = \text{Dev}(\overline{\boldsymbol{\gamma}}) + 2p_D$, where $p_D = \overline{\text{Dev}} - \text{Dev}(\overline{\boldsymbol{\gamma}})$ is the effective number of model parameters. The smaller the DIC value, the better the model fits the data. The posterior means $\overline{\boldsymbol{\gamma}}$ and $\overline{\text{Dev}}$ can be estimated by $\overline{\boldsymbol{\gamma}} = \sum_{j=1}^B \boldsymbol{\gamma}_j/B$ and $\overline{\text{Dev}} = \sum_{j=1}^B \text{Dev}(\boldsymbol{\gamma}_j)/B$, where $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_B$ are samples from the posterior distribution.

The LPML is another useful Bayesian measure of fit, which is defined based on the Conditional Predictive Ordinate (CPO). For the i-th observation, $\text{CPO}_i = \int L(\boldsymbol{\theta}|\boldsymbol{D}_{\text{o},(-i)})$

$\pi(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o},(-i)})d\boldsymbol{\gamma}$, where $\boldsymbol{D}_{\mathrm{o},i}$ is the observed data for the i-th subject, $L(\boldsymbol{\theta}|\boldsymbol{D}_{\mathrm{o},i})$ is the likelihood for the i-th subject, $\boldsymbol{D}_{\mathrm{o},(-i)}$ is the data with $\boldsymbol{D}_{\mathrm{o},i}$ deleted, and $\pi(\boldsymbol{\gamma}|\boldsymbol{D}_{\mathrm{o},(-i)})$ denotes the posterior density of $\boldsymbol{\gamma}$ based on the data $\boldsymbol{D}_{\mathrm{o},(-i)}$. According to Geisser and Eddy (1979) and Gelfand and Dey (1994), a Monte Carlo estimate is given by $\mathrm{LPML} = \sum_{i=1}^{n} \log(\widehat{\mathrm{CPO}}_i)$, where $\widehat{\mathrm{CPO}}_i = [\{\sum_{j=1}^{B} 1/L(\boldsymbol{\theta}_j|\boldsymbol{D}_{\mathrm{o},i})\}/B]^{-1}$. The larger the LPML value, the better the model fits the data. In Section 6, these criteria are computed in order to select the number of intervals J_0, J_1, and J_2 of the baseline hazard functions described in Section 3.1. For Cases C2 and C7, the integrals in (5) and (10) are approximated by Gaussian quadrature.

Bayesian computations using the Gibbs sampler were implemented in the FORTRAN language using IMSL subroutines with double precision arithmetic. An R interface to the code for fitting our proposed model together with an example data set was built and available at the journal website. The convergence of the Gibbs sampler was checked using several diagnostic tools discussed in Robert and Casella (2004, Ch. 12).

5. Simulation Studies

In this section, we conduct simulation studies to examine the empirical performance of the proposed MRS model. In the data generation with $p = 2$ covariates, we first generate n independent $x_{i1} \sim \mathcal{N}(0, 1)$ and given x_{i1}, we sample x_{i2} from the Bernoulli distribution with probability $1/\{1 + \exp(0.5 - 0.1x_{i1})\}$, independently, $i = 1, \ldots, n$. These values remain fixed throughout the 500 repetitions of the simulations. The censoring time for death C_D has a uniform distribution on $(0.9\, t_{\max}, 1.5\, t_{\max})$, with administrative censoring at $t_{\max} = 15$ chosen to control the censoring rate. The censoring time for disease progression C_P is drawn from the uniform distribution on $\{0.5C_D, C_D\}$. For the sake of simplicity, we adopt exponential distributions for the baseline density function in (1) and the baseline hazard functions in (2) and (3). In all simulations, the data are generated under the proposed MRS model. First we sample the frailty w and then the number of latent causes N, T_P, T_G, and T_{D2}. The steps for

data generation are in Table S2. To achieve stability in the computation of the exponential function in (2), imputed times to disease progression in Section 4.2 are truncated at 7.3, which corresponds to the average of the maximum times to disease progression in Cases C3–C5.

Our study comprises two sample sizes (400 and 800) with true values of the parameters shown in Table S3. For Cases C1–C7 in Section 3.1, the average percentages of observations are 11.1, 43.8, 7.9, 12.1, 11.6, 13.2, and 0.3. The average proportion of cured subjects is 0.30. In the Gibbs sampling algorithm, after discarding the first 2000 iterations, we generated 3000 additional samples for posterior inference.

First we assess the frequentist properties of the posterior estimators of the parameters under the MRS model. Some posterior summaries are shown in Table S3. Except for κ, we see the true values of the parameters are well recovered. The averages of the posterior means do not differ too much from the true values. Moreover, except for κ, the averages of the posterior standard deviations (SD's) and the root mean squared errors of the posterior mean (RMSE's) are close, indicating the uncertainty in the estimates is adequately quantified. When $n = 400$, coverage probabilities (CP's) of the 95% highest posterior density (HPD) for κ and λ_{01} are not close to the nominal value. On the other hand, when $n = 800$, CP's range from 0.918 for τ to 0.962 for κ.

Now our interest lies in the CIF's for disease progression, death due to the disease, and death due to other causes. True probabilities are computed as relative frequencies from 1,000,000 runs of the data generation process in Table S2. The arms in Tables 1 and S4 correspond to the values of the binary covariate x_2. Integration with respect to x_1 in (15)–(17) is approximated using the empirical conditional distribution of x_1 given x_2 (for an alternative, see Zeng *et al.*, 2012, Sect. 3.3). For the multisate model, frequentist results

from the semiparametric model with transition-specific covariates in Table 1 were computed with the mstate package (de Wreede *et al.*, 2011) in R.

For both sample sizes, the estimates of the CIF for disease progression from the MRS model in Table 1(a) are very close to the true values, whichever the time and the arm. SD's and RMSE's differ only negligibly. Coverage probabilities of the 95% HPD intervals differ from the nominal value by at most 0.014. On the other hand, estimates from the multistate model are downward biased. In Table S4, estimates from the promotion time cure model ignoring death are markedly biased toward overestimation of the CIF for disease progression, as could be expected in view of Proposition 1, and coverage probabilities of the 95% HPD intervals reach values as low as 0.096.

[Table 1 about here.]

In Table 1(b), we see the multistate and the MRS models perform well in estimating the CIF for death due to the disease. RMSE's are slightly smaller for the MRS model and CP's range from 0.926 to 0.958. Finally, in Table 1(c) the estimates of the CIF for death due to other causes from the multistate model are more biased than the ones from the MRS model, especially when $n = 400$. Similar to the CIF for death due to the disease, RMSE's are slightly smaller for the MRS model and CP's range from 0.934 to 0.960.

6. Analysis of the Prostate Cancer Data

In this section, we conduct a detailed analysis of the data described in Section 2. The number of patients in each one of Cases C1–C7 is shown in Figure 1, noticing that Case C6 was not observed. The vector of baseline covariates (see Section 2) is denoted by $\boldsymbol{x} = (\text{Age, PSA, Vel2, GS7, GS8H, T2})'$, where the covariates Age and PSA are standardized. The time to PSA recurrence T_P in (2) is standardized using observations from Cases C3–C5. Next we present some results obtained from the samples of the posterior distribution. For model selection, in

the Gibbs sampling algorithms, after discarding the first 2000 iterations, every 5^{th} sample from the next 25,000 additional samples is retained, so that the computation of the criteria in Tables S3 and 2 is based on 5000 samples. Henceforth, all the results correspond to the best models in Tables S5 ($J_0 = 20$) and 2 ($J_0 = 20, J_1 = J_2 = 5$).

[Table 2 about here.]

For the MRS model in Table 3, after discarding the first 5000 iterations of the Gibbs samplers, we generated 200,000 additional samples and by taking a spacing of size 40, the summaries are based on 5000 samples. Figures S1–S4 show the trace, histogram, ergodic mean, and autocorrelation plots for κ and τ as well as the trace plots for the remaining parameters under the MRS model. These plots indicate the chains have a good mixing and converge. The impact of the prior on the posterior distribution for some parameters can be assessed in Figure S5. As we can see in Figure S5, the posterior is mostly driven by the data and this behavior is common to all parameters (not shown) under the MRS model.

The point estimate and the HPD interval for the frailty variance τ given in Table 3 suggest a small variability in the frailty variable. Fixing $\tau = 0.1$ and $\tau = 0.7$, the estimates of the criteria (DIC, LPML) are, respectively, (1601.8, -811.4) and (1467.8, -738.4) so that in both cases the results are worse than the best model in Table 2. Hence, a model with a random τ yields a better fit.

A regression coefficient is considered to be significant if its 95% HPD interval does not contain 0. We see in Table 3 PSA velocity greater than 2.0 ng/ml per year (Vel2 = 1) and biopsy Gleason score $\in \{8, 9, 10\}$ (GS8H = 1) are significantly associated with a shorter time to PSA recurrence and a shorter time between disease progression and death due to the disease. These results are not surprising. Age is significantly associated with a shorter time to death due to other causes, while is barely significantly associated with a longer time to PSA recurrence.

[Table 3 about here.]

Based on (15)–(17) and for the two levels of Vel2, we computed the estimates of the PSA recurrence probability, the prostate cancer-specific mortality, and the other causes-specific mortality displayed in Figure 3. The integrals in (15)–(17) are approximated using the empirical conditional distribution of the remaining covariates given Vel2.

The estimates in Table 3 justify the differences between the curves for Vel2 = 0 (PSA velocity $\leqslant$ 2.0 ng/ml per year) and Vel2 = 1 (PSA velocity $>$ 2.0 ng/ml per year). The nonparametric estimates in Figure 3(a) are the Kaplan-Meier estimates, while in Figure 3(b, c) we used the estimators in Fine and Gray (1999). Except for Vel2 = 1 in Figure 3(b), the plots are similar even though the results from the MRS model are adjusted for covariates.

The 5-year estimates of PSA recurrence with 95% HPD intervals under the MRS model are 42.8% (36.1%, 49.7%) and 69.5% (61.8%, 77.3%) for Vel2 = 0 and Vel2 = 1, respectively. For prostate cancer-specific mortality, the corresponding 7-year estimates are 1.3% (0.02%, 4.3%) and 13.0% (1.3%, 33.1%).

For patients with Vel2 = 0, the differences in the estimates (not shown) between the promotion time cure model ignoring death and the MRS model attain a maximum of 18% (see Figure S6). Similar behavior was observed in the simulation study (see Table S4) and this striking difference is not unexpected in light of Proposition 1.

[Figure 3 about here.]

7. Discussion and Concluding Remarks

This work was motivated by a very interesting data set with a complex structure often found in clinical research, which includes one nonterminating event and two terminating events. In practice, either for simplicity or to serve a different purpose, not all these events are considered. For instance, to study the effects of baseline covariates on the disease progression

time, time to death is usually pooled with non-informative censoring time. As an alternative, we connect a cure rate model for the progression of the disease and the models for the times to the terminating events via a frailty term to develop a new model, which is particularly suitable for fitting survival data with semi-competing and competing risks. One of the novelties of the proposed MRS model is to allow the time to death due to other causes to censor both the time to disease progression and the time to death due to the disease such as Cases C2 and C4 in Section 3.1 (see also Figure 1).

The result in Proposition 1 and the simulation study in Section 5 reveal that when estimating the CIF for disease progression, ignoring the information on terminating events can be harmful. On the other hand, for the models and data sets in Conlon *et al.* (2014, Sect. 5.3) and Lee *et al.* (2015, Sect. 3.3), they conclude that simpler models ignoring the time to death lead to small differences in parameters estimates. Moreover, empirical results in Table 1(c) reveal the bias incurred when estimating the CIF for death due to other causes using the multistate model.

Under the multistate structure in Section 2, when fitting the semiparametric model with transition-specific covariates (de Wreede *et al.*, 2011) to the data in Section 2, the likelihood function is monotonic (Bryson and Johnson, 1981). Chen *et al.* (2009) give conditions for the existence of the maximum likelihood estimators. For the data in Section 2, the monotonic likelihood problem happens because GS8H $= 0$ for all 37 patients in Case C2 assigned to transition $1 \rightarrow 4$ (from study entry to death due to other causes). After taking out GS7 and GS8H from the model, we overcome the monotonic likelihood issue. However, the Aalen-Johansen estimator of the transition probability matrix gives negative estimates for elements on the diagonal. Therefore, the MRS model can accommodate more important clinical binary covariates.

It is well known the gamma distribution for the frailty variable imposes a positive associa-

tion among the variables. For Cases C4 and C5 in the data set in Section 2, the estimates of the Spearman correlation coefficient for (T_P, T_{D2}) and (T_P, T_G) are 0.74 and 0.24, respectively. Nonetheless, a more flexible correlation structure might be explored in a future work. Future research might also include model diagnostics and multiple semi-competing risks.

Acknowledgements

We would like to thank the Editor, a Co-Editor, an Associate Editor, and a referee for their valuable comments, which have led to a much improved version of the paper. Dr. de Castro's research was partially supported by CNPq, Brazil. Dr. Chen's research was partially supported by US NIH grants #GM 70335 and #P01 CA142538.

References

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **47**, 501–515.

Beyersmann, J., Schumacher, M., and Allignol, A. (2012). *Competing Risks and Multistate Models with R*. Springer, New York.

Bryson, M. C. and Johnson, M. E. (1981). The incidence of monotone likelihood in the Cox model. *Technometrics*, **23**, 381–383.

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94**, 909–919.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer, New York.

Chen, M.-H., Ibrahim, J. G., and Shao, Q.-M. (2009). Maximum likelihood inference for the Cox regression model with applications to missing covariates. *Journal of Multivariate Analysis*, **100**, 2018–2030.

Conlon, A. S. C., Taylor, J. M. G., and Sargent, D. J. (2014). Multi-state models for colon cancer recurrence and death with a cured fraction. *Statistics in Medicine*, **33**, 1750–1766.

Conlon, A. S. C., Taylor, J. M. G., and Sargent, D. J. (2015). Improving efficiency in clinical trials using auxiliary information: Application of a multi-state cure model. *Biometrics*, **71**, 460–468.

D'Amico, A. V., Renshaw, A. A., Sussman, B., and Chen, M.-H. (2005). Pretreatment PSA velocity and risk of death from prostate cancer following external beam radiation therapy. *Journal of the American Medical Association*, **294**, 440–447.

D'Amico, A. V., Chen, M.-H., Renshaw, A. A., Loffredo, M., and Kantoff, P. W. (2008a). Androgen suppression and radiation vs radiation alone for prostate cancer: a randomized trial. *Journal of the American Medical Association*, **299**, 289–295.

D'Amico, A. V., Halabi, S., Tempany, C., Titelbaum, D., Philips, G. K., Loffredo, M., et al. for the Cancer and Leukemia Group B (2008b). Tumor volume changes on 1.5 tesla endorectal MRI during neoadjuvant androgen suppression therapy for higher-risk prostate cancer and recurrence in men treated using radiation therapy results of the phase II CALGB 9682 study. *International Journal of Radiation Oncology⋆Biology⋆Physics*, **71**, 9–15.

de Castro, M., Chen, M.-H., and Zhang, Y. (2015). Bayesian path specific frailty models for multi-state survival data with applications. *Biometrics*, **71**, 760–771.

de Wreede, L. C., Fiocco, M., and Putter, H. (2011). mstate: An R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, **38**, 1–30.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496–509.

Gaynor, J. J., Feuer, E. J., Tan, C. C., Wu, D. H., Little, C. R., Strauss, D. J. et al. (1993). On the use of cause-specific failure and conditional failure probabilities: examples from

clinical oncology data. *Journal of the American Statistical Association*, **88**, 400–409.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, **74**, 153–160.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, **56**, 501–514.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**, 337–348.

Lee, K. H., Haneuse, S., Schrag, D., and Dominici, F. (2015). Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Applied Statistics*, **64**, 253–273.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, **89**, 958–966.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, **26**, 2389–2430.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association*, **87**, 861–868.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, second edition.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, **64**,

583–639.

Tsodikov, A. D., Ibrahim, J. G., and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*, **98**, 1063–1078.

Zeng, D., Chen, Q., Chen, M.-H., Ibrahim, J. G., and Amgen Research Group (2012). Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study. *Biometrika*, **99**, 167–184.

Zhang, Y., Chen, M.-H., Ibrahim, J. G., Zeng, D., Chen, Q., Pan, Z., et al. (2014). Bayesian gamma frailty models for survival data with semi-competing risks and treatment switching. *Lifetime Data Analysis*, **20**, 76–105.

Supporting Information

Web Appendices, Tables, and Figures referenced in Sections 2–4 are available with this paper at the Biometrics website on Wiley Online Library. Computational code and an example data set are availbale on request. There are many R packages (R Core Team, 2019) for fitting survival data, as can be seen at the CRAN Task View web page `https://cran.r-project.org/web/views/Survival.html`.

Data Availability Statement

The research data cannot be shared. The data set for this article is proprietary and confidential. The authors do not have the permission to publish or share the raw data.

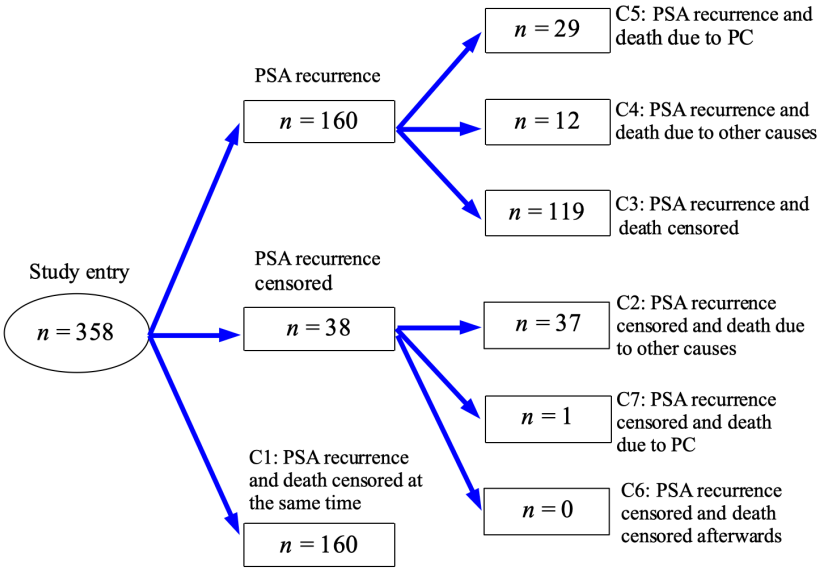


Figure 1. Graphical representation of the PC data. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

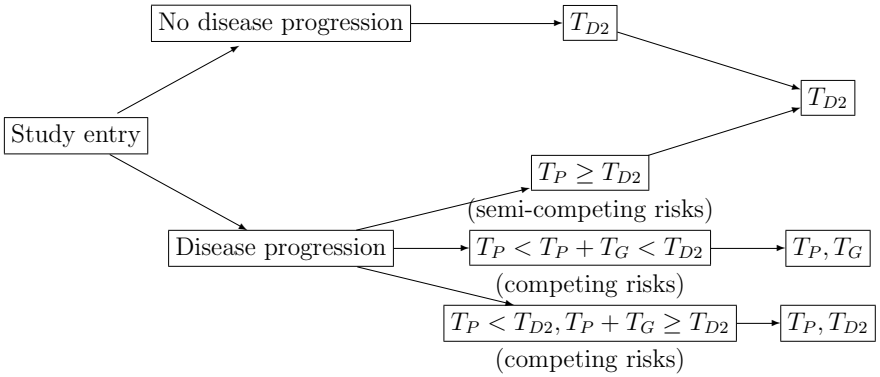
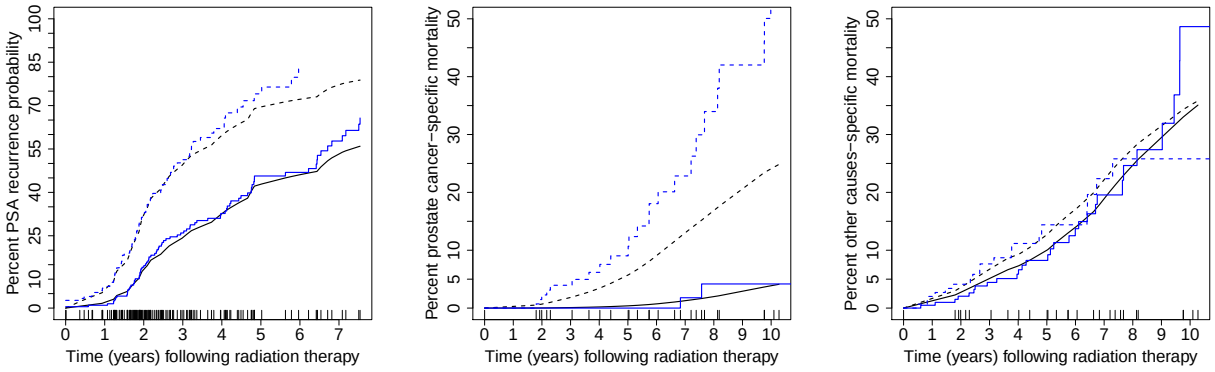


Figure 2. Graphical representation of the relationships among T_P, T_G, and T_{D2}.



(a) (b) (c)

Figure 3. Nonparametric estimate (in blue) and posterior mean from the MRS model (in black), in %, for Vel2 = 0 (solid) and Vel2 = 1 (dashed): (a) PSA recurrence probability, (b) prostate cancer-specific mortality, and (c) other causes specific mortality. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

Table 1

Summaries of the CIF's from 500 replications (True: true probability, Est: average of the estimates, SD*: standard deviation of the estimates, RMSE*: root mean squared error of the estimates, Est: average of the posterior means, SD: average of the posterior standard deviations, RMSE: root mean squared error of the posterior means, and CP: coverage probability of the 95% HPD interval).*

(a) Disease progression

$n = 400$

			Multistate model			Proposed model			
Time	Arm	True	Est*	SD*	RMSE*	Est	SD	RMSE	CP
2	0	0.105	0.094	0.016	0.020	0.107	0.014	0.014	0.960
	1	0.180	0.116	0.022	0.068	0.182	0.024	0.022	0.964
6	0	0.218	0.142	0.023	0.080	0.216	0.023	0.023	0.946
	1	0.323	0.118	0.022	0.205	0.318	0.033	0.031	0.954
12	0	0.289	0.130	0.023	0.161	0.284	0.027	0.028	0.936
	1	0.389	0.074	0.019	0.316	0.381	0.037	0.035	0.954

$n = 800$

			Multistate model			Proposed model			
Time	Arm	True	Est*	SD*	RMSE*	Est	SD	RMSE	CP
2	0	0.105	0.091	0.010	0.017	0.106	0.010	0.010	0.962
	1	0.180	0.120	0.017	0.062	0.184	0.017	0.018	0.938
6	0	0.218	0.141	0.014	0.079	0.217	0.017	0.016	0.958
	1	0.323	0.126	0.018	0.197	0.325	0.024	0.025	0.938
12	0	0.289	0.129	0.016	0.160	0.287	0.020	0.019	0.950
	1	0.389	0.077	0.015	0.312	0.392	0.026	0.027	0.938

(b) Death due to the disease

$n = 400$

			Multistate model			Proposed model			
Time	Arm	True	Est*	SD*	RMSE*	Est	SD	RMSE	CP
2	0	0.005	0.006	0.003	0.003	0.006	0.002	0.002	0.930
	1	0.039	0.037	0.013	0.013	0.041	0.010	0.010	0.954
6	0	0.025	0.028	0.009	0.009	0.025	0.007	0.007	0.926
	1	0.128	0.122	0.025	0.025	0.125	0.022	0.021	0.954
12	0	0.049	0.050	0.014	0.014	0.048	0.012	0.013	0.926
	1	0.189	0.187	0.032	0.032	0.182	0.028	0.027	0.958

$n = 800$

			Multistate model			Proposed model			
Time	Arm	True	Est*	SD*	RMSE*	Est	SD	RMSE	CP
2	0	0.005	0.007	0.002	0.002	0.005	0.001	0.001	0.934
	1	0.039	0.039	0.010	0.010	0.042	0.007	0.007	0.942
6	0	0.025	0.027	0.006	0.006	0.025	0.005	0.005	0.940
	1	0.128	0.128	0.019	0.019	0.130	0.016	0.017	0.942
12	0	0.049	0.049	0.009	0.009	0.049	0.009	0.009	0.946
	1	0.189	0.196	0.026	0.027	0.191	0.020	0.022	0.944

(c) Death due to other causes

$n = 400$

			Multistate model			Proposed model			
Time	Arm	True	Est*	SD*	RMSE*	Est	SD	RMSE	CP
2	0	0.136	0.147	0.020	0.022	0.138	0.013	0.013	0.954
	1	0.210	0.196	0.025	0.029	0.213	0.023	0.023	0.942
6	0	0.321	0.335	0.031	0.034	0.321	0.023	0.023	0.960
	1	0.420	0.399	0.035	0.040	0.421	0.033	0.032	0.946
12	0	0.484	0.542	0.037	0.068	0.482	0.028	0.027	0.960
	1	0.551	0.559	0.038	0.038	0.552	0.035	0.035	0.940

$n = 800$

			Multistate model			Proposed model			
Time	Arm	True	Est*	SD*	RMSE*	Est	SD	RMSE	CP
2	0	0.005	0.007	0.002	0.002	0.005	0.001	0.001	0.934
	1	0.039	0.039	.010	0.010	0.042	0.007	0.007	0.942
6	0	0.025	0.027	0.006	0.006	0.025	0.005	0.005	0.940
	1	0.128	0.128	0.019	0.019	0.130	0.016	0.017	0.942
12	0	0.049	0.049	0.009	0.009	0.049	0.009	0.009	0.946
	1	0.189	0.196	0.026	0.027	0.191	0.020	0.022	0.944

Table 2
DIC and LPML values under the proposed MRS model for the PC data.

J_0	J_1	J_2	DIC	p_D	LPML
5	5	5	1472.5	29.3	-741.5
10	5	5	1466.7	34.7	-738.5
20	5	5	1462.0	47.4	-735.9
20	5	10	1469.9	52.8	-740.3
20	10	5	1464.8	52.5	-738.9

Table 3

Posterior estimates under the MRS model for the PC data (Est: posterior mean, SD: posterior standard deviation, and HPD: 95% HPD interval).

Variable	Est	SD	HPD	Variable	Est	SD	HPD
		T_P model ($\boldsymbol{\beta}_0$)				T_G model ($\boldsymbol{\beta}_1$)	
Age	-0.194	0.094	(-0.381, -0.018)	Age	0.141	0.226	(-0.289, 0.587)
PSA	0.329	0.087	(0.171, 0.509)	PSA	0.159	0.129	(-0.098, 0.418)
Vel2	0.694	0.207	(0.267, 1.083)	Vel2	1.778	0.882	(0.071, 3.449)
GS7	0.103	0.196	(-0.282, 0.480)	GS7	0.627	0.614	(-0.579, 1.835)
GS8H	0.886	0.293	(0.302, 1.461)	GS8H	1.435	0.658	(0.184, 2.713)
T2	0.287	0.188	(-0.064, 0.676)	T2	0.684	0.575	(-0.423, 1.833)
κ	1.263	0.423	(0.637, 2.055)	T_P	-0.413	0.462	(-1.329, 0.459)
		T_{D2} model ($\boldsymbol{\beta}_2$)				Frailty parameter	
Age	0.550	0.167	(0.218, 0.877)	τ	0.310	0.209	(0.021, 0.732)
PSA	-0.547	0.295	(-1.121, 0.004)				
Vel2	0.553	0.357	(-0.159, 1.248)				
GS7	-0.093	0.346	(-0.833, 0.528)				
GS8H	-0.471	0.701	(-1.941, 0.787)				
T2	0.334	0.327	(-0.291, 0.959)				