Supplemental Materials to: "Hierarchical models for semi-competing risks data with application to quality of end-of-life care for pancreatic cancer"

Kyu Ha Lee

Epidemiology and Biostatistics Core, The Forsyth Institute, Cambridge, Massachusetts, U.S.A.

Department of Oral Health Policy and Epidemiology, Harvard School of Dental Medicine, Boston, Massachusetts, U.S.A.

klee@forsyth.org

Francesca Dominici

Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, U.S.A.

Deborah Schrag

Department of Medical Oncology, Dana Farber Cancer Institute, Boston, Massachusetts, U.S.A.

Sebastien Haneuse

Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, U.S.A.

Introduction

This document supplements the main paper titled "Hierarchical models for clustered semi-competing risks data with application to pancreatic cancer". In Section A, we provide technical details regarding the MVN-ICAR specification for baseline hazard functions in PEM models. In Section B, we provide a detailed description of the Metropolis-Hastings-Green algorithm to implement our proposed Bayesian framework (Weibull-MVN, Weibull-DPM, PEM-MVN, PEM-DPM). In Section C, we examine the potential use of methods proposed in Gorfine and Hsu (2011) and Liquet et al. (2012) in the context of the motivating application. In Section D, we provide results from simulation studies that were not presented in the main paper. Finally, Section E supplements analyses of Medicare data from New England in our main paper with i) some additional results, ii) a visual assessment of convergence of the proposed MCMC schemes using potential scale reduction factor, and iii) an extension for the semi-Markov Weibull-MVN in which the proportional hazards assumption is relaxed.

In order to distinguish the two documents, alpha-numeric labels are used for sections, tables, figures, and equations in this document while numeric labels are used in the main paper.

A MVN-ICAR Specification for $\boldsymbol{\lambda}_g$

In MVN-ICAR, the specification of a prior for the components of $\boldsymbol{\lambda}_g$ is considered as a one-dimensional spatial problem. The dependence between neighboring intervals are modeled via a Gaussian intrinsic conditional autoregression (ICAR) (Besag and Kooperberg, 1995). Let $\boldsymbol{\lambda}_g^{(-k)}$ denote the vector given by $\boldsymbol{\lambda}_g$ with the k^{th} element removed. The full conditional prior for $\lambda_{g,k}$ is then taken to be the following normal distribution:

$$\lambda_{g,k}|\boldsymbol{\lambda}_g^{(-k)} \sim \text{Normal}(\nu_{g,k}, \sigma^2_{g,k}), \tag{1}$$

where the conditional mean, $\nu_{g,k} = \mu_{\lambda_g} + \sum_{j\neq k} W^g_{kj}(\lambda_{g,j} - \mu_{\lambda_g})$, is a marginal mean plus a weighted sum of the deviations of the remaining intervals. Let $\bar{\Delta}^g_k = s_{g,k} - s_{g,k-1}$ denote the length of the $I_{g,k}$ interval. We determine the weights for the intervals adjacent to the k^{th} intervals based on these lengths as follows:

$$W^g_{k(k-1)} = \frac{c_{\lambda_g}(\bar{\Delta}^g_{k-1} + \bar{\Delta}^g_k)}{\bar{\Delta}^g_{k-1} + 2\bar{\Delta}^g_k + \bar{\Delta}^g_{k+1}}, \quad W^g_{k(k+1)} = \frac{c_{\lambda_g}(\bar{\Delta}^g_k + \bar{\Delta}^g_{k+1})}{\bar{\Delta}^g_{k-1} + 2\bar{\Delta}^g_k + \bar{\Delta}^g_{k+1}}, \tag{2}$$

where the constant $c_{\lambda_g} \in [0, 1]$ dictates the extent to which $\lambda_{g,k}$ is influenced by adjacent intervals (Haneuse et al., 2008). The remaining weights corresponding to intervals which are not directly adjacent to the k^{th} interval are set to zero. The conditional variance $\sigma^2_{g,k}$ in (1) is given by $\sigma^2_{\lambda_g} Q^g_k$. The $\sigma^2_{\lambda_g}$ is an overall measure of variation across the elements of $\boldsymbol{\lambda}_g$ and the diagonal matrix Q^g_k is given by

$$\frac{2}{\bar{\Delta}^g_{k-1} + 2\bar{\Delta}^g_k + \bar{\Delta}^g_{k+1}}. \tag{3}$$

Given (1), (2), and (3), we can see that $\boldsymbol{\lambda}_g$ jointly follows a (K_g+1)-dimensional multivariate normal (MVN) distribution:

$$\text{MVN}_{K_g+1}(\mu_{\lambda_g}\mathbf{1},\ \sigma^2_{\lambda_g}\Sigma_{\lambda_g}), \tag{4}$$

where μ_{λ_g} is the overall (marginal) mean, $\sigma^2_{\lambda_g}$ the overall variability in elements of $\boldsymbol{\lambda}_g$. The Σ_{λ_g} is given by $(I - \boldsymbol{W}^g)^{-1}\boldsymbol{Q}^g$, where a $(K_g + 1) \times (K_g + 1)$ matrix $\boldsymbol{W}^g_{(k,j)}$=W^g_{kj} and a $(K_g + 1) \times (K_g + 1)$ diagonal matrix $\boldsymbol{Q}^g_{(k,k)}$=Q^g_k.

B Metropolis-Hastings-Green Algorithm

B.1 Weibull models

Let $\Phi_W = \{\alpha_{w,1}, \alpha_{w,2}, \alpha_{w,1}, \kappa_{w,1}, \kappa_{w,2}, \kappa_{w,3}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \vec{\gamma}, \vec{\boldsymbol{V}}\}$ be a set of parameters in the likelihood function of Weibull models. The observed data likelihood $L_W(\mathcal{D}|\Phi_W)$ is given by

$$\prod_{j=1}^{J}\prod_{i=1}^{n_j}\left(\gamma_{ji}\alpha_{w,1}\kappa_{w,1}y_{ji1}^{\alpha_{w,1}-1}\eta_{ji1}\right)^{\delta_{ji1}(1-\delta_{ji2})}\left(\gamma_{ji}^2\alpha_{w,1}\kappa_{w,1}y_{ji1}^{\alpha_{w,1}-1}\eta_{ji1}\alpha_{w,3}\kappa_{w,3}y_{ji2}^{\alpha_{w,3}-1}\eta_{ji3}\right)^{\delta_{ji1}\delta_{ji2}}$$
$$\times\left(\gamma_{ji}\alpha_{w,2}\kappa_{w,2}y_{ji2}^{\alpha_{w,2}-1}\eta_{ji2}\right)^{\delta_{ji2}(1-\delta_{ji1})}\exp\left\{-r_W(y_{ji1},y_{ji2})\right\}, \tag{5}$$

where $\eta_{jig} = \exp\left(\boldsymbol{x}_{jig}^\top\boldsymbol{\beta}_g + V_{jg}\right)$ and

$$r_W(t_{ji1},t_{ji2})$$
$$= \begin{cases}\gamma_{ji}\left\{\kappa_{w,1}t_{ji1}^{\alpha_{w,1}}\eta_{ji1} + \kappa_{w,2}t_{ji1}^{\alpha_{w,2}}\eta_{ji2} + \left(\kappa_{w,3}t_{ji2}^{\alpha_{w,3}} - \kappa_{w,3}t_{ji1}^{\alpha_{w,3}}\right)\eta_{ji3}\right\}, & \text{for Markov model}\\ \gamma_{ji}\left[\kappa_{w,1}t_{ji1}^{\alpha_{w,1}}\eta_{ji1} + \kappa_{w,2}t_{ji1}^{\alpha_{w,2}}\eta_{ji2} + \left\{\kappa_{w,3}(t_{ji2}-t_{ji1})^{\alpha_{w,3}}\right\}\eta_{ji3}\right], & \text{for semi-Markov model}\end{cases}$$

For Weibull models, we use a random scan Gibbs sampling scheme, randomly selecting and updating a (vector of) model parameter at each iteration.

B.1.1 Updating $\boldsymbol{\beta}_g$

Let $\Phi^{-(\beta)}$ denote a set of parameters Φ with β removed. The full conditional posterior distribution of $\boldsymbol{\beta}_1$ can be obtained by

$$\begin{aligned}\pi(\boldsymbol{\beta}_1|\Phi_W^{-(\boldsymbol{\beta}_1)},\theta,\Sigma_V) &\propto L_W(D|\Phi_W).\\ &\propto \prod_{j=1}^{J}\prod_{i=1}^{n_j}\exp\left(\delta_{ji1}\boldsymbol{x}_{ji1}^\top\boldsymbol{\beta}_1 - \gamma_{ji}\kappa_{w,1}y_{ji1}^{\alpha_{w,1}}e^{\boldsymbol{x}_{ji1}^\top\boldsymbol{\beta}_1+V_{j1}}\right).\end{aligned}$$

Analogously, the full conditionals of $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are given by

$$\pi(\boldsymbol{\beta}_2|\Phi_W^{-(\boldsymbol{\beta}_2)},\theta,\Sigma_V) \propto \prod_{j=1}^{J}\prod_{i=1}^{n_j}\exp\left\{\delta_{ji2}(1-\delta_{ji1})\boldsymbol{x}_{ji2}^\top\boldsymbol{\beta}_2 - \gamma_{ji}\kappa_{w,2}y_{ji1}^{\alpha_{w,2}}e^{\boldsymbol{x}_{ji2}^\top\boldsymbol{\beta}_2+V_{j2}}\right\},$$
$$\pi(\boldsymbol{\beta}_3|\Phi_W^{-(\boldsymbol{\beta}_3)},\theta,\Sigma_V) \propto \prod_{j=1}^{J}\prod_{i=1}^{n_j}\exp\left\{\delta_{ji1}\delta_{ji2}\boldsymbol{x}_{ji3}^\top\boldsymbol{\beta}_3 - \gamma_{ji}\kappa_{w,3}\left(y_{ji2}^{\alpha_{w,3}} - y_{ji1}^{\alpha_{w,3}}\right)e^{\boldsymbol{x}_{ji3}^\top\boldsymbol{\beta}_3+V_{j3}}\right\}.$$

Since the full conditionals do not have standard forms, we use Metropolis Hastings (MH) algorithm to update each element of $\boldsymbol{\beta}_g$, $\beta_{g,1}, \ldots, \beta_{g,p_1}$. In our algorithm, the conventional random walk MH is improved in convergence speed by taking some meaningful function of the current value $\beta_{g,k}^{(t-1)}$ for the mean and variance of Normal proposal density. Specifically, let $D_1(\beta_{g,k})$ and $D_2(\beta_{g,k})$ denote the first and second gradients of log-full conditional of $\boldsymbol{\beta}_g$ with respect to $\beta_{g,k}$, then a proposal β^* is drawn from a Normal proposal density that is centered at $\mu(\beta_{g,k}^{(t-1)}) = \beta_{g,k}^{(t-1)} - D_1(\beta_{g,k}^{(t-1)})/D_2(\beta_{g,k}^{(t-1)})$, the updated value from the Newton-Raphson algorithm, with a variance $\sigma^2(\beta_{g,k}^{(t-1)}) = -2.4^2/D_2(\beta_{g,k}^{(t-1)})$, based on the inverse Fisher information evaluated with $\beta_{g,k}^{(t-1)}$ (Roberts et al., 2001; Gelman et al., 2013). Therefore, the acceptance probability for $\beta_{g,k}$ is given by

$$\frac{\pi(\boldsymbol{\beta}_g^*|\Phi_W^{-(\boldsymbol{\beta}_g)}, \theta, \Sigma_V)\text{Normal}\left(\beta_{g,k}^{(t-1)}|\mu(\beta^*), \sigma^2(\beta^*)\right)}{\pi(\boldsymbol{\beta}_g^{(t-1)}|\Phi_W^{-(\boldsymbol{\beta}_g)}, \theta, \Sigma_V)\text{Normal}\left(\beta^*|\mu(\beta_{g,k}^{(t-1)}), \sigma^2(\beta_{g,k}^{(t-1)})\right)}, \tag{6}$$

where $\boldsymbol{\beta}_g^{(t-1)}$ is a sample of $\boldsymbol{\beta}_g$ at current iteration and $\boldsymbol{\beta}_g^*$ is the $\boldsymbol{\beta}_g$ with k-th element replaced by β^*.

B.1.2 Updating $\alpha_{w,g}$

The full conditional posterior distribution of $\alpha_{w,1}$ is given by

$$\begin{aligned}
& \pi(\alpha_{w,1}|\Phi_W^{-(\alpha_{w,1})}, \theta, \Sigma_V) \\
\propto\ & L_W(D|\Phi_W) \times \pi(\alpha_{w,1}) \\
\propto\ & \alpha_{w,1}^{a_{\alpha,1}-1} e^{-b_{\alpha,1}\alpha_{w,1}} \prod_{j=1}^{J}\prod_{i=1}^{n_j} \left(\alpha_{w,1} y_{ji1}^{\alpha_{w,1}}\right)^{\delta_{ji1}} \exp\left(\gamma_{ji}\kappa_{w,1} y_{ji1}^{\alpha_{w,1}} \eta_{ji1}\right).
\end{aligned}$$

Analogously, the full conditionals of $\alpha_{w,2}$ and $\alpha_{w,3}$ are given by

$$\pi(\alpha_{w,2}|\Phi_W^{-(\alpha_{w,2})}, \theta, \Sigma_V) \propto \alpha_{w,2}^{a_{\alpha,2}-1} e^{-b_{\alpha,2}\alpha_{w,2}} \prod_{j=1}^{J}\prod_{i=1}^{n_j} \left(\alpha_{w,2} y_{ji2}^{\alpha_{w,2}}\right)^{\delta_{ji2}(1-\delta_{ji1})} \exp\left(-\gamma_{ji}\kappa_{w,2} y_{ji1}^{\alpha_{w,2}} \eta_{ji2}\right),$$

$$\begin{aligned}
\pi(\alpha_{w,3}|\Phi_W^{-(\alpha_{w,3})}, \theta, \Sigma_V) \propto\ & \alpha_{w,3}^{a_{\alpha,3}-1} e^{-b_{\alpha,3}\alpha_{w,3}} \prod_{j=1}^{J}\prod_{i=1}^{n_j} \left(\alpha_{w,3} y_{ji2}^{\alpha_{w,3}}\right)^{\delta_{ji1}\delta_{ji2}} \\
& \times \exp\left\{-\gamma_{ji}\kappa_{w,3}\left(y_{ji2}^{\alpha_{w,3}} - y_{ji1}^{\alpha_{w,3}}\right)\eta_{ji3}\right\}.
\end{aligned}$$

In MH algorithm to update $\alpha_{w,g}$, we generate a proposal α^* from a Gamma distribution with Gamma $\left((\alpha_{w,g}^{(t-1)})^2/k_0, \alpha_{w,g}^{(t-1)}/k_0\right)$ which corresponds to a distribution with a mean of

$\alpha_{w,g}^{(t-1)}$ and a variance of k_0. The value of k_0 is specified such that the MH step for $\alpha_{w,g}$ achieves an acceptance rate of $25\% \sim 30\%$. Finally the acceptance probability to update $\alpha_{w,g}$ can be written as

$$\frac{\pi(\alpha^*|\Phi_W^{-(\alpha_{w,g})}, \theta, \Sigma_V)\mathcal{G}\left(\alpha_{w,g}^{(t-1)}|(\alpha_{w,g}^*)^2/k_0, \alpha_{w,g}^*/k_0\right)}{\pi(\alpha_{w,g}^{(t-1)}|\Phi_W^{-(\alpha_{w,g})}, \theta, \Sigma_V)\mathcal{G}\left(\alpha_{w,g}^*|(\alpha_{w,g}^{(t-1)})^2/k_0, \alpha_{w,g}^{(t-1)}/k_0\right)}.$$

B.1.3 Updating $\kappa_{w,g}$

The full conditional posterior distribution of $\kappa_{w,g}$ can be obtained by

$$\pi(\kappa_{w,g}|\Phi_W^{-(\kappa_{w,g})}, \theta, \Sigma_V) \propto L_W(D|\Phi_W) \times \pi(\kappa_{w,g}).$$

We see that the full conditionals of $\kappa_{w,g}$ are gamma distributions and the samples can be drawn from following distributions:

$$\begin{aligned}
\kappa_{w,1}|\Phi_W^{-(\kappa_{w,1})}, \theta, \Sigma_V &\sim \text{Gamma}\left(\sum_{j=1}^{J}\sum_{i=1}^{n_j}\delta_{ji1} + a_{\kappa,1},\ \sum_{j=1}^{J}\sum_{i=1}^{n_j}\gamma_{ji}y_{ji1}^{\alpha_{w,1}}\eta_{ji1} + b_{\kappa,1}\right),\\
\kappa_{w,2}|\Phi_W^{-(\kappa_{w,2})}, \theta, \Sigma_V &\sim \text{Gamma}\left(\sum_{j=1}^{J}\sum_{i=1}^{n_j}\delta_{ji2}(1-\delta_{ji1}) + a_{\kappa,2},\ \sum_{j=1}^{J}\sum_{i=1}^{n_j}\gamma_{ji}y_{ji1}^{\alpha_{w,2}}\eta_{ji2} + b_{\kappa,2}\right),\\
\kappa_{w,3}|\Phi_W^{-(\kappa_{w,3})}, \theta, \Sigma_V &\sim \text{Gamma}\left(\sum_{j=1}^{J}\sum_{i=1}^{n_j}\delta_{ji1}\delta_{ji2} + a_{\kappa,3},\ \sum_{j=1}^{J}\sum_{i=1}^{n_j}\gamma_{ji}\left(y_{ji2}^{\alpha_{w,3}} - y_{ji1}^{\alpha_{w,3}}\right)\eta_{ji3} + b_{\kappa,3}\right).
\end{aligned}$$

B.1.4 Updating γ_{ji}

The full conditional posterior distribution of γ_{ji} is given by

$$\begin{aligned}
&\pi(\gamma_{ji}|\Phi_W^{-(\gamma_{ji})}, \theta, \Sigma_V)\\
\propto\ & L_W(D|\Phi_W) \times \pi(\gamma_{ji}|\theta)\\
\propto\ & \gamma_{ji}^{\delta_{ji1}+\delta_{ji2}+\theta^{-1}-1}\exp\left[-r_W(y_{ji1}, y_{ji2}) - \theta^{-1}\gamma_{ji}\right].
\end{aligned}$$

Therefore, we sample γ_{ji} from

$$\text{Gamma}\left(\delta_{ji1} + \delta_{ji2} + \theta^{-1},\ r_W(y_{ji1}, y_{ji2}; \gamma_{ji} = 1) + \theta^{-1}\right).$$

B.1.5 Updating θ

Let $\xi = 1/\theta$ denote the precision parameter of frailty distribution. The full conditional posterior distribution of ξ is given by

$$\begin{aligned}\pi(\xi|\Phi_W, \Sigma_V) &\propto \pi(\xi)\prod_{j=1}^{J}\prod_{i=1}^{n_j}\pi(\gamma_{ji}|\xi) \\ &\propto \frac{\xi^{n\xi+b_\theta-1}e^{-\xi(\sum_{j=1}^{J}\sum_{i=1}^{n_j}\gamma_{ji}+a_\theta)}}{\{\Gamma(\xi)\}^n}\prod_{j=1}^{J}\prod_{i=1}^{n_j}\gamma_{ji}^{\xi-1}.\end{aligned}$$

We revise the traditional random walk MH algorithm for updating ξ as done in Section B.1.1 for $\boldsymbol{\beta}_g$. Let $\mu_\xi(\xi) = \xi - \min\{0, D_{1,\xi}(\xi)/D_{2,\xi}(\xi)\}$ and $\sigma_\xi^2(\xi) = -c_0/D_{2,\xi}(\xi)$, where $D_{1,\xi}(\xi)$ and $D_{2,\xi}(\xi)$ are the first and second gradients of $\log\pi(\xi|\Phi_W^{-(\xi)}, \Sigma_V)$ with respect to ξ. A proposal ξ^* is generated from the following Gamma distribution

$$\text{Gamma}\left(\mu_\xi(\xi^{(t-1)})^2/\sigma_\xi^2(\xi^{(t-1)}),\ \mu(\xi^{(t-1)})/\sigma_\xi^2(\xi^{(t-1)})\right).$$

The value of $c_0 > 0$ is specified such that the algorithm achieve the desired acceptance rate. The acceptance probability to update ξ is then given by

$$\frac{\pi(\xi^*|\Phi_W, \Sigma_V)\text{Gamma}\left(\xi^*|\mu_\xi(\xi^*)^2/\sigma_\xi^2(\xi^*),\ \mu(\xi^*)/\sigma_\xi^2(\xi^*)\right)}{\pi(\xi^{(t-1)}|\Phi_W, \Sigma_V)\text{Gamma}\left(\xi^*|\mu_\xi(\xi^{(t-1)})^2/\sigma_\xi^2(\xi^{(t-1)}),\ \mu(\xi^{(t-1)})/\sigma_\xi^2(\xi^{(t-1)})\right)}.$$

B.1.6 Updating $\boldsymbol{V}_j$ for Weibull-MVN model

The full conditional posterior distribution of V_{j1} can be obtained by

$$\begin{aligned}\pi(V_{j1}|\Phi_W^{-(V_{j1})}, \theta, \Sigma_V) &\propto L_W(D|\Phi_W) \times \pi(\boldsymbol{V}_j|\Sigma_V). \\ &\propto \exp\left\{\sum_{i=1}^{n_j}\left(V_{j1}\delta_{ji1} - \gamma_{ji}\kappa_{w,1}y_{ji1}^{\alpha_{w,1}}\eta_{ji1}\right) - \frac{1}{2}\boldsymbol{V}_j^\top\Sigma_V^{-1}\boldsymbol{V}_j\right\}.\end{aligned}$$

Analogously, the full conditionals of V_{j2} and V_{j3} can be written as

$$\begin{aligned}\pi(V_{j2}|\Phi_W^{-(V_{j2})}, \theta, \Sigma_V) &\propto \exp\left\{\sum_{i=1}^{n_j}\left(V_{j2}\delta_{ji2}(1-\delta_{ji1}) - \gamma_{ji}\kappa_{w,2}y_{ji1}^{\alpha_{w,2}}\eta_{ji2}\right) - \frac{1}{2}\boldsymbol{V}_j^\top\Sigma_V^{-1}\boldsymbol{V}_j\right\}, \\ \pi(V_{j3}|\Phi_W^{-(V_{j3})}, \theta, \Sigma_V) &\propto \exp\left\{\sum_{i=1}^{n_j}\left(V_{j3}\delta_{ji1}\delta_{ji2} - \gamma_{ji}\kappa_{w,3}(y_{ji2}^{\alpha_{w,3}} - y_{ji1}^{\alpha_{w,3}})\eta_{ji3}\right) - \frac{1}{2}\boldsymbol{V}_j^\top\Sigma_V^{-1}\boldsymbol{V}_j\right\}.\end{aligned}$$

As done in Section B.1.1, in a MH step for updating V_{jg}, we sample a proposal V^* from a Normal distribution that is centered at $\mu_V(V_{jg}^{(t-1)}) = V_{jg}^{(t-1)} - D_{1,V}(V_{jg}^{(t-1)})/D_{2,V}(V_{jg}^{(t-1)})$

and has a variance of $\sigma_V^2(V^{(t-1)}) = -2.4^2/D_{2,V}(V^{(t-1)})$, where $D_{1,V}(V_{jg})$ and $D_{2,V}(V_{jg})$ are the first and the second gradients of $\log \pi(V_{jg}|\Phi_W^{-(V_{jg})}, \theta, \Sigma_V)$ with respect to V_{jg}. Finally, the acceptance probability is given by

$$\frac{\pi(V^*|\Phi_W^{-(V_{jg})}, \theta, \Sigma_V)\text{Normal}\left(V_{jg}^{(t-1)}|\mu_V(V^*), \sigma_V^2(V^*)\right)}{\pi(V_{jg}^{(t-1)}|\Phi_W^{-(V_{jg})}, \theta, \Sigma_V)\text{Normal}\left(V^*|\mu_V(V_{jg}^{(t-1)}), \sigma_V^2(V_{jg}^{(t-1)})\right)}.$$

B.1.7 Updating Σ_V for Weibull-MVN model

The full conditional posterior distribution of Σ_V can be written as

$$\begin{aligned}\pi(\Sigma_V|\Phi_W, \theta) &\propto \pi(\Sigma_V)\prod_{j=1}^{J}\pi(\boldsymbol{V}_j|\Sigma_V) \\ &\propto |\Sigma_V|^{-\frac{J+\rho_v+4}{2}}\exp\left\{-\frac{1}{2}\left(\sum_{j=1}^{J}\boldsymbol{V}_j\boldsymbol{V}_j^\top + \Psi_v\right)\Sigma_V^{-1}\right\}.\end{aligned}$$

Therefore, we update Σ_V from the following inverse-Wishart distribution:

$$\Sigma_V|\Phi_W, \theta \sim \text{inverse-Wishart}\left(\sum_{j=1}^{J}\boldsymbol{V}_j\boldsymbol{V}_j^\top + \Psi_v,\ J+\rho_v\right).$$

B.1.8 Updating $\boldsymbol{V}_j$ and Σ_V for Weibull-DPM model

Towards developing this model, suppose that, instead of arising from a single distribution, the $\boldsymbol{V}_j$ are draws from a finite mixture of M multivariate Normal distributions, each with their own mean vector and variance-covariance matrix, $(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$ for $m = 1, \ldots, M$. Let $m_j \in \{1, \ldots, M\}$ denote the specific component or class to which the j^{th} hospital belongs. Since the class-specific $(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$ are not known they are taken to be draws from some distribution, G_0. Furthermore, since the 'true' class memberships are not known, we denote the probability that the j^{th} hospital belongs to any given class by the vector $\boldsymbol{p} = (p_1, \ldots, p_M)$ whose components add up to 1.0. In the absence of prior knowledge regarding the distribution of class memberships for the J hospitals across the M classes, a natural prior for $\boldsymbol{p}$ is the conjugate symmetric Dirichlet$(\tau/M, \ldots, \tau/M)$ distribution; the hyperparameter, τ, is often referred to as the precision parameter (Walker and Mallick,

1997). Jointly, this finite mixture distribution can be summarized by:

$$\begin{aligned} \boldsymbol{V}_j|m_j &\sim MVN(\boldsymbol{\mu}_{m_j}, \Sigma_{m_j}), \\ (\boldsymbol{\mu}_m, \Sigma_m) &\sim G_0, \text{ for } m = 1, \ldots, M, \\ m_j|\boldsymbol{p} &\sim \text{Discrete}(m_j|\ p_1, \ldots, p_M), \\ \boldsymbol{p} &\sim \text{Dirichlet}(\tau/M, \ldots, \tau/M). \end{aligned} \tag{7}$$

Finally, letting $M \to \infty$ the resulting specification is referred to as a Dirichlet process mixture of multivariate Normal distributions (DPM-MVN) (Ferguson, 1973; Bush and MacEachern, 1996). When $M \to \infty$, we cannot explicitly represent the infinite number of $(\boldsymbol{\mu}_m, \Sigma_m)$. Instead, following Neal (2000), we represent and implement the MCMC sampling for only those $(\boldsymbol{\mu}_m, \Sigma_m)$ that are currently associated with some observations at each iteration. In this subsection, we provide a step-by-step detailed description of the MH algorithm to update $\boldsymbol{V}_j$ in Weibull-DPM model.

First, we update a class membership m_j based on $m_j|\boldsymbol{m}_{(-j)}, \boldsymbol{V}_j$, $j = 1, \cdots, J$. Let $\boldsymbol{m}_{-(j)}$ denote a set of all class memberships from clusters except the cluster j. After identifying the "n_m" unique classes of $\boldsymbol{m}_{-(j)}$, we compute the following probabilities for each of the unique values m.

$$P(m_j = m|\boldsymbol{m}_{(-j)}, \boldsymbol{V}_j) = b\frac{n_{-j,m}}{J-1+\tau}\int \text{Normal}(\boldsymbol{V}_j|\boldsymbol{\mu}_{m_j}, \Sigma_{m_j})dH_{-j,m}(\mu, \Sigma), \tag{8}$$

$$P(m_j \neq m_k,\ \forall k \neq j\ |\boldsymbol{m}_{(-j)}, \boldsymbol{V}_j) = b\frac{\tau}{J-1+\tau}\int \text{Normal}(\boldsymbol{V}_j|\boldsymbol{\mu}, \Sigma)dG_0(\mu, \Sigma), \tag{9}$$

where $H_{-j,m}$ is the posterior distribution of (μ, Σ) based on the prior G_0 and $\{\boldsymbol{V}_k : k \neq j, m_k = c\}$. The normalizing constant b makes "$n_m + 1$" probabilities above sum to 1. Let $A = \{j : m_j = m\}$ and H_A be the posterior distribution of (μ, σ) based on the prior G_0 and $\{\boldsymbol{V}_j : j \in A\}$. It can be shown that the H_A is also Normal-inverse Wishart distribution as G_0 is conjugate to multivariate normal distribution:

1. we draw a sample of a class membership.

i) For each m_j, identify the n_m unique values of $\boldsymbol{m}_{(-j)}$.

ii) For each of the unique values m, compute the following probabilities:

$$P(m_j = m|\boldsymbol{m}_{(-j)}, \boldsymbol{V}_j) = b\frac{n_{-j,m}}{J-1+\tau}\int \text{Normal}(\boldsymbol{V}_j|\boldsymbol{\mu}_{m_j}, \Sigma_{m_j})dH_{-j,m}(\mu, \Sigma), \tag{10}$$

$$P(m_j \neq m_k,\ \forall k \neq j\ |\boldsymbol{m}_{(-j)}, \boldsymbol{V}_j) = b\frac{\tau}{J-1+\tau}\int \text{Normal}(\boldsymbol{V}_j|\boldsymbol{\mu}, \Sigma)dG_0(\mu, \Sigma), \tag{11}$$

where $H_{-j,m}$ is the posterior distribution of (μ, Σ) based on the prior G_0 and $\{\boldsymbol{V}_k : k \neq j, m_k = m\}$. The normalizing constant b makes $n_m + 1$ probabilities above sum to 1.0. Let $A = \{j : m_j = m\}$ and H_A be the posterior distribution of (μ, σ) based on the prior G_0 and $\{\boldsymbol{V}_j : j \in A\}$. It can be shown that the H_A is also Normal-inverse Wishart distribution as G_0 is conjugate to multivariate normal distribution:

$$H_A(\mu, \Sigma | \mu_A, \zeta_A, \Psi_A, \rho_A),$$

where

$$\mu_A = \frac{\frac{1}{\zeta_0}\boldsymbol{\mu}_0 + |A|\bar{\boldsymbol{V}}_A}{\frac{1}{\zeta_0} + |A|}, \quad \zeta_A = \left(\frac{1}{\zeta_0} + |A|\right)^{-1}, \quad \rho_A = \rho_0 + |A|,$$

$$\Psi_A = \Psi_0 + \sum_{j \in A} \left(\boldsymbol{V}_j - \bar{\boldsymbol{V}}_A\right)\left(\boldsymbol{V}_j - \bar{\boldsymbol{V}}_A\right)^\top + \frac{\frac{|A|}{\zeta_0}}{\frac{1}{\zeta_0} + |A|}\left(\bar{\boldsymbol{V}}_A - \boldsymbol{\mu}_0\right)\left(\bar{\boldsymbol{V}}_A - \boldsymbol{\mu}_0\right)^\top, \quad (12)$$

with $\bar{\boldsymbol{V}}_A = \frac{1}{|A|}\sum_{k \in A} \boldsymbol{V}_k$. Now we define

$$\begin{aligned} & Q(\boldsymbol{V}_j, \mu_0, \zeta_0, \Psi_0, \rho_0) \\ = & \int f_{\mathcal{N}_3}(\boldsymbol{V}_j | \mu, \Sigma) dF_{NIW}(\mu, \Sigma | \mu_0, \zeta_0, \Psi_0, \rho_0) \\ = & \frac{|\Psi_0|^{\frac{\rho_0}{2}}}{\left|\Psi_0 + \boldsymbol{V}_j\boldsymbol{V}_j^\top + \frac{1}{\zeta_0}\boldsymbol{\mu}_0\boldsymbol{\mu}_0^\top - \left(1 + \frac{1}{\zeta_0}\right)^{-1}\left(\frac{1}{\zeta_0}\boldsymbol{\mu}_0 + \boldsymbol{V}_j\right)\left(\frac{1}{\zeta_0}\boldsymbol{\mu}_0 + \boldsymbol{V}_j\right)^\top\right|^{\frac{\rho_0+1}{2}}} \\ & \times \frac{1}{(\pi\sqrt{2(1+\zeta_0)})^3} \times \frac{\Gamma_{\alpha,3}(\frac{\rho_0+1}{2})}{\Gamma_{\alpha,3}(\frac{\rho_0}{2})} \end{aligned} \quad (13)$$

It follows that the integrals in (10) and (11) are equal to $Q(\boldsymbol{V}_j, \boldsymbol{\mu}_A, \zeta_A, \Psi_A, \rho_A)$ and $Q(\boldsymbol{V}_j, \boldsymbol{\mu}_0, \zeta_0, \Psi_0, \rho_0)$, respectively.

iii) Sample $m_j^{(\text{new})}$ based on the probabilities given in (10) and (11).

2. For all $m \in \{m_1, \ldots, m_J\}$, update (μ_m, Σ_m) using the posterior distribution that is based on $\{\boldsymbol{V}_j : j \in \{k : m_k = m\}\}$.

3. For $j = 1, \ldots, J$, update $\boldsymbol{V}_j$ using its full conditional using Metropolis-Hastings algorithm.

4. We treat τ as random and assign gamma prior Gamma(a_τ, b_τ) for τ. Following Escobar and West (1995), we update τ by

i) sampling an $c \in (0, 1)$ from $Beta(\tau + 1, J)$,

ii) sampling the new τ from the mixture of two gamma distributions:

$$p_c\text{Gamma}(a_\tau + n_m, b_\tau - \log(c)) + (1 - p_c)\text{Gamma}(a_\tau + n_m - 1, b_\tau - log(c)),$$

where the weight p_c is defined such that $p_c/(1 - p_c) = (a_\tau + n_m - 1)/\{J(b_\tau - log(c))\}$.

5. Finally we calculate the total variance-covariance matrix:

$$\Sigma_V = \frac{1}{J}\sum_{j=1}^{J}\left\{\left(\boldsymbol{\mu}_{m_j} - \bar{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}_{m_j} - \bar{\boldsymbol{\mu}}\right)^\top + \Sigma_{m_j}\right\}, \tag{14}$$

where $\bar{\boldsymbol{\mu}} = \sum_{j=1}^{J}\boldsymbol{\mu}_{m_j}/J$.

B.2 PEM models

Let $\Phi_P = \{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \vec{\gamma}, \vec{\boldsymbol{V}}\}$ a set of parameters in the likelihood function of PEM models. The observed data likelihood $L_P(\mathcal{D}|\Phi_P)$ is given by

$$\begin{aligned}
&\prod_{j=1}^{J}\prod_{i=1}^{n_j}\left[\gamma_{ji}\eta_{ji1}\exp\left\{\sum_{k=1}^{K_1+1}\lambda_{1k}I(s_{1,k-1}<y_{ji1}\le s_{1,k})\right\}\right]^{\delta_{ji1}(1-\delta_{ji2})}\\
&\quad\times\left[\gamma_{ji}^2\eta_{ji1}\eta_{ji3}\exp\left\{\sum_{k=1}^{K_1+1}\lambda_{1k}I(s_{1,k-1}<y_{ji1}\le s_{1,k})+\sum_{k=1}^{K_3+1}\lambda_{3k}I(s_{3,k-1}<y_{ji2}\le s_{3,k})\right\}\right]^{\delta_{ji1}\delta_{ji2}}\\
&\quad\times\left[\gamma_{ji}\eta_{ji2}\exp\left\{\sum_{k=1}^{K_2+1}\lambda_{2k}I(s_{2,k-1}<y_{ji2}\le s_{2,k})\right\}\right]^{\delta_{ji2}(1-\delta_{ji1})}\\
&\quad\times\exp\left\{-r_P(y_{ji1},y_{ji2})\right\}, \qquad (15)
\end{aligned}$$

where $\eta_{jig} = \exp\left(\boldsymbol{x}_{jig}^\top\boldsymbol{\beta}_g + V_{jg}\right)$ and

$$\begin{aligned}
&r_P(t_{ji1},t_{ji2})\\
&=\begin{cases}\gamma_{ji}\left(\eta_{ji1}\sum_{k=1}^{K_1+1}e^{\lambda_{1,k}}\Delta_{jik}^1+\eta_{ji2}\sum_{k=1}^{K_2+1}e^{\lambda_{2,k}}\Delta_{jik}^2+\eta_{ji3}\sum_{k=1}^{K_3+1}e^{\lambda_{3,k}}\Delta_{jik}^{*3}\right), & \text{for Markov model}\\ \gamma_{ji}\left(\eta_{ji1}\sum_{k=1}^{K_1+1}e^{\lambda_{1,k}}\Delta_{jik}^1+\eta_{ji2}\sum_{k=1}^{K_2+1}e^{\lambda_{2,k}}\Delta_{jik}^2+\eta_{ji3}\sum_{k=1}^{K_3+1}e^{\lambda_{3,k}}\Delta_{jik}^{*3}\right), & \text{for semi-Markov model}\end{cases}
\end{aligned}$$

$$\begin{aligned}
\Delta_{jik}^g &= \max\Big\{0,\min(y_{ji1},s_{g,k})-s_{g,k-1}\Big\},\\
\Delta_{jil}^{*g} &= \begin{cases}\max\Big\{0,\min(y_{ji2},s_{g,l})-\max(y_{ji1},s_{g,l-1})\Big\}, & \text{for Markov model,}\\ \max\Big\{0,\min(y_{ji2}-y_{ji1},s_{g,l})-s_{g,l-1})\Big\}, & \text{for semi-Markov model.}\end{cases}
\end{aligned}$$

B.2.1 Reversible jump MCMC algorithm

For PEM models, we use a random scan Gibbs sampling scheme, randomly selecting and updating a (vector of) model parameter at each iteration. Let BI_g and DI_g denote a birth and a death of a new time split for transition $g \in \{1,2,3\}$. The probabilities for the update BI_g and DI_g are given by

$$\begin{aligned}
\pi_{BI_g}^{K_g} &= \rho_g\min\Big\{1,\ \frac{\mathrm{Poisson}(K_g+1|\alpha_{K_g})}{\mathrm{Poisson}(K_g|\alpha_{K_g})}\Big\} = \rho_g\min\Big\{1,\ \frac{\alpha_{K_g}}{K_g+1}\Big\},\\
\pi_{DI_g}^{K_g} &= \rho_g\min\Big\{1,\ \frac{\mathrm{Poisson}(K_g-1|\alpha_{K_g})}{\mathrm{Poisson}(K_g|\alpha_{K_g})}\Big\} = \rho_g\min\Big\{1,\ \frac{K_g}{\alpha_{K_g}}\Big\},
\end{aligned}$$

where ρ_g is set such that $\pi_{BI_g}^{K_g} + \pi_{DI_g}^{K_g} < C_g$ and $\sum_{g=1}^{3} C_g < 1$ for $K_g = 1, \ldots, K_{g,\max}$. $K_{g,\max}$ is the preassigned upper limit on the number of time splits for transition g and we set $\pi_{BI_g}^{K_{g,\max}} = 0$. The probabilities of updating other parameters are equally specified from remaining probability $1 - \sum_{g=1}^{3}(\pi_{BI_g}^{K_g} + \pi_{DI_g}^{K_g})$.

B.2.2 Updating $\boldsymbol{\beta}_g$

The full conditional posterior distribution of $\boldsymbol{\beta}_1$ can be obtained by

$$\begin{aligned}
& \pi(\boldsymbol{\beta}_1 | \Phi_P^{-(\boldsymbol{\beta}_1)}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}_\lambda^2, \theta, \Sigma_V) \\
\propto\ & L_P(D|\Phi_P) \\
\propto\ & \prod_{j=1}^{J}\prod_{i=1}^{n_j} \exp\left(\delta_{ji1} \boldsymbol{x}_{ji1}^\top \boldsymbol{\beta}_1 - \gamma_{ji} e^{\boldsymbol{x}_{ji1}^\top \boldsymbol{\beta}_1 + V_{j1}} \sum_{k=1}^{K_1+1} e^{\lambda_{1,k}} \Delta_{jik}^1 \right),
\end{aligned}$$

where $\boldsymbol{\mu}_\lambda = (\mu_{\lambda_1}, \mu_{\lambda_2}, \mu_{\lambda_3})^\top$ and $\boldsymbol{\sigma}_\lambda^2 = (\sigma_{\lambda_1}^2, \sigma_{\lambda_2}^2, \sigma_{\lambda_3}^2)^\top$. Analogously, the full conditionals of $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are given by

$$\begin{aligned}
& \pi(\boldsymbol{\beta}_2 | \Phi_P^{-(\boldsymbol{\beta}_2)}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}_\lambda^2, \theta, \Sigma_V) \\
\propto\ & \prod_{j=1}^{J}\prod_{i=1}^{n_j} \exp\left\{ \delta_{ji2}(1-\delta_{ji1}) \boldsymbol{x}_{ji2}^\top \boldsymbol{\beta}_2 - \gamma_{ji} e^{\boldsymbol{x}_{ji2}^\top \boldsymbol{\beta}_2 + V_{j2}} \sum_{l=1}^{K_2+1} e^{\lambda_{2,l}} \Delta_{jil}^2 \right\}, \\
& \pi(\boldsymbol{\beta}_3 | \Phi_P^{-(\boldsymbol{\beta}_3)}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}_\lambda^2, \theta, \Sigma_V) \\
\propto\ & \prod_{j=1}^{J}\prod_{i=1}^{n_j} \exp\left(\delta_{ji1}\delta_{ji2} \boldsymbol{x}_{ji3}^\top \boldsymbol{\beta}_3 - \gamma_{ji} e^{\boldsymbol{x}_{ji3}^\top \boldsymbol{\beta}_3 + V_{j3}} \sum_{m=1}^{K_3+1} e^{\lambda_{3,m}} \Delta_{jim}^{*3} \right).
\end{aligned}$$

As the full conditionals do not have standard forms, we use MH algorithm to update each element of $\boldsymbol{\beta}_g$. A detailed description of the adapted random walk MH algorithm is provided in Section B.1.1.

B.2.3 Updating $\boldsymbol{\lambda}_g$

The full conditional posterior distribution of $\boldsymbol{\lambda}_1$ is given by

$$\begin{aligned}
& \pi(\boldsymbol{\lambda}_1 | \Phi_P^{-(\boldsymbol{\lambda}_1)}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}_\lambda^2, \theta, \Sigma_V) \\
\propto\ & L_P(D|\Phi_P) \pi(\boldsymbol{\lambda}_1 | \mu_{\lambda_1}, \sigma_{\lambda_1}^2) \\
\propto\ & \prod_{j=1}^{J}\prod_{i=1}^{n_j} \exp\left\{ \delta_{ji1} \lambda_{1k} I(s_{1,k-1} < y_{ji1} \le s_{1,k}) - \gamma_{ji} \Delta_{jik}^1 e^{\lambda_{1k}} \eta_{ji1} \right\} \\
& \times \exp\left\{ -\frac{1}{2\sigma_{\lambda_1}^2} (\boldsymbol{\lambda}_1 - \mu_{\lambda_1}\mathbf{1})^\top \Sigma_{\lambda_1}^{-1} (\boldsymbol{\lambda}_1 - \mu_{\lambda_1}\mathbf{1}) \right\},
\end{aligned}$$

where $\mathbf{1}$ denotes a $K_g + 1$ dimensional vector of 1's. Analogously, the full conditionals of $\boldsymbol{\lambda}_2$ and $\boldsymbol{\lambda}_3$ are given by

$$\begin{aligned}
& \pi(\boldsymbol{\lambda}_2|\Phi_P^{-(\boldsymbol{\lambda}_2)}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}_\lambda^2, \theta, \Sigma_V) \\
\propto & \prod_{j=1}^{J}\prod_{i=1}^{n_j} \exp\left\{\delta_{ji2}(1-\delta_{ji1})\lambda_{2k} I(s_{2,k-1} < y_{ji2} \leq s_{2,k}) - \gamma_{ji}\Delta_{jik}^2 e^{\lambda_{2k}}\eta_{ji2}\right\} \\
& \times \exp\left\{-\frac{1}{2\sigma_{\lambda_2}^2}(\boldsymbol{\lambda}_2 - \mu_{\lambda_2}\mathbf{1})^\top \Sigma_{\lambda_2}^{-1}(\boldsymbol{\lambda}_2 - \mu_{\lambda_2}\mathbf{1})\right\}, \\
& \pi(\boldsymbol{\lambda}_3|\Phi_P^{-(\boldsymbol{\lambda}_3)}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}_\lambda^2, \theta, \Sigma_V) \\
\propto & \prod_{j=1}^{J}\prod_{i=1}^{n_j} \exp\left\{\delta_{ji1}\delta_{ji2}\lambda_{3k} I(s_{3,k-1} < y_{ji2} \leq s_{3,k}) - \gamma_{ji}\Delta_{jik}^{*3} e^{\lambda_{3k}}\eta_{ji3}\right\} \\
& \times \exp\left\{-\frac{1}{2\sigma_{\lambda_3}^2}(\boldsymbol{\lambda}_3 - \mu_{\lambda_3}\mathbf{1})^\top \Sigma_{\lambda_3}^{-1}(\boldsymbol{\lambda}_3 - \mu_{\lambda_3}\mathbf{1})\right\},
\end{aligned}$$

Since the full conditionals do not follow known distributions, MH algorithm is used to update each element of $\boldsymbol{\lambda}_g$. We follow the adapted random walk MH algorithm described in Section B.1.1.

B.2.4 Updating γ_{ji}

The full conditional posterior distribution of γ_{ji} is given by

$$\begin{aligned}
& \pi(\gamma_{ji}|\Phi_P^{-(\gamma_{ji})}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}_\lambda^2, \theta, \Sigma_V) \\
\propto & \; L_P(D|\Phi_P) \times \pi(\gamma_{ji}|\theta) \\
\propto & \; \gamma_{ji}^{\delta_{ji1}+\delta_{ji2}+\theta^{-1}-1} \exp\left[-r_P(y_{ji1}, y_{ji2}) - \theta^{-1}\gamma_{ji}\right].
\end{aligned}$$

Therefore, we sample γ_{ji} from

$$\text{Gamma}\left(\delta_{ji1} + \delta_{ji2} + \theta^{-1},\; r_P(y_{ji1}, y_{ji2}; \gamma_{ji} = 1) + \theta^{-1}\right).$$

B.2.5 Updating (μ_g, σ_g^2)

Full conditional posterior distributions for μ_{λ_g} and $\upsilon_g = 1/\sigma_{\lambda_g}^2$, $g = 1, 2, 3$ are Normal and Gamma distribution, respectively. Therefore, we use Gibbs sampling to update the parameters. We obtain the posterior samples of μ_{λ_g} from

$$\text{Normal}\left(\frac{\mathbf{1}^\top \Sigma_{\lambda_g}^{-1}\boldsymbol{\lambda}_g}{\mathbf{1}^\top \Sigma_{\lambda_g}^{-1}\mathbf{1}},\; \frac{\sigma_{\lambda_g}^2}{\mathbf{1}^\top \boldsymbol{\Sigma}_{\boldsymbol{\lambda}_\mathbf{g}}^{-\mathbf{1}}\mathbf{1}}\right),$$

because the full conditional is given by

$$\begin{aligned}
&\pi(\mu_{\lambda_g}|\Phi_P, \boldsymbol{\mu}_\lambda^{-(\mu_{\lambda_g})}, \boldsymbol{\sigma}^2_\lambda, \theta, \Sigma_V) \propto \pi(\boldsymbol{\lambda}_g|\mu_{\lambda_g}, \sigma^2_{\lambda_g})\pi(\mu_{\lambda_g}) \\
\propto \quad & \exp\left\{\frac{\mathbf{1}^\top\Sigma^{-1}_{\lambda_g}\mathbf{1}}{2\sigma^2_{\lambda_g}}\left(\mu_{\lambda_g} - \frac{\mathbf{1}^\top\Sigma^{-1}_{\lambda_g}\boldsymbol{\lambda}_g}{\mathbf{1}^\top\Sigma^{-1}_{\lambda_g}\mathbf{1}}\right)^2\right\}.
\end{aligned}$$

We update $\upsilon_g = 1/\sigma^{-2}_{\lambda_g}$ from a Gamma distribution given by

$$\text{Gamma}\left(a_{\sigma,g} + \frac{K_g+1}{2},\ b_{\sigma,g} + \frac{1}{2}(\mu_{\lambda_g}\mathbf{1} - \boldsymbol{\lambda}_g)^\top\boldsymbol{\Sigma}^{-1}_{\lambda_g}(\mu_{\lambda_g}\mathbf{1} - \boldsymbol{\lambda}_g)\right),$$

as the full conditional of υ_g is

$$\begin{aligned}
& \pi(\upsilon_g|\Phi_P, \boldsymbol{\mu}_\lambda, (\boldsymbol{\sigma}^2_\lambda)^{-(\sigma^2_{\lambda_g})}, \theta, \Sigma_V) \\
\propto \quad & \pi(\boldsymbol{\lambda}_g|\mu_{\lambda_g}, \sigma^2_{\lambda_g})\pi(\upsilon_g) \\
\propto \quad & (\upsilon_g)^{a_{\sigma,g}+\frac{K_g+1}{2}-1}\exp\left[-\left\{b_{\sigma,g} + \frac{1}{2}(\mu_{\lambda_g}\mathbf{1} - \boldsymbol{\lambda}_g)^\top\boldsymbol{\Sigma}^{-1}_{\lambda_g}(\mu_{\lambda_g}\mathbf{1} - \boldsymbol{\lambda}_g)\right\}\upsilon_g\right].
\end{aligned}$$

B.2.6 Updating θ

Updating the precision parameter $\xi = 1/\theta$ in PEM models requires the exactly same step as that in Weibull models. Therefore, the readers are referred to Section B.1.5 for the full conditional posterior distribution of ξ and the MH algorithm.

B.2.7 Updating $\boldsymbol{V}_j$ for PEM-MVN model

The full conditional posterior distribution of V_{j1} can be obtained by

$$\begin{aligned}
\pi(V_{j1}|\Phi_P^{-(V_{j1})}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}^2_\lambda, \theta, \Sigma_V) \quad &\propto \quad L_P(D|\Phi_P) \times \pi(\boldsymbol{V}_j|\Sigma_V). \\
&\propto \quad \exp\left\{\sum_{i=1}^{n_j}\left(V_{j1}\delta_{ji1} - \gamma_{ji}\eta_{ji1}\sum_{k=1}^{K_1+1}e^{\lambda_{1,k}}\Delta^1_{jik}\right) - \frac{1}{2}\boldsymbol{V}_j^\top\Sigma_V^{-1}\boldsymbol{V}_j\right\}.
\end{aligned}$$

Analogously, the full conditionals of V_{j2} and V_{j3} can be written as

$$\begin{aligned}
\pi(V_{j2}|\Phi_P^{-(V_{j2})}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}^2_\lambda, \theta, \Sigma_V) \quad &\propto \quad \exp\left\{\sum_{i=1}^{n_j}\left(V_{j2}\delta_{ji2}(1-\delta_{ji1}) - \gamma_{ji}\eta_{ji2}\sum_{l=1}^{K_2+1}e^{\lambda_{2,l}}\Delta^2_{jil}\right) - \frac{1}{2}\boldsymbol{V}_j^\top\Sigma_V^{-1}\boldsymbol{V}_j\right\}, \\
\pi(V_{j3}|\Phi_P^{-(V_{j3})}, \boldsymbol{\mu}_\lambda, \boldsymbol{\sigma}^2_\lambda, \theta, \Sigma_V) \quad &\propto \quad \exp\left\{\sum_{i=1}^{n_j}\left(V_{j3}\delta_{ji1}\delta_{ji2} - \gamma_{ji}\eta_{ji3}\sum_{m=1}^{K_3+1}e^{\lambda_{3,m}}\Delta^{*3}_{jim}\right) - \frac{1}{2}\boldsymbol{V}_j^\top\Sigma_V^{-1}\boldsymbol{V}_j\right\}.
\end{aligned}$$

For updating each element of $\boldsymbol{V}_j$, we use the adapted random walk MH algorithm provided in Section B.1.6.

B.2.8 Updating Σ_V for PEM-MVN model

The full conditional posterior distribution of Σ_V in PEM-MVN model is the exactly same as that in Weibull-MVN model. Readers are referred to Section B.1.7 for the Gibbs sampling step for updating Σ_V.

B.2.9 Updating $\boldsymbol{V}_j$ and Σ_V for PEM-DPM model

Updating $\boldsymbol{V}_j$ and Σ_V in PEM-DPM requires the exactly same step as that in Weibull-DPM. Therefore, the readers are referred to Section B.1.8 for detailed algorithm to update $\boldsymbol{V}_j$ and Σ_V. Note that in step 3 of the algorithm, the full conditional of $\boldsymbol{V}_j$ needs to be obtained based on $L_P(D|\Phi_P)$ for PEM-DPM.

B.2.10 Update BI

We specify $\log h_0(t) = \sum_{k=1}^{K_g+1} \lambda_{g,k} I(t \in I_{g,k})$ for the baseline hazard function corresponding to transition g with partition $(K_g,\ \boldsymbol{s}_g)$. Updating $(K_g,\ \boldsymbol{s}_g)$ requires generating a proposal partition and then deciding whether or not to accept the proposal. For update BI (a birth move), we first select a proposal split time s^* uniformly from among the observed event times which are not included in the current partition. Suppose s^* lies between the $(k-1)^{th}$ and k^{th} split times in the current partition. The proposal partition is then defined as

$$\begin{aligned} & (0 = s_{g,0}, ..., s_{g,k-1}, s^*, s_{g,k}, ..., s_{g,K_1+1} = s_{g,\max}) \\ \equiv\ & (0 = s^*_{g,0}, ..., s^*_{g,k-1}, s^*_{g,k}, s^*_{g,k+1}, ..., s^*_{g,K_1+2} = s_{g,\max}). \end{aligned}$$

A height of the two new intervals created by the split at time s^* also needs to be proposed. In order to make the old height be a compromise of the two new heights, the former is taken to be the weighted mean of the latter on the log scale:

$$(s^* - s_{g,k-1})\lambda^*_{g,k} + (s_{g,k} - s^*)\lambda^*_{g,k+1} = (s_{g,k} - s_{g,k-1})\lambda_{g,k}.$$

Defining the multiplicative perturbation $\exp(\lambda^*_{g,j+1})/\exp(\lambda^*_{g,j}) = (1-U)/U$, where $U \sim \text{Uniform}(0,1)$, the new heights are given by

$$\lambda^*_{g,k} \ = \ \lambda_{g,k} \ - \ \frac{s_{g,k} - s^*}{s_{g,k} - s_{g,k-1}} \log\left(\frac{1-U}{U}\right)$$

and

$$\lambda^*_{g,k+1} \;=\; \lambda_{g,k} \;+\; \frac{s^* - s_{g,k-1}}{s_{g,k} - s_{g,k-1}} \log\left(\frac{1-U}{U}\right).$$

The acceptance probability in the Metropolis-Hastings-Green step can be written as the product of the likelihood ratio, prior ratio, proposal ratio, and Jacobian. For $g = 1$, they are given by

$$\begin{aligned}
\textit{likelihood ratio} &= \frac{L_P(D|\Phi^*_P)}{L_P(D|\Phi_P)}, \\
\textit{prior ratio} &= \frac{\text{Poisson}(K_1+1|\alpha_{K_1}) \times \text{MVN}_{K_1+2}(\boldsymbol{\lambda}^*_1|\mu_{\lambda_1}\mathbf{1},\ \sigma^2_{\lambda_1}\Sigma^*_{\lambda_1})}{\text{Poisson}(K_1|\alpha_{K_1}) \times \text{MVN}_{K_1+1}(\boldsymbol{\lambda}_1|\mu_{\lambda_1}\mathbf{1},\ \sigma^2_{\lambda_1}\Sigma_{\lambda_1})} \\
&\quad \times \frac{(2K_1+3)(2K_1+2)(s^* - s_{1,k-1})(s_{1,k} - s^*)}{s^2_{1,\max}(s_{1,k} - s_{1,k-1})}, \\
\textit{proposal ratio} &= \frac{\pi_{DI} \times (1/(K_1+1))}{\pi_{BI} \times (1/\sharp\{y_{ji1} : \delta_{ji1} = 1\}) \times \text{Uniform}(U|0,1)} \\
&= \frac{\rho \min(1, (K_1+1)/\alpha_{K_1}) \sharp\{y_{ji1} : \delta_{ji1} = 1\}}{\rho \min(1, \alpha_{K_1}/(1+K_1))(K_1+1)} = \frac{\sharp\{y_{ji1} : \delta_{ji1} = 1\}}{\alpha_{K_1}}, \\
\textit{Jacobian} &= \left| \begin{array}{cc} d\lambda^*_{1,k}/d\lambda_{1,k} & d\lambda^*_{1,k}/dU \\ d\lambda^*_{1,k+1}/d\lambda_{1,k} & d\lambda^*_{1,k+1}/dU \end{array} \right| = \frac{1}{U(1-U)},
\end{aligned} \tag{16}$$

where Φ^*_P is Φ_P with λ_1 replaced by λ^*_1.

B.2.11 Update DI

For update DI (a death or reverse move), we first sample one of the K_g split times, $s_{g,k}$. The proposal for time splits is given by

$$\begin{aligned}
& (0 = s_{g,0}, ..., s_{g,k-1}, s_{g,k+1}, ..., s_{g,K_g+1} = s_{g,\max}) \\
\equiv\ & (0 = s^*_{g,0}, ..., s^*_{g,k-1}, s^*_{g,k}, ..., s^*_{g,K_g} = s_{g,\max}).
\end{aligned}$$

Following Green (1995):

$$\begin{aligned}
(s_{g,k} - s_{g,k-1})\lambda_{g,k} + (s_{g,k+1} - s_{g,k})\lambda_{g,k+1} &= (s_{g,k+1} - s_{g,k-1})\lambda^*_{g,k}, \\
\textit{perturbation}: \qquad \frac{e^{\lambda_{g,k+1}}}{e^{\lambda_{g,k}}} &= \frac{1-U^*}{U^*}.
\end{aligned}$$

Then the acceptance probability can be obtained as the product of following four components (for $g = 1$):

$$
\begin{aligned}
\textit{likelihood ratio} &= \frac{L_P(D|\Phi_P^*)}{L_P(D|\Phi_P)}, \\
\textit{prior ratio} &= \frac{\text{Poisson}(K_1 - 1|\alpha_{K_1}) \times \text{MVN}_{K_1}(\boldsymbol{\lambda}_1^*|\mu_{\lambda_1}\mathbf{1},\ \sigma_{\lambda_1}^2\Sigma_{\lambda_1}^*)}{\text{Poisson}(K_1|\alpha_{K_1}) \times \text{MVN}_{K_1+1}(\boldsymbol{\lambda}_1|\mu_{\lambda_1}\mathbf{1},\ \sigma_{\lambda_1}^2\Sigma_{\lambda_1})} \\
&\quad \times \frac{s_{1,\max}^2(s_{1,k+1} - s_{1,k-1})}{(2K_1+1)2K_1(s_{1,k} - s_{1,k-1})(s_{1,k+1} - s_{1,k})}, \\
\textit{proposal ratio} &= \frac{\pi_{BI} \times (1/\sharp\{y_{ji1} : \delta_{ji1} = 1\})}{\pi_{DI} \times (1/K_1)} \\
&= \frac{\rho \min(1, \alpha_{K_1}/K_1)K_1}{\rho \min(1, K_1/\alpha_{K_1})\sharp\{y_{ji1} : \delta_{ji1} = 1\}} = \frac{\alpha_{K_1}}{\sharp\{y_{ji1} : \delta_{ji1} = 1\}}, \\
\textit{Jacobian} &= \begin{vmatrix} d\lambda_{1k}/d\lambda_{1k}^* & d\lambda_{1k}/d\lambda_{1,k+1}^* \\ dU^*/d\lambda_{1k}^* & dU^*/d\lambda_{1,k+1}^* \end{vmatrix} = U^*(1 - U^*).
\end{aligned}
$$

C The potential use of existing methods

The methods in the main manuscript were developed specifically for on-going collaboration examining the risk of readmission following a diagnosis of pancreatic cancer. As indicated in the manuscript, the current standard for the analysis of cluster-correlated readmission data is a logisitic-Normal generalized linear mixed model. This model ignores death as a competing risk, however, and, as such, is inappropriate in for the study of pancreatic cancer due to its strong force of mortality.

Viewing the data arising in the pancreatic cancer as *cluster-correlated semi-competing risks data*, the existing literature does have a number of options that could be considered. Here we review two of these options, specifically those proposed in Liquet et al. (2012) and Gorfine and Hsu (2011). For the former, we note that the methods have been implemented in the `frailtypack` package for `R`.

For convenience, expressions (4)-(6) from the main manuscript that the describe the key features of the proposed hierarchical model are repeated here:

$$\begin{aligned}
h_1(t_{ji1};\gamma_{ji},\boldsymbol{X}_{ji1},V_{j1}) &= \gamma_{ji}\; h_{01}(t_{ji1})\; \exp\{\boldsymbol{X}_{ji1}^T\boldsymbol{\beta}_1 \;+\; V_{j1}\}, \qquad t_{ji1}>0 \\
h_2(t_{ji2};\gamma_{ji},\boldsymbol{X}_{ji2},V_{j2}) &= \gamma_{ji}\; h_{02}(t_{ji2})\; \exp\{\boldsymbol{X}_{ji2}^T\boldsymbol{\beta}_2 \;+\; V_{j2}\}, \qquad t_{ji2}>0 \\
h_3(t_{ji2}|t_{ji1};\gamma_{ji},\boldsymbol{X}_{ji3},V_{j3}) &= \gamma_{ji}\; h_{03}(t_{ji2}|t_{ji1})\; \exp\{\boldsymbol{X}_{ji3}^T\boldsymbol{\beta}_3 \;+\; V_{j3}\}, \qquad t_{ji2}>t_{ji1},
\end{aligned}$$

C.1 Liquet et al. (2012)

The R package `frailtypack` provides several classes of frailty models for multivariate survival data including shared frailty models, additive frailty models, nested frailty models, joint frailty models (Rondeau et al., 2012). Among these, the *shared frailty model* and the *joint frailty model* are most relevant the context we consider; additionally, these models form the basis for the analyses presented in Liquet et al. (2012). Here we provide a summary of these two classes using the notation developed in the manuscript, as well as an overview of their drawbacks in regard to the analysis of cluster-correlated semi-competing risks data.

C.1.1 The shared frailty model

In the shared frailty model, the hazard function for the subject i in the cluster j conditional on the cluster-specific shared frailty term $\eta_j = (\eta_{j1}, \eta_{j2}, \eta_{j3})$ is given by

$$\begin{aligned} h_1(t_{ji1}; \boldsymbol{X}_{ji1}, \eta_{j1}) &= \eta_{j1} h_{01}(t_{ji1}) \exp\{\boldsymbol{X}_{ji1}^T \boldsymbol{\beta}_1\}, \qquad t_{ji1} > 0 \\ h_2(t_{ji2}; \boldsymbol{X}_{ji2}, \eta_{j2}) &= \eta_{j2} h_{02}(t_{ji2}) \exp\{\boldsymbol{X}_{ji2}^T \boldsymbol{\beta}_2\}, \qquad t_{ji2} > 0 \\ h_3(t_{ji2}|t_{ji1}; \boldsymbol{X}_{ji3}, \eta_{j3}) &= \eta_{j3} h_{03}(t_{ji2} - t_{ji1}) \exp\{\boldsymbol{X}_{ji3}^T \boldsymbol{\beta}_3\}, \qquad t_{ji2} > t_{ji1}, \end{aligned} \tag{17}$$

Key features of this model, in relation to the proposed framework are:

- Cluster-specific effects are represented via the $(\eta_{j1}, \eta_{j2}, \eta_{j3})$, each of which is assigned an independent univariate parametric distribution (either a log-Normal or a Gamma). As such, the model does not permit the characterization of covariation between the cluster-specific random effects. In contrast, the proposed methods provides analysts with two choices for the joint distribution of the V_j's: a parametric MVN or a non-parametric DPM-MVN.

- There is no patient-specific term, analogous to the γ_{ji} in the proposed model. As such a potentially important source of within-subject correlation between T_1 and T_2 is not accounted for.

- Similar to the propose methods, however, is that the baseline hazard function for $h_3()$ can be specified non-parametrically (via a spline) or parametrically (using the Weibull distribution).

- Although not evident from the model specification, estimation of the shared frailty model is based on three separate fits of the three models. In contrast, because the proposed model considers several components of covariation (i.e. covariation among the V_j's and the patient-specific γ_{ji}'s) we perform estimation/inference using single likelihood. Indeed for the shared frailty model to accommodate these components of covariation, a completely new framework for estimation/inference would need to be developed.

- Estimation of the $(\eta_{j1}, \eta_{j2}, \eta_{j3})$ proceeds using empirical Bayes (after estimation of the remaining components via an integrated likelihood). Uncertainty for these estimates are only available when their distributions are taken to be Gamma distributions.

C.1.2 The joint frailty model

Two variations of a joint frailty model have been implemented in the `frailtypack` package. The first was developed for the analysis of a recurrent non-terminal event and a terminal event and specifies a single hazard function for each. Specifically, the model is given by:

$$\begin{aligned} h_1(t_{ki1}|\omega_i) &= \omega_i r_0(t_{ki1}) \exp\{\boldsymbol{X}_{i1}^T\boldsymbol{\beta}_1\}, \quad \text{for recurrent non-terminal event} \\ h_2(t_{i2}|\omega_i) &= \omega_i^{\alpha} h_0(t_{i2}) \exp\{\boldsymbol{X}_{i2}^T\boldsymbol{\beta}_2\}, \quad \text{for the terminal event} \\ \omega_i &\sim \text{Gamma}(1/\theta, 1/\theta). \end{aligned} \tag{18}$$

where ω_i is a common subject-specific frailty representing unobserved covariates that impact both events. We note that this is specification is similar to the model proposed by Liu et al. (2004).

The second joint frailty model implemented in the `frailtypack` package is for modeling two clustered survival outcomes. Specifically, the model posits that the event-specific hazard functions for the j^{th} cluster are structured as follows:

$$\begin{aligned} h(t_{ji1}|\eta_j) &= h_{01}(t_{ji1}) \exp\{\boldsymbol{X}_{ji1}^T\boldsymbol{\beta}_1 + \eta_j\}, \quad \text{for any event} \\ h(t_{ji2}|\eta_j) &= h_{02}(t_{ji2}) \exp\{\boldsymbol{X}_{ji2}^T\boldsymbol{\beta}_2 + \alpha\eta_j\}, \quad \text{for the terminal event} \\ \eta_j &\sim \text{Normal}(0, \sigma^2) \end{aligned} \tag{19}$$

In relation to the context we consider, the central limitation of these models is that they only consider a single level of the two-level hierarchy inherent to cluster-correlated semi-competing risks data. Specifically, as applied and described in the `fraitypack` package, the first model only considers patient-specific effects while the second model only considers cluster-specific effects. As such neither model would be appropriate for our motivating application since (i) ignoring cluster-specific effects means that one cannot address several of our key scientific questions and (ii) ignoring patient-level effects can result in substantial bias (see the simulation studies in Section 5 of the main manuscript).

We also note that a second limitation is that model (19) does not consider the transition from the non-terminal event to the terminal event; that is there is no analogue for $h_3()$ in the model. This represents a limitation in the sense that information readily available in the data is ignored. In the motivating application in the main manuscript, for example, the fact that the time of death following readmission within 90 days is known for 608 (11.5%)

patients is ignored. Finally, although model (18) does permit a patient to transition from the non-terminal state to the terminal state, this transition is assumed to occur at the same rate at which a patient who is in the initial state transitions directly into the terminal state; that is, in contrast to the proposed model that distinguishes $h_2()$ from $h_3()$, model (18) only has a single hazard for the terminal event.

C.2 Gorfine and Hsu (2011)

Gorfine and Hsu (2011) explicitly consider the related but distinct problem of analyzing *cluster correlated competing risk data* for which T_1 and T_2 are both terminal events (i.e. death due to two causes). Towards analyzing such data, they propose the following hierarchical model:

$$\begin{aligned} h(t_{ji1}|\boldsymbol{X}_{ji}, \epsilon_{j1}(t_{ji1})) &= h_{01}(t_{ji1}) \exp\{\boldsymbol{X}_{ji}^T\boldsymbol{\beta}_1 + \epsilon_{j1}(t_{ji1})\}, \quad \text{for cause 1} \\ h(t_{ji2}|\boldsymbol{X}_{ji}, \epsilon_{j2}(t_{ji2})) &= h_{02}(t_{ji2}) \exp\{\boldsymbol{X}_{ji}^T\boldsymbol{\beta}_2 + \epsilon_{j2}(t_{ji2})\}, \quad \text{for cause 2} \end{aligned} \tag{20}$$

to describe the risk of transitioning into one of the two terminal states for the i^{th} patient in the j^{th} cluster. As part of their development, Gorfine and Hsu (2011) provide a framework within which the distribution of the cluster-specific $\epsilon_{jg}(t)$ terms can be flexibly specified. While this flexibility is very appealing, direct application of this model to our motivating application would be subject to a number of limitations mainly because the method was not designed for the cluster-correlated semi-competing risks setting. Specifically,

- Similar to the joint frailty model given by (19), the application of model (20) means that one would ignore information in the data on the transition from the non-terminal event to the terminal event; that is, there is not analogue of $h_3()$.

- Although model (20) includes cluster-specific random effects, it does not include specification of patient-specific terms analogous to γ_{ji} in the proposed model. As is clear from the simulations presented in Section 5 of the main manuscript, ignoring this component of variation can lead to substantial bias in estimation and poor inferential properties in the cluster-correlated semi-competing risks setting.

D Simulation Results

In order to supplement the results from simulation studies, we provide estimated percent bias, coverage probability, and average relative width of 95% credible/confidence intervals for $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, and θ for our four proposed models and four types of SF models of Liquet et al. (2012) in Table D.1-D.6. We also provide estimated transition-specific baseline survival functions for the models under simulation scenarios 2,3, and 5 in Figure D.1. Note that since results from SF models are almost identical between models that adopt the independent gamma distributions for cluster-specific random effects and those that adopt the independent log-Normal distributions, we only present the results from SF models with the gamma cluster-specific random effects in Figure D.1. We also present Table D.7 that augments Table 6 in the main manuscript by additionally presenting results for the Liquet et al. (2012)'s models that adopt independent log-Normal distributions for the cluster-specific random effects.

The results presented in this section are generally consistent with the conclusions we drew in the main paper: contrary to the existing SF models, our proposed models yielded a small bias and coverage probability estimated to be close to the nominal 0.95 for regression parameters and θ (except scenario 4 for which θ=0); all four of the proposed models estimate the three baseline survival functions very well.

Table D.1: Estimated percent bias and coverage probability for $\boldsymbol{\beta}_1$ and θ for six analyses described in Section 5.2, across six simulation scenarios given in Table 3. Throughout values are based on results from R=500 simulated datasets.

			Percent Bias								Coverage Probability							
Scenario		True	Weibull	Weibull	Weibull	Weibull	PEM	PEM	Spline	Spline	Weibull	Weibull	Weibull	Weibull	PEM	PEM	Spline	Spline
		value	-MVN	-DPM	-SF$_{\mathcal{G}}$[a]	-SF$_{\mathcal{LN}}$[b]	-MVN	-DPM	-SF$_{\mathcal{G}}$	-SF$_{\mathcal{LN}}$	-MVN	-DPM	-SF$_{\mathcal{G}}$	-SF$_{\mathcal{LN}}$	-MVN	-DPM	-SF$_{\mathcal{G}}$	-SF$_{\mathcal{LN}}$
1	β_{11}	0.50	0.1	0.2	-19.8	-21.3	0.4	0.4	-21.0	-20.8	0.96	0.96	0.01	0.01	0.95	0.96	0.00	0.00
	β_{12}	0.80	0.2	0.3	-19.7	-21.3	0.5	0.4	-21.0	-20.8	0.95	0.95	0.00	0.00	0.96	0.97	0.00	0.00
	β_{13}	-0.50	0.3	0.3	-19.8	-18.8	0.3	0.3	-21.2	-20.9	0.97	0.96	0.31	0.31	0.96	0.96	0.25	0.26
	θ	0.50	1.0	1.3			1.4	1.2			0.95	0.95			0.93	0.94		
2	β_{11}	0.50	-0.1	-0.0	-31.8	-33.4	0.1	0.1	-32.8	-32.8	0.94	0.94	0.00	0.00	0.94	0.93	0.00	0.00
	β_{12}	0.80	0.1	0.2	-31.7	-33.3	0.4	0.3	-32.7	-32.7	0.97	0.97	0.00	0.00	0.94	0.95	0.00	0.00
	β_{13}	-0.50	1.2	1.3	-31.1	-29.2	1.1	1.1	-32.2	-32.2	0.94	0.95	0.05	0.05	0.94	0.94	0.04	0.04
	θ	1.00	0.4	0.7			0.7	0.6			0.94	0.95			0.94	0.95		
3	β_{11}	0.50	0.3	0.3	-19.9	-20.7	0.7	0.7	-21.0	-20.9	0.94	0.94	0.00	0.00	0.93	0.94	0.00	0.00
	β_{12}	0.80	0.4	0.4	-19.8	-20.7	0.8	0.8	-20.9	-20.8	0.94	0.94	0.00	0.00	0.94	0.94	0.00	0.00
	β_{13}	-0.50	0.4	0.3	-20.1	-19.7	0.5	0.6	-21.2	-21.2	0.96	0.96	0.31	0.29	0.95	0.96	0.27	0.27
	θ	0.50	2.0	2.1			3.2	3.2			0.96	0.96			0.93	0.95		
4	β_{11}	0.50	3.7	3.7	0.2	-2.9	4.7	4.6	0.3	0.6	0.87	0.86	0.96	0.91	0.81	0.83	0.96	0.95
	β_{12}	0.80	3.6	3.6	-0.0	-3.1	4.5	4.5	0.1	0.4	0.80	0.79	0.95	0.89	0.69	0.70	0.95	0.95
	β_{13}	-0.50	4.0	4.0	0.2	7.3	4.8	4.7	0.2	0.6	0.93	0.94	0.94	0.88	0.93	0.93	0.93	0.94
	θ	0.00																
5	β_{11}	0.50	-0.3	0.1	-20.3	-24.6	0.0	0.3	-21.1	-20.9	0.94	0.95	0.00	0.00	0.96	0.96	0.00	0.00
	β_{12}	0.80	0.0	0.3	-20.0	-24.6	0.3	0.6	-20.9	-20.7	0.95	0.95	0.00	0.00	0.96	0.96	0.00	0.00
	β_{13}	-0.50	-0.2	0.2	-20.4	-13.7	-0.2	0.2	-21.3	-21.1	0.94	0.94	0.29	0.26	0.94	0.94	0.25	0.26
	θ	0.50	-0.2	1.0			0.4	1.3			0.95	0.95			0.95	0.96		
6	β_{11}	0.50	9.3	9.4	-22.1	-23.6	0.4	0.3	-25.9	-25.1	0.58	0.57	0.00	0.00	0.94	0.94	0.00	0.00
	β_{12}	0.80	9.7	9.8	-22.0	-23.5	0.5	0.5	-25.8	-25.0	0.20	0.20	0.00	0.00	0.94	0.95	0.00	0.00
	β_{13}	-0.50	10.2	10.2	-21.6	-18.2	0.8	0.7	-26.1	-24.9	0.81	0.80	0.21	0.21	0.93	0.94	0.10	0.10
	θ	0.50	52.8	53.0			1.8	1.7			0.00	0.00			0.95	0.96		

[a] The SF models that adopt the independent gamma distributions for cluster-specific random effects

[b] The SF models that adopt the independent log-Normal distributions for cluster-specific random effects

Table D.2: Estimated percent bias and coverage probability for $\boldsymbol{\beta}_2$ for six analyses described in Section 5.2, across six simulation scenarios given in Table 3. Throughout values are based on results from R=500 simulated datasets.

			Percent Bias								Coverage Probability							
Scenario		True value	Weibull -MVN	Weibull -DPM	Weibull -SF$_{\mathcal{G}}$[a]	Weibull -SF$_{\mathcal{LN}}$[b]	PEM -MVN	PEM -DPM	Spline -SF$_{\mathcal{G}}$	Spline -SF$_{\mathcal{LN}}$	Weibull -MVN	Weibull -DPM	Weibull -SF$_{\mathcal{G}}$	Weibull -SF$_{\mathcal{LN}}$	PEM -MVN	PEM -DPM	Spline -SF$_{\mathcal{G}}$	Spline -SF$_{\mathcal{LN}}$
1	β_{21}	0.50	-0.1	-0.0	-27.9	-25.9	-0.2	-0.3	-26.1	-26.1	0.93	0.93	0.00	0.00	0.94	0.94	0.00	0.00
	β_{22}	0.80	0.3	0.4	-27.9	-25.6	0.2	0.1	-25.8	-25.7	0.96	0.96	0.00	0.00	0.96	0.96	0.00	0.00
	β_{23}	-0.50	1.3	1.4	-26.5	-21.1	0.9	0.9	-25.2	-25.1	0.94	0.93	0.34	0.35	0.95	0.95	0.34	0.34
2	β_{21}	0.50	-0.1	0.1	-39.2	-39.2	-0.2	-0.2	-39.9	-39.8	0.96	0.96	0.00	0.00	0.96	0.96	0.00	0.00
	β_{22}	0.80	0.2	0.3	-39.1	-39.0	0.0	0.0	-39.6	-39.6	0.96	0.97	0.00	0.00	0.96	0.97	0.00	0.00
	β_{23}	-0.50	1.3	1.4	-38.5	-38.3	0.8	0.8	-39.2	-39.2	0.93	0.94	0.07	0.07	0.93	0.93	0.06	0.06
3	β_{21}	0.50	0.3	0.3	-27.4	-24.9	0.3	0.3	-25.7	-25.6	0.95	0.95	0.01	0.01	0.95	0.96	0.00	0.00
	β_{22}	0.80	0.5	0.5	-27.5	-24.8	0.5	0.5	-25.5	-25.4	0.95	0.95	0.00	0.00	0.94	0.95	0.00	0.00
	β_{23}	-0.50	2.5	2.5	-24.9	-23.2	2.3	2.5	-24.2	-24.1	0.95	0.95	0.37	0.38	0.94	0.94	0.35	0.36
4	β_{21}	0.50	4.5	4.5	-4.9	-4.4	4.7	4.7	-0.6	-0.6	0.89	0.89	0.91	0.90	0.87	0.88	0.96	0.96
	β_{22}	0.80	4.6	4.6	-5.2	-4.6	4.8	4.8	-0.6	-0.5	0.78	0.78	0.88	0.88	0.76	0.79	0.92	0.92
	β_{23}	-0.50	5.4	5.4	25.8	18.1	5.4	5.5	-0.1	-0.0	0.92	0.92	0.88	0.88	0.91	0.93	0.93	0.92
5	β_{21}	0.50	-0.3	0.2	-26.6	-25.1	-0.4	-0.0	-26.1	-25.8	0.94	0.95	0.00	0.00	0.94	0.95	0.00	0.00
	β_{22}	0.80	-0.4	0.1	-26.8	-25.3	-0.5	-0.1	-26.2	-25.9	0.96	0.96	0.00	0.00	0.96	0.96	0.00	0.00
	β_{23}	-0.50	0.1	0.5	-27.2	-24.9	-0.4	0.1	-26.1	-25.8	0.95	0.95	0.33	0.35	0.95	0.95	0.30	0.31
6	β_{21}	0.50	9.2	9.3	-25.1	-23.8	0.1	0.1	-25.0	-25.0	0.70	0.70	0.01	0.01	0.95	0.95	0.00	0.01
	β_{22}	0.80	9.7	9.7	-25.4	-23.8	0.4	0.3	-24.8	-24.8	0.40	0.39	0.00	0.00	0.94	0.95	0.00	0.00
	β_{23}	-0.50	10.4	10.4	-26.6	-13.6	0.8	0.8	-24.5	-24.5	0.85	0.85	0.47	0.47	0.94	0.95	0.36	0.35

[a] The SF models that adopt the independent gamma distributions for cluster-specific random effects

[b] The SF models that adopt the independent log-Normal distributions for cluster-specific random effects

Table D.3: Estimated percent bias and coverage probability for $\boldsymbol{\beta}_3$ for six analyses described in Section 5.2, across six simulation scenarios given in Table 3. Throughout values are based on results from R=500 simulated datasets.

			Percent Bias								Coverage Probability							
Scenario		True value	Weibull -MVN	Weibull -DPM	Weibull -SF$_{\mathcal{G}}$[a]	Weibull -SF$_{\mathcal{LN}}$[b]	PEM -MVN	PEM -DPM	Spline -SF$_{\mathcal{G}}$	Spline -SF$_{\mathcal{LN}}$	Weibull -MVN	Weibull -DPM	Weibull -SF$_{\mathcal{G}}$	Weibull -SF$_{\mathcal{LN}}$	PEM -MVN	PEM -DPM	Spline -SF$_{\mathcal{G}}$	Spline -SF$_{\mathcal{LN}}$
1	β_{31}	1.00	0.4	0.5	-21.8	-12.5	0.7	0.8	-13.3	-13.2	0.94	0.94	0.08	0.09	0.94	0.94	0.06	0.06
	β_{32}	1.00	0.2	0.3	-20.5	-9.0	0.6	0.6	-9.7	-9.7	0.96	0.96	0.28	0.32	0.93	0.94	0.27	0.28
	β_{33}	-1.00	0.2	0.3	44.3	-12.5	0.4	0.4	-13.4	-13.3	0.94	0.94	0.47	0.53	0.94	0.94	0.49	0.50
2	β_{31}	1.00	0.1	0.3	-24.1	-25.3	0.6	0.6	-24.8	-24.7	0.95	0.95	0.00	0.00	0.95	0.95	0.00	0.00
	β_{32}	1.00	0.4	0.5	-22.8	-24.1	0.8	0.8	-23.3	-23.3	0.94	0.94	0.00	0.00	0.95	0.95	0.00	0.00
	β_{33}	-1.00	0.2	0.4	-23.8	-22.6	0.4	0.4	-24.6	-24.5	0.96	0.96	0.08	0.07	0.95	0.95	0.06	0.07
3	β_{31}	1.00	0.6	0.6	-21.4	-14.4	1.1	1.1	-12.6	-12.5	0.96	0.96	0.10	0.14	0.95	0.95	0.09	0.09
	β_{32}	1.00	0.7	0.8	-19.2	-11.2	1.2	1.2	-9.0	-8.9	0.95	0.95	0.35	0.40	0.94	0.94	0.37	0.38
	β_{33}	-1.00	0.3	0.3	13.1	-10.1	0.5	0.6	-12.6	-12.5	0.95	0.95	0.52	0.57	0.93	0.94	0.54	0.54
4	β_{31}	1.00	3.4	3.4	-0.6	11.9	4.1	3.9	11.0	11.3	0.87	0.88	0.11	0.12	0.84	0.86	0.15	0.15
	β_{32}	1.00	3.4	3.5	5.6	19.2	4.2	4.0	18.3	18.6	0.88	0.88	0.00	0.00	0.83	0.86	0.00	0.00
	β_{33}	-1.00	3.0	3.0	20.6	11.6	3.3	3.5	10.3	10.7	0.95	0.95	0.51	0.57	0.94	0.94	0.62	0.61
5	β_{31}	1.00	-0.0	0.5	-22.4	-16.0	0.4	0.8	-14.2	-14.0	0.97	0.96	0.05	0.07	0.96	0.97	0.04	0.05
	β_{32}	1.00	0.0	0.6	-20.2	-12.4	0.4	0.9	-10.4	-10.1	0.96	0.95	0.27	0.31	0.95	0.95	0.25	0.26
	β_{33}	-1.00	0.0	0.5	36.3	-11.4	0.3	0.6	-14.1	-13.9	0.94	0.94	0.44	0.49	0.95	0.94	0.46	0.47
6	β_{31}	1.00	8.4	8.5	-28.5	-28.5	0.5	0.5	-30.4	-30.3	0.31	0.30	0.00	0.00	0.96	0.95	0.00	0.00
	β_{32}	1.00	8.9	9.0	-20.3	-20.3	0.6	0.6	-22.0	-21.8	0.28	0.27	0.00	0.00	0.95	0.95	0.00	0.00
	β_{33}	-1.00	8.8	8.9	-27.9	-27.8	0.4	0.4	-30.2	-30.1	0.67	0.67	0.00	0.00	0.95	0.95	0.00	0.00

[a] The SF models that adopt the independent gamma distributions for cluster-specific random effects

[b] The SF models that adopt the independent log-Normal distributions for cluster-specific random effects

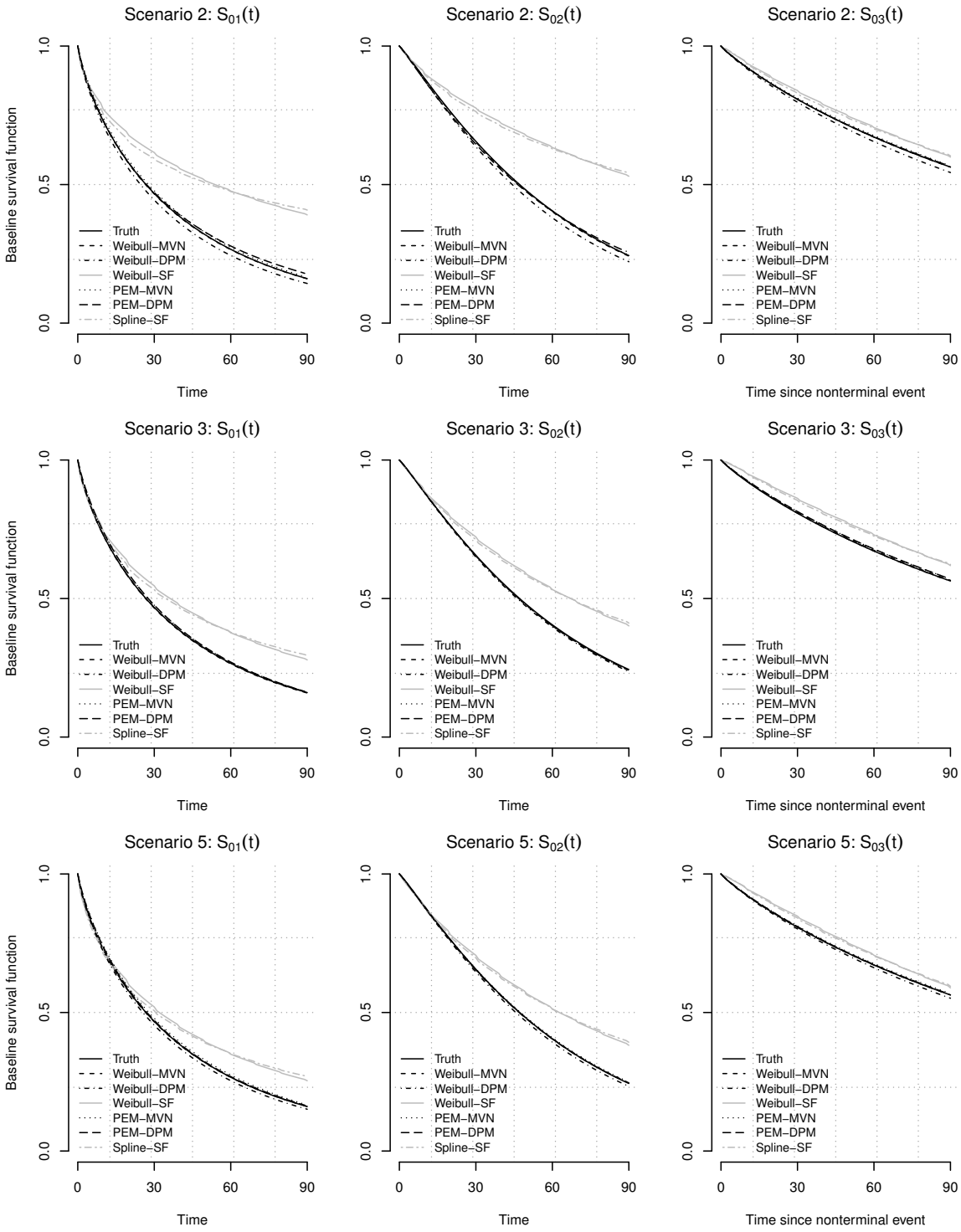

Figure D.1: Estimated transition-specific baseline survival functions, $S_{0g}(\cdot)$=exp$(-H_{0g}(\cdot))$, for each six analyses described in Section 5 under simulation scenarios 2, 3 and 5.

Table D.4: Average relative width of 95% credible/confidence intervals for $\boldsymbol{\beta}_1$ and θ, with the Weibull-MVN model taken as the referent, across six simulation scenarios given in Table 3. Throughout values are based on results from R=500 simulated datasets.

Scenario		Weibull -MVN	Weibull -DPM	Weibull -SF$_{\mathcal{G}}$[a]	Weibull -SF$_{\mathcal{LN}}$[b]	PEM -MVN	PEM -DPM	Spline -SF$_{\mathcal{G}}$	Spline -SF$_{\mathcal{LN}}$
1	β_{11}	1.00	1.00	0.81	0.78	1.02	1.02	0.81	0.81
	β_{12}	1.00	1.00	0.77	0.74	1.04	1.04	0.77	0.77
	β_{13}	1.00	1.00	0.84	0.81	1.00	1.01	0.83	0.84
	θ	1.00	1.00			1.10	1.12		
2	β_{11}	1.00	1.00	0.73	0.70	1.02	1.02	0.73	0.73
	β_{12}	1.00	1.00	0.69	0.66	1.03	1.04	0.69	0.69
	β_{13}	1.00	1.00	0.76	0.72	1.00	1.00	0.76	0.76
	θ	1.00	1.00			1.12	1.14		
3	β_{11}	1.00	1.00	0.81	0.78	1.02	1.02	0.81	0.81
	β_{12}	1.00	1.00	0.76	0.74	1.04	1.04	0.77	0.77
	β_{13}	1.00	1.00	0.83	0.80	1.00	1.01	0.83	0.83
	θ	1.00	1.00			1.10	1.13		
4	β_{11}	1.00	1.00	0.95	0.90	1.02	1.01	0.96	0.96
	β_{12}	1.00	1.00	0.94	0.88	1.03	1.03	0.95	0.95
	β_{13}	1.00	1.00	0.96	0.90	1.01	1.01	0.96	0.96
	θ	1.00	1.00			1.09	1.09		
5	β_{11}	1.00	1.00	0.81	0.74	1.02	1.02	0.81	0.81
	β_{12}	1.00	1.00	0.77	0.70	1.03	1.03	0.77	0.77
	β_{13}	1.00	1.00	0.83	0.76	1.00	1.00	0.83	0.83
	θ	1.00	1.00			1.09	1.09		
6	β_{11}	1.00	1.00	0.74	0.71	0.94	0.95	0.73	0.74
	β_{12}	1.00	1.00	0.72	0.69	0.96	0.97	0.71	0.72
	β_{13}	1.00	1.00	0.76	0.72	0.93	0.93	0.75	0.75
	θ	1.00	1.00			0.89	0.90		

[a] The SF models that adopt the independent gamma distributions for cluster-specific random effects

[b] The SF models that adopt the independent log-Normal distributions for cluster-specific random effects

Table D.5: Average relative width of 95% credible/confidence intervals for $\boldsymbol{\beta}_2$, with the Weibull-MVN model taken as the referent, across six simulation scenarios given in Table 3. Throughout values are based on results from R=500 simulated datasets.

Scenario		Weibull -MVN	Weibull -DPM	Weibull -SF$_{\mathcal{G}}$[a]	Weibull -SF$_{\mathcal{LN}}$[b]	PEM -MVN	PEM -DPM	Spline -SF$_{\mathcal{G}}$	Spline -SF$_{\mathcal{LN}}$
1	β_{21}	1.00	1.00	0.80	0.82	1.01	1.01	0.84	0.84
	β_{22}	1.00	1.00	0.75	0.78	1.02	1.02	0.79	0.79
	β_{23}	1.00	1.00	0.83	0.85	1.00	1.00	0.87	0.87
2	β_{21}	1.00	1.00	0.77	0.77	1.01	1.01	0.78	0.78
	β_{22}	1.00	1.00	0.72	0.72	1.02	1.02	0.73	0.73
	β_{23}	1.00	1.00	0.80	0.80	1.00	1.00	0.80	0.80
3	β_{21}	1.00	1.00	0.80	0.83	1.01	1.01	0.83	0.83
	β_{22}	1.00	1.00	0.75	0.78	1.02	1.03	0.79	0.79
	β_{23}	1.00	1.00	0.82	0.86	1.00	1.00	0.86	0.86
4	β_{21}	1.00	1.00	0.90	0.90	1.01	1.01	0.96	0.96
	β_{22}	1.00	1.00	0.88	0.88	1.01	1.01	0.95	0.95
	β_{23}	1.00	1.00	0.91	0.91	1.00	1.00	0.97	0.97
5	β_{21}	1.00	1.00	0.81	0.83	1.01	1.01	0.84	0.84
	β_{22}	1.00	1.00	0.77	0.78	1.02	1.02	0.79	0.79
	β_{23}	1.00	1.00	0.84	0.86	1.00	1.00	0.86	0.86
6	β_{21}	1.00	1.00	0.78	0.79	0.96	0.96	0.82	0.82
	β_{22}	1.00	1.00	0.75	0.76	0.97	0.97	0.80	0.80
	β_{23}	1.00	1.00	0.79	0.80	0.95	0.95	0.84	0.84

[a] The SF models that adopt the independent gamma distributions for cluster-specific random effects

[b] The SF models that adopt the independent log-Normal distributions for cluster-specific random effects

Table D.6: Average relative width of 95% credible/confidence intervals for $\boldsymbol{\beta}_3$, with the Weibull-MVN model taken as the referent, across six simulation scenarios given in Table 3. Throughout values are based on results from R=500 simulated datasets.

Scenario		Weibull -MVN	Weibull -DPM	Weibull -SF$_{\mathcal{G}}$[a]	Weibull -SF$_{\mathcal{LN}}$[b]	PEM -MVN	PEM -DPM	Spline -SF$_{\mathcal{G}}$	Spline -SF$_{\mathcal{LN}}$
1	β_{31}	1.00	1.00	0.72	0.83	1.01	1.01	0.83	0.83
	β_{32}	1.00	1.00	0.73	0.83	1.01	1.02	0.83	0.83
	β_{33}	1.00	1.00	0.75	0.86	1.00	1.00	0.86	0.86
2	β_{31}	1.00	1.00	0.77	0.75	1.01	1.01	0.77	0.77
	β_{32}	1.00	1.00	0.77	0.75	1.01	1.02	0.77	0.77
	β_{33}	1.00	1.00	0.81	0.79	1.00	1.00	0.81	0.81
3	β_{31}	1.00	1.00	0.73	0.81	1.01	1.01	0.84	0.84
	β_{32}	1.00	1.00	0.74	0.81	1.01	1.02	0.84	0.84
	β_{33}	1.00	1.00	0.76	0.84	1.00	1.00	0.87	0.87
4	β_{31}	1.00	1.00	0.85	0.97	1.01	1.01	0.98	0.98
	β_{32}	1.00	1.00	0.87	0.99	1.01	1.01	1.00	1.00
	β_{33}	1.00	1.00	0.85	0.96	1.00	1.00	0.97	0.97
5	β_{31}	1.00	1.00	0.73	0.79	1.01	1.01	0.83	0.83
	β_{32}	1.00	1.00	0.73	0.80	1.01	1.01	0.83	0.84
	β_{33}	1.00	1.00	0.76	0.82	1.00	1.00	0.86	0.86
6	β_{31}	1.00	1.00	0.69	0.69	0.96	0.96	0.69	0.69
	β_{32}	1.00	1.00	0.72	0.72	0.97	0.97	0.72	0.72
	β_{33}	1.00	1.00	0.73	0.73	0.94	0.94	0.73	0.73

[a] The SF models that adopt the independent gamma distributions for cluster-specific random effects

[b] The SF models that adopt the independent log-Normal distributions for cluster-specific random effects

Table D.7: Mean squared error of prediction ($\times 10^{-2}$) for cluster-specific random effects based on six analyses described in Section 5.2, across six data scenarios given in Table 3. Throughout values are based on results from R=500 simulated datasets.

Scenario		Weibull -MVN	Weibull -DPM	Weibull -SF$_{\mathcal{G}}$[a]	%F	Weibull -SF$_{\mathcal{LN}}$[b]	%F[†]	PEM -MVN	PEM -DPM	Spline -SF$_{\mathcal{G}}$	%F	Spline -SF$_{\mathcal{LN}}$	%F
1	V_1	5.25	5.27	6.40		10355.30		5.27	5.27	6.39		10397.30	
	V_2	7.66	7.70	8.70	17.8	11199.69	6.6	7.67	7.72	8.68	0.2	11131.71	0.8
	V_3	9.91	9.95	12.13		11221.04		9.91	9.96	12.11		11214.38	
2	V_1	6.36	6.41	8.10		10535.62		6.37	6.41	8.09		10476.70	
	V_2	8.76	8.85	10.23	10.4	11328.83	7.8	8.77	8.86	10.20	0.0	11208.54	0.6
	V_3	11.13	11.19	13.85		11417.49		11.13	11.19	13.91		11449.00	
3	V_1	5.03	5.04	6.27		10398.49		5.04	5.04	6.22		10329.00	
	V_2	6.34	6.34	8.28	15.8	11357.53	9.2	6.36	6.36	8.24	0.0	11348.16	0.6
	V_3	7.55	7.49	11.66		10932.18		7.57	7.55	11.69		10895.77	
4	V_1	3.84	3.85	4.99		9798.48		3.87	3.87	5.01		9765.63	
	V_2	6.25	6.27	7.19	12.8	11076.72	12.4	6.25	6.27	7.12	0.4	11102.66	5.4
	V_3	7.89	7.90	9.57		10893.62		7.90	7.91	9.52		10886.70	
5	V_1	6.95	6.26	10.87		10005.24		6.96	6.27	10.86		9869.41	
	V_2	11.52	10.50	14.95	12.8	11090.98	78.4	11.50	10.52	14.92	0.2	10976.78	76.4
	V_3	15.46	14.66	25.04		11156.06		15.46	14.72	24.94		11073.46	
6	V_1	5.05	5.01	6.34		9670.25		4.89	4.85	6.26		9804.42	
	V_2	7.58	7.55	8.60	5.4	11259.19	9.2	7.41	7.39	8.49	1.4	11272.07	0.4
	V_3	6.72	6.65	13.42		10095.61		6.44	6.40	13.70		10233.22	

[a] The SF models that adopt the independent gamma distributions for cluster-specific random effects

[b] The SF models that adopt the independent log-Normal distributions for cluster-specific random effects

[†] % of times SF models yield at least one of $\hat{\boldsymbol{V}}_j$ being $-\infty$, resulting in MSEP being ∞

E Application to Medicare data from New England

E.1 Additional Results

In our main paper, posterior summaries for hazard ratio (HR) parameters for readmission, $\exp(\boldsymbol{\beta}_1)$, from models for which a semi-Markov specification was adopted for $h_{03}(\cdot)$ are presented. In Table E.1-E.4, we provide posterior summaries for HR parameters for death without readmission, $\exp(\boldsymbol{\beta}_2)$, and those for death following readmission, $\exp(\boldsymbol{\beta}_3)$, from both Markov and semi-Markov models. We also provide the posterior estimates of $\exp(\boldsymbol{\beta}_1)$ from Markov models in Table E.5.

From Table 4 (in the main paper) and Table E.1-E.5, we see the little difference in posterior estimates of HR parameters between Markov and semi-Markov models in this particular application. Therefore, our analyses in this document mainly focus on HR parameters for death ($\exp(\boldsymbol{\beta}_2)$ and $\exp(\boldsymbol{\beta}_3)$) under semi-Markov models. As seen in Table E.1 and Table E.2, our proposed framework show how risk for death changes depending on whether or not a patient experiences the readmission. For example, whereas there is evidence of an increased risk of death for long hospital stay among individuals who have not been readmitted (HR 1.10; 95% CI 1.04, 1.18 in PEM-DPM), the same conclusion cannot be drawn for individuals who have been readmitted (HR 0.98; 95% CI 0.87, 1.09 in PEM-DPM). In addition, the association between death and two of covariates (age and Charlson/Deyo score) is stronger in this magnitude (i.e. farther from zero) while the association between death and some other covariates (sex, source of entry to initial hospitalization, length of stay, discharged location, and whether or not patients underwent a procedure during the hospitalization) is weakened among patient who have been readmitted. In general, our analyses show evidence of increased risk for death for patients with male gender, older age, initially hospitalized via some route other than ER, higher comorbidity score, a procedure during the hospitalization, a discharge to a place other than home (without care).

Table E.1: Posterior medians (PM) and 95% credible intervals (CI) for hazard ratio parameters for death without readmission ($\exp(\boldsymbol{\beta}_2)$) from semi-competing risks data analyses based on semi-Markov models.

	Weibull-MVN PM (95%CI)	Weibull-DPM PM (95%CI)	PEM-MVN PM (95%CI)	PEM-DPM PM (95%CI)
Sex				
Male	1.00	1.00	1.00	1.00
Female	0.69 (0.60, 0.79)	0.69 (0.61, 0.78)	0.75 (0.67, 0.83)	0.75 (0.67, 0.83)
Age*	1.09 (1.04, 1.13)	1.09 (1.04, 1.14)	1.07 (1.03, 1.11)	1.07 (1.03, 1.11)
Race				
White	1.00	1.00	1.00	1.00
Non-white	0.93 (0.70, 1.22)	0.93 (0.70, 1.22)	0.94 (0.74, 1.17)	0.94 (0.75, 1.18)
Source of entry to initial hospitalization				
Emergency room	1.00	1.00	1.00	1.00
Other facility	1.61 (1.41, 1.85)	1.61 (1.41, 1.86)	1.50 (1.33, 1.70)	1.49 (1.32, 1.68)
Charlson/Deyo score				
≤ 1	1.00	1.00	1.00	1.00
> 1	1.40 (1.12, 1.71)	1.39 (1.13, 1.73)	1.26 (1.08, 1.50)	1.27 (1.06, 1.51)
Procedure during hospitalization				
No	1.00	1.00	1.00	1.00
Yes	0.09 (0.07, 0.12)	0.09 (0.07, 0.12)	0.13 (0.10, 0.16)	0.13 (0.10, 0.16)
Length of stay**	1.15 (1.07, 1.24)	1.15 (1.06, 1.24)	1.10 (1.04, 1.18)	1.10 (1.04, 1.18)
Care after discharge				
Home	1.00	1.00	1.00	1.00
Home with care	2.41 (2.00, 2.91)	2.45 (2.02, 2.94)	2.22 (1.85, 2.63)	2.21 (1.90, 2.61)
Hospice	22.99 (18.08, 30.16)	23.71 (18.28, 31.20)	13.94 (11.22, 17.43)	13.85 (11.33, 17.08)
ICF/SNF	5.22 (4.29, 6.39)	5.33 (4.32, 6.45)	4.25 (3.57, 5.06)	4.25 (3.66, 5.01)
Other	4.81 (3.48, 6.70)	4.93 (3.58, 6.84)	3.79 (2.91, 4.98)	3.81 (2.94, 4.91)

* standardized so that 0 corresponds to an age of 77 years and so that one unit increment corresponds to 10 years

** standardized so that 0 corresponds to 10 days and so that one unit increment corresponds to 7 days

Table E.2: Posterior medians (PM) and 95% credible intervals (CI) for hazard ratio parameters for death following readmission ($\exp(\boldsymbol{\beta}_3)$) from semi-competing risks data analyses based on semi-Markov models.

	Weibull-MVN PM (95%CI)	Weibull-DPM PM (95%CI)	PEM-MVN PM (95%CI)	PEM-DPM PM (95%CI)
Sex				
Male	1.00	1.00	1.00	1.00
Female	0.81 (0.66, 1.00)	0.81 (0.66, 0.98)	0.84 (0.70, 1.00)	0.84 (0.71, 1.01)
Age*	1.10 (1.02, 1.19)	1.10 (1.02, 1.19)	1.09 (1.02, 1.17)	1.09 (1.02, 1.17)
Race				
White	1.00	1.00	1.00	1.00
Non-white	1.15 (0.77, 1.67)	1.14 (0.79, 1.65)	1.12 (0.79, 1.54)	1.11 (0.78, 1.54)
Source of entry to initial hospitalization				
Emergency room	1.00	1.00	1.00	1.00
Other facility	1.54 (1.25, 1.90)	1.55 (1.25, 1.91)	1.42 (1.17, 1.72)	1.42 (1.16, 1.72)
Charlson/Deyo score				
≤ 1	1.00	1.00	1.00	1.00
> 1	1.51 (1.11, 2.06)	1.52 (1.11, 2.07)	1.41 (1.06, 1.85)	1.40 (1.05, 1.84)
Procedure during hospitalization				
No	1.00	1.00	1.00	1.00
Yes	0.21 (0.15, 0.29)	0.21 (0.15, 0.29)	0.28 (0.20, 0.39)	0.28 (0.21, 0.39)
Length of stay**	1.01 (0.89, 1.13)	1.01 (0.89, 1.13)	0.98 (0.88, 1.09)	0.98 (0.87, 1.09)
Care after discharge				
Home	1.00	1.00	1.00	1.00
Home with care	1.44 (1.13, 1.81)	1.44 (1.13, 1.82)	1.35 (1.08, 1.68)	1.34 (1.08, 1.65)
Hospice	10.23 (4.66, 22.01)	10.43 (4.83, 22.33)	6.46 (3.33, 12.58)	6.35 (3.30, 12.29)
ICF/SNF	2.54 (1.87, 3.45)	2.57 (1.87, 3.46)	2.08 (1.52, 2.77)	2.07 (1.56, 2.76)
Other	2.78 (1.64, 4.49)	2.72 (1.61, 4.44)	2.24 (1.40, 3.48)	2.25 (1.41, 3.43)

* standardized so that 0 corresponds to an age of 77 years and so that one unit increment corresponds to 10 years

** standardized so that 0 corresponds to 10 days and so that one unit increment corresponds to 7 days

Table E.3: Posterior medians (PM) and 95% credible intervals (CI) for hazard ratio parameters for death without readmission ($\exp(\boldsymbol{\beta}_2)$) from semi-competing risks data analyses based on Markov models.

	Weibull-MVN PM (95%CI)	Weibull-DPM PM (95%CI)	PEM-MVN PM (95%CI)	PEM-DPM PM (95%CI)
Sex				
Male	1.00	1.00	1.00	1.00
Female	0.69 (0.60, 0.79)	0.69 (0.60, 0.79)	0.75 (0.67, 0.84)	0.75 (0.67, 0.83)
Age*	1.08 (1.04, 1.13)	1.08 (1.04, 1.13)	1.07 (1.03, 1.11)	1.07 (1.03, 1.11)
Race				
White	1.00	1.00	1.00	1.00
Non-white	0.92 (0.69, 1.21)	0.92 (0.70, 1.22)	0.94 (0.75, 1.18)	0.94 (0.75, 1.16)
Source of entry to initial hospitalization				
Emergency room	1.00	1.00	1.00	1.00
Other facility	1.61 (1.42, 1.84)	1.62 (1.42, 1.86)	1.49 (1.32, 1.67)	1.49 (1.33, 1.69)
Charlson/Deyo score				
≤ 1	1.00	1.00	1.00	1.00
> 1	1.40 (1.12, 1.73)	1.39 (1.12, 1.72)	1.26 (1.06, 1.49)	1.27 (1.06, 1.51)
Procedure during hospitalization				
No	1.00	1.00	1.00	1.00
Yes	0.09 (0.07, 0.12)	0.09 (0.07, 0.12)	0.13 (0.10, 0.16)	0.13 (0.11, 0.17)
Length of stay**	1.15 (1.07, 1.23)	1.15 (1.07, 1.23)	1.10 (1.04, 1.17)	1.10 (1.04, 1.17)
Care after discharge				
Home	1.00	1.00	1.00	1.00
Home with care	2.44 (2.04, 2.91)	2.42 (2.02, 2.93)	2.20 (1.86, 2.62)	2.20 (1.90, 2.61)
Hospice	23.52 (17.98, 30.13)	23.55 (18.05, 30.53)	13.72 (11.22, 17.49)	13.78 (11.10, 17.1)
ICF/SNF	5.30 (4.36, 6.43)	5.29 (4.38, 6.47)	4.23 (3.61, 5.16)	4.25 (3.62, 5.02)
Other	4.88 (3.59, 6.72)	4.87 (3.54, 6.73)	3.78 (2.93, 4.99)	3.82 (2.92, 4.97)

* standardized so that 0 corresponds to an age of 77 years and so that one unit increment corresponds to 10 years

** standardized so that 0 corresponds to 10 days and so that one unit increment corresponds to 7 days

Table E.4: Posterior medians (PM) and 95% credible intervals (CI) for hazard ratio parameters for death following readmission ($\exp(\boldsymbol{\beta}_3)$) from semi-competing risks data analyses based on Markov models.

	Weibull-MVN PM (95%CI)	Weibull-DPM PM (95%CI)	PEM-MVN PM (95%CI)	PEM-DPM PM (95%CI)
Sex				
Male	1.00	1.00	1.00	1.00
Female	0.81 (0.66, 0.98)	0.81 (0.67, 0.99)	0.84 (0.71, 1.01)	0.85 (0.71, 1.02)
Age*	1.11 (1.02, 1.19)	1.10 (1.03, 1.20)	1.10 (1.03, 1.18)	1.09 (1.02, 1.18)
Race				
White	1.00	1.00	1.00	1.00
Non-white	1.14 (0.78, 1.64)	1.15 (0.79, 1.67)	1.13 (0.81, 1.55)	1.14 (0.79, 1.58)
Source of entry to initial hospitalization				
Emergency room	1.00	1.00	1.00	1.00
Other facility	1.58 (1.28, 1.97)	1.58 (1.28, 1.97)	1.44 (1.18, 1.75)	1.46 (1.21, 1.77)
Charlson/Deyo score				
≤ 1	1.00	1.00	1.00	1.00
> 1	1.53 (1.12, 2.11)	1.53 (1.12, 2.11)	1.40 (1.02, 1.84)	1.40 (1.06, 1.86)
Procedure during hospitalization				
No	1.00	1.00	1.00	1.00
Yes	0.20 (0.14, 0.28)	0.20 (0.14, 0.28)	0.27 (0.19, 0.37)	0.27 (0.19, 0.36)
Length of stay**	1.00 (0.89, 1.13)	1.01 (0.89, 1.13)	0.98 (0.88, 1.09)	0.98 (0.88, 1.08)
Care after discharge				
Home	1.00	1.00	1.00	1.00
Home with care	1.44 (1.15, 1.82)	1.44 (1.13, 1.81)	1.32 (1.06, 1.63)	1.33 (1.07, 1.66)
Hospice	11.81 (5.18, 25.66)	11.6 (5.08, 24.49)	6.95 (3.49, 12.75)	6.79 (3.24, 13.28)
ICF/SNF	2.70 (1.96, 3.68)	2.69 (1.99, 3.61)	2.12 (1.59, 2.81)	2.17 (1.63, 2.87)
Other	2.92 (1.74, 4.77)	2.89 (1.74, 4.68)	2.32 (1.46, 3.65)	2.36 (1.47, 3.67)

* standardized so that 0 corresponds to an age of 77 years and so that one unit increment corresponds to 10 years

** standardized so that 0 corresponds to 10 days and so that one unit increment corresponds to 7 days

Table E.5: Posterior medians (PM) and 95% credible intervals (CI) for hazard ratio parameters for readmission ($\exp(\boldsymbol{\beta}_1)$) from semi-competing risks data analyses based on Markov models.

	Weibull-MVN PM (95%CI)	Weibull-DPM PM (95%CI)	PEM-MVN PM (95%CI)	PEM-DPM PM (95%CI)
Sex				
Male	1.00	1.00	1.00	1.00
Female	0.79 (0.70, 0.91)	0.80 (0.70, 0.91)	0.85 (0.76, 0.95)	0.85 (0.76, 0.95)
Age*	0.90 (0.86, 0.95)	0.90 (0.86, 0.95)	0.91 (0.87, 0.94)	0.91 (0.87, 0.95)
Race				
White	1.00	1.00	1.00	1.00
Non-white	1.11 (0.86, 1.45)	1.11 (0.86, 1.44)	1.12 (0.89, 1.40)	1.11 (0.89, 1.38)
Source of entry to initial hospitalization				
Emergency room	1.00	1.00	1.00	1.00
Other facility	1.18 (1.03, 1.35)	1.19 (1.03, 1.36)	1.12 (0.99, 1.26)	1.12 (0.99, 1.27)
Charlson/Deyo score				
≤ 1	1.00	1.00	1.00	1.00
> 1	1.49 (1.19, 1.84)	1.50 (1.19, 1.85)	1.40 (1.15, 1.68)	1.39 (1.15, 1.68)
Procedure during hospitalization				
No	1.00	1.00	1.00	1.00
Yes	0.45 (0.37, 0.53)	0.45 (0.37, 0.53)	0.57 (0.49, 0.66)	0.57 (0.48, 0.66)
Length of stay**	1.15 (1.07, 1.23)	1.15 (1.07, 1.23)	1.12 (1.05, 1.19)	1.12 (1.05, 1.19)
Care after discharge				
Home	1.00	1.00	1.00	1.00
Home with care	0.95 (0.82, 1.11)	0.95 (0.82, 1.11)	0.89 (0.78, 1.02)	0.89 (0.78, 1.01)
Hospice	0.39 (0.23, 0.62)	0.38 (0.22, 0.64)	0.27 (0.16, 0.42)	0.27 (0.16, 0.43)
ICF/SNF	0.88 (0.73, 1.06)	0.88 (0.73, 1.07)	0.76 (0.63, 0.90)	0.76 (0.64, 0.90)
Other	1.05 (0.77, 1.43)	1.04 (0.77, 1.45)	0.89 (0.68, 1.18)	0.89 (0.67, 1.18)

* standardized so that 0 corresponds to an age of 77 years and so that one unit increment corresponds to 10 years

** standardized so that 0 corresponds to 10 days and so that one unit increment corresponds to 7 days

E.2 Convergence diagnostics

For our proposed models, we assess the convergence of our MCMC scheme by evaluating the potential scale reduction factor (PSRF) of Gelman et al. (2013). The potential problem with PSRF is that it has not converged but happens to be close to 1 by chance even though the PSRF is actually fluctuating. Therefore, for each parameter, the PSRF was calculated at several points in time with the first half discarded as burn-in. Then, we summarize the results using mean, maximum, and minimum value of PSRF for all model parameters at different iterations. The results are shown in Figure E.1. As the number of MCMC iterations increases, the mean PSRF converges toward 1 and the maximum of PSRF is less than 1.05 indicating that all model parameters have converged well.

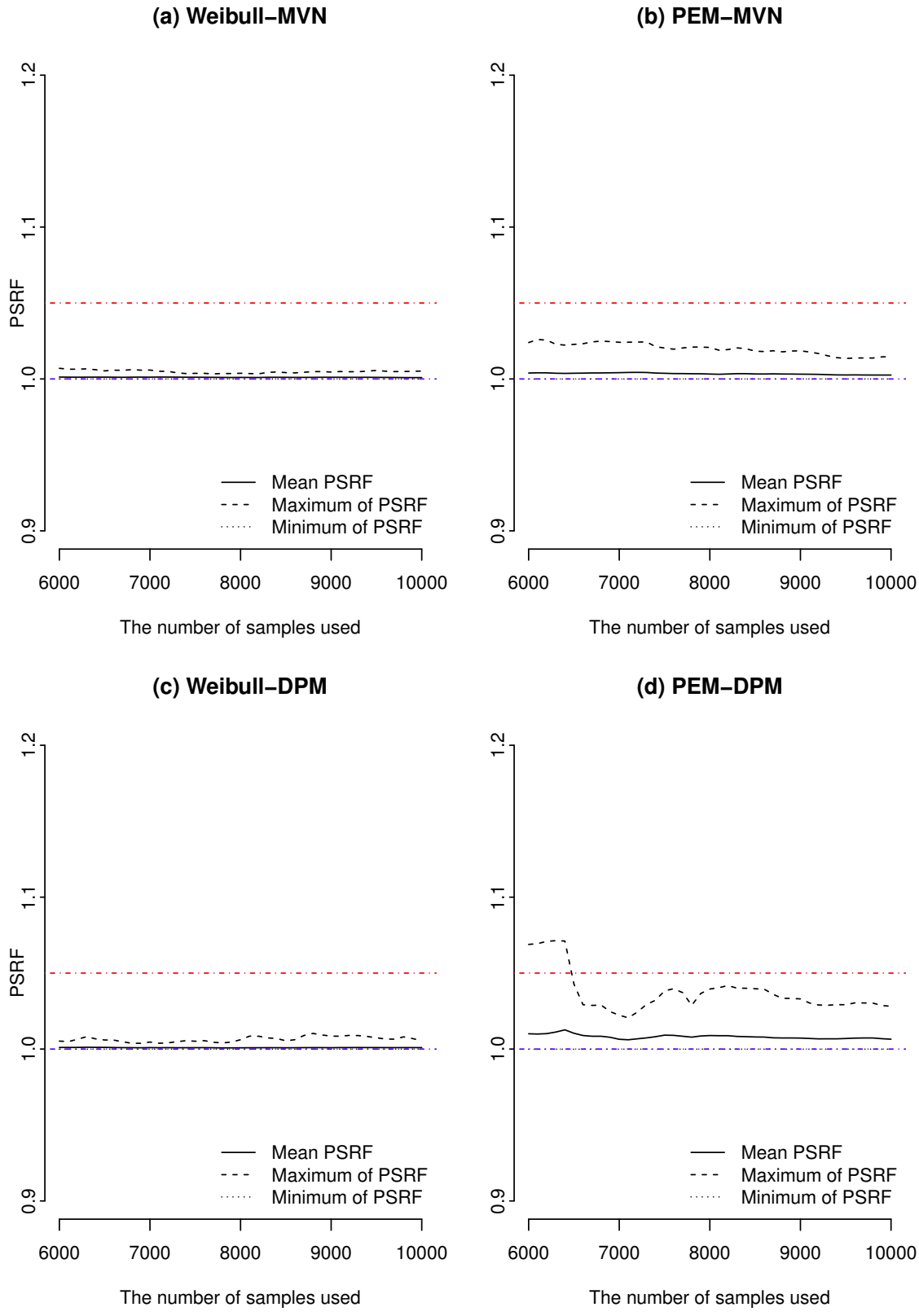

Figure E.1: The mean, maximum, and minimum value of the potential scale reduction factor (PSRF) of all model parameters from the analysis of Medicare data.

E.3 Checking the Proportional Hazards Assumption

The proposed hierarchical models assume constant hazard ratios over time conditional on the cluster-specific and patient-specific random effects. We can check the proportional hazards assumption for our proposed models by adopting the heteroscedastic hazards regression model (Hsieh, 2001; Nikulin et al., 2006). This approach permits the shape parameters for each of the three Weibull baseline hazard functions to depend upon covariate values. Therefore, we consider the following semi-Markov heteroscedastic Weibull-MVN model:

$$\begin{aligned}
h_1(t_{ji1};\gamma_{ji},\boldsymbol{X}_{ji1},\boldsymbol{Z}_{ji1},V_{j1}) &= \gamma_{ji}\ \alpha_1\kappa_1 e^{\boldsymbol{Z}_{ji1}^T\boldsymbol{\beta}_1^*} t_{ji1}^{\alpha_1 e^{\boldsymbol{Z}_{ji1}^T\boldsymbol{\beta}_1^*}-1}\ \exp\{\boldsymbol{X}_{ji1}^T\boldsymbol{\beta}_1^0+V_{j1}\}, \quad t_{ji1}>0,\\
h_2(t_{ji2};\gamma_{ji},\boldsymbol{X}_{ji2},\boldsymbol{Z}_{ji2},V_{j2}) &= \gamma_{ji}\ \alpha_2\kappa_2 e^{\boldsymbol{Z}_{ji2}^T\boldsymbol{\beta}_2^*} t_{ji2}^{\alpha_2 e^{\boldsymbol{Z}_{ji2}^T\boldsymbol{\beta}_2^*}-1}\ \exp\{\boldsymbol{X}_{ji2}^T\boldsymbol{\beta}_2^0+V_{j2}\}, \quad t_{ji2}>0,\\
h_3(t_{ji2}|t_{ji1};\gamma_{ji},\boldsymbol{X}_{ji3},\boldsymbol{Z}_{ji3},V_{j3}) &= \gamma_{ji}\ \alpha_3\kappa_3 e^{\boldsymbol{Z}_{ji3}^T\boldsymbol{\beta}_3^*}(t_{ji2}-t_{ji1})^{\alpha_3 e^{\boldsymbol{Z}_{ji3}^T\boldsymbol{\beta}_3^*}-1}\ \exp\{\boldsymbol{X}_{ji3}^T\boldsymbol{\beta}_3^0+V_{j3}\}, t_{ji2}>t_{ji1},
\end{aligned} \tag{21}$$

where $\boldsymbol{Z}_{jig}$ is a vector of covariates for the i^{th} patient in the j^{th} hospital. Note that when $\boldsymbol{\beta}_g^*$=0, this model reduces to our proposed Weibull-MVN model. Therefore, the proportional hazards assumption can be tested under the nested Weibull-MVN model (H_0: β_{kg}^*=0 for the k^{th} covariate) within the model (21) (H_1: $\beta_{kg}^* \neq 0$) while setting $\boldsymbol{X}_{jig}$=$\boldsymbol{Z}_{jig}$. We conducted two sets of analyses: one by including all of the covariates at the same time (multivariable analysis) and the other by including one covariate at a time (one-covariate analysis). We present the estimates of $\boldsymbol{\beta}_g^*$ and $\boldsymbol{\beta}_g^0$ from the multivariate analysis in Table E.6 and the estimated hazard ratios over time in Figure E.2-E.4. It appears that the proportional hazards assumption for death with and without readmission holds for all of the covariates except whether or not the patient underwent a procedure during the hospitalization. Interestingly, a number of covariates did exhibit non-proportionality in their impact on the risk of readmission including source of entry and whether or not the patient underwent a procedure during their hospitalization. We also provide the estimates of $\boldsymbol{\beta}_g^*$ and $\boldsymbol{\beta}_g^0$ from one-covariate analysis in Table E.7.

Table E.6: Posterior medians (PM) and 95% credible intervals (CI) for $\boldsymbol{\beta}_g^*$ and $\boldsymbol{\beta}_g^0$ from the multivariable analysis using the heteroscedastic Weibull-MVN model.

	$e^{\boldsymbol{\beta}_1^*}$	$e^{\boldsymbol{\beta}_2^*}$	$e^{\boldsymbol{\beta}_3^*}$	$e^{\boldsymbol{\beta}_1^0}$	$e^{\boldsymbol{\beta}_2^0}$	$e^{\boldsymbol{\beta}_3^0}$
	PM (95%CI)	PM (95%CI)	PM (95%CI)	PM (95%CI)	PM (95%CI)	PM (95%CI)
Sex: Female	0.99 (0.91, 1.06)	1.00 (0.94, 1.07)	0.90 (0.78, 1.03)	0.82 (0.61, 1.13)	0.66 (0.48, 0.89)	1.12 (0.70, 1.86)
Age	0.98 (0.96, 1.01)	1.00 (0.98, 1.02)	0.99 (0.95, 1.05)	0.96 (0.86, 1.07)	1.08 (0.97, 1.20)	1.12 (0.93, 1.33)
Race: Non-white	1.07 (0.92, 1.24)	0.98 (0.85, 1.13)	0.96 (0.73, 1.24)	0.85 (0.43, 1.61)	1.03 (0.50, 1.92)	1.31 (0.52, 2.92)
Entry: Others	1.15 (1.06, 1.25)	1.04 (0.97, 1.11)	1.10 (0.96, 1.26)	0.73 (0.51, 1.03)	1.43 (1.03, 2.00)	1.18 (0.74, 1.89)
Deyo: > 1	1.14 (1.00, 1.28)	0.98 (0.88, 1.08)	1.11 (0.90, 1.36)	0.93 (0.53, 1.59)	1.63 (1.02, 2.63)	1.14 (0.51, 2.30)
Procedure: Yes	0.67 (0.60, 0.75)	0.69 (0.58, 0.81)	0.93 (0.72, 1.20)	1.78 (1.16, 2.60)	0.42 (0.21, 0.83)	0.25 (0.10, 0.57)
Length of stay	1.03 (0.99, 1.07)	0.99 (0.96, 1.03)	1.01 (0.94, 1.09)	1.03 (0.87, 1.20)	1.19 (1.01, 1.38)	0.97 (0.74, 1.25)
Discharge: Home with care	0.95 (0.86, 1.04)	1.10 (0.97, 1.21)	1.00 (0.84, 1.19)	1.21 (0.84, 1.76)	1.57 (0.95, 2.94)	1.48 (0.81, 2.67)
Discharge: Hospice	1.35 (1.02, 1.74)	1.05 (0.93, 1.20)	1.07 (0.69, 1.56)	0.17 (0.04, 0.54)	23.25 (13.64, 41.08)	10.95 (2.96, 35.14)
Discharge: ICF/SNF	1.12 (0.99, 1.24)	1.03 (0.91, 1.14)	1.01 (0.81, 1.27)	0.62 (0.39, 1.01)	4.90 (3.03, 9.25)	2.70 (1.29, 5.63)
Discharge: Other	1.14 (0.95, 1.35)	0.96 (0.80, 1.15)	1.19 (0.83, 1.67)	0.68 (0.31, 1.42)	6.00 (2.71, 13.24)	1.69 (0.41, 5.67)

Table E.7: Posterior medians (PM) and 95% credible intervals (CI) for $\boldsymbol{\beta}_g^*$ and $\boldsymbol{\beta}_g^0$ from one-covariate analyses using the heteroscedastic Weibull-MVN model.

	$e^{\boldsymbol{\beta}_1^*}$	$e^{\boldsymbol{\beta}_2^*}$	$e^{\boldsymbol{\beta}_3^*}$	$e^{\boldsymbol{\beta}_1^0}$	$e^{\boldsymbol{\beta}_2^0}$	$e^{\boldsymbol{\beta}_3^0}$
	PM (95%CI)	PM (95%CI)	PM (95%CI)	PM (95%CI)	PM (95%CI)	PM (95%CI)
Sex: Female	0.98 (0.91, 1.06)	1.09 (1.02, 1.18)	0.91 (0.79, 1.05)	0.83 (0.60, 1.14)	0.63 (0.47, 0.84)	1.22 (0.76, 1.93)
Age	1.04 (1.01, 1.06)	1.06 (1.03, 1.08)	0.99 (0.94, 1.05)	0.81 (0.73, 0.90)	1.07 (0.97, 1.18)	1.23 (1.02, 1.47)
Race: Non-white	1.10 (0.94, 1.29)	1.00 (0.85, 1.15)	1.03 (0.79, 1.32)	0.88 (0.44, 1.68)	1.25 (0.68, 2.27)	1.34 (0.54, 2.99)
Entry: Others	1.37 (1.26, 1.48)	1.17 (1.09, 1.26)	1.09 (0.95, 1.26)	0.47 (0.33, 0.66)	1.72 (1.27, 2.31)	1.89 (1.17, 3.05)
Deyo: > 1	1.24 (1.08, 1.40)	1.02 (0.91, 1.14)	1.12 (0.89, 1.38)	0.76 (0.42, 1.34)	1.83 (1.17, 2.81)	1.31 (0.59, 2.72)
Procedure: Yes	0.62 (0.56, 0.68)	0.72 (0.61, 0.86)	0.97 (0.76, 1.21)	2.57 (1.80, 3.65)	0.32 (0.16, 0.57)	0.25 (0.11, 0.52)
Length of stay	0.97 (0.92, 1.01)	0.95 (0.92, 0.99)	1.00 (0.93, 1.07)	1.13 (0.95, 1.33)	1.17 (1.03, 1.33)	0.90 (0.70, 1.13)
Discharge: Home with care	0.79 (0.72, 0.86)	1.18 (1.09, 1.28)	0.97 (0.83, 1.12)	2.15 (1.54, 3.01)	0.30 (0.20, 0.44)	0.93 (0.56, 1.52)
Discharge: Hospice	1.47 (1.08, 1.90)	1.21 (1.10, 1.32)	1.16 (0.74, 1.69)	0.16 (0.04, 0.52)	8.67 (6.07, 12.32)	9.79 (2.56, 29.95)
Discharge: ICF/SNF	1.16 (1.05, 1.28)	1.17 (1.08, 1.25)	0.97 (0.81, 1.15)	0.48 (0.31, 0.74)	1.10 (0.80, 1.51)	1.64 (0.92, 2.85)
Discharge: Other	1.10 (0.91, 1.31)	0.99 (0.85, 1.15)	1.18 (0.84, 1.58)	0.72 (0.32, 1.53)	1.49 (0.80, 2.67)	0.85 (0.23, 2.55)

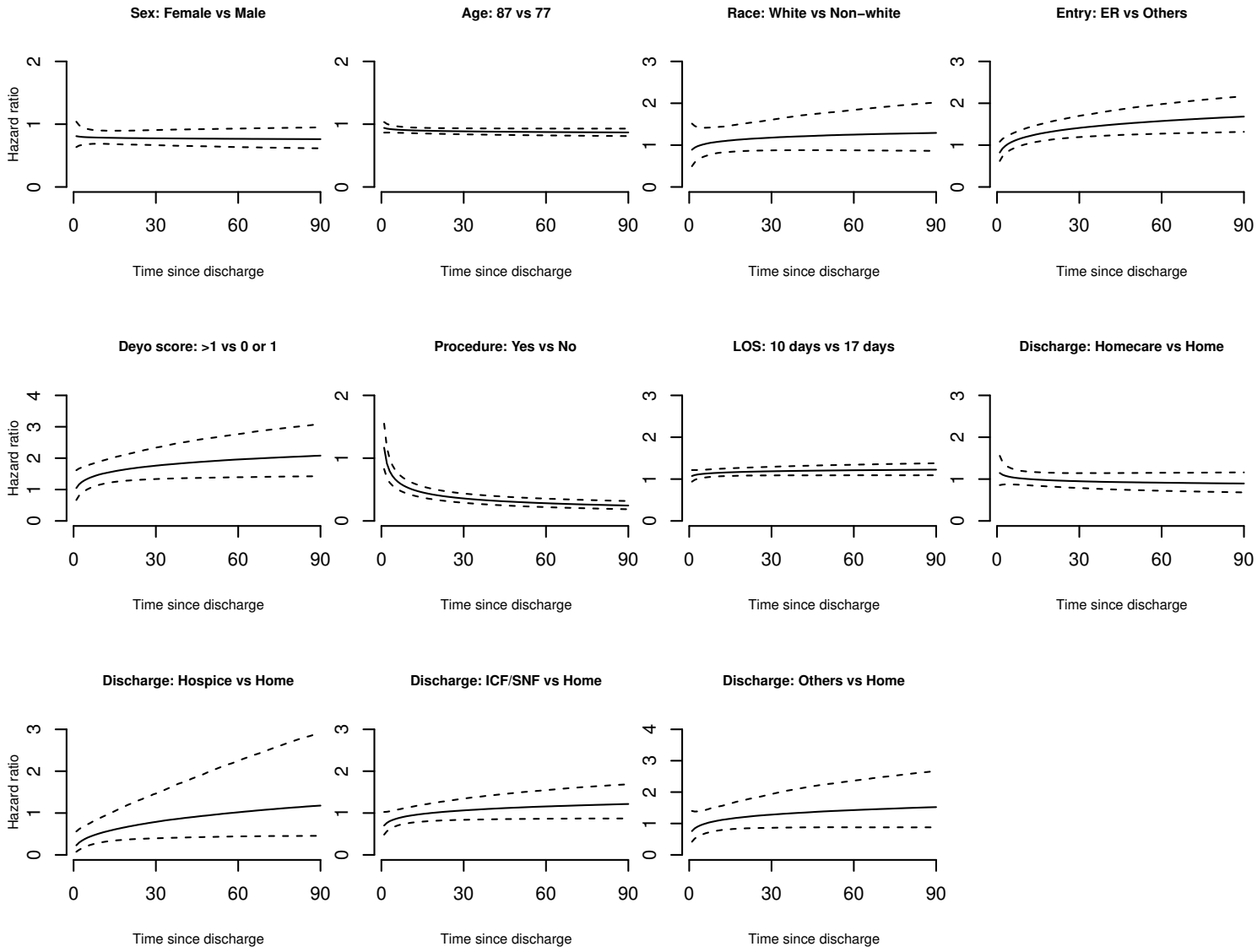

Figure E.2: Pointwise posterior median and 95% CIs for the hazard ratio associated with each of the covariates for readmission from the analysis of the New England Medicare data using the heteroscedastic Weibull-MVN model.

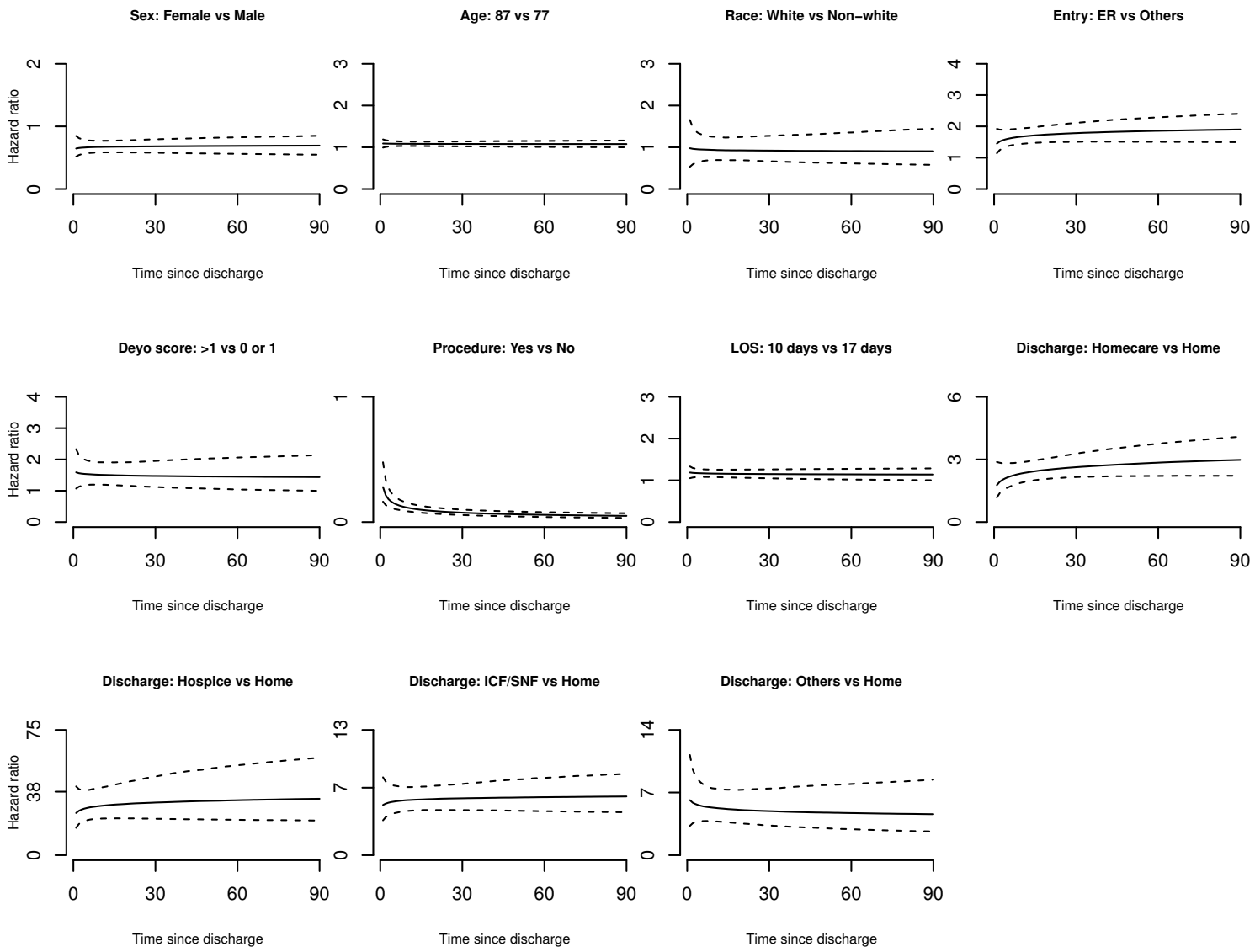

Figure E.3: Pointwise posterior median and 95% CIs for the hazard ratio associated with each of the covariates for death without readmission from the analysis of the New England Medicare data using the heteroscedastic Weibull-MVN model.

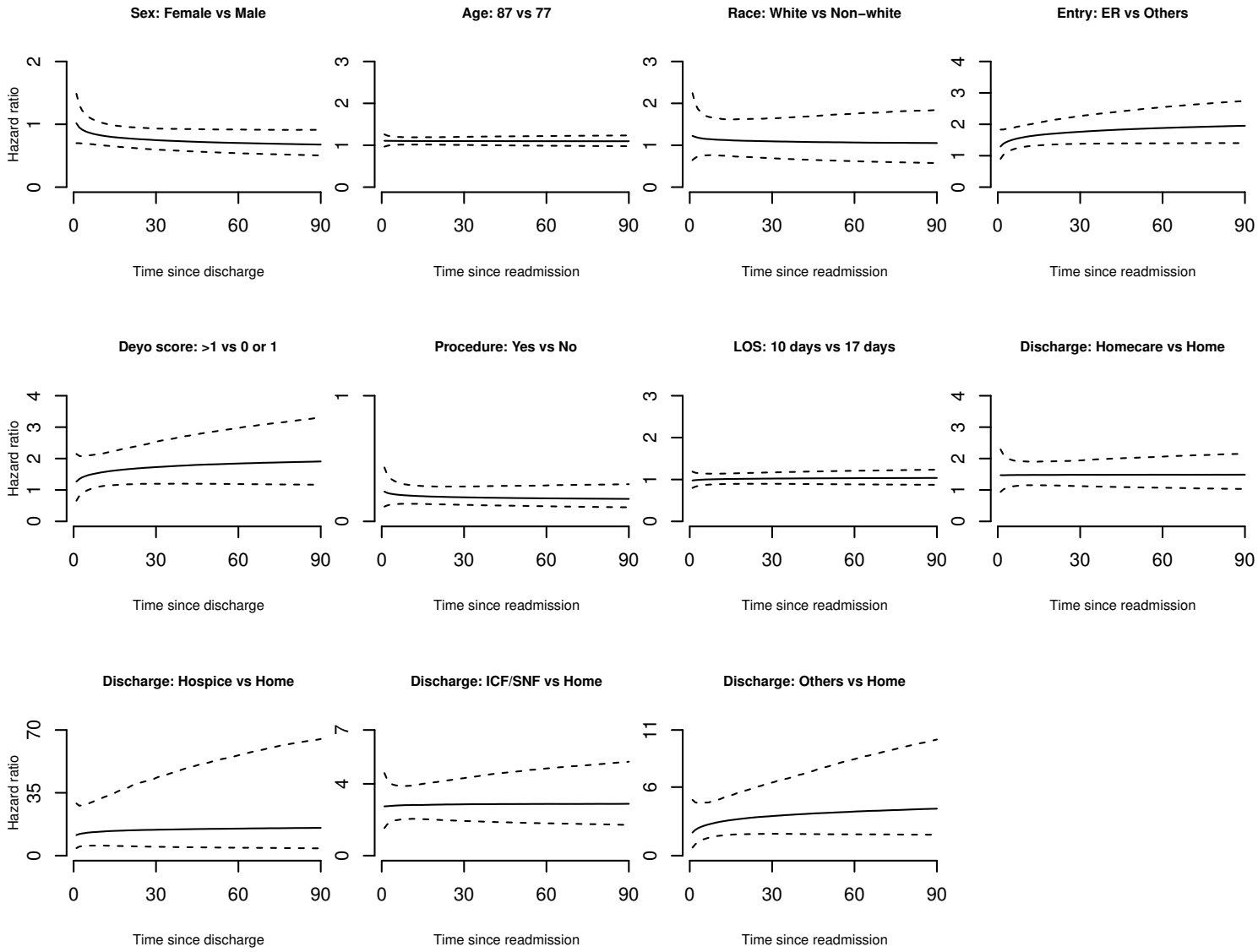

Figure E.4: Pointwise posterior median and 95% CIs for the hazard ratio associated with each of the covariates for death following readmission from the analysis of the New England Medicare data using the heteroscedastic Weibull-MVN model.

References

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.

Bush, C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika*, 83(2):275–285.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC Boca Raton.

Gorfine, M. and Hsu, L. (2011). Frailty-based competing risks model for multivariate survival data. *Biometrics*, 67(2):415–426.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.

Haneuse, S. J.-P., Rudser, K., and Gillen, D. (2008). The separation of timescales in Bayesian survival modeling of the time-varying effect of a time-dependent exposure. *Biostatistics*, 9(3):400–410.

Hsieh, F. (2001). On heteroscedastic hazards regression models: theory and application. *Journal of the Royal Statistical Society: Series B*, 63(1):63–79.

Liquet, B., Timsit, J.-F., and Rondeau, V. (2012). Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. *BMC Medical Research Methodology*, 12(1):79.

Liu, L., Wolfe, R., and Huang, X. (2004). Shared frailty models for recurrent events and terminal events. *Biometrics*, 60:747–756.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265.

Nikulin, M. S., Commenges, D., and Huber, C. (2006). *Probability, Statistics, and Modelling in Public Health*. Springer.

Roberts, G. O., Rosenthal, J. S., et al. (2001). Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, 16(4):351–367.

Rondeau, V., Mazroui, Y., and Gonzalez, J. R. (2012). frailtypack: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, 47(4):1–28.

Walker, S. G. and Mallick, B. K. (1997). Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society: Series B*, 59(4):845–860.