Bayesian variable selection for a semi-competing risks model with three hazard functions

Andrew G. Chapple, Marina Vannucci, Peter F. Thall, Steven Lin

A variable selection procedure is developed for a semi-competing risks regression model with three hazard functions that uses spike-and-slab priors and stochastic search variable selection algorithms for posterior inference. A rule is devised for choosing the threshold on the marginal posterior probability of variable inclusion based on the Deviance Information Criterion (DIC) that is examined in a simulation study. The method is applied to data from esophageal cancer patients from the MD Anderson Cancer Center, Houston, TX, where the most important covariates are selected in each of the hazards of effusion, death before effusion, and death after effusion. The DIC procedure that is proposed leads to similar selected models regardless of the choices of some of the hyperparameters. The application results show that patients with intensity-modulated radiation therapy have significantly reduced risks of pericardial effusion, pleural effusion, and death before either effusion type.

© 2017 Elsevier B.V. All rights reserved.
3D-CRT. Due to the relationship between increased dosage and effusion rates, IMRT potentially could result in fewer incidences of pleural and pericardial effusion in esophageal cancer patients compared to standard 3D-CRT treatment. Assessing the impact of IMRT on time to effusion is more complicated than assessing the impact of IMRT on overall survival time, even when pleural and pericardial effusion are considered separately. When the survival time of interest is a non-terminal event such as effusion, death is commonly assumed to be a non-informative independent censoring event (Shirai et al., 2011). This assumption is invalid because both pleural and pericardial effusion could lead to death, which means that death may indicate that a patient experienced effusion, which is informative censoring. Another complication with this data structure is that patients can experience effusion followed by death, but not death first and effusion afterwards. Administrative right censoring could occur before a patient experiences either event type or after a patient has effusion. Due to these complications, this data structure must be analyzed with a semi-competing risks model. This model has different hazards for three events: a given non-terminal event, death before the non-terminal event and death after the non-terminal event.

Lee et al. (2015) developed a novel Bayesian semi-parametric semi-competing risks regression model for a non-terminal event and death. Their motivating non-terminal event of interest was hospital readmission for patients diagnosed with advanced pancreatic cancer. Since pancreatic cancer has high mortality rates, they were concerned with end of life care and keeping patients comfortable at home during their final days. They considered three different hazard functions: the hazard of a non-terminal event, the hazard of death without a non-terminal event, and the hazard of death after a non-terminal event. Each of these three hazard functions resembled a Cox-type regression including a baseline hazard function which was assumed to be piecewise exponential, individual patient frailty parameters, and a linear combination of patient covariates. They used the posterior sample of the beta coefficients in the three hazards for inference on what types of homecare affected the hazard that a patient would return to the hospital, the hazard of death before returning to the hospital, and the hazard of death after patients were readmitted to the hospital. They implemented their algorithm in the package SemiCompRisks (Lee and Haneuse, 2015).

We initially aimed at implementing the semi-competing risks model of Lee et al. (2015) to analyze the effects of IMRT on effusion and overall survival for an observational study consisting of 470 patients at The University of Texas M.D. Anderson Cancer Center in Houston, TX, treated between January 1998 and April 2012 (He et al., 2016). However, it was unclear which baseline covariates should be included in the model for analyzing the IMRT effect, particularly because of the correlation between treatment group assignment and the baseline covariates, which could affect clinical conclusions. Consequently, in this paper we develop a variable selection procedure for the semi-competing risks model of Lee et al. that uses spike-and slab priors and stochastic search variable selection (SSVS) algorithms for posterior inference. The proposed procedure performs variable selection for each of the linear terms in the three hazard functions. Furthermore, we devise a protocol to choose the threshold on the marginal posterior probability of variable inclusion based on the Deviance Information Criterion (DIC). The code for the described methodology can be found in the R package SCRSELECT (Chapple, 2016). In the application to the data from the esophageal cancer patients we do not perform variable selection on the IMRT status. To correct for some of the bias introduced in a nonrandomized observational study, we estimate the probability of receiving IMRT for each patient as a function of other covariates and include this propensity score in each hazard function. This allows us to compare the effects of IMRT on effusion and death while correcting for bias for non-randomization in every potential model. We present results from analyses done separately for pleural and pericardial effusion, where we show how the DIC procedure we propose leads to similar selected models, regardless of the choice of some of the hyperparameters. We find that patients with IMRT radiation have significantly reduced risks of pericardial effusion, pleural effusion and death before either effusion type. The rest of the paper is organized as follows: In Section 2, we describe the Bayesian semi-parametric semi-competing risks model, the variable selection priors and the Markov Chain Monte Carlo procedure for posterior inference. We also present the DIC-based procedure that we propose for the final covariate selection. In Section 3, we perform a simulation study to assess our proposed DIC-based procedure. In Section 4, we describe the case study data and discuss results and sensitivity to hyperparameter choices. Section 5 concludes the paper with a discussion.

2. Methods

2.1. Semi-parametric semi-competing risks model

Let $T_{1i}$ denote the time to a non-terminal event and $T_{2i}$ be the time to death for patient $i$. Lee et al. (2015) model covariate effects in the three hazard functions in the following manner. They denote $h_1$, the hazard of a non-terminal event, $h_2$ the hazard of a terminal event when the non-terminal event has not occurred and $h_3$, the hazard of a terminal event after the non-terminal event has occurred. Let $x_i$ denote the vector of patient covariates and $\beta_1, \beta_2, \beta_3$ denote the three coefficient vectors associated with $x_i$ in hazards 1, 2, and 3, respectively. They list the functional forms of the three hazards for the semi-Markov model as

$$h_1(T_{1i}|\gamma_1, \beta_1, x_i) = \gamma_1 h_{01}(T_{1i}) \exp(x_i^T \beta_1),$$  \hspace{1cm} (1)

$$h_2(T_{2i}|\gamma_1, \beta_2, x_i) = \gamma_1 h_{02}(T_{2i}) \exp(x_i^T \beta_2)$$  \hspace{1cm} (2)

and

$$h_3(T_{2i}|T_{1i}, \gamma_1, \beta_3, x_i) = \gamma_1 h_{03}(T_{2i} - T_{1i}) \exp(x_i^T \beta_3).$$  \hspace{1cm} (3)
The basic idea is to introduce a latent binary random vector. **2.2. Variable selection priors**

Values as the split points. We also adopt this prior formulation, which limits the number of split point intervals that have a time between effusion and death in interval \((s_{j-1}, s_j]\) for event \(g\). The observed events are \(Y_i = \min(T_i, \tau_i, C_i)\), where \(\delta_i = \mathbb{I}[Y_i \leq \min(T_i, C_i)]\) and \(\delta_i = \mathbb{I}[Y_i > C_i]\). These are realizations of \(T_i\), the time to non-terminal event, \(\tau_i\), the time to death, and \(C_i\), the independent censoring time for patient \(i\). The likelihood for this parameterization of the baseline hazard functions and given hazard functions (1)–(3) is

\[
L(Y_1, Y_2, \delta_1, \delta_2, X; \beta, \beta_2, \beta_3, \gamma, s_1, s_2, s_3, \lambda_1, \lambda_2, \lambda_3) = \prod_{j=1}^{S_{g,j}} \prod_{l=1}^{I_{g,j}} \exp[\lambda_j d_{lj} - \exp(\lambda_j)] \sum_{m \in M} \Delta_{mj}^{l} \gamma_m \exp(x'_m \beta_1) \times \exp(\lambda_j d_{lj} - \exp(\lambda_j)] \sum_{q \in Q_{g,j}} \Delta_{qj}^{l} \gamma_q \exp(x'_q \beta_2) \\
\times \prod_{m' \in D_g} \gamma_{m'} \exp(x'_{m'} \beta_3) \prod_{q' \in Q_{g,j}} \gamma_{q'} \exp(x'_{q'} \beta_2) \prod_{r' \in R_{g,j}} \gamma_{r'} \exp(x'_{r'} \beta_3),
\]

where \(d_{lj}\) is the number of patients who experienced a non-terminal event in the interval \((s_{j-1}, s_j]\), \(d_{lj}\) is the number of patients who experienced a terminal event but did not previously experience a non-terminal event in the interval \((s_{j-1}, s_{j+1}]\), and \(d_{lj} = \mathbb{I}[i : s_{j-1} < s_j = 1, \delta_i = 1]\) is the number of patients who have a time between effusion and death in interval \((s_{j-1}, s_j]\). \(\Delta_{lj} = \max(0, \min(y_i, \tau_i) - s_{j-1})\) for \(g = 1, 2, 3\) and \(\Delta_{lj} = \max(\min(y_i - \tau_i, s_j), \tau_i - s_{j-1})\). We denote by \(R_{g}\) the risk set (the set of patients who have neither experienced event \(g\) nor been censored by time \(s_{j-1}\)) for interval \(I_{g,j}\) and let \(D_{g,j}\) denote the set of patients who experienced event \(g\) in this interval. We follow Lee et al. and assign priors for \(g = 1, 2, 3\)

\[
\lambda_g \mid b_g, \mu_{g,k}, \sigma_{g,k}^2 \sim N_{b_g+1}((\mu_{g,k}, 1, \sigma_{g,k}^2 \Sigma_{g,k}), \\
\sigma_{g,k}^{-2} \sim \text{Gamma}(a_{g,k}, b_g), \\
\pi(\mu_{g,k}) \sim 1, \\
\gamma \sim \text{Gamma}(\theta-1, 1), \\
\theta^{-1} \sim \text{Gamma}(\psi, \omega).
\]

This prior formulation has hyperparameters \(a_g, b_g, \psi, \omega\). Here \(\Sigma_{g,k}\) is a function of the current partition \(s_g\) corresponding to a Multivariate Normal intrinsic conditional autoregression model (ICAR) as formulated by Besag and Kooperberg (1995).

Lee et al. (2015) consider the number of split points \(I_g\) to be a Poisson random variable with mean \(a_g\). They place a prior distribution on the partition \(s_g\) by drawing \(2I_g + 1\) uniform random variables on \([0, s_{g,\text{max}}]\) and take the even indexed values as the split points. We also adopt this prior formulation, which limits the number of split point intervals that have no events.

### 2.2. Variable selection priors

Bayesian variable selection methods were introduced by George and McCulloch (1993, 1997) for normal linear models. The basic idea is to introduce a latent binary random vector \(\eta = (\eta_1, \ldots, \eta_p)\) with \(\eta_k = 0\) indicating that the variable \(x_{ik}\) should be excluded from the model and \(\eta_k = 1\) otherwise. Generalizing this notation to three hazard functions, we introduce three latent vectors \(\eta_k = (\eta_{g,k}, \ldots, \eta_{g,p})\) one for each hazard function \(g = 1, 2, 3\), where \(\eta_{g,k} = 1\) indicates that the variable \(x_{ik}\) is important in hazard \(g\) and \(\eta_{g,k} = 0\) otherwise. The indicator \(\eta_{g,k}\) is included in the prior distribution of \(\beta_{g,k}\) to define the mixture

\[
\beta_{g,k} \sim N(0, \lambda^2_{g,k}) + (1 - \eta_{g,k}) \delta_{g,k}(\beta_{g,k}),
\]

where \(\delta_{g,k}\) is the point mass distribution at 0. Mixture priors of type (4) are known as spike-and-slab priors in the Bayesian variable selection literature. Here we choose \(\lambda^2_{g,k}\) in (4) as the kth diagonal element of \(c \left(X'_i \eta_k X_{i}(\eta_k)\right)^{-1}\) with \(X_{i}(\eta_k)\) denoting a matrix of the columns of \(X\) corresponding to \(\eta_k = 1\), obtaining a prior resembling Zelnner’s g-prior (Zelnner and Goel, 1986). This prior mimics the correlation structure of the data. Denote by \(\beta_{\eta_k}\) the coefficient vector corresponding to the entries of \(\eta_k = 1\), and \(\beta_{\eta_k}\) the coefficient vector corresponding to the entries of \(\eta_k = 0\). Then the prior distribution of \(\beta_{\eta_k}\) can be written as

\[
\pi(\beta_{\eta_k}) \sim N(0, c \left(X'_i \eta_k X_{i}(\eta_k)\right)^{-1}),
\]

where \(c\) is a constant for interval \(I_{g,j}\) and \(D_{g,j}\) denote the set of patients who experienced event \(g\) in this interval. We follow Lee et al. and assign priors for \(g = 1, 2, 3\)
and that \( P[\beta(g_k) = 0] = 1 \). We assume \( \eta_{g,k} \sim \text{Bernoulli}(w_g) \) for all \( k = 1, \ldots, p \) and \( g = 1, 2, 3 \). Formally, the prior of \( \eta_g \) is
\[
\pi(\eta_g | w_g) = \prod_{k=1}^{p} w_{g,k}^{\eta_{g,k}} (1 - w_{g,k})^{1 - \eta_{g,k}} = w_g^{\eta_g} (1 - w_g)^{p - \eta_g},
\]
where \( t_g = \sum_{k=1}^{p} \eta_{g,k} \) is the number of \( \eta_{g,k} = 1 \) in hazard \( g \). We assume a beta prior distribution for each of the \( w_g \)s with parameters \( z_{g1} \) and \( z_{g2} \). Following Brown et al. (1998), we integrate out \( w_g \) to get the marginal prior for \( \pi(\eta_g) \) as
\[
\pi(\eta_g) = \int \pi(\eta_g | w_g) \pi(w_g) dw_g = \frac{\text{Beta}(t_g + z_{g1}, p - t_g + z_{g2})}{\text{Beta}(z_{g1}, z_{g2})},
\]
where \( \text{Beta}(. \ , .) \) is the beta function.

### 2.3. Markov chain Monte Carlo

For posterior inference, we implement a Markov Chain Monte Carlo (MCMC) sampling scheme with stochastic search variable selection (SVSS) applied on the three hazard functions that use add, delete, and swap moves. We employ some of the same sampling schemes of Lee et al. but use different algorithms in our sampling of \( s_g, \lambda_g, \mu_{g,k}, \sigma_{g,k} \) and the sequence of the Gibbs sampler. Lee et al. (2015) used a random scan Gibbs sampler where the probability of adding an additional split point or deleting one of the current split points was a function of the current number of split points and hyperparameters. The remaining moves are the frailty sampler, the hierarchical frailty parameter, 3 baseline hazards, 3 hierarchical baseline hazard parameters and 3 regression parameters, which are assigned equal probabilities from what is left after accounting for birth and death probabilities for the three different baseline hazard functions. Our approach differs in that we do not randomly select what move to perform at each iteration, and instead do all the moves for each of the three hazard functions consecutively. In summary, a generic iteration of the MCMC sampler does the following:

- **Update** \( (\eta_g, \beta_g) \) jointly via a Metropolis step. This is done through add, delete, and swap moves. If \( t_g = 0 \), add one variable automatically and if \( t_g = p \) delete one variable automatically. Otherwise, with probability \( \phi \) perform a swap move. With probability \( 1 - \phi \) perform a Add/Delete move and randomly select one entry of \( \eta_g \), say \( \eta_{g,k} \), and if \( \eta_{g,k} = 1 \), perform a delete move, otherwise perform an add move. The details of these three moves are described below:

1. **Add** if \( \eta_{g,k} = 0 \), set \( \eta_{g,k}^{*} = 1 \) and sample \( \beta_{g,k} | \beta_{g,(-k)}', \eta_{g} \) from a normal distribution. Denoting \( \Sigma = \text{c}(X_{10g}^{T}X_{10g}^{-1}) \), then the proposal distribution has mean \( \mu_{\text{new}} = \Sigma_{k,(-k)}^{-1} \Sigma_{(-k),k}(\eta_{g})_{(-k)} \) and a variance of \( \sigma_{\text{new}}^2 = \Sigma_{k,(-k)}^{-1} \Sigma_{(-k),(-k)} \). Denote \( \Sigma_{k,(-k)} \) as the kth diagonal element of \( \Sigma \). Then set \( \Sigma_{k,(-k)}^{*} \) as the kth row without the kth column entry and \( \Sigma_{(-k),(-k)} \) as the submatrix without row and column k of \( \Sigma \). The proposal \( (\eta_g, \beta_g^{*}) \) is accepted jointly with probability
   \[
   \min \left[ \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_g^{*}, \ldots)N(\beta_{g,k}|\mu_{\text{new}}, \sigma_{\text{new}}^2)\text{Beta}(t_g + 1 + z_{g1}, m_g - t_g - 1 + z_{g2})}{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_g, \ldots)\text{Beta}(t_g + z_{g1}, m_g - t_g + z_{g2})}, 1 \right].
   \]

2. **Delete** if \( \eta_{g,k} = 1 \), set \( \eta_{g,k}^{*} = 0 \) and \( \beta_{g,k}^{*} = 0 \). Denote \( R = \text{c}(X_{10g}^{T}X_{10g}^{-1}) \), \( \mu_{\text{old}} = R_{k,(-k)}^{-1}(\eta_{g})_{(-k)} \) and variance \( \sigma_{\text{old}}^2 = R_{k,k} - R_{k,(-k)}R_{(-k),(-k)}^{-1}R_{(-k),k} \). Then the proposal \( (\eta_g, \beta_g^{*}) \) is accepted with probability
   \[
   \min \left[ \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_g^{*}, \ldots)\text{Beta}(t_g - 1 + z_{g1}, m_g - t_g + 1 + z_{g2})}{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_g, \ldots)\text{Beta}(t_g + z_{g1}, m_g - t_g + z_{g2})}, 1 \right].
   \]

3. **Swap** randomly select one \( \eta_{g,k} = 1 \) and one \( \eta_{g,j} = 0 \) and swap their values, setting \( \eta_{g,k}^{*} = 0 \) and \( \eta_{g,j}^{*} = 1 \). Then set \( \beta_{g,k}^{*} = 0 \) and sample \( \beta_{g,k}^{*} | \eta_{g,k}^{*}, \beta_{g,(-k)}' \) from a normal distribution with mean \( \mu_{\text{new}} \) and variance \( \sigma_{\text{new}} \). These proposed values \( (\eta_g, \beta_g^{*}) \) are accepted with probability
   \[
   \min \left[ \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_g^{*}, \ldots)N(\beta_{g,k}^{*}|\mu_{\text{new}}, \sigma_{\text{new}}^2)}{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_g, \ldots)N(\beta_{g,k}^{*}|\mu_{\text{old}}, \sigma_{\text{old}}^2)}, 1 \right].
   \]

- **Update** \( \beta(g_k) \) via a Metropolis step. \( \beta(g_k) \) is first updated conditionally by sampling \( \beta(g_k) | \eta_g, \beta(g_{(-k)}) \) from a \( N(\mu_{\text{old}}, \sigma_{\text{old}}) \) for all \( k = 1, \ldots, p \) variables currently in hazard \( g \). Then \( \beta(g_k) \) is updated jointly by sampling a proposal \( \beta(g_k) \sim N\left(0, c(X_{10g}^{T}X_{10g}^{-1})\right) \). This provides better mixing in our Markov Chain by moving entries of \( \beta(g_k) \) further from or closer to zero which affects future \( (\beta_g, \eta_g) \) updates.

- **Update** \( \sigma_{g,k}^{2} \) via a Metropolis step in the same manner listed in the supplemental material of Lee et al.

- **Update** \( \gamma_l \) for \( i = 1, \ldots, n \) via a Gibbs step in the same manner listed in the supplemental material of Lee et al.

- **Update** \( \mu_{3g} \) and \( \sigma_{3g}^{2} \) via a Gibbs step in the same manner listed in the supplemental material of Lee et al. for \( g = 1, 2, 3 \).

The posteriors for each are normal and inverse-gamma distributed, respectively.
• Update $\lambda_{g,j}|\lambda_{g,(-j)}$, $s_g$ for $j = 1, \ldots, J_g + 1$ via a Metropolis step. Lee et al. use the first and second derivatives of $\pi(\lambda_g|\text{Data})$ in the $\lambda_g|s_g, \mu_{\lambda_k}, \sigma^2_{\lambda_k}$ proposal distribution. These derivatives involve $\beta_g$, which often has entries that change drastically in magnitude in the SVSS procedure, making the sampling scheme of Lee et al. extremely inefficient. Furthermore, the tuning parameter they use in the proposal distribution for $\lambda_g$ must be tuned for each hazard $g = 1, 2, 3$ extensively to get good Metropolis–Hastings acceptance rates. To avoid these issues, we sampled $\lambda_{g,k}^*|\lambda_{g,(-k)}$ from a $U(\lambda_{g,k}^* - c_g, \lambda_{g,k} + c_g)$ distribution where $\lambda_{g,k}$ is the previous sampled value. This follows the approach of Haneuse et al. (2008), and we use their default value $c_g = .25$ for each $g$, which gives good acceptance rates within our MCMC.

Since the proposal ratio is $\frac{2c_g}{2\hat{s}_g} = 1$, the proposed value $\lambda_{g,k}^*$ is accepted with probability

$$\min \left[ \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|\lambda_g^*, \ldots)N_{J_g+1}(\lambda_g^*|\mu_{\lambda_k}, \sigma^2_{\lambda_k}\Sigma_{\lambda_k})}{L(Y_1, Y_2, \delta_1, \delta_2, X|\lambda_g, \ldots)N_{J_g+1}(\lambda_g|\mu_{\lambda_k}, \sigma^2_{\lambda_k}\Sigma_{\lambda_k})}, 1 \right].$$

• Add a split point to $s_g$ and update $(s_g, \lambda_g)$ jointly via a Metropolis–Hastings–Green step. Propose a new split point on $s_g^* \sim U[0, s_{g,\max}]$ then the new $\lambda_{g,k}^*$ heights created by adding this split point are based on a multiplicative perturbation like in Green (1995) and Lee et al. (2015). That is, if $s^*$ is proposed such that it is in the interval $[s_{g,j-1}, s_{g,j}]$ then the new $\lambda_{g}^*$ values are

$$\lambda_{g,j}^* = \lambda_{g,j} - \frac{s_{g,j} - s^*}{s_{g,j} - s_{g,j-1}} \log \left( \frac{1 - U_g}{U_g} \right),$$

and

$$\lambda_{g,j+1}^* = \lambda_{g,j} + \frac{s_{g,j} - s^*}{s_{g,j} - s_{g,j-1}} \log \left( \frac{1 - U_g}{U_g} \right).$$

$U_g \sim U[0, 1]$ is drawn at every iteration. This is the only difference in our sampler from the sampler of Lee et al. for the Metropolis–Hastings–Green step, as they set $U_g$ as a tuning parameter. This move is accepted with probability

$$\min \left[ \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|s_g^*, \lambda_g^*, \ldots)P(U_g | 1|\alpha)N_{J_g+1}(\lambda_g^*|\mu_{\lambda_k}, \sigma^2_{\lambda_k}\Sigma_{\lambda_k})}{L(Y_1, Y_2, \delta_1, \delta_2, X|s_g, \lambda_g, \ldots)P(U_g | 1|\alpha)N_{J_g+1}(\lambda_g|\mu_{\lambda_k}, \sigma^2_{\lambda_k}\Sigma_{\lambda_k})}, 1 \right].$$

• Delete a split point from $s_g$ and update $(s_g, \lambda_g)$ jointly via a Metropolis–Hastings–Green step. We randomly select one of the current split points that are not $0$ or $s_{g,\max}$ with equal probability and delete it. Assume we delete split point $s_{g,j}$. Following the methods outlined by Green (1995), we have that the multiplicative perturbation is $\phi_{s_{g,j+1}} - \lambda_{g,j} = \frac{1 - U_g}{U_g}$ and that the new height of the interval created by deleting a split point is a compromise of the previous two heights over this interval, defined as

$$\lambda_{g,j}^* = \frac{(s_{g,j} - s_{g,j-1})\lambda_{g,j} + (s_{g,j+1} - s_{g,j})\lambda_{g,j+1}}{s_{g,j+1} - s_{g,j-1}}.$$

We draw a new $U_g \sim U[0, 1]$ as in Green (1995), rather than setting it to be a hyperparameter as in Lee et al. Now we accept the vector $(s_g^*, \lambda_g^*)$ jointly with probability

$$\min \left[ \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|s_g^*, \lambda_g^*, \ldots)P(U_g | 1|\alpha)N_{J_g+1}(\lambda_g^*|\mu_{\lambda_k}, \sigma^2_{\lambda_k}\Sigma_{\lambda_k})}{L(Y_1, Y_2, \delta_1, \delta_2, X|s_g, \lambda_g, \ldots)P(U_g | 1|\alpha)N_{J_g+1}(\lambda_g|\mu_{\lambda_k}, \sigma^2_{\lambda_k}\Sigma_{\lambda_k})}, 1 \right].$$

### 2.4. Model determination

To determine the final model on which to draw inference on our treatment and covariate effects on survival, we calculated the marginal posterior probabilities of inclusion for each variable $k = 1, \ldots, p$ in hazard function $g$ as the proportion of $\eta_{g,k} = 1$ in the posterior sample. We then selected the variables in hazard $g$ with marginal posterior probabilities of inclusion (PPI) greater than $\tau_g \in (0, 1)$. Formally, variable $k = 1, \ldots, p$ in hazard $g = 1, 2, 3$ was included in the final model if

$$\Pr[\eta_{g,k} = 1|\text{Data}] > \tau_g.$$

(6)

Thus, one must specify $\tau_g$ for each of the three hazard functions. We decided the optimum $\tau^* = (\tau_1^*, \tau_2^*, \tau_3^*)$ vector based on the deviance information via the DIC–$\tau_g$ procedure, defined as follows. Recall that the deviance information criterion of Spiegelhalter et al. (2002) for $\beta$ is

$$\text{DIC} = -2 \log(L(\hat{\beta})) + 2p_{\text{DIC}},$$

(7)
where \( p_{DIC} = 2 \left( \log(L(\hat{\beta})) - \frac{1}{B} \sum_{b=1}^{B} \log(L(\hat{\beta}^b)) \right) \), \( \hat{\beta}^b \) is the sampled value of \( \beta \) for iteration \( b \) and \( \hat{\beta} \) is the posterior mean of \( \beta = (\beta_1, \beta_2, \beta_3) \). Since our likelihood contains many nuisance parameters \( (\gamma, j_b, s_g, \lambda_g, \ldots) \), Spiegelhalter et al. (2002) recommend to plug-in posterior means of these parameters in determining the DIC for a given \( \beta \) vector. We use the DIC to select the optimal model via the following algorithm, which we call the DIC-\( \tau_g \) procedure.

1. Calculate the posterior mode or median of \( J_g \) for \( g = 1, 2, 3 \), depending on the shape of the posterior distribution. The posterior mode is calculated if the mode had a greater posterior density than all other possible values for \( J_g \) while the posterior median of \( J_g \) is calculated if there were several values of \( J_g \) that occurred with about the same frequency.

2. Compute the posterior mean nuisance parameters \( \gamma = (\gamma_1, \ldots, \gamma_\nu) \), \( s_g | j_b \) and \( \lambda_g | j_b \) for \( g = 1, 2, 3 \).

3. Perform a three-dimensional grid search for the optimum \( \tau = (\tau_1, \tau_2, \tau_3) \) over values \( \tau_g = 0.05, 0.1, \ldots, 0.9 \) that give different models based on the threshold criterion (6) for \( g = 1, 2, 3 \).

   (a) For a given \( \tau = (\tau_1, \tau_2, \tau_3) \), find the \( \eta^\tau = (\eta_1^\tau, \eta_2^\tau, \eta_3^\tau) \) that satisfies (6).

   (b) Sample \( \beta(\eta^\tau) \) 10,000 times via a Metropolis–Hastings algorithm using the prior distribution (5), the same \( c \) value used in the variable selection sampler and the posterior quantities calculated in steps 2 and 3.

   (c) Discard the first half of the sample and save the DIC.

4. The vector \( \tau^\tau = (\tau_1^\tau, \tau_2^\tau, \tau_3^\tau) \) chosen by the DIC-\( \tau_g \) procedure is the \( \tau^\tau \) that produces the smallest DIC. For models with DIC values that differ by 1 or less from the model with the smallest DIC, the most parsimonious model was selected as the final model. A small decrease in DIC here indicates that including more variables does not increase the information criterion significantly over the more parsimonious model.

5. The final model includes the variables that satisfy (6) for the optimal \( \tau^\tau = (\tau_1^\tau, \tau_2^\tau, \tau_3^\tau) \).

After the DIC-\( \tau_g \) procedure selects \( \eta^\tau = (\eta_1^\tau, \eta_2^\tau, \eta_3^\tau) \) based on \( \tau^\tau = (\tau_1^\tau, \tau_2^\tau, \tau_3^\tau) \), we resample \( \beta(\eta^\tau) \) 100,000 times using the prior (5) \( \beta(\eta^\tau) \sim N \left( 0, c \left( X_{(\eta^\tau)}^T X_{(\eta^\tau)} \right)^{-1} \right) \) and do posterior inference after discarding the first 50,000 MCMC samples. While we could compute the DIC for all \( 19^3 \) different \( \tau \) vectors, small changes in \( \tau_g \) of .05 or .10 may not add any additional variables to hazard \( g \), which means that two different \( \tau_g \) vectors could lead to the same \( \eta_g \). This reduces the number of \( \tau \) values that we need to try based on our marginal posterior probabilities of inclusion. On the other hand, increasing \( \tau_g \) by .05 could add more than one variable to hazard \( g \). Since our case study has 11 variables, there are at most \( 11^3 \) different \( \tau \) vectors that choose a unique subset of variables. We also did not consider a finer grid \( \tau_g \) .

3. Simulation study

We performed a simulation study under six different scenarios to see how well our method selected important variables in the posterior probabilities of inclusion, and how accurately the DIC-\( \tau_g \) method selected the final model. We used the simID function from the SemiCompRisks package to simulate patient data under the Weibull semi-competing risks model, to examine how well our method performs when the baseline hazards are not truly piecewise exponential. For all simulation studies, we used \( \kappa_1 = .05, \kappa_2 = .01, \kappa_3 = \kappa_1, \alpha_1 = .8, \alpha_2 = 1.1, \alpha_3 = .9 \) and \( \theta = .5 \) which are the values given in the example for the simID documentation. Additionally, we set the censoring time for patients at 2000 days. For each simulated data set, we used the actual patient covariate matrix from our data to mimic a correlated data structure, rather than generating patient covariates independently. We performed the simulation study as follows for each simulation:

1. Run two MCMC chains using the package SCRSELECT (Chapple, 2016) for 100,000 iterations with dispere starting values. One chain starts with all the variables included, the other starts with no variables included.

2. Discard the first 40,000 samples and combine the two chains, saving the marginal posterior probabilities of inclusion as well as the posterior means of nuisance parameters for the baseline hazard and frailties \( (s_g | j_b, \lambda_g | j_b, \gamma) \), as outlined in the DIC-\( \tau_g \) procedure.

3. For \( \tau_g = (.05, .1, \ldots, .9) \) and \( g = 1, 2, 3 \) run an MCMC as described in Section 2.4 for the DIC-\( \tau_g \) procedure using the quantities from step 2 while skipping any \( (\tau_1, \tau_2, \tau_3) \) vectors that do not change the variables included in each hazard.

4. Find the combination \( \tau_1, \tau_2, \tau_3 \) that produces the smallest DIC. When there are DIC values within 1 of each other for different models, take the one with less variables included (higher \( \tau \) values).

5. Assess the results of the simulation.

We present results averaged over 100 replicated data sets for each simulation scenario (Table 1). For each of the seven scenarios we kept the last two entries of the \( \beta_g \) vectors constant and did not allow these to be selected from the model. These correspond to the propensity score and treatment effect of the XRT modality, respectively. These last two entries are set to \((.15, -.15)\) for the hazard of effusion, \((.3, -.2)\) for the hazard of death before effusion, and \((.01, -.05)\) for the hazard of death after effusion. These values were selected in part to mimic the case study results. This allowed selection on 11 out
of the 13 covariates in each hazard. Table 1 displays a summary of the six simulation scenarios that we considered for the DIC-τg method, where we only list the true coefficients of variables that were considered for variable selection.

Since the gth hazard of the model includes \( e^{\psi g} \beta_g \), entries of magnitude greater than 1 would be too unrealistic and easy to pick out in simulations. Instead we focused on coefficients having magnitudes of .3 to .9, with a few challenging coefficient values having magnitude less than .3. We used the same hyperparameter settings used in the application for the simulations for comparison. We set the variable selection hyperparameters \( c = 20 \) in (5) and \((z_{g1}, z_{g2}) = (4, 1.6)\) for \( g = 1, 2, 3 \). We set \((a_g, b_g) = (\psi, \omega) = (7, 7)\) as a non-informative prior for \( \sigma_{g2}^{-2} \) and \( \theta^{-1} \). We set \( a_g = 3 \) corresponding to 3 expected split points in each baseline hazard, and the maximum possible number of split points to 10.

Before examining how well the DIC-τg procedure picked out the final model, we examined marginal posterior probability of inclusion results for the simulations in terms of included variables and excluded variables. In most hazards and scenarios, the mean marginal probability of inclusion for the included variables is greater than .9. There are several exceptions to this that can be explained by the simulation settings. In scenario two hazard three, the mean marginal probability of inclusion for the two variables is .797, but the median is 1.00. This is due to the small true simulation value for variable 8 of .2, which is much harder to detect in the SVSS.

In Table 2, we display a summary of the results of the DIC-τg procedure showing the mean number of false positives, false negatives, and the mean probability of a correct decision. We denote NFP= mean number of false positives and \( N_0 = \) total number of variables with true \( \beta_{g,k} \neq 0 \) for all three hazards combined. Similarly, NFN= mean number of false negatives and \( N_+ = \) total number of variables with true \( \beta_{g,k} = 0 \) for all three hazards combined. We denote PCD as the probability of a correct decision of whether or not a variable is included. We also computed a statistic resembling the area under the curve (AUC) where we plotted the true positive and false positive rates of all 100 simulations, connected these points in the same manner as the ROC curve and estimated the area under this curve. This statistic, which we denote AUC* operates in the same manner as the traditional AUC statistic where larger values indicate a better classifier and values near .5 indicate a poor classification method. We used this since it evaluated the DIC-τg procedure rather than just the SVSS portion like the traditional AUC score would. The AUC* and AUC statistics were undefined and omitted for scenario 7 since there were no true positives.

In all seven scenarios the DIC-τg method correctly determined variable inclusion on average with probability .76 to .82 and the AUC* scores for the simulations were all greater than .88, showing that the DIC-τg procedure performs well in determining variable inclusion. The DIC-τg method correctly identified the true model in at least one simulation for scenarios 2, 3 and 7. The majority of the incorrect decisions about variable inclusion were false positives, and false negatives did not even occur in scenarios 3 or 7 which are the sparse and null scenarios. Each scenario had about 6 false positives on average, with the second scenario having the highest mean false positive rate. The false positives are partially due to correlations among the covariates, and not imposing enough separation in the marginal posterior probabilities of inclusion, which can be mitigated through careful adjustments to the hyperparameter \( c \).

4. Application

4.1 Data

This observational data set came from M.D. Anderson Cancer Center in Houston, TX. It consists of 470 esophageal cancer patients who had one of two XRT modalities for radiation: 3-Dimensional Conformal Radiation Therapy (3D-CRT) or Intensity Modulated Radiation Therapy (IMRT). Some patients also received induction chemotherapy. Patients were followed from the end of radiation therapy until death or censoring. The dates pleural or pericardial effusion occurred were recorded during this time. The two path diagrams in Fig. 1 describe the semi-competing risks structure and enumerate the different patient outcomes. In each of these paths, we display the total number of patients who experienced each event followed by the numbers of patients who received 3D-CRT and IMRT, respectively, in each transition. Pleural effusion occurs much more frequently in these patients than pericardial effusion. Patients who experienced either Pericardial or Pleural effusion died afterwards in over 75% of all cases (no censoring).

We did separate analyses for the pleural effusion and pericardial effusion path structures shown in Fig. 1. Each patient had 20 baseline covariates, including individual characteristics and characteristics related to their tumor. Several patient
Table 2
Simulation study of the DIC-τg procedure.

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>NFP/N0</th>
<th>NFN/N0+</th>
<th>PCD</th>
<th>AUC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.92/20</td>
<td>0.48/13</td>
<td>0.806</td>
<td>0.972</td>
</tr>
<tr>
<td>2</td>
<td>7.29/23</td>
<td>0.50/10</td>
<td>0.764</td>
<td>0.957</td>
</tr>
<tr>
<td>3</td>
<td>5.94/28</td>
<td>0.00/5</td>
<td>0.820</td>
<td>1.000</td>
</tr>
<tr>
<td>4</td>
<td>5.75/21</td>
<td>0.28/12</td>
<td>0.817</td>
<td>0.935</td>
</tr>
<tr>
<td>5</td>
<td>6.21/24</td>
<td>0.45/9</td>
<td>0.798</td>
<td>0.928</td>
</tr>
<tr>
<td>6</td>
<td>5.66/13</td>
<td>1.98/20</td>
<td>0.769</td>
<td>0.881</td>
</tr>
<tr>
<td>7</td>
<td>5.88/33</td>
<td>0.00/0</td>
<td>0.822</td>
<td>–</td>
</tr>
</tbody>
</table>

Fig. 1. Path diagrams for the three event types along with how many patients (Total, 3D-CRT, IMRT) experienced each for pleural and pericardial effusion. 212 patients received 3D-CRT and 258 patients received IMRT.

Covariates had high pairwise correlations or a small number of patients. Because this caused severe MCMC convergence problems, these variables were excluded from the analysis. The covariates considered were XRT modality, induction chemotherapy, age, BMI, asthma, diabetes, smoking status, and also binary variables for whether or not a patient had an adeno histology, a good KPS performance score, stage 3–4 cancer, and for tumor location 3 (lower/distal) or tumor location 2 (middle) vs tumor location 1 (upper). Because some patients did not have smoking status, BMI, or tumor histology information, they were removed from the data, leaving 470 patients. Clinicians were primarily interested in comparing the effects of the two modalities on pleural effusion, pericardial effusion, and death, but it was unclear what covariates should be included in each hazard function. We addressed this concern through our variable selection method for the three hazard functions. We included IMRT status in each hazard to evaluate radiation therapy modality effects on each of the three hazard functions. Because the data were observational and not randomized, we included the propensity score of each patient as a covariate always in each hazard to correct for some of the bias introduced. The propensity score was estimated by

\[ \text{logit}^{-1}(Pr(\text{IMRT} | X)) = \beta_0 + \beta_1 \cdot \text{Hypertension} + \beta_2 \cdot \text{Differentiation} + \beta_3 \cdot \text{AdenoHistology}, \]

the fitted logistic regression model of these covariates on IMRT status. We did not allow the propensity score or IMRT status to be removed from the model. Thus, \( X_0(\eta_g) \) always contains a column with individual propensity scores and XRT status. Likewise, \( \beta_0(\eta_g) \) always contains an entry corresponding to the estimated coefficient for the propensity score and XRT status in each hazard function.

4.2. Hyperparameter settings

For the hyperparameters, we set \( c = 20 \) in (5) for both case studies. We found that varying \( c \) greatly changed the posterior inclusion probabilities, with larger values of \( c \) forcing more variables out of the model and smaller values keeping more variables in the model. Smith and Kohn (1996) recommend a value between 10 and 100. We obtained vague priors on \( w_g \) as in Brown et al. (1998) by imposing the constraint \( z_g1 + z_g2 = 2 \), corresponding to a prior effective sample size of 2 for the beta prior of the covariate inclusion probability, with some desirable mean percentage of inclusion. We set \( (z_g1, z_g2) = (4.6, 1.6) \) for \( g = 1, 2, 3 \), corresponding to an expectation of 2 to 3 variables being included in each hazard function. We set \( (\alpha, \beta_g) = (\psi, \omega) = (.7, .7) \) as a non-informative prior for \( \sigma_{\eta_g}^{-2} \) and \( \theta^{-1} \). This corresponds to .025 and .975 quantiles of .23 and 155.61 for \( \sigma_{\eta_g}^{-2} \) and \( \theta \). We set \( \alpha_0 = 3 \) corresponding to 3 expected split points in each baseline hazard, and the maximum possible number of split points to 10, which neither data set reached for any hazard. Hyperparameter settings for the ICAR formulation are discussed in the Appendix. For both data sets, two chains were run with two different starting values for the three hazard \( \beta_g \) vectors; the first with all \( \eta_{g,k} = 1 \) and \( \beta_{g,k} = 1 \) for all \( k = 1, \ldots, p \) and \( g = 1, 2, 3 \) the second with \( \eta_{1,1} = 1, \eta_{2,5} = 1 \) and \( \eta_{3,7} = 1 \) (these entries were randomly chosen) and all non-zero coefficients set to \( \beta_{g,k} = -1 \).

4.3. Case study results: Pleural effusion

The two chains were run for 100,000 iterations with the first 40,000 discarded as burn-in. The scale reduction factors for the \( \beta_g \) coefficients had estimates and upper confidence interval estimates below 1.01, which indicates good convergence.
Fig. 2. Pleural effusion results: Combined marginal posterior probabilities of inclusion and the DIG-$\tau$ cutoff thresholds $\tau^*_g$.

for both chains along with the trace plots for each (Gelman and Rubin, 1992). Furthermore, the posterior probabilities of inclusion for the two chains did not differ substantially, with the biggest difference of 4.46%. The correlation of the marginal posterior probabilities of inclusion for the two chains was all above .99 for all three hazards. The posterior samples of both chains were then combined and the resulting marginal posterior probabilities of inclusion can be seen graphically in Fig. 2 and numerically in Table 5.

The optimal model based on the DIC-$\tau_g$ had a DIC of 1050.845 and indicated that $(\tau^*_1, \tau^*_2, \tau^*_3) = (.45, .75, .45)$ was the appropriate vector of upper bounds to produce the optimum model. This model had 2, 7 and 3 variables in hazards 1, 2 and 3, respectively. After finding the optimal model, the Gibbs sampler was rerun using 100,000 iterations. Table 3 displays the fitted optimal model with 95% posterior credible intervals for the hazard ratio $\exp[\beta_{g,k}]$ and $P = P[\beta_{g,k} > 0 | \text{Data}]$ for the $k$th variable included in each hazard. That is, $P$ is the posterior probability that the effect of covariate $k$ increases the hazard of event $g$, so that large values of $P$, above .95 or .99, correspond to a harmful effect of the covariate.

We see from the final model chosen by the DIC-$\tau_g$ procedure that patients with IMRT radiation had significantly reduced hazards of pleural effusion and death before pleural effusion compared to 3D-CRT ($P < .01$). Older age greatly increased the hazard of pleural effusion ($P = .99$) and patients with Stage 3–4 cancer had an increased hazard of death after pleural effusion ($P = .99$). Patients with diabetes and adeno histology had significantly increased hazards of death before pleural effusion ($P < .99$) and patients with a good KPS score had a significantly reduced hazard of death before pleural effusion ($P = .97$) but patients with stage 3–4 cancer had a significantly increased risk of death before pericardial effusion (Table 4). This could be attributed to the number of patients who died without pericardial effusion ($n = 285$) compared to those who died without pleural effusion ($n = 197$).

4.4. Case study results: Pericardial effusion

The two chains were run for 100,000 iterations with the first 40,000 discarded as burn-in. The scale reduction factors for the $\beta_g$ coefficients had estimates and upper confidence interval estimates below 1.01, which indicates good convergence for both chains along with the trace plots for each (Gelman and Rubin, 1992). The posterior probabilities of inclusion for the two chains did not differ substantially, with the biggest difference 2.80%. The correlation of the marginal posterior probabilities of inclusion for the two chains was .994, .995 and .972 for the three hazards, respectively. The posterior samples of both
Table 3
Pleural effusion case study: Posterior quantities for the model chosen by DIC-$\tau_g$ procedure. Notation: $P = P[\beta_{g,k} > 0|\text{Data}]$, HR (CI) = posterior hazard ratio and 95% credible interval. $g = 1, 2, 3$ correspond to the three hazard functions and $k = 1, \ldots, p$ index to the covariates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$h_1$: Effusion</th>
<th>$h_2$: Death w/o Effusion</th>
<th>$h_3$: Death after Effusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (CI)</td>
<td>$P$</td>
<td>HR (CI)</td>
</tr>
<tr>
<td>Asthma</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Diabetes</td>
<td>–</td>
<td>–</td>
<td>1.18 (1.03–1.33)</td>
</tr>
<tr>
<td>Stage 3–4</td>
<td>–</td>
<td>–</td>
<td>1.13 (1.00–1.29)</td>
</tr>
<tr>
<td>Location 2</td>
<td>–</td>
<td>–</td>
<td>.86 (1.74–1.01)</td>
</tr>
<tr>
<td>Location 3</td>
<td>–</td>
<td>–</td>
<td>.87 (1.73–1.04)</td>
</tr>
<tr>
<td>Age</td>
<td>1.18 (1.04–1.36)</td>
<td>.99</td>
<td>–</td>
</tr>
<tr>
<td>Smoker</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>BMI</td>
<td>.90 (.78–1.02)</td>
<td>.05</td>
<td>.88 (1.76–1.00)</td>
</tr>
<tr>
<td>Induction Chemo</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Good KPS</td>
<td>–</td>
<td>–</td>
<td>.80 (1.71–.91)</td>
</tr>
<tr>
<td>Adeno Histology</td>
<td>–</td>
<td>–</td>
<td>1.21 (1.02–1.42)</td>
</tr>
<tr>
<td>Propensity Score</td>
<td>.97 (.86–1.12)</td>
<td>.34</td>
<td>.92 (1.81–1.05)</td>
</tr>
<tr>
<td>IMRT</td>
<td>.84 (.74–.96)</td>
<td>&lt;.01</td>
<td>.85 (1.75–.97)</td>
</tr>
</tbody>
</table>

Fig. 3. Pericardial Effusion results: Combined marginal posterior probabilities of inclusion and the DIG-$\tau_g$ cutoff thresholds $\tau_g^*$. Chains were combined and the resulting marginal posterior probabilities of inclusion can be seen graphically in Fig. 3 and numerically in Table 7.

Variables above the thresholds in Fig. 3 are included in the final model. The optimal model based on the DIC-$\tau_g$ procedure had a DIC of 895.37, with $\tau^* = (.55, .65, .45)$. This model produced 7, 6, and 2 variables in hazards 1, 2, and 3, respectively. After finding the optimal model, the Gibbs sampler was rerun using 100,000 iterations and the first 50,000 samples were discarded as burn-in. Table 4 displays the fitted optimal model with 95% posterior credible intervals for the hazard ratio $\exp[\beta_{g,k}]$ and $P = P[\beta_{g,k} > 0|\text{Data}]$, the posterior probability that a larger value of a covariate is more hazardous, for the $k$th variable included in each hazard $g$. We used the same hyperparameter settings for this model as for the pleural effusion model, namely $c = 20$, $z_g^1 = .4$ and $z_g^2 = 1.6$.

Table 4 shows that patients with IMRT had significantly decreased hazards of pericardial effusion and death before pericardial effusion, compared to those who received 3D-CRT ($P < .01$ for both hazards). All other 95% credible intervals for
the hazard of pericardial effusion contain 1, so even though Age and Adeno histology had coefficients that indicated the hazard decreased with these variables ($P = .03$ and $.04$) they only decreased the hazard of pericardial effusion slightly. The final model for the hazard of death before pericardial effusion has clearer trends, namely patients with Stage 3–4 cancer, Diabetes, or an Adeno Histology had significantly increased hazards of death without prior pericardial effusion ($P > .99$), evidenced by the 95% credible intervals. IMRT was associated with a significant reduction in the hazard of death prior to pericardial effusion as did patients with an increased BMI ($P < .01$). No variables included in the hazard of death after pericardial effusion changed the hazard significantly, indicating that patient baseline covariates and radiation therapy have little effect on survival following pericardial effusion.

### 4.5. Sensitivity analyses

We assessed how sensitive the marginal posterior probabilities of inclusion and model selected via the DIC-$\tau_g$ procedure were to the hyperparameters $c$, $z_{g1}$, and $z_{g2}$. Because we wanted to distinguish among variables, we did not consider $c$ and ($z_{g1}$, $z_{g2}$) values that force all variables in or out of the model. For $c$, first we found a range of appropriate values for the data that did not include or exclude all the variables. Varying $c$ within this range of appropriate values did not have a large impact on the final models. We tested sensitivity of our model to $c$ by looking at $c = 20$ and $c = 30$, in conjunction with two $\psi_g$ hyperparameter combinations of ($z_{g1}$, $z_{g2}$) = (4, 1.6) as in the primary analysis and ($z_{g1}$, $z_{g2}$) = (1.1, 1.9) corresponding to an expectation of 1 or 2 variables included in each hazard. We found that our method was sensitive to the choices of $c$ and ($z_{g1}$, $z_{g2}$) in shrinking the marginal posterior probabilities of inclusion, but that the final model chosen by the DIC−$\tau_g$ procedure did not differ much. This helps address concerns about the sensitivity of the model to final inferences based on treatments and covariates. For sensitivity to the survival hyperparameters, we refer to the supplemental material of Lee et al. (2015).

To assess sensitivity to the pleural effusion case study, Table 5 shows the marginal posterior probability of inclusion in each hazard function for each of the four different settings of $c$ and $z_{g1}$. For each hazard function, increasing $c$ had a greater shrinkage impact than decreasing $z_{g1}$ for the pleural effusion data.

In general, the ordering of importance of the variables by marginal Posterior Probabilities of Inclusion (PPIs) does not change much. The primary difference in ordering comes for variables that have lower marginal PPIs relative to the variables with the highest PPIs in each hazard, because these variables tend to have marginal PPIs very close to each other, which increases the likelihood of different orderings by chance alone. As mentioned previously, different values of $c$ and $z_{g1}$ induce different degrees of sparsity, so we wanted to know if our final model selected via the DIC−$\tau_g$ procedure was sensitive to $c$ and $z_{g1}$. Table 6 displays the final models selected by the DIC−$\tau_g$ procedure for each of the four models considered. The same two variables, Age and BMI, were chosen in the final model for all but one hyperparameter setting. In all four settings, the hazard of pleural effusion was decreased significantly for patients who had IMRT radiation ($P < .01$) and increased for patients with older age ($P \geq .99$). For the hazard of death before pleural effusion, the variables Diabetes, Stage 3–4, Location 2, Location 3, BMI, good KPS and adeno Histology were included for all four hyperparameter settings. In all four settings, the hazard of death before pleural effusion was reduced significantly for patients with IMRT radiation ($P \leq .01$). All four hyperparameter settings had Stage 3–4 in the final model for the hazard of death after pleural effusion, and in each model, having stage 3–4 cancer significantly increased a patient's hazard of death after pleural effusion ($P \geq .99$). In models including more than just Stage 3–4 cancer in $h_3$, all other credible intervals contained 1, indicating that Stage 3–4 cancer is the key driver for this hazard.

Next, we similarly assessed the sensitivity of the pericardial effusion data set to the variable selection hyperparameters. For the first two hazard functions, increasing $c$ caused more shrinkage than decreasing $z_{g1}$ while the third hazard function achieved more shrinkage by lowering $z_{g1}$. The four different settings considered had good agreement for the order of

---

**Table 4**

Percardial Effusion case study: Posterior quantities for the model chosen by DIC-$\tau_g$ procedure. Notation: \( P = P(\beta_{g,k} > 0 \mid \text{Data}) \), HR (CI) = posterior hazard ratio and 95% credible interval. \( g = 1, 2, 3 \) correspond to the three hazard functions and \( k = 1, \ldots, p \) index to the covariates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$h_1$: Effusion HR (CI)</th>
<th>$h_2$: Death w/o Effusion HR (CI)</th>
<th>$h_3$: Death after Effusion HR (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$P$</td>
<td>$P$</td>
</tr>
<tr>
<td>Asthma</td>
<td>.91 (.75–1.09)</td>
<td>.15</td>
<td>.11 (.90–1.36)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>.85 (.72–1.06)</td>
<td>.03</td>
<td>.85</td>
</tr>
<tr>
<td>Stage 3–4</td>
<td>.16 (.95–1.42)</td>
<td>.93</td>
<td>.93</td>
</tr>
<tr>
<td>Location 2</td>
<td>.12 (.94–1.54)</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>Location 3</td>
<td>.85 (.76–1.09)</td>
<td>.16</td>
<td>.85 (.76–.95)</td>
</tr>
<tr>
<td>Age</td>
<td>.85 (.72–1.06)</td>
<td>.03</td>
<td>.85</td>
</tr>
<tr>
<td>Smoker</td>
<td>.89 (.74–.106)</td>
<td>.09</td>
<td>.89</td>
</tr>
<tr>
<td>BMI</td>
<td>.91 (.76–1.09)</td>
<td>.16</td>
<td>.85 (.76–.95)</td>
</tr>
<tr>
<td>Good KPS</td>
<td>.83 (.68–1.03)</td>
<td>.04</td>
<td>.83 (.75–.92)</td>
</tr>
<tr>
<td>Adeno Histology</td>
<td>1.07 (.89–1.30)</td>
<td>.75</td>
<td>1.00 (.87–1.17)</td>
</tr>
<tr>
<td>IMRT</td>
<td>.73 (.62–.87)</td>
<td>&lt;.01</td>
<td>.80 (.72–.89)</td>
</tr>
</tbody>
</table>

---

**Table 5**

Pericardial Effusion case study: Sensitivity of the DIC−$\tau_g$ procedure to the posterior quantities for the model chosen by DIC−$\tau_g$ procedure. Notation: $P(\beta_{g,k} > 0 \mid \text{Data})$, HR (CI) = posterior hazard ratio and 95% credible interval. \( g = 1, 2, 3 \) correspond to the three hazard functions and \( k = 1, \ldots, p \) index to the covariates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$h_1$: Effusion HR (CI)</th>
<th>$h_2$: Death w/o Effusion HR (CI)</th>
<th>$h_3$: Death after Effusion HR (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P$</td>
<td>$P$</td>
<td>$P$</td>
</tr>
<tr>
<td>Asthma</td>
<td>.91 (.75–1.09)</td>
<td>.15</td>
<td>.11 (.90–1.36)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>.85 (.72–1.06)</td>
<td>.03</td>
<td>.85</td>
</tr>
<tr>
<td>Stage 3–4</td>
<td>.16 (.95–1.42)</td>
<td>.93</td>
<td>.93</td>
</tr>
<tr>
<td>Location 2</td>
<td>.12 (.94–1.54)</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>Location 3</td>
<td>.85 (.76–1.09)</td>
<td>.16</td>
<td>.85 (.76–.95)</td>
</tr>
<tr>
<td>Age</td>
<td>.85 (.72–1.06)</td>
<td>.03</td>
<td>.85</td>
</tr>
<tr>
<td>Smoker</td>
<td>.89 (.74–.106)</td>
<td>.09</td>
<td>.89</td>
</tr>
<tr>
<td>BMI</td>
<td>.91 (.76–1.09)</td>
<td>.16</td>
<td>.85 (.76–.95)</td>
</tr>
<tr>
<td>Good KPS</td>
<td>.83 (.68–1.03)</td>
<td>.04</td>
<td>.83 (.75–.92)</td>
</tr>
<tr>
<td>Adeno Histology</td>
<td>1.07 (.89–1.30)</td>
<td>.75</td>
<td>1.00 (.87–1.17)</td>
</tr>
<tr>
<td>IMRT</td>
<td>.73 (.62–.87)</td>
<td>&lt;.01</td>
<td>.80 (.72–.89)</td>
</tr>
</tbody>
</table>
Table 5
Sensitivity analysis of prior hyperparameter effects on marginal PPI for the Pleural Effusion case study. Notation: \( c \) is a parameter in the prior of \( \beta_{\eta_k} \), given in (5) and \((z_{g1}, z_{g2})\) are the prior hyperparameters on \( w_g \), the probability that \( \eta_{g,k} = 1 \).

<table>
<thead>
<tr>
<th>Variable</th>
<th>( c = 20 )</th>
<th>( c = 30 )</th>
<th>( c = 20 )</th>
<th>( c = 30 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((z_{g1}, z_{g2}) =)</td>
<td>(4, 1.6)</td>
<td>(1.1, 1.9)</td>
<td>(1.1, 1.9)</td>
<td>(1.1, 1.9)</td>
</tr>
<tr>
<td>( h_1: Effusion ) before Death</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>.378</td>
<td>.184</td>
<td>.207</td>
<td>.080</td>
</tr>
<tr>
<td>Diabetes</td>
<td>.373</td>
<td>.174</td>
<td>.199</td>
<td>.073</td>
</tr>
<tr>
<td>Stage 3–4</td>
<td>.352</td>
<td>.168</td>
<td>.191</td>
<td>.071</td>
</tr>
<tr>
<td>Location 2</td>
<td>.369</td>
<td>.174</td>
<td>.201</td>
<td>.077</td>
</tr>
<tr>
<td>Location 3</td>
<td>.426</td>
<td>.209</td>
<td>.227</td>
<td>.091</td>
</tr>
<tr>
<td>Age</td>
<td>.759</td>
<td>.629</td>
<td>.509</td>
<td>.361</td>
</tr>
<tr>
<td>Smoker</td>
<td>.373</td>
<td>.181</td>
<td>.194</td>
<td>.074</td>
</tr>
<tr>
<td>BMI</td>
<td>.600</td>
<td>.378</td>
<td>.343</td>
<td>.165</td>
</tr>
<tr>
<td>Induction Chemo</td>
<td>.345</td>
<td>.166</td>
<td>.191</td>
<td>.071</td>
</tr>
<tr>
<td>Good KPS</td>
<td>.430</td>
<td>.253</td>
<td>.242</td>
<td>.111</td>
</tr>
<tr>
<td>Adeno Histology</td>
<td>.364</td>
<td>.176</td>
<td>.197</td>
<td>.071</td>
</tr>
<tr>
<td>( h_2: Death ) w/o Effusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>.889</td>
<td>.760</td>
<td>.779</td>
<td>.543</td>
</tr>
<tr>
<td>Stage 3–4</td>
<td>.833</td>
<td>.712</td>
<td>.731</td>
<td>.507</td>
</tr>
<tr>
<td>Location 2</td>
<td>.801</td>
<td>.631</td>
<td>.686</td>
<td>.421</td>
</tr>
<tr>
<td>Location 3</td>
<td>.760</td>
<td>.583</td>
<td>.638</td>
<td>.400</td>
</tr>
<tr>
<td>Age</td>
<td>.643</td>
<td>.479</td>
<td>.562</td>
<td>.330</td>
</tr>
<tr>
<td>Smoker</td>
<td>.599</td>
<td>.412</td>
<td>.500</td>
<td>.272</td>
</tr>
<tr>
<td>BMI</td>
<td>.805</td>
<td>.621</td>
<td>.667</td>
<td>.400</td>
</tr>
<tr>
<td>Induction Chemo</td>
<td>.607</td>
<td>.419</td>
<td>.497</td>
<td>.270</td>
</tr>
<tr>
<td>Good KPS</td>
<td>.968</td>
<td>.922</td>
<td>.896</td>
<td>.713</td>
</tr>
<tr>
<td>Adeno Histology</td>
<td>.821</td>
<td>.623</td>
<td>.717</td>
<td>.466</td>
</tr>
<tr>
<td>( h_3: Death ) ) after Effusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>.393</td>
<td>.273</td>
<td>.296</td>
<td>.169</td>
</tr>
<tr>
<td>Diabetes</td>
<td>.361</td>
<td>.223</td>
<td>.251</td>
<td>.136</td>
</tr>
<tr>
<td>Stage 3–4</td>
<td>.577</td>
<td>.451</td>
<td>.443</td>
<td>.292</td>
</tr>
<tr>
<td>Location 2</td>
<td>.459</td>
<td>.321</td>
<td>.345</td>
<td>.195</td>
</tr>
<tr>
<td>Location 3</td>
<td>.476</td>
<td>.325</td>
<td>.347</td>
<td>.189</td>
</tr>
<tr>
<td>Age</td>
<td>.418</td>
<td>.293</td>
<td>.317</td>
<td>.186</td>
</tr>
<tr>
<td>Smoker</td>
<td>.359</td>
<td>.238</td>
<td>.261</td>
<td>.146</td>
</tr>
<tr>
<td>BMI</td>
<td>.392</td>
<td>.254</td>
<td>.277</td>
<td>.149</td>
</tr>
<tr>
<td>Induction Chemo</td>
<td>.357</td>
<td>.225</td>
<td>.263</td>
<td>.134</td>
</tr>
<tr>
<td>Good KPS</td>
<td>.417</td>
<td>.283</td>
<td>.301</td>
<td>.168</td>
</tr>
<tr>
<td>Adeno Histology</td>
<td>.405</td>
<td>.264</td>
<td>.291</td>
<td>.151</td>
</tr>
</tbody>
</table>

Table 6
Sensitivity Analysis of prior hyperparameter effects on the model selected in the Pleural Effusion case study. Notation: \( c \) is a dispersion parameter in the prior of \( \beta_{\eta_k} \), given in (5) and \((z_{g1}, z_{g2})\) are the beta prior hyperparameters on \( w_g \), the probability that \( \eta_{g,k} = 1 \) for \( k = 1, \ldots, p \).

<table>
<thead>
<tr>
<th>( c )</th>
<th>((z_{g1}, z_{g2}))</th>
<th>Included Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>(.4, 1.6)</td>
<td>Age, BMI</td>
</tr>
<tr>
<td>30</td>
<td>(.4, 1.6)</td>
<td>Age, BMI</td>
</tr>
<tr>
<td>20</td>
<td>(.1, 1.9)</td>
<td>Age</td>
</tr>
<tr>
<td>30</td>
<td>(.1, 1.9)</td>
<td></td>
</tr>
</tbody>
</table>

\( h_1: Effusion \) before Death

\( h_2: \) Death w/o Effusion

\( h_3: \) Death after Effusion

Importance based on marginal posterior probability of inclusion for the second and third hazard functions. Table 7 shows that Stage 3–4, good KPS and Adeno histology had the three highest marginal posterior probabilities of inclusion (in order)
Table 7
Sensitivity analysis of prior hyperparameter effects on marginal PPI for the Pericardial Effusion case study. Notation: $c$ is a parameter in the prior of $\beta_{\eta_k}$, given in (5) and $(z_{g1}, z_{g2})$ are the prior hyperparameters on $w_g$, the probability that $\eta_{g,k} = 1$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$c = 20$</th>
<th>$c = 30$</th>
<th>$c = 20$</th>
<th>$c = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(z_{g1}, z_{g2})$</td>
<td>$(.4, 1.6)$</td>
<td>$(.4, 1.6)$</td>
<td>$(1, 1.9)$</td>
<td>$(1, 1.9)$</td>
</tr>
<tr>
<td>Asthma</td>
<td>.556</td>
<td>.277</td>
<td>.338</td>
<td>.130</td>
</tr>
<tr>
<td>Diabetes</td>
<td>.615</td>
<td>.399</td>
<td>.418</td>
<td>.192</td>
</tr>
<tr>
<td>Stage 3–4</td>
<td>.538</td>
<td>.305</td>
<td>.357</td>
<td>.137</td>
</tr>
<tr>
<td>Location 2</td>
<td>.658</td>
<td>.410</td>
<td>.439</td>
<td>.199</td>
</tr>
<tr>
<td>Location 3</td>
<td>.624</td>
<td>.363</td>
<td>.410</td>
<td>.167</td>
</tr>
<tr>
<td>Age</td>
<td>.762</td>
<td>.551</td>
<td>.525</td>
<td>.276</td>
</tr>
<tr>
<td>Smoker</td>
<td>.639</td>
<td>.394</td>
<td>.435</td>
<td>.191</td>
</tr>
<tr>
<td>BMI</td>
<td>.631</td>
<td>.412</td>
<td>.436</td>
<td>.199</td>
</tr>
<tr>
<td>Induction Chemo</td>
<td>.508</td>
<td>.388</td>
<td>.329</td>
<td>.126</td>
</tr>
<tr>
<td>Good KPS</td>
<td>.521</td>
<td>.440</td>
<td>.347</td>
<td>.130</td>
</tr>
<tr>
<td>Adeno Histology</td>
<td>.731</td>
<td>.405</td>
<td>.508</td>
<td>.257</td>
</tr>
</tbody>
</table>

$h_1$: Effusion before Death

| Asthma | .622 | .460 | .568 | .400 |
| Diabetes | .839 | .745 | .820 | .703 |
| Stage 3–4 | .993 | .993 | .990 | .975 |
| Location 2 | .705 | .562 | .688 | .538 |
| Location 3 | .630 | .485 | .579 | .425 |
| Age | .601 | .462 | .565 | .413 |
| Smoker | .630 | .479 | .577 | .423 |
| BMI | .847 | .759 | .831 | .673 |
| Induction Chemo | .686 | .563 | .638 | .490 |
| Good KPS | .946 | .923 | .929 | .880 |
| Adeno Histology | .898 | .839 | .867 | .789 |

$h_2$: Death w/o Effusion

| Asthma | .413 | .242 | .222 | .122 |
| Diabetes | .488 | .315 | .271 | .172 |
| Stage 3–4 | .385 | .216 | .204 | .110 |
| Location 2 | .422 | .259 | .227 | .135 |
| Location 3 | .457 | .282 | .251 | .142 |
| Age | .378 | .215 | .206 | .109 |
| Smoker | .397 | .231 | .218 | .116 |
| BMI | .412 | .236 | .218 | .119 |
| Induction Chemo | .388 | .225 | .212 | .111 |
| Good KPS | .440 | .270 | .244 | .134 |
| Adeno Histology | .405 | .237 | .218 | .119 |

$h_3$: Death after Effusion

for the hazard of death before pericardial effusion in all four settings considered. Diabetes, Location 3 and a good KPS score had the highest three marginal posterior probabilities of inclusion for the hazard of death after pericardial effusion for all four models considered. There was, however, more sensitivity to $c$ and $z_{g1}$ for the hazard of pericardial effusion. While all four settings identified Age as the variable with highest marginal posterior probability of inclusion for hazard 1, three out of the four models identified Adeno Histology and Location 3 and the second and third most important variables based on the marginal PPI. The model that did not identify this trend, with $c = 30$ and $(z_{g1}, z_{g2}) = (.4, 1.6)$, had four variables with marginal PPI values between .40 and .44, which included the variables associated with Adeno Histology and a good KPS score. This discrepancy in ordering could be due, in part, to these four marginal PPI values being close to each other. In Table 8, we see the final models selected by the DIC-$\tau_g$ procedure for each of the four models considered.

The models chosen for the pericardial effusion data do not differ much for the four $c$ and $z_{g1}$ values considered. The hazard of pericardial effusion included Diabetes, Location 2, Location 3, Age, Smoking Status, BMI and Adeno histology in all four models considered. All four models included Diabetes, Stage 3–4, BMI, a good KPS score and Adeno Histology in the hazard of death before pericardial effusion. The variables selected for the hazard of death after pericardial effusion showed some sensitivity to $c$ and $z_{g1}$ choices, but all four models contained Diabetes as a covariate. Additionally, the conclusion that no covariates impact the hazard of death after pericardial effusion did not change for any of the four models.

5. Discussion

We have developed a variable selection procedure for a three-hazard model for semi-competing risks data using spike-and-slab priors and the Stochastic Variable Selection Search (SVSS) algorithm. We devised a criterion, DIC-$\tau_g$, for choosing the threshold on the marginal posterior probability of variable inclusion based on the Deviance Information Criteria. As seen in the sensitivity analyses, the DIC-$\tau_g$ procedure led to similar selected models, regardless of the choices of the
hyperparameters $c, \gamma_1$ and $\gamma_2$. The hazard of death before effusion of either type chose the same set of variables for all hyperparameter choices in both data sets. The hazard of pericardial effusion had the same set of variables for all four settings and the remaining hazard functions had minor differences in variables chosen. Clearly, there is a substantial effect of $c$ and $(\gamma_1, \gamma_2)$ on decisions about variable inclusion in the final model when choosing an arbitrary $\tau_g$. The DIC-$\tau_g$ procedure mitigates much of this sensitivity.

Our simulation study showed that the DIC-$\tau_g$ procedure performs well in determining which variables should be included in the model. This method can display a substantial improvement by tuning the sparsity parameter $c$ to separate the marginal posterior probabilities of inclusion of the variables.

In an application to data from esophageal cancer patients, we were able to use this procedure to select the important covariates in each of the three hazards and use the final variables included to evaluate the treatment and covariate effects on the hazards of effusion, death before effusion, and death after effusion. In our selected models, we saw the same medical conclusion, there was strong evidence that patients with IMRT radiation had significantly reduced risks of pericardial effusion, pleural effusion, and death before either effusion type. Further evidence of this treatment effect was shown in the SVSS posteriors with $P[\beta_{1,\text{IMRT}} > 0|\text{Data}] = .02$ and $P[\beta_{2,\text{IMRT}} > 0|\text{Data}] = .04$ for pleural effusion and $P[\beta_{1,\text{IMRT}} > 0|\text{Data}] < .01$ and $P[\beta_{2,\text{IMRT}} > 0|\text{Data}] < .01$ for pericardial effusion. For the models determined by the SVSS algorithm, patients who received IMRT had significantly reduced risks of pericardial and pleural effusion as well as death before effusion type. This agrees with results from several previous studies which showed that IMRT increased patient survival compared to 3D-CRT (Liu et al., 2012, 2016).

The proposed method provides a flexible, practical variable selection procedure for semi-competing risks. The SVSS algorithm for semi-competing risks is implemented in the package SCRSELECT (Chapple, 2016), which is now available on CRAN. This function computes and saves all posterior quantities and returns the marginal posterior probabilities of inclusion for each hazard function. The code allows for $0, 1, \ldots, p - 2$ variables to be excluded from the selection procedure, while we kept two variables out of the selection procedure to evaluate the treatment effect in our application. This program takes between two and four hours to run 100,000 iterations, depending on the number of split points in each hazard during the MCMC. Additionally, there is a program that runs the SVSS algorithm on two disperse chains followed by the DIC-$\tau_g$ procedure which performs a grid search through all the possible models for given vectors of marginal posterior probabilities of inclusion. This takes six to twelve hours depending on the separation of the variables in terms of marginal posterior probabilities of inclusion and returns these posterior probabilities, the threshold of inclusion and the final model selected by the DIC-$\tau_g$ procedure.

Acknowledgments

The authors thank the two referees for their constructive comments on an earlier version of this paper. This study was performed at the M.D. Anderson Cancer Center in Houston, Tx. Peter Thall’s research was supported by NIH R01 CA 83932. Andrew Chapple was partially supported by the NIH grant 5T32-CA096520-07.

Appendix. MCMC algorithm details on $(\eta_g, \beta_g)$ and $\lambda_g$ updates

We report additional details of the $(\eta_g, \beta_g)$ and $\lambda_g$ Markov Chain Monte Carlo samplers here.

Table 8
Sensitivity analysis of prior hyperparameter effects on the model selected in the Pericardial Effusion case study. Notation: $c$ is a dispersion parameter in the prior of $\beta_{\eta_k}$, given in (5) and $(\gamma_1, \gamma_2)$ are the beta prior hyperparameters on $u_k$, the probability that $\eta_{g,k} = 1$ for $k = 1, \ldots, p$.

<table>
<thead>
<tr>
<th>c</th>
<th>$(\gamma_g, \beta_g)$</th>
<th>Included Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>(2.1, 1.6)</td>
<td>Diabetes, Location 2, Location 3, Age, Smoking Status, BMI, Adeno Histology</td>
</tr>
<tr>
<td>30</td>
<td>(2.1, 1.6)</td>
<td>Diabetes, Location 2, Location 3, Age, Smoking Status, BMI, Adeno Histology</td>
</tr>
<tr>
<td>20</td>
<td>(1.1, 1.9)</td>
<td>Diabetes, Location 2, Location 3, Age, Smoking Status, BMI, Adeno Histology</td>
</tr>
<tr>
<td>30</td>
<td>(1.1, 1.9)</td>
<td>Diabetes, Location 2, Location 3, Age, Smoking Status, BMI, Adeno Histology</td>
</tr>
</tbody>
</table>

$h_1$: Effusion before Death

$h_2$: Death w/o Effusion

$h_3$: Death after Effusion
• Update \((\eta_{g}, \beta_{g})\) via a Metropolis step. With probability \(\phi\) we randomly choose one entry of \(\eta_{g}\) and if it is 0, set it to 1, while if it is 1, set it to 0. With probability \(1 - \phi\), we randomly choose a new entry that is a 0 and a new entry that is a 1 and change their values. If for some reason there are no values of 1 or 0 in the previous iteration, we can only perform an add or delete step and randomly choose one entry to change. We then update \(\beta_{g}\) in the following way under 3 different scenarios. If \(\eta_{g,k}\) is ”deleted”, we set \(\beta_{g,k} = 0\) and \(\eta_{g,k} = 0\). If \(\eta_{g,k}\) is ”added”, we sample \(\beta_{g,k}\) from \(\beta_{g,k}|\eta_{g,k}, \beta_{g}(-k)\) and set \(\eta_{g,k} = 1\). Finally if \(\eta_{g,k} = 0\) and \(\eta_{g,j} = 1\) are swapped then we set \(\eta_{g,j} = 0, \eta_{g,k} = 1, \beta_{g,j} = 0\) and we sample \(\beta_{g,k}\) from the distribution of \(\beta_{g,k}|\eta_{g,k}^*, \beta_{g}(-k)\). The vectors \(\eta_{g}^*\) and \(\beta_{g}^*\) are jointly accepted over the previous values \((\eta_{g}, \beta_{g})\) with probability \(\alpha_1^*, \alpha_2^*, \alpha_3^* = \min(\alpha_1, 1), \min(\alpha_2, 1), \min(\alpha_3, 1)\), respectively for the three moves listed above.

1. Delete Move

For a delete move, the acceptance probability is

\[
\alpha_1 = \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_{g}^*, \ldots) \pi(\beta_{g}^*, \ldots) \pi(\eta_{g}^* \mid \eta_{g}^*)}{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_{g}^*, \ldots) \pi(\beta_{g,j}^*, \beta_{g}(-j), \ldots) \pi(\eta_{g}^*)} \times \text{Beta}(t_g + 1 - \delta_g, 1, m_g - t_g + 1 + \delta_g, 2),
\]

where \(\mu_{\text{old}} = R_{k,-(k)}\) and \(\sigma_{\text{old}} = R_{k,-(k)} - R_{k,-(k)} R_{(-k),(-k)} R_{(-k),(-k)}, R = c \left( x_{(t_g)}^I, x_{(t_g)}^I \right)^{-1}\).

2. Add move

For the add step we set \(\eta_{g,k} = 1\) and sample \(\beta_{g,k}^*\) corresponding to \(\eta_{g,k} = 1\) from \(\pi(\beta_{g,k}^*, \beta_{g}(-k), \ldots)\) and accept \(\eta_{g,k}^*, \beta_{g,k}^*\) with probability \(\alpha_2^* = \min(\alpha_2, 1)\) where

\[
\alpha_2 = \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_{g}^*, \ldots) \pi(\beta_{g}^*, \ldots) \pi(\eta_{g}^*)}{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_{g}^*, \ldots) \pi(\beta_{g,j}^*, \beta_{g}(-j), \ldots) \pi(\eta_{g}^*)} \times \text{Beta}(t_g + 1 + \delta_g, 1, m_g - t_g - 1 + \delta_g, 2).
\]

where we denote \(\mu_{\text{new}} = \Sigma_{k,-(k)}^{-1} \Sigma_{(-k),(-k)} \beta_{(t_g),(-k)}\) and \(\sigma_{\text{new}}^2 = \Sigma_{k,k} - \Sigma_{k,-(k)} \Sigma_{(-k),(-k)}^{-1} \Sigma_{(-k),(-k)}\) with \(\Sigma = c \left( x_{(t_g)}^I, x_{(t_g)}^I \right)^{-1}\).

3. Swap Move

For the swap step, where entry \(j\) is set to 0 and \(k\) sampled for \(\beta_{g,j}^*\), we jointly accept \(\eta_{g,k}^*, \beta_{g,k}^*, \beta_{g,j}^*\) with probability \(\alpha_3^* = \min(1, \alpha_3)\) where

\[
\alpha_3 = \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_{g}^*, \ldots) \pi(\beta_{g}^*, \ldots) \pi(\eta_{g}^*)}{L(Y_1, Y_2, \delta_1, \delta_2, X|\beta_{g}^*, \ldots) \pi(\beta_{g,k}^*, \beta_{g}(-k), \ldots) \pi(\eta_{g}^*)} \times \text{Beta}(t_g + 1 - \delta_g, 1, m_g - t_g + 1 + \delta_g, 2).
\]

where we denote \(\mu_{\text{new}} = \Sigma_{k,-(k)}^{-1} \Sigma_{(-k),(-k)} \beta_{(t_g),(-k)}\) and \(\sigma_{\text{new}}^2 = \Sigma_{k,k} - \Sigma_{k,-(k)} \Sigma_{(-k),(-k)}^{-1} \Sigma_{(-k),(-k)}\) with \(\Sigma = c \left( x_{(t_g)}^I, x_{(t_g)}^I \right)^{-1}\). Likewise, we define \(\mu_{\text{old}} = R_{j,-(j)} R_{(-j),(-j)} \beta_{(t_g),(-j)}\) and \(\sigma_{\text{old}}^2 = R_{j,j} - R_{j,-(j)} R_{(-j),(-j)}\) with \(R = c \left( x_{(t_g)}^I, x_{(t_g)}^I \right)^{-1}\).

After we jointly accept or reject \((\eta_{g,k}^*, \beta_{g,j}^*)\), we resample each entry of \(\beta_{g,k}^*\) conditional on the other non-zero entries and then sample \(\beta_{g,j}\) jointly for better mixing of our posterior distribution.

• \(\lambda_g\) update

We used the same sampler as \(\text{Lee et al. (2015)}\) for all samplers except the birth and death of a split point in hazard \(g = 1, 2, 3\) along with using a different proposal distribution for \(\lambda|\eta_g, \mu_{\beta g}, \sigma_{\beta g}^2, \Sigma_{\beta g}\). Recall the MVN-ICAR specification of the prior for \(\lambda_g\) that was employed in \(\text{Lee et al. which sets up two matrices based on distances between adjacent split points (Besag and Kooperberg, 1995; Lee et al., 2015). Formally, define } \hat{\Delta}\)
Our sampler differs from Lee et al. (2015) in that we sample our proposal \( \lambda_{g,k}^* \) from a uniform distribution on \([\lambda_{g,k} - c_g, \lambda_{g,k} + c_g]\) where \(c_g\) is a tuning parameter for \(k = 1, \ldots, J_g + 1\). We set \(c_g = .25\) as in Haneuse et al. (2008). This proposal distribution causes the proposal ratio to be 1 which leads to an acceptance probability of this move to be the minimum of 1 and

\[
\alpha = \frac{L(Y_1, Y_2, \delta_1, \delta_2, X|\lambda_{g,k}^*, \ldots)N_{J_g+1}(\lambda_{g,k}^*|\hat{\mu}_{\lambda_{g,k}}, \sigma_{\lambda_{g,k}}^2 \Sigma_{\lambda_{g,k}})}{L(Y_1, Y_2, \delta_1, \delta_2, X|\lambda_{g,k}, \ldots)N_{J_g+1}(\lambda_{g,k}|\hat{\mu}_{\lambda_{g,k}}, \sigma_{\lambda_{g,k}}^2 \Sigma_{\lambda_{g,k}})}.
\]

References


Chapple, A., 2016. Package: Scrselect. CRAN.


