

# Bayesian Semiparametric Joint Regression Analysis of Recurrent Adverse Events and Survival in Esophageal Cancer Patients

Juhee Lee \*

Department of Applied Mathematics and Statistics, University California Santa Cruz, Santa Cruz, CA

Peter F. Thall

Department of Biostatistics, M.D. Anderson, Huston, TX

Steven H. Lin

Department of Radiation Oncology, M.D. Anderson, Huston, TX

## Abstract

We propose a Bayesian semiparametric joint regression model for a recurrent event process and survival time. Assuming independent latent subject frailties, we define marginal models for the recurrent event process intensity and survival distribution as functions of the subject's frailty and baseline covariates. A robust Bayesian model, called Joint-DP, is obtained by assuming a Dirichlet process for the frailty distribution. We present a simulation study that compares posterior estimates under the Joint-DP model to a Bayesian joint model with lognormal frailties, a frequentist joint model, and marginal models for either the recurrent event process or survival time. The simulations show that the Joint-DP model does a good job of correcting for treatment assignment bias, and has favorable estimation reliability and accuracy compared with the alternative models. The Joint-DP model is applied to analyze an observational dataset from esophageal cancer patients treated with chemo-radiation, including the times of recurrent effusions of fluid to the heart or lungs, survival time, prognostic covariates, and radiation therapy modality.

*Keywords:* Accelerated Failure Time, Bayesian nonparametrics, Chemoradiation, Dirichlet process, Esophageal cancer, Joint model, Nonhomogeneous Point process.

---

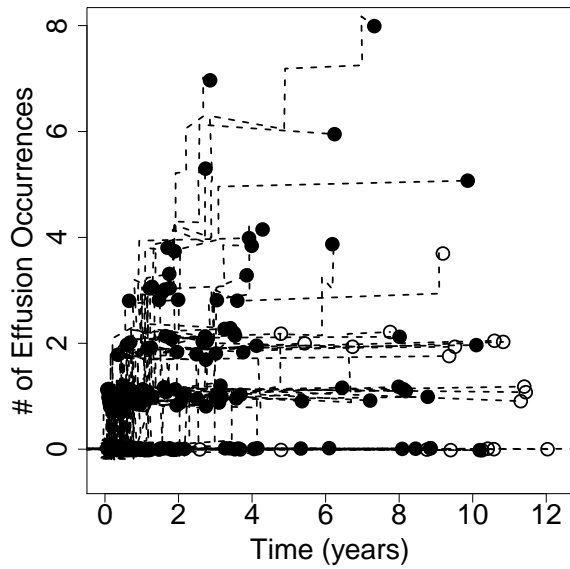
\*Address for Correspondence: Applied Mathematics and Statistics, Baskin School of Engineering, University of California 1156 High Street Mail Stop SOE2 Santa Cruz, CA 95064 USA. E-mail: juheele@soe.ucsc.edu.

# 1 Introduction

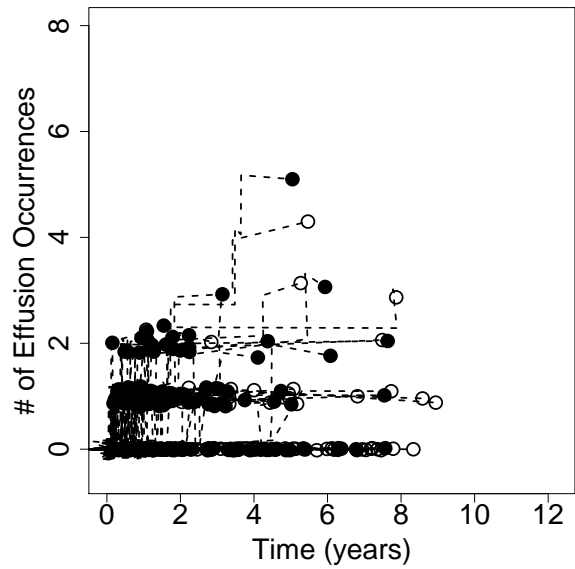
Esophageal cancer is the eighth most common cancer in the world, with over 16,000 new cases annually in the United States (Torre *et al.*, 2015). For patients whose disease cannot be removed by surgical resection, concurrent chemo-radiation therapy (CRT) is the established standard treatment. While survival time for these patients has increased over the years with advances in CRT delivery, the treatment process may take many weeks or months, and late onset toxicities remain a serious problem. One of the most troubling adverse events associated with esophageal RT is abnormal accumulation of fluid around the heart (pericardial effusion) or lungs (pleural effusion). A patient may experience effusions for many months or possibly several years from the start of RT, and some patients suffer multiple occurrences. While effusion events are not immediately fatal, it is believed that a higher overall effusion rate may increase the risk of death by impairing heart or lung function.

The RT modality used most commonly worldwide is three-dimensional conformal radiation therapy (3DCRT), which uses computed tomography for anatomic visualization so that the radiation may be targeted to reduce exposure of normal anatomical structures, especially the heart and lungs, located near the tumor. A newer modality is intensity modulated radiation therapy (IMRT), which uses multiple radiation beams designed to focus the delivered radiation more closely to the targeted cancer, and reduce exposure to surrounding organs. Recent comparisons of IMRT to 3DCRT have focused on overall survival (OS), local recurrence, and distant recurrence (Lin *et al.*, 2012), and the semi-competing risks (SCR) structure of the times to first effusion and death, since death may censor first effusion time but not conversely. The analyses reported by He *et al.* (2016) applied the SCR regression model of Lee *et al.* (2015), which includes three hazard functions, one for time to a given non-fatal event, one for time to death following the non-fatal event, and one for time to death without the non-fatal event. Chapple *et al.* (2017) refined the analyses of He *et al.* (2016) by performing Bayesian variable selection in the linear component of each hazard function. He *et al.* (2016) and Chapple *et al.* (2017) found that IMRT decreased the hazards of time to first effusion, time to death without effusions, and time to death after first effusion, compared to 3DCRT, under their semi-competing risks models.

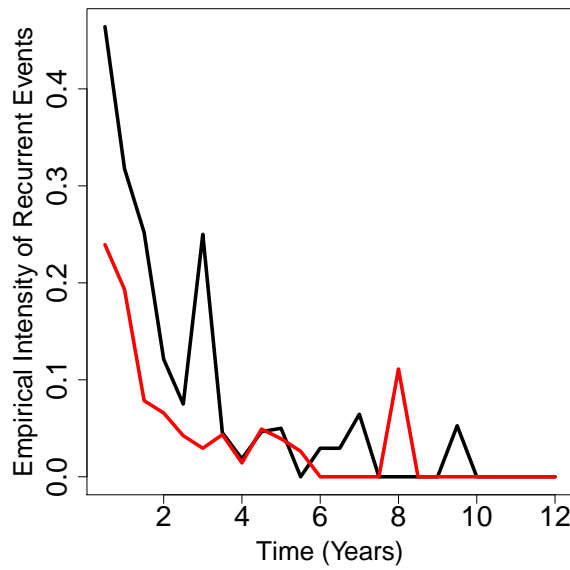
Our motivating dataset includes 468 esophageal cancer patients who received either 3DCRT or IMRT. The patients were followed from the start of RT to the time of death or administrative right censoring. Since the overall occurrence rate of each type of effusion, pericardial or pleural, was low, and moreover some patients experienced both types of effusions, to obtain reasonably reliable inferences we combine them, and hereafter refer to either type as “effusion.” A key difference between our analyses of the esophageal cancer dataset and those reported by He *et al.* (2016) and Chapple *et al.* (2017) is that we account for the occurrence times of all effusions experienced by each patient, rather than only the first effusion time. Figures 1(a) and (b) illustrate the observed recurrent effusion event processes for each RT modality. Each line ends with either the symbol  $\bullet$  at the time of death or  $\circ$  at the time of administrative right censoring. The plots show that, for some patients, effusions continued to recur for many years, while a substantial number of patients died or were right censored without any effusions. Empirical intensities of recurrent effusion events



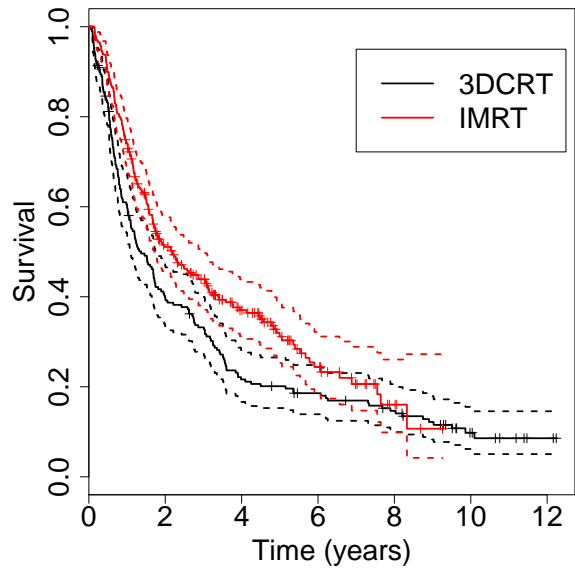
(a) 3DCRT



(b) IMRT



(c) Empirical Intensity of Recurrent Events



(d) Kaplan-Meier curves

Figure 1: [Esophageal Cancer] (a) and (b) give plots of counts of recurrent effusion events over time with the two radiation therapy modalities for the esophageal cancer patients, with  $\bullet$  and  $\circ$  representing patients with observed and censored survival outcomes, respectively. For the two radiation therapy modalities, 3DCRT (black) and IMRT (red), empirical intensities of effusion events for each modality are plotted in (c), and Kaplan-Meier curves are given in (d) with 95% confidence intervals. Observed  $T_{i,j}$  and  $N_{i,j}$  are jittered for better presentation.

are plotted in Figure 1(c) for each RT modality group (3DRT in black and IMRT in red). The plot shows that the effusion rate decreased over time, and was slightly lower for IMRT. Figure 1(d) gives Kaplan-Meier curves of survival times for the RT modality groups, with 95% confidence bands. In this dataset, patients were not randomized between the two RT modalities. Consequently, apparent differences in the plots may be due to actual differences between the RT modalities, or imbalances in known covariates or unknown latent variables related to effusion or survival. Thus, possible bias may arise for any posterior comparison of the RT modalities, and must be accounted for in the analyses.

Initially, the methodology described here was motivated by the desire to compare the two RT modality effects on both the effusion process and survival in the esophageal cancer dataset, while accounting for effects of patient baseline prognostic covariates, including age, body mass index (BMI), performance status, tumor histology, and disease stage. After formulating a robust model, described below, for the frailties that induce association between the recurrent effusion event process and survival, we were further motivated to compare the performance of this new model to established joint recurrent event-survival models, and to simpler marginal regression models for either the recurrent event process or survival time.

A central methodological issue is that, if the effusion process and the risk of death are related, then death informatively censors the effusion process. In this sense, the effusion point process plays a role similar to that of a longitudinal real-valued process observed with the possibility of termination by death or informative drop-out, a data structure that has received a great deal of attention. Joint models for a longitudinal process with an informative terminating event have been proposed by De Gruttola and Tu (1994); Faucett (1996); Wulfsohn and Tsiatis (1997); Song *et al.* (2002); Henderson *et al.* (2000); Ye *et al.* (2008), among many others. The four chapters in Part IV of Fitzmaurice *et al.* (2009) provide reviews of a variety of joint longitudinal-survival models. Bayesian approaches for joint modeling of longitudinal outcomes and survival time have been proposed by Brown and Ibrahim (2003), Hatfield *et al.* (2011), and Hatfield *et al.* (2014). While a stochastic point process is structurally different from a real-valued longitudinal process, since both types of processes may be right censored informatively by death, or some other terminal event such as drop out, in general a given method for constructing a joint longitudinal-survival time model may be adapted for constructing a joint point process-survival model, or conversely.

Various frequentist and Bayesian joint models for a recurrent event process and a terminal event also have been proposed. A common approach, which we take here, is to assume iid subject-specific random effects (frailties)  $\{\gamma_1, \dots, \gamma_n\}$ , with  $\gamma_i$  appearing in both the intensity function of the recurrent event process and the survival distribution of the  $i^{th}$  subject. Conditional on  $\gamma_i$ , a point process model is assumed for the recurrent event process hazard, and a hazard function is assumed for survival time. See, for example, Ghosh and Lin (2000); Huang and Wang (2004); Liu *et al.* (2004); Kalbfleisch *et al.* (2013); Liu and Huang (2009); Sinha *et al.* (2008); Ouyang *et al.* (2013); Wen *et al.* (2016) for examples. Cook and Lawless (2002) gave a thorough account of models and methods for non-fatal recurrent events. Liu *et al.* (2004) developed a likelihood-based, semiparametric joint model by assuming nonparametric baseline intensity and hazard functions for the recurrent event process

and survival time, assuming a gamma distribution for shared frailties and incorporating covariates through a Cox-type model. Xu *et al.* (2017) developed a general likelihood-based, semiparametric joint model by allowing an arbitrary distribution for the frailties and non-parametric baseline functions. They assumed accelerated failure time (AFT)-type models to estimate covariate effects for recurrent psychiatric admissions and survival times of patients with schizophrenia. Sinha *et al.* (2008) provided an extensive review of Bayesian approaches for joint modeling. Ouyang *et al.* (2013) investigated a set of Bayesian models by assuming different combinations of prior distributions for the baseline intensity and hazard functions, with a gamma frailty distribution. One case assumes a gamma process on the baseline cumulative intensity and a Weibull for the baseline hazard function, with gamma frailties in the recurrent event intensity function and survival time hazard function. Wen *et al.* (2016) developed a Bayesian joint model for multivariate recurrent events with a terminal event, assuming a bivariate normal distribution for the frailties.

In this paper, motivated by the esophageal cancer dataset, we propose a Bayesian semi-parametric approach to jointly model recurrent events and survival, accounting for baseline covariates. Like many existing methods, our modeling framework assumes iid frailties that appear in both the recurrent event intensity and survival hazard. For each subject, we assume conditional independence of the recurrent event process and survival time, given the subject’s frailty. The main departure from existing Bayesian approaches is that we assume a Bayesian nonparametric (BNP) prior distribution, specifically a Dirichlet process (DP), for the patient-specific frailties. We call this a *Joint-DP model*. Since the patients in our motivating dataset were not randomized between the two RT modalities, adjustment for possible bias is needed. Our approach is to view patient frailties as latent variables that may account for such biasing effects, as well as heterogeneity between patients. The DP prior relaxes commonly used distributional assumptions on the frailty distribution, and allows the frailty distribution to have features, such as skewness or multimodality, that may not be captured by particular parametric forms, such as a gamma or lognormal distribution. Accommodating such irregular features may help with bias correction due to lack of randomization in treatment assignment, and also may improve estimation of covariate effects and the baseline functions.

From our exploratory analysis of the esophageal cancer dataset, it was observed that patients had on average 0.82 recurrent events, and 1.93 and 4.88 years, respectively, for observed and censored survival times. Due to sparse information on recurrent events in the motivating data, our model formulation is intended to provide simplicity and flexibility. Each patient’s recurrent event time sequence is assumed to be a realization of a nonhomogeneous Poisson process having intensity function that is stochastic due to the patient-specific frailties. This is an example of a Cox process (Cox, 1955), also known as a doubly stochastic Poisson process (Grandell, 1976). We model covariate effects on the Poisson process through a regression structure in the log intensity process (Lawless, 1987). We assume an accelerated failure time (AFT) model to describe the regression of survival time on covariates and the frailty, with the particular AFT survival distribution chosen through preliminary goodness-of-fit analyses. Using this approach, the Joint-DP model is applied to analyze the

esophageal cancer data. For this dataset, assuming these parametric baseline functions helps with practical estimation issues due to the limited amount of recurrent event information.

The remainder of the paper is organized as follows. Section 2 describes the proposed Bayesian semiparametric joint model. Section 3 reports simulation studies to evaluate the Joint-DP model and compare it to simpler Bayesian regression models and to the likelihood-based joint model of Xu *et al.* (2017). Section 4 summarizes analyses of the esophageal cancer dataset, and we close with a discussion in Section 5.

## 2 Probability Model

### 2.1 Sampling Model

For subjects  $i = 1, \dots, n$ , let  $N_i(t)$  denote the number of recurrent events in the time interval  $(0, t]$ , denote survival time by  $D_i$ , and let  $C_i$  denote the right censoring time. The final follow up time is  $\tilde{D}_i = \min\{D_i, C_i\}$ , and we denote the binary indicator of censoring by  $\delta_i = 1$  if  $\tilde{D}_i = C_i < D_i$ , and  $\delta_i = 0$  if  $\tilde{D}_i = D_i$ . Censoring is assumed to be independent of the recurrent events, survival, and covariates. Denote the vector of baseline covariates and treatment variables by  $\mathbf{Z}_i$ , with  $\mathbf{Z}_{Ri} = (Z_{Ri1}, \dots, Z_{Rip})'$  and  $\mathbf{Z}_{Di} = (Z_{Di1}, \dots, Z_{Diq})'$  the subvectors of  $\mathbf{Z}_i$  that affect the recurrent event process and survival, respectively.

For the recurrent event data, let  $T_{i,j}$  denote the time of  $j^{th}$  occurrence for patient  $i$ . Note that  $\{N_i(t), t > 0\}$  and  $\{T_{i,1}, T_{i,2}, \dots\}$  carry equivalent information about the recurrent event process. A time-varying pattern in the empirical intensity is suggested by Figure 1(c), and we assume a nonhomogeneous Poisson process (NHPP) with power law intensity (also known as a Weibull process) for the recurrent effusions. The power law intensity function has been used widely to flexibly accommodate time-varying patterns of recurrent events (Guida *et al.*, 1989; Kuo and Yang, 1996). Let  $\gamma_1, \dots, \gamma_n$  denote iid patient-specific random frailties that characterize intrinsic patient variability not explained by the observed covariates. For patient  $i$ , the time-varying conditional intensity (hazard) function of the Poisson recurrent event process  $[N_i(t) \mid \gamma_i, \mathbf{Z}_{Ri}, \boldsymbol{\beta}_R, \xi]$  is given by

$$\psi_R(t, \mathbf{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R, \xi) = \gamma_i \xi t^{\xi-1} \exp(\boldsymbol{\beta}'_R \mathbf{Z}_{Ri}). \quad (1)$$

The parameter  $\xi$  describes the time-varying pattern of the intensity function. If  $\xi < 1$  ( $> 1$ ), the intensity decreases (increases) over time. The latent frailty  $\gamma_i$  plays the role of a baseline intensity that varies from patient to patient, but it is assumed to be independent of time and the covariate parameter vector  $\boldsymbol{\beta}_R$ . Under the model in (1), the cumulative intensity function is  $\Psi_R(t, \mathbf{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R, \xi) = \int_0^t \psi_R(u, \mathbf{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R, \xi) du = \gamma_i t^\xi \exp(\boldsymbol{\beta}'_R \mathbf{Z}_{Ri})$ , which implies that the covariate effects are multiplicative in the cumulative intensity. The Poisson process assumption implies that, given  $\gamma_i$ , time gaps between recurrent events for each patient are independent. If this assumption is believed to be violated, alternative models may be considered, such as a Poisson cluster process or Markovian process. Note that a scale parameter for the intensity function is omitted due to potential identifiability issues. This is further discussed in §2.2.

For brevity, we denote  $\mathbf{N} = \{(N_i(\tilde{D}_i))_{i=1}^n\}$ ,  $\tilde{\mathbf{D}} = \{\tilde{D}_i\}_{i=1}^n$  and  $\boldsymbol{\psi}_R = \{\psi_R(t, \mathbf{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R, \xi)\}_{i=1}^n$ . Assuming independence between patients given  $\boldsymbol{\psi}_R$ , the likelihood of the observed recurrent event process given the latent frailties  $\gamma_1, \dots, \gamma_n$  and the truncation times  $\tilde{\mathbf{D}}$  is

$$\begin{aligned} p(\mathbf{N} \mid \boldsymbol{\psi}, \tilde{\mathbf{D}}) &= \prod_{i=1}^n \left\{ \prod_{j=1}^{N_i(\tilde{D}_i)} \psi_R(T_{i,j}, \mathbf{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R) \right\} \exp \left\{ - \int_0^{\tilde{D}_i} \psi_R(t, \mathbf{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R) dt \right\} \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^{N_i(\tilde{D}_i)} \gamma_i \xi T_{i,j}^{\xi-1} \exp(\boldsymbol{\beta}'_R \mathbf{Z}_{Ri}) \right\} \exp(-\gamma_i \tilde{D}_i^\xi e^{\boldsymbol{\beta}'_R \mathbf{Z}_{Ri}}). \end{aligned} \quad (2)$$

We assume conditional independence of  $N_i(t)$  and  $D_i$  given  $\gamma_i$ , and assume an accelerated failure-time (AFT) model for the failure time distribution given  $\gamma_i$ . Specifically, denoting  $\eta_i = \log(\gamma_i)$ , we assume that

$$\log(D_i) = \mu + \zeta \eta_i + \boldsymbol{\beta}'_D \mathbf{Z}_{Di} + \sigma W_i, \quad (3)$$

where the  $W_i$ 's are iid following a distribution with cdf  $F_W$ . Any tractable parametric or nonparametric model can be used for  $F_W$ , e.g., see Ghosh and Ghosal (2006); Walker and Mallick (1999) for nonparametric models. In preliminary goodness-of-fit analyses of survival time in the esophageal cancer data, we explored a set of distributions for  $F_W$ , and chose the extreme value distribution, characterized by density function  $f_W(w) = \exp(w - e^w)$ ,  $-\infty < w < \infty$ . The model selection procedure for  $W_i$  in (3) can be considered part of the methodology. Details of the model selection procedure and the extended model with model selection are described in § 2.3. With the choice of the extreme value distribution, the induced cdf  $F_{D,i}(t) = F_D(t \mid \mathbf{Z}_{Di}, \gamma_i, \mu, \zeta, \boldsymbol{\beta}_D)$  for  $t > 0$  of  $D_i$  is the Weibull with shape parameter  $1/\sigma$  and scale parameter  $\exp\{-(\mu + \zeta \eta_i + \boldsymbol{\beta}'_D \mathbf{Z}_{Di})/\sigma\}$ . This yields the survival hazard function

$$\begin{aligned} \psi_D(t, \mathbf{Z}_{Di}, \gamma_i, \mu, \zeta, \boldsymbol{\beta}_D, \sigma) &= \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \exp \left( - \frac{\mu + \zeta \eta_i + \boldsymbol{\beta}'_D \mathbf{Z}_{Di}}{\sigma} \right) \\ &= \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \gamma_i^{-\frac{\zeta}{\sigma}} \exp \left( - \frac{\mu + \boldsymbol{\beta}'_D \mathbf{Z}_{Di}}{\sigma} \right). \end{aligned} \quad (4)$$

For each subject  $i = 1, \dots, n$ , the random patient frailty,  $\gamma_i$ , is the key element that creates the Bayesian joint model from its two main components, the recurrent event process intensity function  $\psi_R(t, \mathbf{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R, \xi)$  in (1) and the survival process hazard function  $\psi_D(t, \mathbf{Z}_{Di}, \gamma_i, \mu, \zeta, \boldsymbol{\beta}_D)$  in (4). The hazard function  $\psi_D$  has a common factor  $\exp(-\mu/\sigma)$  and subject specific factors  $\gamma_i^{-\zeta/\sigma}$ . Compared to the intensity function  $\psi_R$  in (1) that is proportional to  $\gamma_i$ , the effect of  $\gamma_i$  on  $D_i$  is adjusted through the term  $\gamma_i^{-\zeta/\sigma}$ . This allows  $\gamma_i$  to have different effect sizes on the recurrent event intensity and the survival distribution hazard. Another way to view  $\gamma_i$  is that the patient's survival data is used to adjust for possible informative censoring in the intensity of the recurrent event process. For example,

if  $\zeta < 0$ , then a small value of  $\tilde{D}_i$  for a patient with larger frailty  $\gamma_i$  implies, on average, larger  $N_i(t)$  (more recurrent events). Additionally, from the practical viewpoint of obtaining a fitted joint model, the conditional independence of  $N_i(t)$  and  $D_i$  given  $\gamma_i$  greatly simplifies the MCMC computations required for posterior inference.

## 2.2 Prior

As described in Section 1, we utilize random frailties to account for between-patient heterogeneity and to adjust for potential biasing effects due to the lack of randomization in treatments as in our motivating dataset. We do not attempt to model possible systematic differences in patients between the two treatment groups due to the lack of randomization, as is done with propensity score based methods such as inverse probability of treatment weighting (IPTW) (Robins *et al.*, 2000; Austin, 2013) or doubly robust (augmented) IPTW (Robins *et al.*, 2000; Bang and Robins, 2005). Instead, we account for possible bias through a flexible modeling strategy for the distribution of the  $\gamma_i$ 's. To do this, we take a BNP approach by assuming a Dirichlet process (DP) (Ferguson, 1973) for the distribution of  $\gamma_i$  in Eqs (1) and (3). This adjusts for possible biases in the posterior estimates of the covariate effects,  $\beta_R$  and  $\beta_D$ , which quantify covariate and treatment effects on the recurrent event process and survival time, respectively. The basic idea is that, accounting robustly for irregular effects, possibly due to treatment selection bias, in the underlying patient-specific baseline recurrent event intensity and survival hazard avoids or mitigates these effects being reflected incorrectly in the estimates of  $\beta_R$  and  $\beta_D$ .

BNP models, including the DP or its many variants, have been applied successfully to capture nearly any pattern in data for a wide variety of applications. They express a full range of uncertainty about the form of the distribution function for parameters in a model by defining a distribution on the space of probability distribution functions. This enables one to obtain consistent estimation of essentially any distribution. See Müller and Rodriguez (2013) for a recent review of DP models and computational methods. For our application, because the DP provides a general form for the distribution of the  $\gamma_i$ 's, it allows us to focus on the distribution from which the random effects are drawn, instead of the individual random effects themselves. Bush and MacEachern (1996) used a BNP model for randomized block designs in a similar way. Recalling that  $\eta_i = \log(\gamma_i)$ , we assume

$$\eta_i \stackrel{iid}{\sim} G, \quad G \sim \text{DP}(\alpha G_0), \quad (5)$$

where  $\alpha$  is the total mass parameter and  $G_0$  is the baseline distribution of the DP. We assume  $G_0 = \text{N}(\bar{\eta}, v^2)$ .  $G_0$  determines the features of  $G$  such as shape, spread and location, and  $\alpha$  expresses the prior strength of belief in  $G_0$ . The DP prior in (5) does not hold the mean of  $G$  at zero even with  $\bar{\eta} = 0$ , but does not create an identifiability issue. A scale parameter is omitted in (1), and no fixed effect is paired with the random effect, while  $\zeta$  and  $\mu$  are common for all subjects in (3) (Li *et al.*, 2011). Following (Escobar and West, 1995), we assume that  $\alpha$  is random with prior  $\alpha \sim \text{Ga}(a_\alpha, b_\alpha)$  Using the stick-breaking representation



(Sethuraman, 1994),  $G$  can be expressed as an infinite mixture,

$$G = \sum_{k=1}^{\infty} v_k \prod_{\ell=1}^{k-1} (1 - v_{\ell}) \delta(\eta_k^*), \quad (6)$$

where  $v_k \mid \alpha \stackrel{iid}{\sim} \text{Be}(1, \alpha)$ ,  $\eta_k^* \stackrel{iid}{\sim} G_0$  and  $\delta(\eta^*)$  is a unit point mass at  $\eta^*$ . As Eq (6) shows,  $G$  is almost surely discrete and it allows ties in the  $\eta_i$ 's with positive probabilities, resulting in clustering of the  $\eta_i$ 's into a random number of subgroups. The following remark describes the dependence structure between recurrent events and survival time induced under the proposed model given by (1), (4), and (5).

**Remark 2.1** Let  $\mathcal{H}_i^- = \{N_i(u), 0 < u < t\}$  and  $N_i(t^-)$  denote the recurrent event history and the number of recurrent events in time interval  $(0, t)$ . Assume a NHPP for the recurrent events with the conditional intensity in (1). Let  $\psi'_{Ri}(t) = \xi t^{\xi-1} \exp(\beta'_R \mathbf{Z}_{Ri})$  and  $\Psi'_{Ri}(t) = \int_0^t \psi'_{Ri}(u) du$ . Similarly, let  $\psi'_{Di}(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \exp\left(-\frac{\mu + \beta'_D \mathbf{Z}_{Di}}{\sigma}\right)$  and  $\Psi'_{Di}(t) = \int_0^t \psi'_{Di}(u) du$ . Then,

$$\begin{aligned} \psi_{Ri}(t \mid \mathcal{H}_i(t^-), D_i \geq t) &\equiv \psi_{Ri}(t \mid N_i(t^-), D_i \geq t) = \psi'_{Ri}(t) \frac{E\left[\gamma_i^{N_i(t^-)+1} \exp\{-\gamma_i \Psi'_{Ri}(t) - \gamma_i^{-\zeta/\sigma} \Psi'_{Di}(t)\}\right]}{E\left[\gamma_i^{N_i(t^-)} \exp\{-\gamma_i \Psi'_{Ri}(t) - \gamma_i^{-\zeta/\sigma} \Psi'_{Di}(t)\}\right]}, \\ \psi_{Di}(t \mid \mathcal{H}_i(t^-)) &\equiv \psi_{Di}(t \mid N_i(t^-)) = \psi'_{Di}(t) \frac{E\left[\gamma_i^{N_i(t^-)-\zeta/\sigma} \exp\{-\gamma_i \Psi'_{Ri}(t) - \gamma_i^{-\zeta/\sigma} \Psi'_{Di}(t)\}\right]}{E\left[\gamma_i^{N_i(t^-)} \exp\{-\gamma_i \Psi'_{Ri}(t) - \gamma_i^{-\zeta/\sigma} \Psi'_{Di}(t)\}\right]}, \end{aligned}$$

where the expectations are taken with respect to the distribution  $G$  of  $\gamma_i$  in (5).

Remark 2.1 is an extension of the results in Sinha *et al.* (2008) for a random frailty distribution. Similar to their results, the shared frailty model with the NHPP in (1) implies that given  $D_i > t$  the conditional risks of a new recurrent event and survival depend on the recurrent event history through the number  $N(t^-)$  of past recurrent events. The value  $\zeta = 0$  implies independence of the recurrent event process and survival time. For given  $\sigma$ ,  $\zeta$  moderates the effect of survival time. In contrast with Sinha *et al.* (2008), who assume a gamma frailty distribution, our model in (5) assumes the frailty distribution itself to be random and the dependence structure between the recurrent event process and survival time remains unspecified.

Our use of a BNP prior for the random effects is similar in spirit to the approach of Xu *et al.* (2016) who used a BNP model for transition time distributions in a setting involving multi-stage dynamic treatment regimes and nonrandomized data. Their BNP model, dubbed a ‘‘dependent Dirichlet process with a Gaussian Process (DDP-GP)’’ assumed a GP prior on the means of the Gaussian mixture components of the DDP. They showed, by simulation, that the DDP-GP model yielded substantially better bias correction and better reliability for estimating overall mean survival, compared with conventional frequentist IPTW and augmented IPTW. Aside from the fact that our data structure is very different from that considered by Xu *et al.* (2016), our modeling approach differs from theirs in that we assume

a DP prior for the distribution of the patient random frailties, which in our setting are included primarily to obtain a joint model for the recurrent event process and survival time. Our simulation study, reported in Section 3, below, shows that this approach does a good job of reliably correcting for bias in the posterior estimates of both  $\beta_R$  and  $\beta_D$ .

For priors on  $\beta_R$ ,  $\beta_D$ ,  $\xi$ ,  $\mu$ , and  $\zeta$ , we assume  $\beta_R \sim N_p(\bar{\beta}_R, \Sigma_R)$ ,  $\beta_D \sim N_q(\bar{\beta}_D, \Sigma_D)$ ,  $\xi \sim \text{Gamma}(a_\xi, b_\xi)$ ,  $\zeta \sim N(\bar{\zeta}, \omega^2)$  and  $\mu \sim N(\bar{\mu}, \tau^2)$ . We complete the prior model specification by assuming  $\sigma \sim \text{IG}(a_\sigma, b_\sigma)$ . Specification of numerical hyperparameter values is discussed in § 3.

We use standard Markov chain Monte Carlo (MCMC) methods to implement posterior inference on the parameters  $\beta_R$ ,  $\beta_D$ ,  $\{\eta_i\}$  (equivalently,  $\{\gamma_i\}$ ),  $\xi$ ,  $\mu$ ,  $\zeta$  and  $\sigma$ . Usual MCMC posterior simulation proceeds by iteratively updating each of the parameters conditional on the currently computed values of all other parameters. It is possible to improve the mixing of the Markov chain, however, by jointly updating some parameters by means of a Metropolis-Hastings transition probability that proposes changes in those parameters. For our model, joint updating of  $\zeta$  and  $\mu$  may greatly improve the mixing. For updating  $\gamma_i$ , the algorithm in MacEachern and Müller (1998) is implemented to speed convergence of the Markov chain. Details of posterior computation are described in the Supplementary Material. We diagnose convergence and mixing of the described posterior MCMC simulation using trace plots and autocorrelation plots of imputed parameters. For both the upcoming simulation examples and the data analysis, we found no evidence of practical convergence problems. The posterior simulation is implemented in R and C and for the esophageal cancer data in § 4 simulation of 17000 Monte Carlo samples takes less than one minute on a 3.33 GHz CPU. An R package of the code used for simulations and the analysis of the esophageal cancer dataset in the following sections is available from the authors’ website <https://users.soe.ucsc.edu/~juheele/>.

## 2.3 Comparators

For comparison, we include five different Bayesian models, and also the frequentist joint scale-change model (JSCM) proposed by Xu *et al.* (2017). We first consider a Bayesian joint model with log-normal distributions for frailties,  $\gamma_i \stackrel{iid}{\sim} \text{logNormal}(0, \tilde{u}^2)$  with fixed  $\tilde{u}^2$  while assuming the same baseline functions in (1) and (4), called the “Joint-logNormal” model. We also consider alternative marginal models, for the recurrent event data while treating death as independent right censoring, called “R” models, or for survival time while ignoring the recurrent event data entirely, called “S” models. For each marginal Bayesian model, we either model frailty distributions nonparametrically through the DP or parametrically through a lognormal distribution. This gives four different models, “R-logNormal”, “R-DP”, “S-logNormal” and “S-DP.” For the “S” models, we assume that survival time follows an AFT distribution, given by

$$\log(D_i) = \mu_i^S + \beta_D' \mathbf{Z}_{Di} + \sigma W_i. \quad (7)$$

That is, the ‘‘S’’ models include a component with both a subject-specific frailty and covariate effects, similar to (3). We assume either  $\mu_i^S \stackrel{iid}{\sim} N(\bar{\mu}^S, (\nu^S)^2)$  with fixed  $\bar{\mu}^S$  and  $(\nu^S)^2$ , or  $\mu_i^S \stackrel{iid}{\sim} G_{\mu^S}$  and  $G_{\mu^S} \sim DP(\alpha_{\mu^S} \cdot G_{\mu^S_0})$ , where  $\alpha_{\mu^S} \sim Ga(a_\alpha^S, b_\alpha^S)$  and  $G_{\mu^S_0} = N(\bar{\mu}^S, (\nu^S)^2)$  with fixed  $\bar{\mu}^S$  and  $(\nu^S)^2$ . We assume the same prior distributions for  $\beta_D$  and  $\sigma$  as those under the Joint-DP model. Similarly, we define the ‘‘R’’ models for the recurrent event process only, that assume the likelihood (2). The R-logNormal model assumes a lognormal prior for  $\gamma_i$ , and the R-DP model assumes a DP in (5) for the prior distribution of  $\eta_i = \log(\gamma_i)$ . Both R models treat  $\tilde{D}_i$  as a censoring time that is independent of  $\mathbf{N}_i$ , and ignore the possibility that death informatively censors the recurrent event process.

Lastly, we include as a comparator the JSCM given by Xu *et al.* (2017) implemented in a R function, *reReg*, in the library *reReg*. The JSCM assumes AFT models for both recurrent events and failure times and relates recurrent events and survival to covariates through the regressions  $\log(T_{i,j}) \propto -\beta'_R \mathbf{Z}_{Ri}$  and  $\log(D_i) \propto -\beta'_D \mathbf{Z}_{Di}$ , respectively. In particular, the intensity of the recurrent events takes the form  $\lambda_i(t) = \gamma_i \lambda_0(t \exp(\beta'_R \mathbf{Z}_{Ri})) \exp(\beta'_R \mathbf{Z}_{Ri})$  and the hazard function of the survival time  $h_i(t) = \gamma_i h_0(t \exp(\beta'_D \mathbf{Z}_{Di})) \exp(\beta'_D \mathbf{Z}_{Di})$  for patient  $i$  at time  $t$ .  $\lambda_0$  and  $h_0$  are arbitrary but absolutely continuous baseline intensity and hazard functions, respectively. While the models for survival time are similar, the AFT model for recurrent events in JSCM differs from the Poisson process model assumed in our proposed models. Under the AFT model, the covariate effects in the original time scale change the time scale by a factor  $\exp(\beta'_R \mathbf{Z}_{Ri})$ , that is, for a patient with covariate  $\mathbf{Z}_{Ri}$ , the cumulative intensity is  $\Lambda_i(t) = \int_0^t \lambda_i(u) du = \gamma_i \Lambda_0(t \exp(\beta'_R \mathbf{Z}_{Ri}))$ , where  $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ . The AFT model for recurrent events includes a homogenous Poisson process as a special case, that is, the model in (1) with  $\xi = 1$ , but does not accommodate NHPP with  $\xi \neq 1$  in (1). Importantly, in contrast with (4), in the JSCM the frailty  $\gamma_i$  appears multiplicatively in  $h_i(t)$  without any adjustment, and the distribution of  $\gamma_i$  is unspecified, with  $E(\gamma | \mathbf{Z}, \beta_D) = m_\gamma$ . The parameter  $\beta_D$  in the JSCM has the opposite sign of the parameter  $\beta_D$  in (3) and (7), so to facilitate fair comparison, negative values of the JSCM-based estimates are used for the illustration in §3 and 4. The JSCM computes parameter estimates for the two processes sequentially, first estimating parameters for the recurrent event process including the frailties, then plugging in the frailty estimates and estimating the covariate effects on survival, similarly to Huang and Wang (2004). A resampling method is used to estimate the variances of the estimates, and asymptotic normality is used as a basis for a test and confidence intervals.

While we use a parameteric distribution of  $W_i$  in (3) to keep the model simple due to sparse recurrent event information in the motivating dataset, we empirically select a distribution of  $W_i$  for the Joint-DP model that better fits the data using the S-DP model. In particular, we fit the S-DP model to the survival data, using each of several distributions for  $W$ , and choose the distribution that gives the best fit in terms of smallest deviance information criterion (DIC). For the esophageal cancer data, we fit the survival data in § 4 with each of the standard extreme value, normal, and logistic distributions for  $W$ , which correspond to Weibull, lognormal, and log-logistic distributions for  $F_D$ , respectively. Comparison of the three distributions based on DIC values showed that the extreme value distribution for  $W$

gave the best fit to the survival data in § 4. In general, one may choose a distribution of  $W_i$  based on some other reasonable model comparison procedures using either the survival data only or the entire data (e.g. see Chapter 7 of Robert (2007)), or use a nonparametric model for  $W_i$  similar to the models in Sinha *et al.* (2008) and Ouyang *et al.* (2013). The Joint-DP model can be extended by including this model selection procedure to choose a distribution for  $W_i$ . We examined robustness of this extended approach empirically through simulation studies in §2.3 of the Supplementary Materials.

### 3 Simulation Study

In this section, we assess the performance of the Joint-DP model and compare it to the alternative models described above through a simulation study. For each dataset, we simulated a sample of  $n = 500$  patients with three covariates in both the recurrent event intensity and survival distribution ( $p_R = p_D = 3$ ,  $\mathbf{Z}_{Ri} = \mathbf{Z}_{Di} = \mathbf{Z}_i$ ). We simulated each patient's covariates by letting  $Z_{1i} \stackrel{iid}{\sim} 0.5N(-0.5, 0.25^2) + 0.5N(0.5, 0.25^2)$  and  $Z_{2i} \stackrel{iid}{\sim} \text{Unif}(-\sqrt{3}, \sqrt{3})$ . We introduced bias in the treatment assignment by simulating the treatment indicator  $Z_{3i} \in \{0, 1\}$  using the probability

$$w_i = \begin{cases} 0.05 & \text{if } 1/\{1 + \exp(-2Z_{1i})\} < 0.05, \\ 1/\{1 + \exp(-2Z_{1i})\} & \text{if } 0.05 \leq 1/\{1 + \exp(-2Z_{1i})\} \leq 0.95, \\ 0.95 & \text{if } 1/\{1 + \exp(-2Z_{1i})\} > 0.95. \end{cases}$$

We assigned each patient to treatment  $Z_{3i} = 0$  with probability  $w_i$  and treatment  $Z_{3i} = 1$  with probability  $1 - w_i$ . This randomization mimics physician behavior in which  $Z_{1i}$  is used for choosing treatment, with a stochastic component to reflect between-physician variability in this covariate-based treatment selection process. For example,  $w = 0.26$  if  $Z_1 = -0.5$ , and  $w = 0.73$  if  $Z_1 = 0.5$ . Since  $w_i$  is increasing in  $Z_{1i}$ , patients with larger  $Z_{1i}$  are more likely to be given treatment  $Z_{3i} = 0$ . We simulated patient frailties  $\eta^{\text{TR}} = \log(\gamma^{\text{TR}})$  from a mixture distribution with  $K = 5$  components,  $p(\eta_i^{\text{TR}}) = \sum_{k=1}^K \pi_k \phi(\eta_i^{\text{TR}} | \eta_k^{\text{TR}}, 0.25)$ , where  $\phi(\cdot | \eta^*, v^2)$  is the normal distribution with mean  $\eta^*$  and variance  $v^2$ . We fixed the mixture component means  $(\eta_1^{\text{TR}}, \dots, \eta_5^{\text{TR}}) = (-1.5, -0.4, 0.0, 0.4, 0.6)$  and simulated the component weights from  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_5) \sim \text{Dir}(3, 3, 3, 3, 3)$ , where  $\text{Dir}(\mathbf{a})$  denotes the Dirichlet distribution with parameter  $\mathbf{a}$ . We set the covariate effects  $\boldsymbol{\beta}_R^{\text{TR}} = (0.30, 0.25, -0.30)'$ ,  $\boldsymbol{\beta}_D^{\text{TR}} = (-0.30, -0.25, 0.30)'$ ,  $\xi^{\text{TR}} = 0.8$ ,  $\zeta^{\text{TR}} = -1.2$ ,  $\mu^{\text{TR}} = 0.5$  and  $\sigma^{\text{TR}} = 0.7$ . We simulated the death time  $D_i$  from (3) under a Weibull model with the above assumed true parameter values. We let  $\delta_i = 0$  (the survival outcome is observed) or 1 (the survival outcome is censored) with probabilities 0.75 and 0.25, respectively, set  $\tilde{D}_i = D_i$  for patients with  $\delta_i = 0$ , and for patients with  $\delta_i = 1$  we assumed uniform censoring time  $\tilde{D}_i \sim \text{Unif}(0, D_i)$ .  $N_i(t)$  was simulated from a nonhomogeneous Poisson process with intensity function  $\gamma_i^{\text{TR}} \xi^{\text{TR}} t^{\xi^{\text{TR}} - 1} \exp(\boldsymbol{\beta}_R^{\text{TR}'} \mathbf{Z}_i)$ , with  $\gamma_i^{\text{TR}} = \exp(\eta_i^{\text{TR}})$  and censored at  $t = \tilde{D}_i$ . A total of 1000 datasets were simulated under this set-up.

For the Joint-DP model prior hyperparameters, we used large values for prior variances

to express weak prior information. We set  $\bar{\beta}_R = \bar{\beta}_D = \bar{\eta} = \bar{\mu} = 0$ , with  $\Sigma_R = \Sigma_D = \text{diag}(10^2)$  and  $\tau^2 = 10^2$ ,  $a_\sigma = b_\sigma = 1$ . We fixed  $a_\alpha = b_\alpha = 3$  for the DP total mass parameter  $\alpha$ , which implies prior mean 6.79 and standard deviation 3.74 for the number of clusters. We did this using the R function `DPelicit` of `DPpackage` (Jara *et al.*, 2018). We let  $v^2 = 1$  for the base distribution  $G_0$ . For the priors of  $\xi$  and  $\zeta$ , we set  $a_\xi = b_\xi = 1$ ,  $\bar{\zeta} = 0$ , and  $\omega^2 = 10^2$ . To run the MCMC simulation, we initialized the parameters at their prior means. For  $\eta_i$ , we let each  $\eta_i$  have its own cluster and  $\eta_k^*$  be a random draw from  $G_0$ . We then implemented posterior inference using MCMC simulation over 10,000 iterations, discarding the first 4,000 iterations as burn-in and choosing every other sample as thinning.

The simulation results are summarized for  $\beta$  in Table 1 and for  $\xi$  and  $\sigma$  in Table 2. For each of the 1000 simulated datasets, we computed estimates of posterior means and 95% credible intervals (CIs) of the parameters, and used them to evaluate four quantities, average of the posterior means (Mean), average difference between each posterior mean and the truth (Bias), average lengths of the CIs (Ave. CI length), and proportion of the CIs containing the truth (Coverage). For all  $\beta$ 's, the Joint-DP model produces very small biases and the CIs capture their true values with proportions close to 95%, but slightly less than 95%. Although bias is introduced by using  $X_1$  to assign  $X_3$  to either of the values 0 and 1, the model yields estimation biases close to zero, and their coefficients are well estimated with tight intervals. The model also provides reasonable inference for the power parameter of the baseline intensity function  $\xi$  and the shape parameter of the baseline hazards function  $\sigma$ .

For comparison, we fit each of the five models described in § 2.3. Under the Bayesian models, we specified their hyperparameters and ran MCMC posterior simulations similarly to the Joint-DP model. The results under the comparators also are summarized in Tables 1 and 2. In the tables, the fitted joint models utilizing both the recurrent event data and the survival time data are given on the top, followed by fits of the models for recurrent event outcomes only (R models) and the models for survival time only (S models).

We first compare the three joint models, Joint-DP, Joint-logNormal, and JSCM, given in the top three boxes in each of Tables 1 and 2. Compared to the Joint-DP model, the Joint-logNormal model yields larger Bias, wider CIs, and smaller Coverage for all  $\beta$ , especially for treatment effects  $\beta_{R3}$  and  $\beta_{D3}$ . For example, (Bias, Ave. CI length, Coverage) are (-0.322, 0.392, 28.8%) for  $\beta_{R3}$  and (0.216, 0.509, 60.6%) for  $\beta_{D3}$  under the Joint-logNormal model. This comparatively poor performance in estimation of  $\beta_1$  and  $\beta_3$  for both recurrent events and survival may be due to the simulated treatment assignment bias, and also possibly mis-specification of the frailty distribution. The Joint-logNormal model also produces poor estimation of the baseline function parameters  $\xi$  and  $\sigma$  compared to the Joint-DP model. The JSCM produces larger bias and larger confidence intervals for both  $\beta_R$  and  $\beta_D$ , although the estimates of  $\beta_{R1}$  and  $\beta_{R3}$  are better than those under the Joint-logNormal model. Since the JSCM assumes AFT models for recurrent events and survival times, the assumption under the JSCM is satisfied only for survival times in the simulation truth and their estimates of  $\beta_D$  only are comparable to the simulation truth. The coverage probabilities of the confidence intervals under JSCM are above 99% for  $\beta_D$ , possibly leading to the conclusion of no

		Recurrent Events			Survival		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP</b> (Joint with DP)	Mean	0.288	0.247	-0.312	-0.291	-0.251	0.311
	Bias	<b>-0.012</b>	<b>-0.003</b>	<b>-0.012</b>	<b>0.009</b>	<b>-0.001</b>	<b>0.011</b>
	Ave. CI length	<b>0.419</b>	<b>0.208</b>	0.464	<b>0.432</b>	<b>0.213</b>	0.483
	Coverage	93.1%	92.6%	93.2%	<b>94.1%</b>	93.5%	<b>94.3%</b>
<b>Joint-logNormal</b> (Joint with logNormal Frailty)	Mean	0.174	0.261	-0.622	-0.202	-0.255	0.516
	Bias	-0.126	0.011	-0.322	0.098	-0.005	0.216
	Ave. CI length	0.473	0.246	<b>0.392</b>	0.504	0.246	0.509
	Coverage	75.0%	<b>95.3%</b>	28.8%	85.3%	<b>94.4%</b>	60.6%
<b>JSCM</b>	Mean	0.365	0.310	-0.365	-0.225	-0.202	0.244
	Bias	0.065	0.060	-0.065	-0.075	-0.048	0.056
	Ave. CI length	0.760	0.381	0.868	1.512	0.745	1.690
	Coverage	92.3%	88.4%	94.0%	99.5%	99.6%	99.6%
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.260	0.223	-0.276			
	Bias	-0.040	-0.027	0.024			
	Ave. CI length	0.446	0.221	0.496			
	Coverage	<b>93.6%</b>	91.4%	<b>94.8%</b>			
<b>R-logNormal</b> (Recurrent events only with logNormal Frailty)	Mean	0.016	0.222	-0.856			
	Bias	-0.284	-0.028	-0.556			
	Ave. CI length	0.490	0.260	<b>0.406</b>			
	Coverage	38.9%	94.1%	3.3%			
<b>S-DP</b> (Survival only with DP)	Mean				-0.276	-0.251	0.339
	Bias				0.024	-0.001	0.039
	Ave. CI length				0.457	0.224	0.507
	Coverage				94.0%	93.7%	93.3%
<b>S-logNormal</b> (Survival only with logNormal Frailty)	Mean				0.121	-0.253	1.270
	Bias				0.421	-0.003	0.970
	Ave. CI length				0.454	0.233	<b>0.425</b>
	Coverage				16.6%	92.0%	0.7%

Table 1: Simulation results for covariate effect estimation. The values are based on 1000 simulated datasets, each fit using each model. The proposed joint model assuming a DP prior for the frailty distribution is labeled Joint-DP. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

		Baseline Intensity( $\xi$ )	Baseline Hazards ( $\sigma$ )
Simulation Truth		0.800	0.700
<b>Joint-DP</b> (Joint with DP)	Mean	0.796	0.702
	Bias	<b>-0.004</b>	<b>0.002</b>
	Ave. CI length	0.122	<b>0.202</b>
	Coverage	<b>96.1%</b>	97.1%
<b>Joint-logNormal</b> (Joint with logNormal Frailty)	Mean	0.848	0.617
	Bias	0.048	-0.083
	Ave. CI length	0.117	0.264
	Coverage	58.8%	74.9%
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.686	
	Bias	-0.114	
	Ave. CI length	0.105	
	Coverage	3.3%	
<b>R-logNormal</b> (Recurrent events only with logNormal Frailty)	Mean	0.695	
	Bias	-0.105	
	Ave. CI length	<b>0.099</b>	
	Coverage	11.0%	
<b>S-DP</b> (Survival only with DP)	Mean		0.736
	Bias		0.036
	Ave. CI length		0.265
	Coverage		<b>95.4%</b>
<b>S-logNormal</b> (Survival only with logNormal Frailty)	Mean		0.608
	Bias		-0.092
	Ave. CI length		0.291
	Coverage		42.4%

Table 2: Simulation results for parameters in baseline functions. Results are based on 1000 simulated datasets, each fit with each model. The parameters  $\xi$  and  $\sigma$  control time-varying patterns in the recurrent event baseline intensity and survival hazard functions, respectively. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length) and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

significant effect on survival for truly significant variables. This may be due to the violation of the assumption for recurrent events, the two-step estimation method, or the modeling assumption that the frailties appear in the hazard functions without adjustment.

Another interesting comparison is between the joint models and the models with only one outcome. Table 1 shows that improvement in inference about  $\beta_R$  and  $\beta_D$  through

combining information from two sources via the Joint-DP model is minimal. In contrast, Table 2 shows that the Joint-DP model greatly improves estimation of the baseline intensity and hazard function parameters  $\xi$  and  $\sigma$  compared to the R-DP and S-DP models. Although the performance of the Joint-logNormal is inferior to that of the Joint-DP model, the Joint-logNormal model gives much more reliable estimates of  $\beta_R$  and  $\beta_D$  compared to the R-logNormal and S-logNormal models. Considered together, these results imply that combining the two sources of information through joint modeling improves covariate effect estimation, and greatly improves estimation of the baseline intensity and hazard functions, compared to fitting separate models for the recurrent event process and survival. Moreover, the latter improvement is largest when a BNP model is assumed for the frailty distribution.

We further examined the performance of the Joint-DP model through additional simulation studies, Simulations 2-8 in §2 of the Supplementary Materials. In these simulations, we kept most of the simulation set-up used in Simulation 1, including biased generation of  $Z_3$ , and focused on comparisons of the Joint-DP model to particular sets of models. In Simulations 2 and 3, we simulated the frailties from a lognormal distribution, that is, the frailty distribution is parametric in the simulation truth and from a mixture of gamma distributions, respectively. The performance of the models with the DP for the frailty distribution is almost the same as in Simulation 1, or is slightly improved, with posterior coverage probabilities closer to 95%. Interestingly, the models with lognormal frailty distribution perform poorly in Simulation 2, especially for  $\beta_1$  and  $\beta_3$  estimation for the recurrent events. This comparison implies that, even when the frailty distribution is correctly specified, biased treatment assignment may severely deteriorate inference, especially for parameters related to treatment and treatment assignment. In Simulations 4 and 5, we generated recurrent events from a homogeneous Poisson process and a Poisson cluster process, respectively, for further comparisons of the Joint-DP to JSCM. Recall that JSCM assumes an AFT model for recurrent events, and the NHPP in Simulation 1 violates the assumptions under JSCM. From Simulation 4, when the assumption under JSCM is not violated, JSCM yields on average unbiased estimates, but with larger interval estimates and larger coverage probabilities than the Joint-DP model. Simulation 5 shows that the Joint-DP produces reasonable inference on  $\beta_R$  and  $\beta_D$  even when the recurrent event model is not correctly specified. Simulation 6 is conducted to study robustness of the extended Joint-DP that includes the model selection procedure to misspecification of the distribution of  $W_i$ . In this simulation, inference on  $\beta$  under the extended model is minimally affected compared to the Joint-DP that assumes the true distribution of  $W_i$  known, and is better than JSCM that assumes a nonparametric distribution for  $W_i$ . Lastly, in Simulation 6, we further compared the Joint-DP model to the single outcome R-DP and S-DP models. The results show that when survival times are more heavily censored, the Joint-DP significantly improves inference by borrowing strength between the recurrent event and survival time outcomes. A more detailed summary of the additional simulations is given in §2 of the Supplementary Materials.



## 4 Analysis of the Esophageal Cancer Data

In this section, we summarize our analyses of the esophageal cancer dataset described in §1. Recall that the primary goal is to compare effects of the RT modalities, 3DRT and IMRT, on recurrent effusion occurrences and survival, while accounting for effects of age, BMI, KPS score, histology, and cancer stage. The continuous variables age and BMI were standardized to have mean 0 and variance 1. The variables RT modality, KPS score, histology and cancer stage are binary, with RT modality=1 for IMRT, KPS score=1 for good, histology = 1 for adenocarcinoma, and cancer stage = 1 for stage 3–4 (advanced) cancer. The RT modality and covariate parameter vectors thus have dimension  $p_R = p_D = 6$ .

Model	Covariates	$\beta_R$		$\beta_D$	
<b>Joint-DP</b> (Joint with DP)	IMRT	<b>-0.82</b>	<b>(-1.07, -0.56)</b>	<b>0.45</b>	<b>(0.21, 0.68)</b>
	Age	0.08	(-0.05, 0.22)	0.01	(-0.12, 0.14)
	BMI	-0.13	(-0.27, 0.00)	<b>0.16</b>	<b>(0.04, 0.29)</b>
	KPS Score	-0.22	(-0.64, 0.11)	<b>0.65</b>	<b>(0.26, 1.06)</b>
	Adeno Histology	-0.18	(-0.46, 0.10)	-0.16	(-0.44, 0.10)
	Cancer Stage	0.06	(-0.19, 0.34)	<b>-0.46</b>	<b>(-0.73, -0.21)</b>
<b>Joint-logNormal</b> (Joint with logNormal Frailty)	IMRT	<b>-0.88</b>	<b>(-1.15, -0.62)</b>	<b>0.45</b>	<b>(0.18, 0.66)</b>
	Age	0.08	(-0.06, 0.22)	-0.01	(-0.15, 0.12)
	BMI	-0.12	(-0.28, 0.03)	<b>0.15</b>	<b>(0.01, 0.28)</b>
	KPS Score	<b>-0.36</b>	<b>(-0.61, -0.10)</b>	<b>0.68</b>	<b>(0.38, 0.93)</b>
	Adeno Histology	-0.10	(-0.39, 0.20)	<b>-0.31</b>	<b>(-0.56, -0.02)</b>
	Cancer Stage	0.08	(-0.16, 0.33)	<b>-0.47</b>	<b>(-0.68, -0.22)</b>
<b>JSCM</b>	IMRT	<b>-0.85</b>	<b>(-1.45, -0.25)</b>	-0.22	(-0.98, 0.54)
	Age	<b>0.24</b>	<b>(0.06, 0.41)</b>	0.07	(-0.32, 0.46)
	BMI	-0.16	(-0.41, 0.09)	0.07	(-0.25, 0.38)
	KPS Score	-0.54	(-1.15, 0.06)	0.70	(-0.20, 1.60)
	Adeno Histology	-0.29	(-0.89, 0.30)	-0.26	(-1.29, 0.77)
	Cancer Stage	0.11	(-0.21, 0.43)	<b>-0.84</b>	<b>(-1.64, 0.04)</b>

Table 3: Fits of the Esophageal Cancer Data under the Joint Models. Point estimates of  $\beta_R$  and  $\beta_D$  are given with their 95% credible intervals in parentheses under the proposed joint model with DP, Joint-logNormal, and with confidence intervals under the frequentist JSCM. Covariates having a statistically significant effect are given in boldface.

We specified hyperparameters similar to those in the simulations for the Bayesian models. The MCMC simulation was run over 17,000 iterations, with the first 8,000 iterations discarded as burn-in and every third sample kept as thinning and used for inference. Posterior inferences are summarized in Tables 3 and 4. When interpreting the numerical parameter estimates in Table 3, it is important to bear in mind that  $\beta_{Rj} < 0$  corresponds to a lower recurrent effusion rate, while  $\beta_{Dj} > 0$  corresponds to a lower death rate, that is, longer survival. Under the Joint-DP model, the posterior means of the coefficients for the RT modality

Model	Baseline Intensity( $\xi$ )	
<b>Joint-DP</b>	0.84	(0.76, 0.92)
<b>Joint-logNormal</b>	0.90	(0.83, 0.98)
<b>R-DP</b>	0.68	(0.62, 0.74)
<b>R-logNormal</b>	0.70	(0.64, 0.77)
	Baseline Hazards ( $\sigma$ )	
<b>Joint-DP</b>	0.60	(0.45, 0.76)
<b>Joint-logNormal</b>	0.43	(0.30, 0.61)
<b>S-DP</b>	0.63	(0.49, 0.78)
<b>S-logNormal</b>	0.58	(0.44, 0.71)

Table 4: Baseline Function Parameter Estimates of the Esophageal Cancer Data. Point estimates of  $\xi$  and  $\sigma$  are given with their 95% credible intervals in parentheses under Joint-DP, Joint-logNormal, R-DP, R-logNormal, S-DP and S-logNormal.

IMRT are  $\hat{\beta}_{R1} = -0.82$  and  $\hat{\beta}_{D1} = 0.45$ , with respective 95% credible intervals  $(-1.07, -0.56)$  and  $(0.21, 0.69)$ . On average, IMRT decreases the effusion rate by a multiplicative factor of  $\exp(-0.82) = 0.442$ , and decreases the hazard of death by a multiplicative factor of 0.468. One may infer that IMRT provides both a significantly lower rate of effusion occurrences and longer survival compared to 3DRT. These inferences agree with previous findings in Chuong *et al.* (2016) that IMRT is significantly associated with better clinical outcomes, such as decreased toxicity in the lungs and heart and longer survival. Higher BMI, higher KPS score, and lower cancer stage also were significantly associated with longer survival, while no other covariate besides RT modality had a significant effect on the effusion occurrence intensity. While our inference that higher BMI is positively associated with longer survival in RT patients may seem spurious or counterintuitive, it agrees with previously published results (e.g., Zhang *et al.* (2013); Ji *et al.* (2016)). Recalling the forms (1) of the effusion intensity function and (4) of the hazard of death, Table 4 shows that the effusion intensity decreased over time ( $\hat{\xi} = 0.84$ ) and the hazard of death increased over time ( $\hat{\sigma} = 0.60$ ). It is also reflected in the posterior estimates of the functions  $t^\xi$  and  $t^{1/\sigma-1}$  that are proportional to the cumulative baseline intensity and hazard functions  $\int_0^t \psi_R(u, \mathbf{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R, \xi) du$  and  $\int_0^t \psi_D(u, \mathbf{Z}_{Di}, \gamma_i, \mu, \zeta, \boldsymbol{\beta}_R, \sigma) du$  in Figure 2(a) and (d), where the black solid and dashed lines represent posterior means and 95% pointwise credible intervals, respectively. Figure 2 of the Supplementary Materials illustrates the posterior distributions of all parameters. From panel (o) of the figure, the posterior mean of  $\zeta$  is -1.32, implying that the intensity of the effusion occurrence process and the hazard of death are positively associated. The posterior mean of  $-\zeta/\sigma$  is 2.25, implying that the patient random effects are scaled up for survival. In particular, a simpler model with  $\zeta \equiv 1$  would miss the fact that the magnitudes of the impacts of  $\gamma_i$  on the two hazard functions differ. Panel (a) of Figure 3 shows a kernel density estimate of the posterior means  $\hat{\gamma}_i$  of  $\gamma_i$ . A very interesting result is that the posterior distribution of  $\gamma_i$  is bimodal, which is revealed by the DP distribution due to its flexibility.

This suggests that, because a simpler, unimodal assumed distribution for the  $\gamma_i$ 's would miss this bimodality, this in turn would affect the posterior estimates of  $\beta_R$  and  $\beta_D$ .

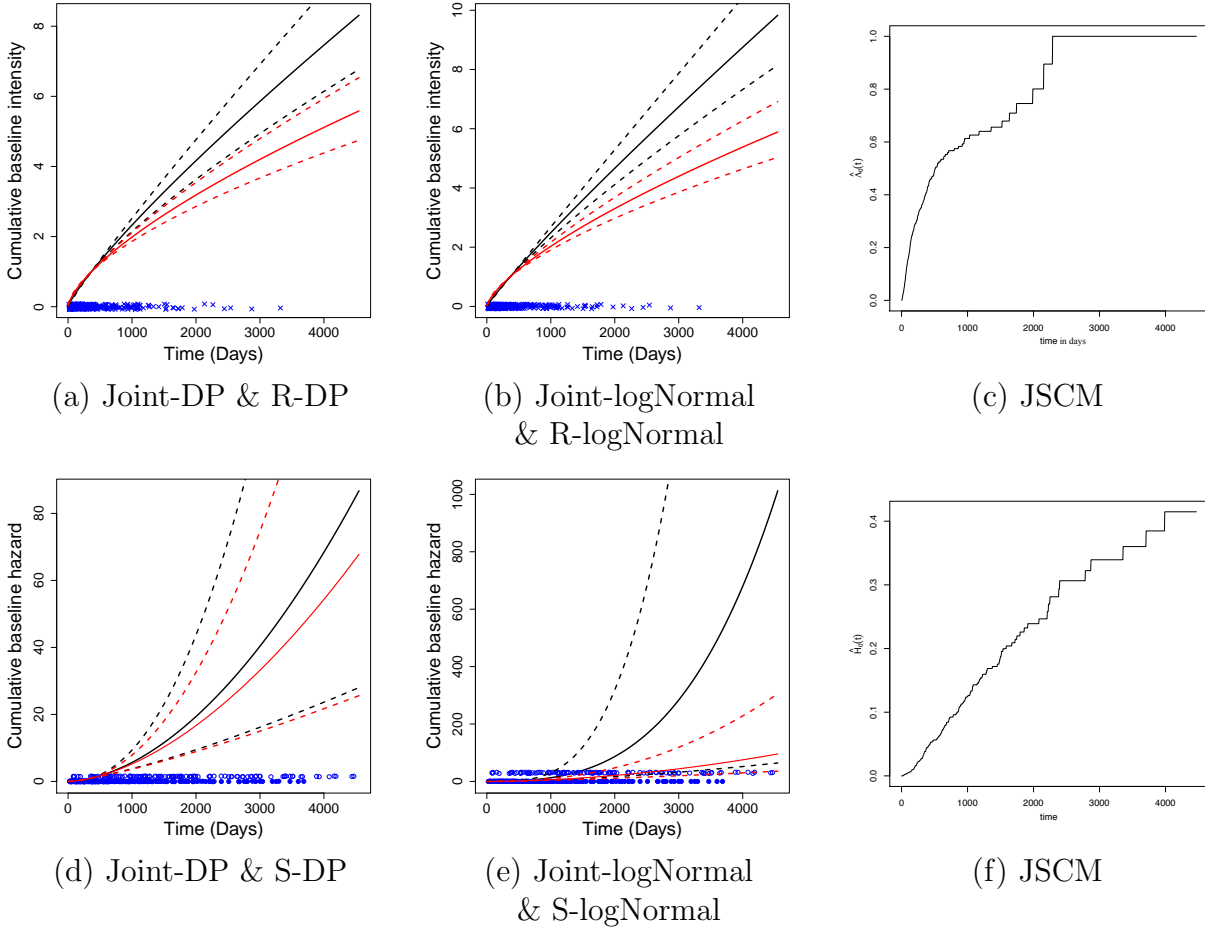
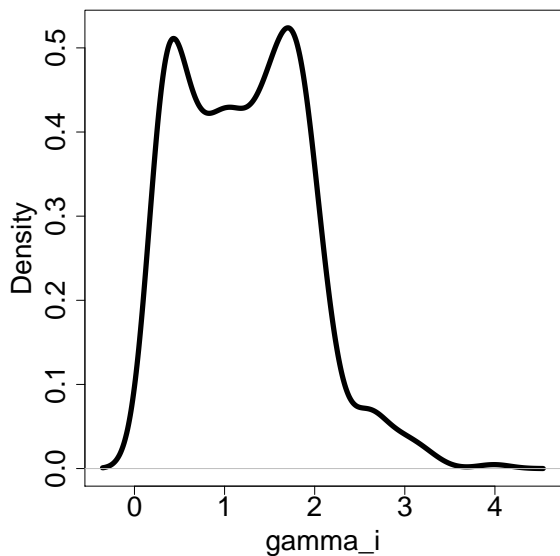
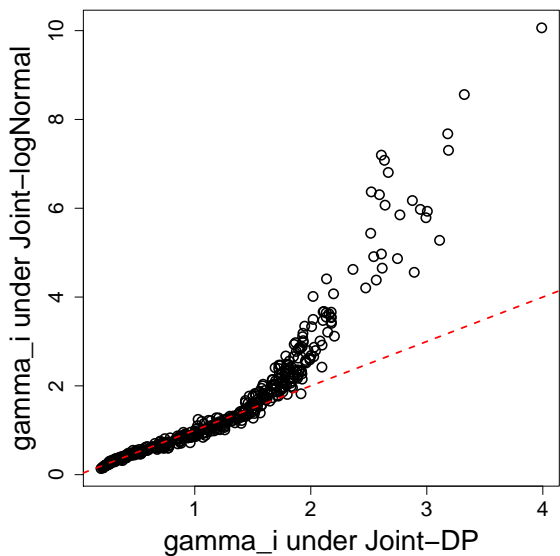


Figure 2: Baseline Functions for the Esophageal Cancer Data. In panels (a)-(c), inferred baseline intensity functions under different models are illustrated. Panels (d)-(f) has plots of inferred baseline hazard functions. Solid lines and dotted lines are posterior means and 95% credible intervals. The joint models are in black and the models with single outcomes in red. Blue crosses in (a)-(c) denote observed  $T_{i,j}$ . Open and closed dots in (d)-(f) represent observed  $\tilde{D}_i$  with  $\delta_i = 1$ (censored) and  $\delta_i = 0$ (observed), respectively. Note that the intensity and hazard functions of the Bayesian models including Joint-DP omits scale parameters to avoid potential identifiability issues and are not directly comparable to those under JSCM and between the models.

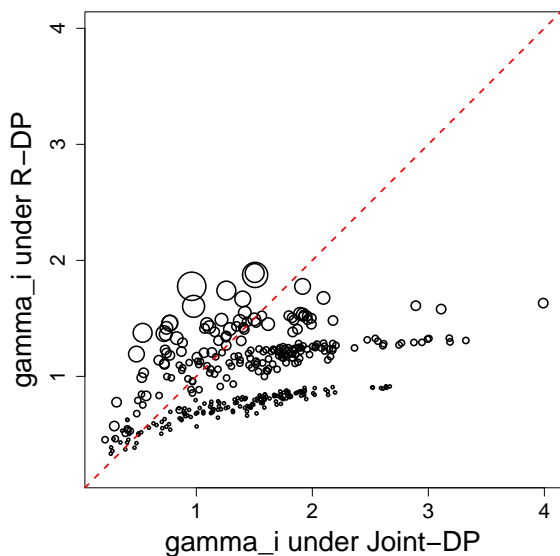
Posterior inferences on  $\beta_R$  and  $\beta_D$  under the comparators also are summarized in Tables 3, 5 and 4. Table 3 shows that the two joint models Joint-logNormal and JSCM identify different sets of significant covariates. For example, the Joint-logNormal model concludes that all covariates but Age are statistically significant for survival. The Joint-logNormal



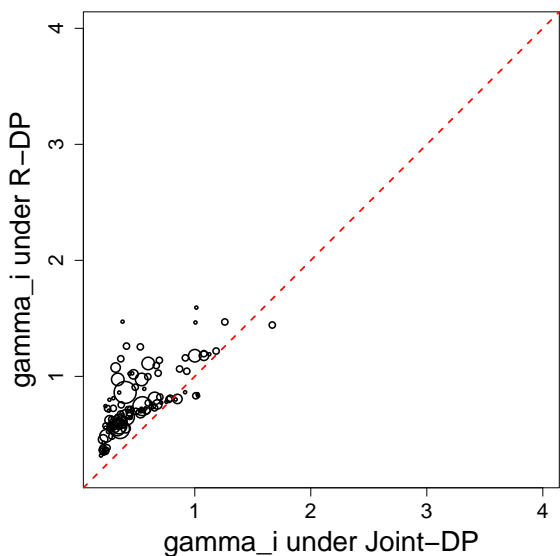
(a)  $\hat{\gamma}_i$  under Joint-DP



(b)  $\hat{\gamma}_i$  under Joint-DP vs Joint-logNormal



(c)  $\hat{\gamma}_i$  under Joint-DP vs R-DP  
for subjects with observed survival



(d)  $\hat{\gamma}_i$  under Joint-DP vs R-DP  
for subjects with censored survival

Figure 3: [Esophageal Cancer Data – Frailties] Posterior distributions of frailties  $\{\gamma_i\}$ . A kernel density estimate of the posterior means of the frailties  $\{\gamma_i\}$  is given in (a). A scatterplot of  $\hat{\gamma}_i$  pairs obtained from Joint-DP and Joint-logNormal is shown in (b). Scatterplots of  $\hat{\gamma}_i$  pairs obtained from Joint-DP and R-DP are shown in (c) and (d) for subjects with  $\delta_i = 0$  (observed survival) and  $\delta_1$  (censored survival), respectively.

model gives slightly larger  $\hat{\xi}$  and smaller  $\hat{\sigma}$  compared to the estimates under the Joint-DP

model, which is similar to the comparison in Simulation 1. We compared the posterior frailty estimates under the models Joint-DP and Joint-logNormal graphically, shown in Figure 3(b). The figure shows that frailty estimates are similar under the two models when they are small, but the Joint-logNormal model tends to produce much more extreme estimates than Joint-DP for larger frailties. We also compare the Joint-DP and Joint-logNormal based on model assessment metrics, DIC (Spiegelhalter *et al.*, 2002) and the log-pseudo marginal likelihood statistic (LPML, Gelfand *et al.* (1992); Gelfand and Dey (1994)). DIC and LPML are commonly used for model comparison in the Bayesian paradigm. DIC measures posterior prediction error based on deviance penalized by model complexity. LPML is a metric based on cross validated posterior predictive probability, and is defined as the sum of the logarithms of subject specific conditional predictive ordinates (CPOs, Geisser (1993)). Sinha *et al.* (2008) and Ouyang *et al.* (2013) used CPO for model comparison and diagnostics for their Bayesian shared frailty models. We integrate out subject level latent parameters (frailties) and compute DIC and LPML based on the partially marginalized likelihoods for more reliable comparison (Millar, 2009; Lee *et al.*, 2016). Details of the DIC and LPML computation are discussed in §3 of the Supplementary Materials. Table 6 provides DIC and LPML for the model fits. Note that a model with a smaller DIC and/or a larger LPML indicates a model for a better fit of the data. Both criteria indicate that the Joint-DP gives a substantially better fit to the data. JSCM produces larger confidence intervals, especially for  $\beta_D$ . In sharp contrast with both of the Bayesian joint models, JSCM indicates that there is no statistically significant effect of IMRT on survival. Point estimates of the cumulative baseline intensity and hazard functions are illustrated in Figure 2(c) and (f), respectively. Recall that JSCM does not impose any functional form for the baseline functions. While the estimates under the Joint-DP, Joint-logNormal and JSCM are not directly comparable due to the difference in the formulation, the estimated patterns in the baseline functions between the Joint-DP and JSCM are closer than those between the Joint-logNormal and JSCM.

The separate models with DP in Table 5 identify the same set of significant covariates as the Joint-DP model, although their point estimates are slightly different. The estimates of the baseline function parameters are compared in Table 4. The posterior estimates of  $\xi$  are considerably different under the Joint-DP and R-DP models, with much smaller  $\hat{\xi}$  under the R-DP model. The estimates of the  $\gamma_i$ 's under the Joint-DP and R-DP models are compared graphically in Figure 3(c) and (d) for patients with  $\delta = 0$ (observed) and 1 (censored), respectively. The dot sizes are proportional to observed effusion event counts  $N_i(\tilde{D}_i)$ . The scatter plots show that  $\hat{\gamma}_i$  tends to be larger under the R-DP model than under the Joint-DP model for patients with large  $N_i(\tilde{D}_i)$  (big symbols) and/or censored survival (in panel (d) ). Also, patients with  $N_i(\tilde{D}_i) = 0$  and observed survival tend to have larger  $\hat{\gamma}_i$  under the Joint-DP model (see a band of small black circles below the 45-degree line in panel (c)). This illustrates that the Joint-DP model properly accounts for informative censoring on effusion events by death and yields  $\hat{\gamma}_i$  values different from those obtained under the R-DP model. The patterns of the estimated cumulative baseline intensity and hazard functions under the joint models and the single outcome models are compared in Figure 2(a) and (d) where the back and red colors are for the joint models and the single outcome models,

Model	Covariates	$\beta_R$		$\beta_D$	
<b>R-DP</b> (Recurrent events only with DP)	IMRT	<b>-0.73</b>	<b>(-0.98, -0.50)</b>		
	Age	0.09	(-0.05, 0.23)		
	BMI	-0.07	(-0.21, 0.07)		
	KPS Score	-0.13	(-0.45, 0.18)		
	Adeno Histology	-0.19	(-0.46, 0.07)		
	Cancer Stage	-0.01	(-0.26, 0.26)		
<b>R-logNormal</b> (Recurrent events only with logNormal Frailty)	IMRT	<b>-0.85</b>	<b>(-1.14, -0.54)</b>		
	Age	0.09	(-0.07, 0.24)		
	BMI	-0.10	(-0.27, 0.06)		
	KPS Score	<b>-0.45</b>	<b>(-0.74, -0.13)</b>		
	Adeno Histology	-0.26	(-0.56, 0.04)		
	Cancer Stage	-0.09	(-0.37, 0.18)		
<b>S-DP</b> (Survival only with DP)	IMRT			<b>0.49</b>	<b>(0.27, 0.74)</b>
	Age			0.01	(-0.12, 0.15)
	BMI			0.13	(-0.01, 0.26)
	KPS Score			<b>0.74</b>	<b>(0.40, 1.07)</b>
	Adeno Histology			-0.13	(-0.43, 0.13)
	Cancer Stage			<b>-0.42</b>	<b>(-0.70, -0.13)</b>
<b>S-logNormal</b> (Survival only with logNormal Frailty)	IMRT			<b>0.58</b>	<b>(0.38, 0.79)</b>
	Age			0.03	(-0.10, 0.14)
	BMI			0.09	(-0.03, 0.23)
	KPS Score			<b>1.00</b>	<b>(0.75, 1.24)</b>
	Adeno Histology			-0.10	(-0.37, 0.11)
	Cancer Stage			-0.25	(-0.48, 0.01)

Table 5: Fits of the Esophageal Cancer Data under the Single Outcome Models. Point estimates of  $\beta_R$  and  $\beta_D$  are given with their 95% credible intervals in parentheses under R-DP, R-logNormal, S-DP and S-logNormal. Covariates having a statistically significant effect are given in boldface.

Model	DIC	LPML
Joint-DP	5654.21	-4113.72
Joint-logNormal	14206.75	-4639.29
R-DP	1501.91	-764.98
R-logNormal	1541.62	-779.88
S-DP	2167.01	-2036.54
S-logNormal	3116.00	-2523.84

Table 6: DIC and LPML under Bayesian models for the Esophageal Cancer Data.

respectively. While the estimates are not directly comparable, the difference between the joint model and the single outcome models is smaller under the models with DP than those with log normal. Similar to comparison of the Joint-DP to the Joint-logNormal, DIC and LPML indicate that the single outcome models with DP provide a better fit to the data than those with lognormal as shown in Table 6.

## 5 Discussion

We have presented a semiparametric Bayesian approach to jointly model recurrent events and a terminal event through shared latent frailties. An empirical Bayes approach is used to establish numerical prior hyperparameter values. All parameters in the proposed joint model are estimated simultaneously based on both the recurrent events and terminal events, through Bayesian hierarchical modeling. By assuming that the patient-specific random effects follow a DP prior, the proposed model propagates uncertainties at all levels and provides valid statistical inferences on covariate effects.

Our simulation studies and analysis of the esophageal cancer show that utilizing all sources of information in the data to estimate patient frailties yields good performance in posterior inferences for covariate and treatment effects on both the recurrent event process and survival. In the simulations, our proposed Joint-DP model shows robustness to a mild violation of the modeling assumption on the shared frailties. The Joint-DP model compared quite favorably to either of two BNP models that consider one outcome only, and to the likelihood-based method of Xu *et al.* (2017). Based on our simulations, the joint model and a similarly constructed survival (S-only) model appear to reliably estimate treatment effects in the absence of randomization, when treatment assignment is covariate-dependent. That is, the joint model and R-only model including latent patient-specific frailties assumed to follow a DP distribution both do a good job of bias correction, provided that the covariates used for treatment assignment are available.

The proposed joint model may be extended to accommodate more complex data structures, such as time-varying covariates or multivariate recurrent event processes. Also, different baseline intensity functions for the Poisson process of recurrent events may be assumed. This may be done, for example, by considering the set of time-varying intensity functions that have been applied successfully under various Bayesian frameworks (Kuo and Yang, 1996). These are potential areas for future research.

## Acknowledgment

Juhee Lee's research was supported by NSF grant DMS-1662427. Peter Thall's research was supported by NIH grant NIH R01 CA 83932.

## References

- Austin, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in Medicine* **32**, 16, 2837–2849.
- Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 4, 962–973.
- Brown, E. R. and Ibrahim, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* **59**, 2, 221–228.
- Bush, C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 2, 275–285.
- Chapple, A. G., Vannucci, M., Thall, P. F., and Lin, S. (2017). Bayesian variable selection for a semi-competing risks model with three hazard functions. *Computational statistics & data analysis* **112**, 170–185.
- Chuong, M. D., Hallemeier, C. L., Jabbour, S. K., Yu, J., Badiyan, S., Merrell, K. W., Mishra, M. V., Li, H., Verma, V., and Lin, S. H. (2016). Improving outcomes for esophageal cancer using proton beam therapy. *International Journal of Radiation Oncology\* Biology\* Physics* **95**, 1, 488–497.
- Cook, R. J. and Lawless, J. F. (2002). Analysis of repeated events. *Statistical Methods in Medical Research* **11**, 2, 141–166.
- Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)* **17**, 2, 129–164.
- De Gruttola, V. and Tu, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 4, 1003–1014.
- Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the american statistical association* **90**, 430, 577–588.
- Faucett, C. L. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* **15**, 15, 1663–1685.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 2, 209–230.
- Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G., eds. (2009). *Longitudinal Data Analysis*. Chapman & Hall/CRC Press Handbooks of Modern Statistical Methods.
- Geisser, S. (1993). *Predictive inference*. 55. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, New York, NY.



- Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**, 3, 501–514.
- Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. Tech. rep., STANFORD UNIV CA DEPT OF STATISTICS.
- Ghosh, D. and Lin, D. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 2, 554–562.
- Ghosh, S. K. and Ghosal, S. (2006). Semiparametric accelerated failure time models for censored data. *Bayesian Statistics and Its Applications* **15**, 213–229.
- Grandell, J. (1976). *Doubly Stochastic Poisson Processes: Lecture notes on Mathematics No. 529*. Springer-Verlag, New York.
- Guida, M., Calabria, R., and Pulcini, G. (1989). Bayes inference for a non-homogeneous Poisson process with power intensity law [reliability]. *IEEE Transactions on Reliability* **38**, 5, 603–609.
- Hatfield, L. A., Boye, M. E., and Carlin, B. P. (2011). Joint modeling of multiple longitudinal patient-reported outcomes and survival. *Journal of Biopharmaceutical Statistics* **21**, 5, 971–991.
- Hatfield, L. A., Hodges, J. S., and Carlin, B. P. (2014). Joint models: when are treatment estimates improved? *Statistics and Its Interface* **7**, 4, 439–453.
- He, L., Chapple, A., Liao, Z., Komaki, R., Thall, P. F., and Lin, S. H. (2016). Bayesian regression analyses of radiation modality effects on pericardial and pleural effusion and survival in esophageal cancer. *Radiotherapy and Oncology* **121**, 1, 70–74.
- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 4, 465–480.
- Huang, C.-Y. and Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99**, 468, 1153–1165.
- Jara, A., Hanson, T., Quintana, F., Mueller, P., and Rosner, G. (2018). Package dppackage.
- Ji, W., Zheng, W., Li, B., Cao, C., and Mao, W. (2016). Influence of body mass index on the long-term outcomes of patients with esophageal squamous cell carcinoma who underwent esophagectomy as a primary treatment: A 10-year medical experience. *Medicine* **95**, 29, e4204.

- Kalbfleisch, J. D., Schaubel, D. E., Ye, Y., and Gong, Q. (2013). An estimating function approach to the analysis of recurrent and terminal events. *Biometrics* **69**, 2, 366–374.
- Kuo, L. and Yang, T. Y. (1996). Bayesian computation for nonhomogeneous Poisson processes in software reliability. *Journal of the American Statistical Association* **91**, 434, 763–773.
- Lawless, J. F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association* **82**, 399, 808–815.
- Lee, K. H., Dominici, F., Schrag, D., and Haneuse, S. (2016). Hierarchical models for semicompeting risks data with application to quality of end-of-life care for pancreatic cancer. *Journal of the American Statistical Association* **111**, 515, 1075–1095.
- Lee, K. H., Haneuse, S., Schrag, D., and Dominici, F. (2015). Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64**, 2, 253–273.
- Li, Y., Müller, P., and Lin, X. (2011). Center-adjusted inference for a nonparametric Bayesian random effect distribution. *Statistica Sinica* **21**, 3, 1201–1223.
- Lin, S. H., Wang, L., Myles, B., Thall, P. F., Hofstetter, W. L., Swisher, S. G., Ajani, J. A., Cox, J. D., Komaki, R., and Liao, Z. (2012). Propensity score-based comparison of long-term outcomes with 3-dimensional conformal radiotherapy vs intensity-modulated radiotherapy for esophageal cancer. *International Journal of Radiation Oncology\* Biology\* Physics* **84**, 5, 1078–1085.
- Liu, L. and Huang, X. (2009). Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **58**, 1, 65–81.
- Liu, L., Wolfe, R. A., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 3, 747–756.
- MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 2, 223–238.
- Millar, R. B. (2009). Comparison of hierarchical bayesian models for overdispersed count data using dic and bayes’ factors. *Biometrics* **65**, 3, 962–969.
- Müller, P. and Rodriguez, A. (2013). *Nonparametric Bayesian Inference*, vol. 9 of *NSF-CBMS Regional Conference Series in Probability and Statistics*, chap. Chapter 3: Dirichlet Process, 23–41. Beachwood, Ohio, USA; and Alexandria, Virginia, USA: Institute of Mathematical Statistics and American Statistical Association.

- Ouyang, B., Sinha, D., Slate, E. H., and Van Bakel, A. B. (2013). Bayesian analysis of recurrent event with dependent termination: an application to a heart transplant study. *Statistics in Medicine* **32**, 15, 2629–2642.
- Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Texts in Statistics, Springer Science & Business Media, 2nd edn.
- Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 5, 550–560.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 2, 639–650.
- Sinha, D., Maiti, T., Ibrahim, J. G., and Ouyang, B. (2008). Current methods for recurrent events data with dependent termination: a Bayesian perspective. *Journal of the American Statistical Association* **103**, 482, 866–878.
- Song, X., Davidian, M., and Tsiatis, A. A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 4, 742–753.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 4, 583–639.
- Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA: a cancer journal for clinicians* **65**, 2, 87–108.
- Walker, S. and Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics* **55**, 2, 477–483.
- Wen, S., Huang, X., Frankowski, R. F., Cormier, J. N., and Pisters, P. (2016). A Bayesian multivariate joint frailty model for disease recurrences and survival. *Statistics in Medicine* **35**, 26, 4794–4812.
- Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 1, 330–339.
- Xu, G., Chiou, S. H., Huang, C.-Y., Wang, M.-C., and Yan, J. (2017). Joint scale-change models for recurrent events and failure time. *Journal of the American Statistical Association* **112**, 518, 794–805.
- Xu, Y., Müller, P., Wahed, A. S., and Thall, P. F. (2016). Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *Journal of the American Statistical Association* **111**, 515, 921–950.

- Ye, W., Lin, X., and Taylor, J. M. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach. *Biometrics* **64**, 4, 1238–1246.
- Zhang, S., Yang, H., Luo, K., Huang, Q., Chen, J., Yang, F., Cai, X., Xie, X., Liu, Q., Bella, A., *et al.* (2013). The impact of body mass index on complication and survival in resected oesophageal cancer: a clinical-based cohort and meta-analysis. *British Journal of Cancer* **109**, 11, 2894–2903.

# Supplementary Materials: Joint Bayesian Semiparametric Regression Analysis of Recurrent Adverse Events and Survival in Esophageal Cancer Patients

## 1 MCMC Algorithm

Let  $\boldsymbol{\theta} = (\xi, \boldsymbol{\gamma}, \boldsymbol{\beta}_R, \mu, \zeta, \boldsymbol{\beta}_D, \sigma)$  denote the vector of all unknown parameters. The joint probability model of data and parameters under the Joint-DP model is

$$p(\mathbf{N}, \tilde{\mathbf{D}}, \boldsymbol{\delta}, \boldsymbol{\theta}) = p(\mathbf{N}, \tilde{\mathbf{D}}, \boldsymbol{\delta} \mid \boldsymbol{\theta})p(\xi)p(\boldsymbol{\beta}_R)p(\boldsymbol{\beta}_D)p(\boldsymbol{\gamma})p(\mu)p(\zeta)p(\sigma).$$

In (5) of the main text, we defined a probability distribution  $\eta_i \stackrel{iid}{\sim} G$  for  $\eta_i = \log(\gamma_i)$ , and let  $G \sim \text{DP}(\alpha G_0)$  where  $G_0 = \text{N}(\bar{\eta}, v^2)$ . We use the Gibbs sampling scheme in MacEachern and Müller (1998) to simulate  $\eta = \log(\gamma)$ . The random measure  $G$  in the assumed model is a.s. discrete and yields a positive probability of ties among the  $\eta_i$ 's. We let  $\eta_k^*$ ,  $k = 1, \dots, K$ , denote the  $K \leq n$  unique values among the  $\eta_i$ 's. We also introduced a  $n$ -dimensional cluster membership vector  $\mathbf{s} = (s_1, \dots, s_n)$  by letting  $s_i = k$  if and only if  $\eta_i = \eta_k^*$ . Given partition  $\mathbf{s}$ ,  $c_k$  denotes the size of cluster  $k$ , and we denote  $\mathbf{c} = (c_1, \dots, c_K)$  for the sizes of the  $K$  clusters. The distributions of  $\mathbf{s}$  and  $\boldsymbol{\eta}^* = (\eta_1^*, \dots, \eta_K^*)$  implied by the Dirichlet process (DP) are

$$p(\mathbf{s} \mid \alpha) = \alpha^K \frac{\prod_{k=1}^K \Gamma(c_k)}{\prod_{i=1}^n (\alpha + i - 1)} \quad (1)$$

and given  $\mathbf{s}$ ,  $\eta_k^* \mid \mathbf{s} \stackrel{iid}{\sim} G_0 = \text{N}(\bar{\eta}, v^2)$ ,  $k = 1, \dots, K$ . The Pólya urn can be defined as the distribution of  $\mathbf{s}$  in (1) (Blackwell and MacQueen, 1973). Assuming the exchangeability of subjects, the complete conditional prior probability of  $s_i$  can be described using the Pólya urn scheme as follows: let  $\mathbf{s}^-$ ,  $\mathbf{c}^-$  and  $K^-$  denote  $\mathbf{s}$ ,  $\mathbf{c}$  and  $K$  without subject  $i$ . The prior probability that subject  $i$  joins in existing cluster  $k$ ,  $k = 1, \dots, K^-$  is  $p(s_i = k \mid \mathbf{s}^-) = c_k^- / (\alpha + n - 1)$  and the conditional prior probability of subject  $i$  being in a new singleton cluster,  $s_i = K^- + 1$  is  $p(s_i = K^- + 1 \mid \mathbf{s}^-) = \alpha / (\alpha + n - 1)$ .

From the relationship between  $\boldsymbol{\eta}$ ,  $\boldsymbol{\eta}^*$ , and  $\mathbf{s}$ , sampling  $\eta_i$  is equivalent to sampling  $\boldsymbol{\eta}^*$  and  $\mathbf{s}$ . We denote  $\tilde{\boldsymbol{\theta}} = (\xi, \boldsymbol{\eta}^*, \mathbf{s}, \boldsymbol{\beta}_R, \mu, \zeta, \boldsymbol{\beta}_D, \sigma)$ . The parameters are estimated by iteratively

drawing samples from the full conditional posterior distributions given the data and the other parameters.

1. Update  $\eta_k^*$ .

Recall  $\gamma_k^* = \exp(\eta_k^*)$ . Let  $\boldsymbol{\theta}_{-\gamma} = (\xi, \boldsymbol{\beta}_R, \mu, \zeta, \boldsymbol{\beta}_D, \sigma)$  be all unknown parameters except  $\gamma$ . For  $k = 1, \dots, K$ ,

$$\begin{aligned} p(\eta_k^* | \mathbf{N}, \tilde{\mathbf{D}}, \boldsymbol{\delta}, \mathbf{s}, \boldsymbol{\theta}_{-\gamma}) &\propto p(\eta_k^*) \prod_{i=1|s_i=k}^n p(N_i, \tilde{D}_i, \delta_i | \gamma_k^*, \xi, \boldsymbol{\beta}_R, \mu, \zeta, \boldsymbol{\beta}_D, \sigma) \\ &\propto \exp\left\{-\frac{(\eta_k^* - \bar{\eta})^2}{2v^2}\right\} \prod_{i=1|s_i=k}^n (\gamma_k^*)^{N_i} \exp\left(-\gamma_k^* \tilde{D}_i^\xi e^{\boldsymbol{\beta}'_R \mathbf{Z}_{Ri}}\right) \\ &\quad \left\{\frac{\lambda_i}{\sigma} \tilde{D}_i^{1/\sigma-1} \exp\left(-\lambda_i \tilde{D}_i^{1/\sigma}\right)\right\}^{1-\delta_i} \\ &\quad \left\{\exp\left(-\lambda_i \tilde{D}_i^{1/\sigma}\right)\right\}^{\delta_i}, \end{aligned}$$

where  $\lambda_i = \exp\left(-\frac{\mu + \zeta \eta_k^* + \boldsymbol{\beta}'_D \mathbf{Z}_{Di}}{\sigma}\right)$ . We use a random-walk Metropolis-Hastings algorithm to simulate  $\eta_k^*$  from the full conditional. For the esophageal cancer data analysis, we generated a proposal  $\eta_k^{*'} = \eta_k^* + \epsilon$ ,  $\epsilon \sim N(0, 0.2^2)$ .

2. Update  $\mathbf{s}$ .

We sequentially updated a cluster label for subject  $i$ ,  $i = 1, \dots, n$  as follows: For  $k = 1, \dots, K^-$ ,

$$\begin{aligned} p(s_i = k | N_i, \tilde{D}_i, \delta_i, \boldsymbol{\theta}_{-\gamma}, \boldsymbol{\eta}^{*-}, \mathbf{s}^-) &\propto c_k^- \cdot p(N_i, \tilde{D}_i, \delta_i | \gamma_k^*, \boldsymbol{\theta}_{-\gamma}) \\ &\propto c_k^- (\gamma_k^*)^{N_i} \exp\left(-\gamma_k^* \tilde{D}_i^\xi e^{\boldsymbol{\beta}'_R \mathbf{Z}_{Ri}}\right) \\ &\quad \left\{\frac{\lambda_i}{\sigma} \tilde{D}_i^{1/\sigma-1} \exp\left(-\lambda_i \tilde{D}_i^{1/\sigma}\right)\right\}^{1-\delta_i} \\ &\quad \left\{\exp\left(-\lambda_i \tilde{D}_i^{1/\sigma}\right)\right\}^{\delta_i}. \end{aligned}$$

$$\begin{aligned} p(s_i = K^- + 1 | N_i, \tilde{D}_i, \delta_i, \boldsymbol{\theta}_{-\gamma}, \boldsymbol{\eta}^{*-}, \mathbf{s}^-) &\propto \alpha \cdot p(N_i, \tilde{D}_i, \delta_i | \gamma^*, \boldsymbol{\theta}) \\ &\propto \alpha \int (\gamma^*)^{N_i} \exp\left(-\gamma^* D_i e^{\boldsymbol{\beta}'_R \mathbf{Z}_{Ri}}\right) \\ &\quad \left\{\frac{\lambda_i}{\sigma} \tilde{D}_i^{1/\sigma-1} \exp\left(-\lambda_i \tilde{D}_i^{1/\sigma}\right)\right\}^{1-\delta_i} \\ &\quad \left\{\exp\left(-\lambda_i \tilde{D}_i^{1/\sigma}\right)\right\}^{\delta_i} g_0(\eta^*) d\eta^*, \end{aligned}$$

where  $g_0$  is the probability density function of  $G_0$ . The integral can be evaluated numerically (MacEachern and Müller, 1998). We sample  $k$  from  $\{1, \dots, K^-, K^- + 1\}$  with probabilities proportional to  $p(s_i | N_i, \tilde{D}_i, \delta_i, \boldsymbol{\theta}_{-\gamma}, \boldsymbol{\eta}^{*-})$ .

3. Update  $\xi$ .

$$\begin{aligned} p(\xi | \mathbf{N}, \tilde{\mathbf{D}}, \boldsymbol{\theta}) &\propto p(\xi) \prod_{i=1}^n p(N_i(\tilde{D}_i) | \boldsymbol{\theta}) \\ &\propto \xi^{a_\xi - 1} \exp(-b_\xi \xi) \prod_{i=1}^n \left\{ \prod_{j=1}^{N_i(\tilde{D}_i)} \xi t_{ij}^{\xi - 1} \right\} \exp\left(-\gamma_i \tilde{D}_i^\xi e^{\boldsymbol{\beta}'_R \mathbf{Z}_{Ri}}\right). \end{aligned}$$

We used a random-walk Metropolis-Hastings algorithm to simulate  $\xi$  from the full conditional. For the esophageal cancer data analysis, we generated a proposal  $\xi' = \xi + \epsilon$ ,  $\epsilon \sim \text{N}(0, 0.2^2)$ .

4. Update  $\boldsymbol{\beta}_R$ .

$$\begin{aligned} p(\boldsymbol{\beta}_R | \mathbf{N}, \tilde{\mathbf{D}}, \boldsymbol{\theta}) &\propto p(\boldsymbol{\beta}_R) \prod_{i=1}^n p(N_i(\tilde{D}_i) | \boldsymbol{\theta}) \\ &\propto \exp\left\{-\frac{(\boldsymbol{\beta}_R - \bar{\boldsymbol{\beta}}_R)' \Sigma_R^{-1} (\boldsymbol{\beta}_R - \bar{\boldsymbol{\beta}}_R)}{2}\right\} \prod_{i=1}^n \exp(N_i \boldsymbol{\beta}'_R \mathbf{Z}_{Ri}) \exp\left(-\gamma_i \tilde{D}_i^\xi e^{\boldsymbol{\beta}'_R \mathbf{Z}_{Ri}}\right). \end{aligned}$$

We used a random-walk Metropolis-Hastings algorithm to simulate  $\boldsymbol{\beta}_R$  from the full conditional. For better acceptance, we updated an element of  $\boldsymbol{\beta}_R$  at a time given the others in  $\boldsymbol{\beta}_R$ . Alternatively, the entire vector  $\boldsymbol{\beta}_D$  can be updated with a Metropolis-Hastings jump in a block. For the esophageal cancer data analysis, we generated a proposal  $\boldsymbol{\beta}'_{Rp} = \boldsymbol{\beta}_{Rp} + \epsilon$ ,  $\epsilon \sim \text{N}(0, 0.5^2)$ .

5. Update  $\boldsymbol{\beta}_D$ .

$$\begin{aligned} p(\boldsymbol{\beta}_D | \tilde{\mathbf{D}}, \boldsymbol{\delta}, \boldsymbol{\theta}) &\propto p(\boldsymbol{\beta}_D) \prod_{i=1}^n p(\tilde{D}_i, \delta_i | \boldsymbol{\theta}) \\ &\propto \exp\left\{-\frac{(\boldsymbol{\beta}_D - \bar{\boldsymbol{\beta}}_D)' \Sigma_D^{-1} (\boldsymbol{\beta}_D - \bar{\boldsymbol{\beta}}_D)}{2}\right\} \\ &\quad \times \prod_{i=1}^n \left\{ \frac{\lambda_i}{\sigma} \tilde{D}_i^{1/\sigma - 1} \exp\left(-\lambda_i \tilde{D}_i^{1/\sigma}\right) \right\}^{1 - \delta_i} \left\{ \exp\left(-\lambda_i \tilde{D}_i^{1/\sigma}\right) \right\}^{\delta_i}. \end{aligned}$$

We used a random-walk Metropolis-Hastings algorithm to simulate  $\boldsymbol{\beta}_D$  from the full conditional. We may update one element of  $\boldsymbol{\beta}_D$  at a time given the others in  $\boldsymbol{\beta}_D$ . Alternatively, the entire vector  $\boldsymbol{\beta}_D$  can be updated with a Metropolis-Hastings jump in a block. For the esophageal cancer data analysis, we generated a proposal  $\boldsymbol{\beta}'_{Dp} = \boldsymbol{\beta}_{Dp} + \epsilon$ ,  $\epsilon \sim \text{N}(0, 0.5^2)$ .

6. Update  $\mu$ .

$$\begin{aligned} p(\mu \mid \tilde{\mathbf{D}}, \boldsymbol{\delta}, \boldsymbol{\theta}) &\propto p(\mu) \prod_{i=1}^n p(N_i, \tilde{D}_i, \delta_i \mid \boldsymbol{\theta}) \\ &\propto \exp\left\{-\frac{(\mu - \bar{\mu})^2}{2\nu^2}\right\} \prod_{i=1}^n \left\{\frac{\lambda_i \tilde{D}_i^{1/\sigma-1} \exp(-\lambda_i \tilde{D}_i^{1/\sigma})}{\sigma}\right\}^{1-\delta_i} \left\{\exp(-\lambda_i \tilde{D}_i^{1/\sigma})\right\}^{\delta_i}. \end{aligned}$$

We used a random-walk Metropolis-Hastings algorithm to simulate  $\mu$  from the full conditional. For the esophageal cancer data analysis, we generated a proposal  $\mu' = \mu + \epsilon$ ,  $\epsilon \sim N(0, 0.8^2)$ .

7. Update  $\zeta$ .

$$\begin{aligned} p(\zeta \mid \tilde{\mathbf{D}}, \boldsymbol{\delta}, \boldsymbol{\theta}) &\propto p(\zeta) \prod_{i=1}^n p(\tilde{D}_i, \delta_i \mid \boldsymbol{\theta}) \\ &\propto \exp\left\{-\frac{(\zeta - \bar{\zeta})^2}{2\omega^2}\right\} \\ &\quad \prod_{i=1}^n \left\{\frac{\lambda_i \tilde{D}_i^{1/\sigma-1} \exp(-\lambda_i \tilde{D}_i^{1/\sigma})}{\sigma}\right\}^{1-\delta_i} \times \left\{\exp(-\lambda_i \tilde{D}_i^{1/\sigma})\right\}^{\delta_i}. \end{aligned}$$

We used a random-walk Metropolis-Hastings algorithm to simulate  $\zeta$  from the full conditional. For the esophageal cancer data analysis, we generated a proposal  $\zeta' = \zeta + \epsilon$ ,  $\epsilon \sim N(0, 0.5^2)$ .

8. Update  $\sigma$ .

$$\begin{aligned} p(\sigma \mid \tilde{\mathbf{D}}, \boldsymbol{\delta}, \boldsymbol{\theta}) &\propto p(\sigma) \prod_{i=1}^n p(N_i, \tilde{D}_i, \delta_i \mid \boldsymbol{\theta}) \\ &\propto (\sigma)^{a_\sigma-1} \exp\left(-\frac{b_\sigma}{\sigma}\right) \\ &\quad \prod_{i=1}^n \left\{\frac{\lambda_i \tilde{D}_i^{1/\sigma-1} \exp(-\lambda_i \tilde{D}_i^{1/\sigma})}{\sigma}\right\}^{1-\delta_i} \left\{\exp(-\lambda_i \tilde{D}_i^{1/\sigma})\right\}^{\delta_i}. \end{aligned}$$

We used a random-walk Metropolis-Hastings algorithm to simulate  $\sigma$  from the full conditional. For the esophageal cancer data analysis, we generated a proposal  $\sigma' = \sigma + \epsilon$ ,  $\epsilon \sim N(0, 0.3^2)$ .

## 2 Additional Simulations

### 2.1 Frailty Distribution: Simulations 2 and 3

We conducted simulations for further examination of the proposed joint model with the DP for the frailty distribution, the Joint-DP model. We compared the Joint-DP model to the



		Recurrent Events			Survival		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP</b> (Joint with DP)	Mean	0.299	0.251	-0.307	-0.298	-0.252	0.305
	Bias	<b>-0.001</b>	<b>0.001</b>	-0.007	<b>0.002</b>	<b>-0.002</b>	<b>0.005</b>
	Ave. CI length	<b>0.354</b>	<b>0.176</b>	<b>0.391</b>	<b>0.318</b>	<b>0.156</b>	0.355
	Coverage	94.3%	94.4%	<b>95.2%</b>	94.6%	<b>94.9%</b>	<b>95.0%</b>
<b>Joint-logNormal</b> (Joint with logNormal Frailty)	Mean	0.172	0.264	-0.655	-0.276	-0.253	0.351
	Bias	-0.128	0.014	-0.355	0.024	-0.003	0.051
	Ave. CI length	0.497	0.259	0.404	0.329	0.157	0.369
	Coverage	89.3%	98.7%	3.1%	94.5%	94.7%	91.6%
<b>JSCM</b>	Mean	0.365	0.310	-0.382	-0.237	-0.200	0.241
	Bias	0.065	0.060	-0.082	-0.063	-0.050	0.059
	Ave. CI length	0.645	0.316	0.718	1.616	0.785	1.795
	Coverage	89.7%	86.1%	91.4%	99.8%	99.3%	99.5%
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.293	0.246	-0.302			
	Bias	-0.007	-0.004	<b>-0.002</b>			
	Ave. CI length	0.356	<b>0.176</b>	0.397			
	Coverage	<b>94.9%</b>	<b>94.9%</b>	<b>94.7%</b>			
<b>R-logNormal</b> (Recurrent events only with logNormal Frailty)	Mean	0.130	0.251	-0.714			
	Bias	-0.170	<b>0.001</b>	-0.414			
	Ave. CI length	0.493	0.262	0.402			
	Coverage	79.4%	98.8%	0.1%			
<b>S-DP</b> (Survival only with DP)	Mean				-0.294	-0.252	0.312
	Bias				0.006	<b>-0.002</b>	0.012
	Ave. CI length				0.318	<b>0.156</b>	<b>0.354</b>
	Coverage				<b>94.9%</b>	94.0%	94.8%
<b>S-logNormal</b> (Survival only with logNormal Frailty)	Mean				-0.143	-0.268	0.678
	Bias				0.157	-0.018	0.378
	Ave. CI length				0.316	0.166	0.291
	Coverage				49.8%	82.6%	0.0%

Table 1: [Simulation 2] Simulation results for covariate effect estimation. The values are based on 1000 simulated datasets, each fit using each model. The lognormal distribution is used to simulate the true frailties. The proposed joint model assuming a DP prior for the frailty distribution is labeled Joint-DP. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

		Baseline Intensity( $\xi$ )	Baseline Hazards ( $\sigma$ )
Simulation Truth		0.800	0.700
<b>Joint-DP</b> (Joint with DP)	Mean	0.808	0.692
	Bias	<b>0.008</b>	<b>-0.008</b>
	Ave. CI length	0.123	0.174
	Coverage	<b>94.3%</b>	<b>95.1%</b>
<b>Joint-logNormal</b> (Joint with logNormal Frailty)	Mean	0.855	0.712
	Bias	0.055	0.012
	Ave. CI length	0.150	<b>0.136</b>
	Coverage	69.0%	93.5%
<b>R- DP</b> (Recurrent events only with DP)	Mean	0.788	
	Bias	-0.012	
	Ave. CI length	<b>0.111</b>	
	Coverage	93.8%	
<b>R-logNormal</b> (Recurrent events only with logNormal Frailty)	Mean	0.809	
	Bias	0.009	
	Ave. CI length	0.121	
	Coverage	96.6%	
<b>S-DP</b> (Survival only with DP)	Mean		0.674
	Bias		-0.026
	Ave. CI length		0.224
	Coverage		97.5%
<b>S-logNormal</b> (Survival only with logNormal Frailty)	Mean		0.185
	Bias		-0.515
	Ave. CI length		0.177
	Coverage		0.0%

Table 2: [Simulation 2] Simulation results for parameters in baseline functions. Results are based on 1000 simulated datasets, each fit with each model. The lognormal distribution is used to simulate the true frailties. The parameters  $\xi$  and  $\sigma$  control time-varying patterns in the recurrent event baseline intensity and survival hazard functions, respectively. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length) and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

six different models, including a frequentist method Joint Scale-Change Model (JSCM), that are described in §2.3 of the main text. We kept most of the simulation set-up in Simulation 1, except the distributions used to simulate the frailties. For Simulation 2, we generated  $\eta_i^{\text{TR}} = \log(\gamma_i^{\text{TR}})$  from a single lognormal distribution,  $\eta_i^{\text{TR}} \stackrel{iid}{\sim} N(0, 0.25)$ . For Simulation 3, we

used a mixture of gamma distributions as follows: We simulated patient frailties  $\gamma_i^{\text{TR}}$  from a mixture distribution with  $K = 5$  components,  $p(\gamma_i^{\text{TR}}) = \sum_{k=1}^K \pi_k \text{Gamma}(\gamma_i^{\text{TR}} | 14.3, b_k^{\text{TR}})$ , where  $\text{Gamma}(\cdot | a, b)$  is the gamma distribution with mean  $a/b$  and variance  $a/b^2$ . We fixed the mixture component means  $(b_1^{\text{TR}}, \dots, b_5^{\text{TR}}) = (64.0, 21.3, 14.3, 9.6, 7.8)$  and simulated the component weights from  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_5) \sim \text{Dir}(3, 3, 3, 3, 3)$ , where  $\text{Dir}(\mathbf{a})$  denotes the Dirichlet distribution with parameter  $\mathbf{a}$ . We specified the hyperparameters and initialized the parameters for the Bayesian models as in Simulation 1.

The results from Simulation 2 are summarized in Tables 1 and 2. The Joint-DP model performs very well for estimation of  $\boldsymbol{\beta}$ ,  $\xi$  and  $\sigma$ . The bias due to the treatment assignment using  $Z_3$  is well corrected and the model produces very tiny biases in the estimation of all  $\boldsymbol{\beta}$ . Surprisingly, the Joint-logNormal model does not produce good inference for  $\boldsymbol{\beta}$ , despite the fact that the assumed model for the frailties is the same as that used for the simulation. In particular, estimates of  $\beta_1$  and  $\beta_3$  under the Joint-logNormal model have larger biases and wider intervals with smaller coverage probabilities than those under the Joint-DP model. For example, Bias, Ave. CI length and Coverage for  $\beta_{R3}$  are (-0.004, 0.383, 93.6%) under the Joint-DP model versus (-0.355, 0.404, 3.1%) under the Joint-logNormal model. Similar to the results in Simulation 1, the JSCM produces slightly biased estimates with wider confidence intervals, especially for  $\boldsymbol{\beta}_D$ . This possibly is due to the two-step estimation method for  $\boldsymbol{\beta}$  and  $\gamma$ , or the assumption that frailties are included in the two hazard functions, of the recurrent event process and the survival time distribution, without any adjustment to allow their effects to have different magnitudes. Comparisons of the Joint-DP model to the R-DP and S-DP models show that the performance of each one outcome model with DP is very close to that of the Joint-DP model. On the other hand, the improvement by joint modeling remains critical for the models with lognormal frailty distribution, although the frailty model is correctly specified. In particular, the estimates of  $\beta_1$  and  $\beta_3$  are significantly improved under the Joint-logNormal model compared to the R-logNormal and S-logNormal models.

Tables 3 and 4 summarize the results from Simulation 3. The Joint-DP model yields on average unbiased estimates of  $\boldsymbol{\beta}_R$  and  $\boldsymbol{\beta}_D$ . Comparing to the results from Simulation 1, where the true frailties are simulated from a mixture of lognormal distributions, the performance of the Joint-DP model remains approximately the same, implying that the model captures the pattern in the true frailty distribution reasonably well. Similarly, the performance of the single outcome models with DP is about the same as in Simulation 1. Estimates of  $\boldsymbol{\beta}_R$  and  $\boldsymbol{\beta}_D$  under the single outcome models tend to be slightly more biased with wider interval estimates than those under the Joint-DP model. The Joint-DP model produces great improvements in estimation of  $\xi$  and  $\sigma$  compared to the R-DP and S-DP models. The models with lognormal frailty distributions, the Joint-logNormal, R-logNormal and S-logNormal models do not perform well, especially for estimation of coefficients related to treatment and prognostic factor,  $\beta_{R1}$ ,  $\beta_{R3}$ ,  $\beta_{D1}$  and  $\beta_{D3}$  and produce large biases. The JSCM also produces large biases, wider interval estimates, and coverage probabilities greater than 95%, even for  $\boldsymbol{\beta}_D$  although the assumption under the JSCM for survival times is satisfied in the simulation truth.

		Recurrent Events			Survival		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP</b> (Joint with DP)	Mean	0.293	0.248	-0.313	-0.300	-0.252	0.307
	Bias	<b>-0.007</b>	<b>-0.002</b>	<b>-0.013</b>	<b>0.000</b>	<b>-0.002</b>	<b>0.007</b>
	Ave. CI length	<b>0.422</b>	<b>0.209</b>	0.467	<b>0.440</b>	<b>0.217</b>	0.490
	Coverage	<b>94.5%</b>	93.6%	94.0%	<b>95.2%</b>	94.8%	<b>94.5%</b>
<b>Joint-logNormal</b> (Joint with logNormal Frailty)	Mean	0.162	0.260	-0.657	-0.198	-0.255	0.537
	Bias	-0.138	0.010	-0.357	0.102	-0.005	0.237
	Ave. CI length	0.472	0.245	<b>0.394</b>	0.506	0.247	0.512
	Coverage	75.4%	<b>95.7%</b>	23.2%	87.6%	95.8%	54.7%
<b>JSCM</b>	Mean	0.369	0.306	-0.366	0.237	0.197	-0.226
	Bias	0.069	0.056	-0.066	-0.063	-0.053	0.074
	Ave. CI length	0.782	0.383	0.871	1.517	0.744	1.695
	Coverage	92.9%	90.4%	93.5%	99.6%	99.3%	99.4%
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.262	0.222	-0.281			
	Bias	-0.038	-0.028	0.019			
	Ave. CI length	0.447	0.221	0.496			
	Coverage	93.6%	91.7%	<b>94.7%</b>			
<b>R-logNormal</b> (Recurrent events only with logNormal Frailty)	Mean	0.002	0.221	-0.892			
	Bias	-0.298	-0.029	-0.592			
	Ave. CI length	0.490	0.259	0.407			
	Coverage	35.1%	94.0%	1.7%			
<b>S-DP</b> (Survival only with DP)	Mean				-0.286	-0.252	0.338
	Bias				0.014	<b>-0.002</b>	0.038
	Ave. CI length				0.462	0.227	0.512
	Coverage				94.0%	<b>94.6%</b>	93.8%
<b>S-logNormal</b> (Survival only with logNormal Frailty)	Mean				0.140	-0.252	1.329
	Bias				0.440	<b>-0.002</b>	1.029
	Ave. CI length				0.458	0.236	<b>0.429</b>
	Coverage				13.7%	90.7%	0.4%

Table 3: [Simulation 3] Simulation results for covariate effect estimation. The values are based on 1000 simulated datasets, each fit using each model. A mixture of gamma distributions is used to simulate the true frailties. The proposed joint model assuming a DP prior for the frailty distribution is labeled Joint-DP. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

		Baseline Intensity( $\xi$ )	Baseline Hazards ( $\sigma$ )
Simulation Truth		0.800	0.700
<b>Joint-DP</b> (Joint with DP)	Mean	0.795	0.703
	Bias	<b>-0.005</b>	<b>0.003</b>
	Ave. CI length	0.122	<b>0.205</b>
	Coverage	<b>95.3%</b>	96.5%
<b>Joint-logNormal</b> (Joint with logNormal Frailty)	Mean	0.846	0.611
	Bias	0.046	-0.089
	Ave. CI length	0.116	0.271
	Coverage	59.6%	74.4%
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.684	
	Bias	-0.116	
	Ave. CI length	0.104	
	Coverage	2.2%	
<b>R-logNormal</b> (Recurrent events only with logNormal Frailty)	Mean	0.691	
	Bias	-0.109	
	Ave. CI length	<b>0.098</b>	
	Coverage	9.7%	
<b>S-DP</b> (Survival only with DP)	Mean		0.743
	Bias		0.043
	Ave. CI length		0.269
	Coverage		<b>94.3%</b>
<b>S-logNormal</b> (Survival only with logNormal Frailty)	Mean		0.630
	Bias		-0.070
	Ave. CI length		0.294
	Coverage		42.2%

Table 4: [Simulation 3] Simulation results for parameters in baseline functions. Results are based on 1000 simulated datasets, each fit with each model. A mixture of gamma distributions is used to simulate the true frailties. The parameters  $\xi$  and  $\sigma$  control time-varying patterns in the recurrent event baseline intensity and survival hazard functions, respectively. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length) and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

## 2.2 Point Processes for Recurrent Events: Simulations 4 & 5

We simulated recurrent events from point processes different from the processes used in Simulations 1-3, and compared the performance of the proposed Joint-DP to that of the JSCM. In particular, recurrent events were simulated under a homogeneous Poisson process

(that is,  $\xi^{\text{TR}} = 1$  in (2) of the main text) and under a Poisson cluster process for Simulations 4 and 5, respectively. The Joint-DP model and JSCM include homogeneous Poisson processes as a special case and for Simulation 4, both models are expected to produce on average unbiased estimates of  $\beta$ . Poisson cluster processes may be generated by first simulating a Poisson or non-homogeneous Poisson process of parent points, also called immigrants, and then for each parent point generating a random number of offspring points, also called children, that are distributed in some random fashion around the parent point. This may be done in one-dimensional real time or more generally in a suitable space of any finite dimension. However, in general it is not known whether an observed point is an immigrant or offspring. A key point is that, in real time, the time gaps between successive observed points (event times) of a Poisson cluster process are not independent. The remaining simulation set-up is kept the same as in Simulation 1 for both simulations. Note that survival times were simulated from a Weibull distribution, which is the model assumed for the Joint-DP model, and is also a special case under the JSCM.

For Simulation 4, the results of estimation of  $\beta$  are summarized in Table 5(a). Biases are close to zero under both models, implying that on average both models recover the true values of  $\beta$ . However, the average lengths of interval estimates under the JSCM are greater for both  $\beta_R$  and  $\beta_D$  than those under the Joint-DP model. Most notably, interval estimates of  $\beta_D$  under the JSCM are approximately four times wider and their coverage probabilities are greater than 95%. This possibly is due to the assumption under the JSCM that frailties have the same effect on the baseline intensity and survival hazard functions. Table 5(b) illustrates estimation of  $\xi$  and  $\sigma$ , the time-varying parameters for the intensity and hazard functions, respectively, for the Joint-DP model. Both  $\xi$  and  $\sigma$  are estimated reasonably, while the coverage probability for  $\xi$  is less than 95%.

For Simulation 5, we generated recurrent event times  $T_{i,j}$  using a nonhomogeneous Poisson cluster process. Specifically, we used a branch structure representation of a Poisson cluster process and simulated  $T_{i,j}$  as follows:

1. Given observed survival time  $\tilde{D}_i$ , we generated immigrants (parent points)  $N'_i$  and  $T'_{i,j'}$ ,  $j' = 1, \dots, N'_i$  from the Poisson process with intensity  $\lambda_i^{\text{TR}}(t) = \gamma_i^{\text{TR}} \xi^{\text{TR}} t^{\xi^{\text{TR}} - 1} \exp(\beta_R^{\text{TR},j'} \mathbf{Z}_i)$ . Frailties  $\gamma_i^{\text{TR}}$  and covariates  $\mathbf{Z}_i$  were simulated the same as in Simulation 1, with  $\xi^{\text{TR}}$  and  $\beta_R^{\text{TR}}$  fixed at the same values.
2. For subjects with  $N'_i > 0$ , we generated offspring (children points) for immigrants,  $N''_{i,j'}$ ,  $j' = 1, \dots, N'_i$  from a uniform distribution over the integers  $\{1, 2, 3\}$  and let  $T''_{i,j',m} = T'_{i,j'} + e_{i,j',m}$ ,  $m = 1, \dots, N''_{i,j'}$  with  $e_{i,j',m} \stackrel{iid}{\sim} N(0, 0.2^2)$ .
3. We then let the number of recurrent event occurrences  $N(\tilde{D}_i)_i = \sum_{j'=1}^{N'_i} N''_{i,j'}$  and their occurrence times  $T_{i,j} = \{T''_{i,j',m}, j' = 1, \dots, N'_i, m = 1, \dots, N''_{i,j'}\}$ .

Table 6(a) summarizes the results for estimation of  $\beta$ . From the table, the JSCM produces slightly larger biases than the Joint-DP model. Similar to Simulation 4, interval estimation and coverage probabilities are greatly inferior under the JSCM. Interval estimates are wider

		Recurrent Events			Survival		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP</b> (Joint with DP)	Mean	0.283	0.245	-0.312	-0.291	-0.252	0.309
	Bias	-0.017	<b>-0.005</b>	-0.012	0.009	-0.002	0.009
	Ave. CI length	<b>0.396</b>	<b>0.197</b>	<b>0.439</b>	<b>0.430</b>	<b>0.211</b>	<b>0.478</b>
	Coverage	93.1%	95.7%	<b>94.5%</b>	<b>94.2%</b>	<b>94.0%</b>	<b>94.8%</b>
<b>JSCM</b>	Mean	0.298	0.241	-0.290	0.300	0.249	-0.300
	Bias	<b>-0.002</b>	-0.009	<b>0.010</b>	<b>0.000</b>	<b>0.001</b>	<b>0.000</b>
	Ave. CI length	0.699	0.344	0.793	1.752	0.856	1.958
	Coverage	<b>94.2%</b>	<b>95.6%</b>	88.4%	99%	99.4%	97.2%

(a) Estimation of  $\beta$ 

		Baseline Intensity( $\xi$ )	Baseline Hazards ( $\sigma$ )
Simulation Truth		1.000	0.700
<b>Joint-DP</b> (Joint with DP)	Mean	0.978	0.702
	Bias	-0.022	0.002
	Ave. CI length	0.134	0.198
	Coverage	89.5%	95.5%

(b) Estimation of  $\xi$  and  $\sigma$  under Joint-DP

Table 5: [Simulation 4] Simulation results for covariate effect estimation in (a) and parameters  $\xi$  and  $\sigma$  in the baseline function in (b). A homogeneous Poisson process is used to simulate recurrent events. The parameters  $\beta_R$  and  $\beta_D$  quantifies the effects of covariates on the recurrent event baseline intensity and survival hazard functions, respectively. The parameters  $\xi$  and  $\sigma$  control time-varying patterns in the recurrent event baseline intensity and survival hazard functions, respectively. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

and coverage probabilities are less than 95% for  $\beta_{R2}$  and greater than 95% for  $\beta_D$ . Estimation of  $\xi$  and  $\sigma$  under the Joint-DP model is illustrated in Table 6(b). Note that  $\xi^{\text{TR}}$  is used to generate immigrants only. The parameter  $\xi$  under the Joint-DP model has a different interpretation from  $\xi^{\text{TR}}$  and is not directly comparable. Inference on  $\sigma$  under the Joint-DP model is greatly affected. Bias is larger than that in the previous simulations including Simulation 4 and the coverage probability is much smaller than 95%.

		Recurrent Events			Survival		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP</b> (Joint with DP)	Mean	0.302	0.254	-0.298	-0.289	-0.245	0.291
	Bias	<b>0.002</b>	<b>0.004</b>	<b>0.002</b>	<b>0.011</b>	<b>0.005</b>	<b>-0.009</b>
	Ave. CI length	<b>0.455</b>	<b>0.226</b>	<b>0.503</b>	<b>0.464</b>	<b>0.228</b>	<b>0.517</b>
	Coverage	93.6%	<b>92.9%</b>	92.4%	<b>93.9%</b>	<b>93.1%</b>	<b>94.2%</b>
<b>JSCM</b>	Mean	0.324	0.281	-0.343	-0.217	-0.187	0.231
	Bias	0.024	0.031	-0.043	0.083	0.063	-0.069
	Ave. CI length	0.800	0.392	0.893	1.567	0.761	1.746
	Coverage	<b>94.3%</b>	91.7%	<b>94.2%</b>	99.7%	99.1%	99.8%

(a) Estimation of  $\beta$ 

		Baseline Intensity( $\xi$ )	Baseline Hazards ( $\sigma$ )
Simulation Truth		0.800	0.700
<b>Joint-DP</b> (Joint with DP)	Mean	0.874	0.896
	Bias	0.074	0.196
	Ave. CI length	0.098	0.183
	Coverage	24.6%	1.5%

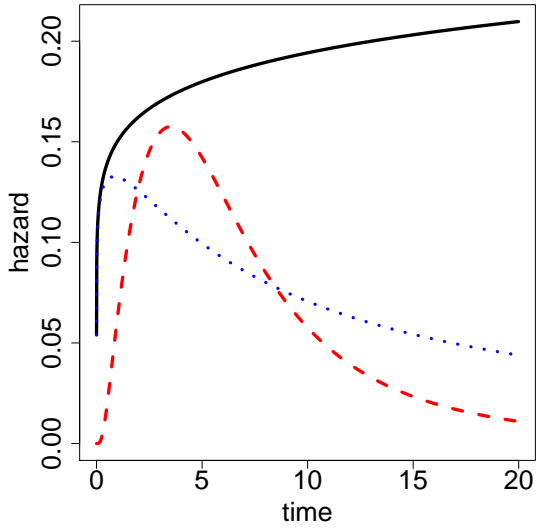
(b) Estimation of  $\xi$  and  $\sigma$  under Joint-DP

Table 6: [Simulation 5] Simulation results for covariate effect estimation in (a) and parameters  $\xi$  and  $\sigma$  in the baseline function in (b). A Poisson cluster process is used to simulate recurrent events. The parameters  $\beta_R$  and  $\beta_D$  quantifies the effects of covariates on the recurrent event baseline intensity and survival hazard functions, respectively. The parameters  $\xi$  and  $\sigma$  control time-varying patterns in the recurrent event baseline intensity and survival hazard functions, respectively. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

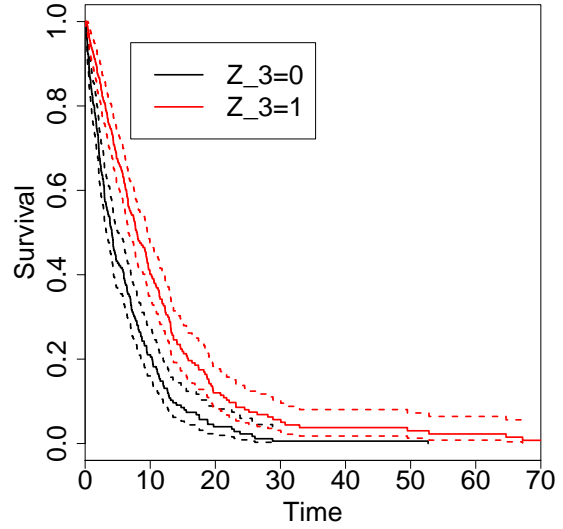
### 2.3 Survival Distributions: Simulations 6A-6C

We next examined robustness of the proposed approach for joint analysis to misspecification of the distribution  $F_W$  of  $W_i$ . For the esophageal cancer data, we chose a distribution for survival time based on empirical model fit comparison. Specifically, we considered AFT models with three different distributions, the standard extreme, normal and logistic distributions for  $W_i$  in (3) of the main text. The three distributions correspond to Weibull, log-normal, log-logistic distributions for  $D_i$ , respectively. We fit the S-DP model and chose the distribution that produced the smallest DIC for joint analysis. To assess robustness of the above proce-

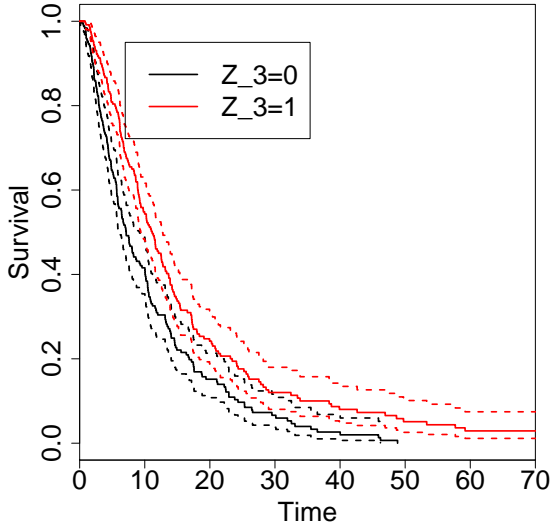




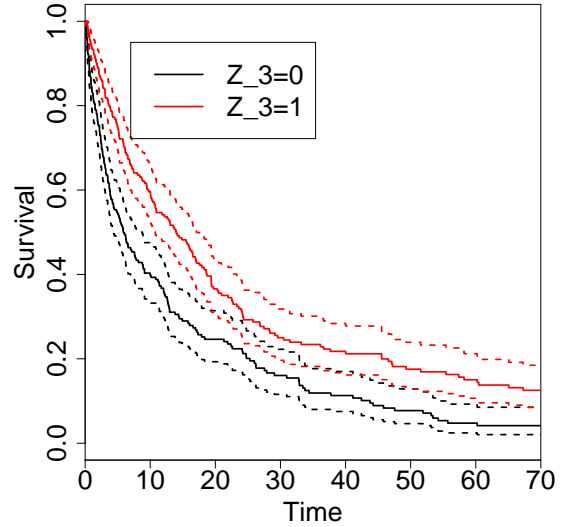
(a) Assumed baseline hazard



(b) Standard extreme



(c) Standard normal



(d) Standard logistic

Figure 1: [Simulation 6] In panel (a), baseline hazard function  $h_0(t)$  under the standard extreme (black solid line), standard normal (red dashed line), standard logistic (blue dotted line) distributions for  $W_i$ . They correspond to Weibull, log-normal, log-logistic distributions, respectively.  $\sigma^{\text{TR}} = 1$  is assumed. In panels (b)-(d), Kaplan-Meier curves are given with a simulated dataset under the three different distributions assumed for  $W_i$  for illustration.

ture, we included the model selection procedure for  $W_i$  and extended the Joint-DP model, called Joint-DP with Model Selection. We then conducted simulation studies by generating  $W_i$  from several different distributions.

		Recurrent Events			Survival		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP with True <math>F_W</math></b>	Mean	0.294	0.249	-0.303	-0.294	-0.250	0.304
	Bias	<b>-0.006</b>	<b>-0.001</b>	<b>-0.003</b>	0.006	<b>0.000</b>	<b>0.004</b>
	Ave. CI length	<b>0.311</b>	<b>0.153</b>	<b>0.345</b>	<b>0.388</b>	<b>0.190</b>	<b>0.433</b>
	Coverage	<b>94.2%</b>	94.8%	<b>95.2%</b>	<b>95.1%</b>	<b>95.2%</b>	94.6%
<b>Joint-DP with Model Selection</b>	Mean	0.294	0.249	-0.303	-0.295	-0.251	0.304
	Bias	<b>-0.006</b>	<b>-0.001</b>	<b>-0.003</b>	<b>0.005</b>	-0.001	<b>0.004</b>
	Ave. CI length	<b>0.311</b>	<b>0.153</b>	<b>0.345</b>	0.393	0.192	0.437
	Coverage	93.8%	<b>95.1%</b>	95.5%	94.6%	<b>95.2%</b>	<b>95.1%</b>
<b>JSCM</b>	Mean	0.37	0.31	-0.37	0.30	0.25	-0.30
	Bias	0.07	0.06	-0.07	0.00	0.00	0.00
	Ave. CI length	0.05	0.02	0.05	0.80	0.40	0.90
	Coverage	10.4%	6.6%	11.5%	99.9%	99.9%	100.0%

Table 7: [Simulation 6A: Standard Extreme (Weibull)] Simulation results for covariate effect estimation. An AFT model with the standard extreme distribution (Weibull) is used to simulate survival times. The proposed joint model is extended to include the model selection based on DIC using S-DP model and labeled Joint-DP with Model Selection. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

In Simulation 6A, we used the same simulation set up for recurrent events and frailties as in Simulation 1. For survival times, we fixed  $\mu^{\text{TR}} = 1.8$ ,  $\zeta^{\text{TR}} = -0.5$ ,  $\sigma^{\text{TR}} = 0.9$ . We simulated  $W_i$  from the standard extreme distribution and let  $\log(D_i) = \mu^{\text{TR}} + \zeta\gamma_i^{\text{TR}} + \mathbf{Z}_i^{\text{TR}}\boldsymbol{\beta}_D^{\text{TR}} + \sigma W_i$ . We used the same censoring procedure used in Simulation 1 to generate censored time  $\tilde{D}_i$ . For Simulations 6B and 6C, we kept the same simulation set up used in Simulation 6A except the distribution for  $W_i$ . We used the standard normal and logistic distributions for Simulations 6B and 6C, respectively. Figure 1(a) illustrates the assumed hazard functions under the distributions of  $W_i$ . Figure 1(b)-(d) illustrates Kaplan-Meier curves for datasets simulated under the standard extreme, normal and logistic distributions, respectively. For each simulation, we fit the following models for comparison: (1) the Joint-DP model that assumes the true distribution  $F_W$  of  $W_i$  is known, (2) Joint-DP with Model Selection and (3) JSCM.

The results of estimation of  $\boldsymbol{\beta}_R$  and  $\boldsymbol{\beta}_D$  under 6A-6C are summarized in Tables 7-9. Table 10 shows how often each of the distributions is chosen by the Joint-DP with Model Selection. Performance of the Joint-DP with Model Selection for inference on  $\boldsymbol{\beta}$  is very similar to that of the Joint-DP model with the true  $F_W$ . This may be because the true

		Recurrent Events			Survival		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP with True <math>F_W</math></b>	Mean	0.299	0.253	-0.305	-0.301	-0.252	0.302
	Bias	<b>-0.001</b>	<b>0.003</b>	<b>-0.005</b>	-0.001	<b>-0.002</b>	<b>0.002</b>
	Ave. CI length	<b>0.281</b>	<b>0.139</b>	<b>0.313</b>	<b>0.382</b>	<b>0.187</b>	0.427
	Coverage	<b>93.8%</b>	93.1%	<b>93.4%</b>	96.1%	96.6%	95.9%
<b>Joint-DP with Model Selection</b>	Mean	0.299	0.253	-0.305	-0.300	-0.252	0.302
	Bias	<b>-0.001</b>	<b>0.003</b>	<b>-0.005</b>	<b>0.000</b>	<b>-0.002</b>	<b>0.002</b>
	Ave. CI length	<b>0.281</b>	<b>0.139</b>	<b>0.313</b>	<b>0.382</b>	<b>0.187</b>	<b>0.426</b>
	Coverage	<b>93.8%</b>	<b>93.4%</b>	93.2%	<b>95.8%</b>	<b>96.0%</b>	<b>95.5%</b>
<b>JSCM</b>	Mean	0.37	0.32	-0.38	0.30	0.25	-0.30
	Bias	0.07	0.07	-0.08	0.00	0.00	0.00
	Ave. CI length	0.03	0.02	0.04	0.91	0.45	1.02
	Coverage	8.0%	5.4%	8.7%	99.9%	100.0%	100.0%

Table 8: [Simulation 6B: Standard Normal (Log-Normal)] Simulation results for covariate effect estimation. An AFT model with the standard normal distribution (log-normal) is used to simulate survival times. The proposed joint model is extended to include the model selection procedure based on DIC using S-DP model, and is labeled Joint-DP with Model Selection. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

distribution of  $W_i$  is chosen most of time, as shown in Table 10. The JSCM assumes an arbitrary distribution for  $W_i$  in an AFT framework and accommodates all three distributions used for Simulation 6 as special cases. Since the NHPP is not nested in the JSCM, we focus on the estimation of  $\beta_D$ . The JSCM performs poorly compared to both the Joint-DP model and Joint-DP with Model Selection. The JSCM produces larger biases with larger interval estimates for  $\beta_D$  than the Joint-DP with Model Selection. The performance for Simulation 6C is worse under all models, but the JSCM is affected more severely than either the Joint-DP model or Joint-DP with Model Selection.

## 2.4 Additional Comparisons to Single Outcome Models: Simulations 7 & 8

We performed simulation studies, Simulations 7 and 8 to focus specifically on how the Joint-DP model may improve inferences compared to the single outcome models, R-DP and S-DP models. For Simulation 7, we kept the simulation set up as in Simulation 1, but increased the probability of censoring; we let  $D_i$  be censored with probability 0.8. Note that the probability

		Recurrent Events			Survival		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP with True <math>F_W</math></b>	Mean	0.300	0.252	-0.299	-0.293	-0.249	0.302
	Bias	<b>0.000</b>	<b>0.002</b>	<b>0.001</b>	0.007	0.001	<b>0.002</b>
	Ave. CI length	0.268	<b>0.131</b>	0.295	<b>0.645</b>	<b>0.315</b>	<b>0.719</b>
	Coverage	89.2%	<b>90.4%</b>	<b>90.0%</b>	<b>95.4%</b>	<b>95.0%</b>	<b>95.4%</b>
<b>Joint-DP with Model Selection</b>	Mean	0.301	0.253	-0.299	-0.294	-0.250	0.302
	Bias	0.001	0.003	<b>0.001</b>	<b>0.006</b>	<b>0.000</b>	<b>0.002</b>
	Ave. CI length	<b>0.267</b>	<b>0.131</b>	<b>0.294</b>	0.650	0.319	0.723
	Coverage	<b>89.3%</b>	88.8%	89.5%	94.6%	94.7%	94.3%
<b>JSCM</b>	Mean	0.20	0.17	-0.20	0.11	0.09	-0.11
	Bias	-0.10	-0.08	0.10	-0.19	-0.16	0.19
	Ave. CI length	0.08	0.04	0.08	1.59	0.78	1.77
	Coverage	11.7%	8.8%	10.8%	99.4%	96.4%	99.9%

Table 9: [Simulation 6C: Standard Logistic (Log-Logistic)] Simulation results for covariate effect estimation. An AFT model with the standard logistic distribution (log-logistic) distribution is used to simulate survival times. The proposed joint model is extended to include the model selection procedure based on DIC using the S-DP model, and labeled Joint-DP with Model Selection. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible (confidence) intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

True Model	Selected Model		
	Standard Extreme	Standard Normal	Standard Logistic
Standard Extreme	0.921	0.039	0.040
Standard Normal	0.000	0.967	0.033
Standard Logistic	0.000	0.352	0.648

Table 10: [Simulation 6A-6C: Model Selection] Survival time is simulated under three different models, standard extreme (Weibull), standard normal (log-normal), standard logistic (log-logistic) models. The Joint-DP is extended to include the model selection based on DIC using DP-S. The percentage of selected models under each true distribution is tabulated.

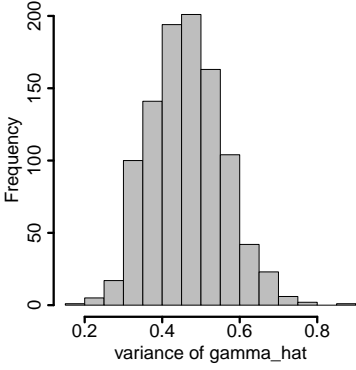
of censoring is 0.25 in Simulation 1. Due to the larger probability of censoring, observed survival or censoring time  $\tilde{D}_i$  and the number of observed recurrent events  $N_i(\tilde{D}_i)$  are smaller than those in Simulation 1. Thus, inference in Simulation 7 is more challenging than that in Simulation 1. The simulation results under the three models are summarized in Table 11. Comparing of the Joint-DP model to the R-DP model, biases and average CI lengths are

larger under the R-DP for the estimation of  $\beta_R$ . Inference on  $\xi$ , the parameter that captures the time-varying pattern in the baseline intensity function is very different under the Joint-DP and R-DP models. The Joint-DP model provides more reasonable inference on  $\xi$ . Most notably, the Joint-DP model clearly outperforms the S-DP model even when the recurrent event is rarely occurred (the average number of recurrent events per subject in the simulated dataset is less than one for most of simulated datasets). The performance of the S-DP model is greatly affected by larger prevalence of censoring. Recall that  $Z_{1i}$  is used as a prognostic factor to determine treatment  $Z_{3i} \in \{0, 1\}$  to reflect a physician’s covariate-based treatment selection process in the simulation. The table shows that while estimation of  $\beta_{D2}$  under the S-DP model is comparable to that under the Joint-DP model, estimation of  $\beta_{D1}$  and  $\beta_{D3}$  is much worsened under the S-DP model for all criteria, bias, average CI length and coverage.

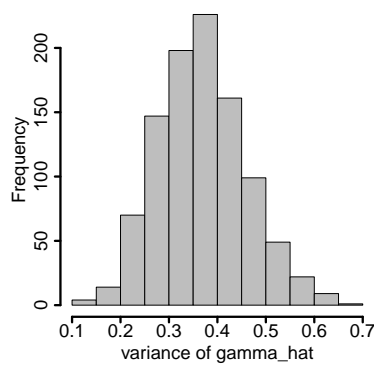
For Simulation 8, we simulated recurrent events from a NHPP with larger intensities, that is, larger  $\eta_k^{*TR}$ , while keeping a large probability for survival times being censored similar to Simulation 7. We simulated patient frailties  $\eta^{TR} = \log(\gamma^{TR})$  from a mixture distribution with  $K = 5$  components,  $p(\eta_i^{TR}) = \sum_{k=1}^K \pi_k \phi(\eta_i^{TR} | \eta_k^{*TR}, 0.25)$ , where the mixture component means  $(\eta_1^{*TR}, \dots, \eta_5^{*TR}) = (0.0, 1.1, 1.5, 2.1, 2.5)$ . The component weights are simulated from a Dirichlet distribution  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_5) \sim \text{Dir}(3, 3, 3, 3, 3)$ . The distribution used to simulate  $\gamma_i^{TR}$  has a larger variance than those used in the other simulations, such as Simulations 1 and 7. We set the covariate effects  $\beta_R^{TR} = (0.30, 0.25, -0.30)'$ ,  $\beta_D^{TR} = (-0.30, -0.25, 0.30)'$ ,  $\xi^{TR} = 0.8$ ,  $\zeta^{TR} = -1.2$  and  $\sigma^{TR} = 0.7$ , the same as in Simulation 1. But we fixed,  $\mu^{TR} = 2.5$ . We simulated the death time  $D_i$  from a Weibull model in Eq (4) of the main text with the above assumed true parameter values. We let  $\delta_i = 0$  (the survival outcome is observed) or 1 (the survival outcome is censored) with probabilities 0.25 and 0.75, respectively, set  $\tilde{D}_i = D_i$  for patients with  $\delta_i = 0$ , and for patients with  $\delta_i = 1$  we assumed uniform censoring time  $\tilde{D}_i \sim \text{Unif}(0, D_i)$ . Table 12 summarizes the results of Simulation 8. Similar to the results of Simulation 7, the Joint-DP model improves estimation of all parameters including  $\beta_R$  and  $\beta_D$  compared to those under the single outcome models. Especially for  $\beta_D$ ,  $\beta_{R1}$  and  $\beta_{R2}$  that are related to biased treatment assignment and treatment, we observe better performance of the Joint-DP. Due to larger  $\gamma_i$ , estimation of  $\xi$  is improved but estimation of  $\sigma$  is deteriorated under the Joint-DP and S-DP models. For both parameters, the Joint-DP model produces better inference on average. Figure 2 illustrates histograms of variances of the posterior means of random frailties in simulated datasets. For each simulated dataset, we computed variances of posterior means of  $\gamma_i$  for the Joint-DP and R-DP models and of  $\mu_i$  for the S-DP model. Note that  $\gamma_i$  under the Joint-DP model and  $\mu_i$  under the S-DP model are not directly comparable, while  $\gamma_i$ 's under the Joint-DP and R-DP models are comparable.

### 3 Additional Results of the Esophageal Cancer Data Analysis

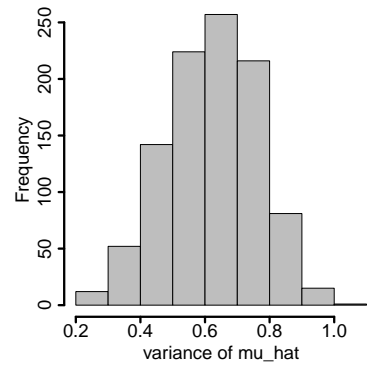
We illustrate the posterior distributions of parameters including regression coefficients  $\beta_R$  and  $\beta_D$  in Figure 3. The vertical dashed lines represents the posterior means with 95%



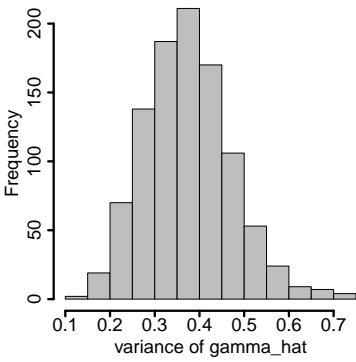
(a) Simulation 1 & J-DP



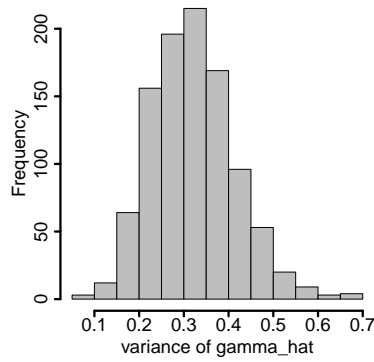
(b) Simulation 1 & R-DP



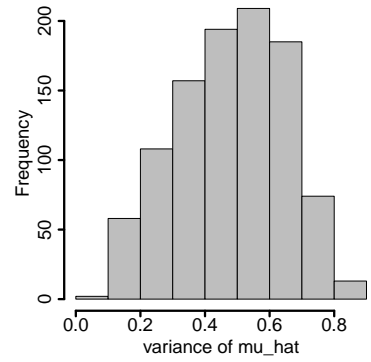
(c) Simulation 1 & S-DP



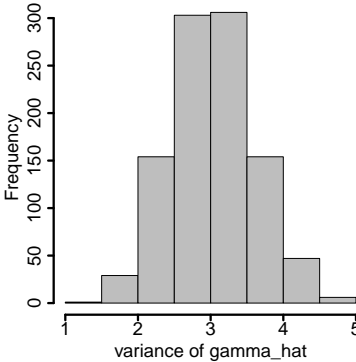
(d) Simulation 7 & J-DP



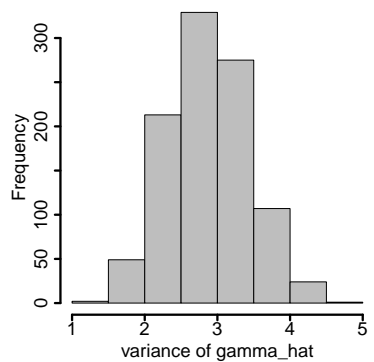
(e) Simulation 7 & R-DP



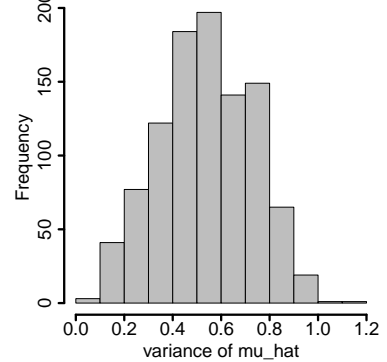
(f) Simulation 7 & S-DP



(g) Simulation 8 & J-DP



(h) Simulation 8 & R-DP



(i) Simulation 8 & S-DP

Figure 2: [Simulations 1, 7 & 8] Histograms of variances of posterior means of random frailties. For each simulated dataset, the variance of the posterior means of  $\gamma_i$  for the J-DP and R-DP models (similarly, the variance of posterior means of  $\mu_i$  for the S-DP model) is computed. Each histogram is based on 1000 simulated datasets.

credible intervals. Posterior means and 95% credible interval estimates of  $\beta_R$ ,  $\beta_D$ ,  $\xi$  and  $\sigma$  are also summarized in Tables 3 and 4 of the main text.

In §4 of the main text, we compute the deviance information criterion (DIC, Spiegelhalter *et al.* (2002)) and the log-pseudo marginal likelihood statistic (LPML, Gelfand *et al.* (1992); Gelfand and Dey (1994)). Since  $\beta_R$  and  $\beta_D$  are the parameters of primary interest, we calculate the metrics using the partially marginalized likelihood by integrating out subject level latent variables  $\gamma_i$  for more reliable model comparison (Millar, 2009; Lee *et al.*, 2016);

$$p(\mathcal{D} \mid \boldsymbol{\theta}') = \int p(\mathcal{D} \mid \boldsymbol{\theta}', \boldsymbol{\gamma}) p(\boldsymbol{\gamma}) d\boldsymbol{\gamma}, \quad (2)$$

where  $\mathcal{D} = \{\mathbf{N}, \tilde{D}, \boldsymbol{\delta}\}$  is all data, and  $\boldsymbol{\theta}'$  all parameters but frailties  $\boldsymbol{\gamma}$ . The corresponding DIC is

$$\text{DIC} = D(\bar{\boldsymbol{\theta}}') + 2p_D,$$

where  $D(\boldsymbol{\theta}') = \sum_{i=1}^n -2 \log(\mathcal{D}_i \mid \boldsymbol{\theta}')$  with  $\mathcal{D}_i = \{N_i, \tilde{D}_i, \delta_i\}$  and  $\bar{\boldsymbol{\theta}}'$  is the posterior mean of  $\boldsymbol{\theta}'$ . The effective sample size is  $p_D = \bar{D}(\boldsymbol{\theta}') - D(\boldsymbol{\theta}')$ , where  $\bar{D}(\boldsymbol{\theta}')$  denotes the posterior mean of  $D(\boldsymbol{\theta}')$ . The partially marginalized density in (2) is not available in closed form and we use numerical approximation using the posterior distribution  $G$  of  $\eta = \log(\boldsymbol{\gamma})$ . We evaluate DIC using posterior simulation of  $\boldsymbol{\theta}$  and select the model with the smallest value of DIC. To compute LPML, we compute the conditional predictive ordinate (CPO) of  $\mathcal{D}_i$  given the data with the observation of subject  $i$  deleted  $\mathcal{D}_{-i} = \{\mathbf{N}, \tilde{D}, \boldsymbol{\delta}\}_{-i}$ ;

$$p(\mathcal{D}_i \mid \mathcal{D}_{-i}) = \int p(\mathcal{D}_i \mid \boldsymbol{\theta}') p(\boldsymbol{\theta}' \mid \mathcal{D}_{-i}) d\boldsymbol{\theta}'.$$

Larger values of CPO indicate a better fit of the model to the data. A Monte Carlo estimate of  $p(\mathcal{D}_i \mid \mathcal{D}_{-i})$  can be obtained based on the MCMC samples of  $\boldsymbol{\theta}'$  from  $p(\boldsymbol{\theta}' \mid \mathcal{D})$  (Chen *et al.*, 2012). We evaluate the LPML defined as  $\sum_{i=1}^n \log p(\mathcal{D}_i \mid \mathcal{D}_{-i})$  and select the model with the largest value of LPML.

## References

- Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* 353–355.
- Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.
- Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**, 3, 501–514.
- Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. Tech. rep., STANFORD UNIV CA DEPT OF STATISTICS.

- Lee, K. H., Dominici, F., Schrag, D., and Haneuse, S. (2016). Hierarchical models for semicompeting risks data with application to quality of end-of-life care for pancreatic cancer. *Journal of the American Statistical Association* **111**, 515, 1075–1095.
- MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 2, 223–238.
- Millar, R. B. (2009). Comparison of hierarchical bayesian models for overdispersed count data using dic and bayes' factors. *Biometrics* **65**, 3, 962–969.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 4, 583–639.



		Recurrent Events ( $\beta_R$ )			Survival ( $\beta_D$ )		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP</b> (Joint with DP)	Mean	0.273	0.234	-0.289	-0.291	-0.256	0.318
	Bias	<b>-0.027</b>	<b>-0.016</b>	<b>0.011</b>	<b>0.009</b>	-0.006	<b>0.018</b>
	Ave. CI length	<b>0.489</b>	<b>0.242</b>	<b>0.542</b>	<b>0.742</b>	<b>0.364</b>	<b>0.828</b>
	Coverage	92.7%	<b>94.7%</b>	<b>94.7%</b>	<b>94.4%</b>	94.6%	<b>95.2%</b>
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.263	0.225	-0.276			
	Bias	-0.037	-0.025	0.024			
	Ave. CI length	0.509	0.251	0.564			
	Coverage	<b>93.0%</b>	93.8%	94.5%			
<b>S-DP</b> (Survival only with DP)	Mean				-0.239	-0.253	0.425
	Bias				0.061	<b>-0.003</b>	0.125
	Ave. CI length				0.811	0.398	0.900
	Coverage				93.6%	<b>94.7%</b>	92.1%

(a) Covariate Effect ( $\beta_R$  and  $\beta_D$ )

		Baseline Intensity( $\xi$ )	Baseline Hazards ( $\sigma$ )
Simulation Truth		0.800	0.700
<b>Joint-DP</b> (Joint with DP)	Mean	0.731	0.695
	Bias	<b>-0.069</b>	<b>-0.005</b>
	Ave. CI length	0.125	<b>0.304</b>
	Coverage	<b>43.5%</b>	95.7%
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.688	
	Bias	-0.112	
	Ave. CI length	<b>0.119</b>	
	Coverage	8.0%	
<b>S-DP</b> (Survival only with DP)	Mean		0.734
	Bias		0.034
	Ave. CI length		0.447
	Coverage		<b>95.6%</b>

(b) Parameters in Baseline Functions ( $\xi$  and  $\sigma$ )

Table 11: [Simulation 7] Simulation results for covariate effect estimation in (a) and estimation of parameters in baseline functions in (b). The values are based on 1000 simulated datasets, each fit using each model. Survival times are censored with probability 0.8. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

		Recurrent Events ( $\beta_R$ )			Survival ( $\beta_D$ )		
Parameters		$\beta_{R1}$	$\beta_{R2}$	$\beta_{R3}$	$\beta_{D1}$	$\beta_{D2}$	$\beta_{D3}$
Simulation Truth		0.300	0.250	-0.300	-0.300	-0.250	0.300
<b>Joint-DP</b> (Joint with DP)	Mean	0.299	0.242	-0.265	-0.302	-0.252	0.289
	Bias	<b>-0.001</b>	<b>-0.008</b>	<b>0.035</b>	<b>-0.002</b>	<b>-0.002</b>	<b>-0.011</b>
	Ave. CI length	<b>0.303</b>	<b>0.149</b>	<b>0.336</b>	<b>0.589</b>	<b>0.288</b>	<b>0.656</b>
	Coverage	<b>92.5%</b>	<b>92.2%</b>	<b>92.3%</b>	<b>94.1%</b>	96.1%	<b>93.8%</b>
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.291	0.238	-0.260			
	Bias	-0.009	-0.012	0.040			
	Ave. CI length	0.306	0.150	0.339			
	Coverage	92.2%	92.1%	90.0%			
<b>S-DP</b> (Survival only with DP)	Mean				-0.241	-0.252	0.417
	Bias				0.059	<b>-0.002</b>	0.117
	Ave. CI length				0.787	0.385	0.871
	Coverage				<b>93.3%</b>	<b>95.5%</b>	91.1%

(a) Covariate Effect ( $\beta_R$  and  $\beta_D$ )

		Baseline Intensity( $\xi$ )	Baseline Hazards ( $\sigma$ )
Simulation Truth		0.800	0.700
<b>Joint-DP</b> (Joint with DP)	Mean	0.773	0.647
	Bias	<b>-0.027</b>	<b>-0.053</b>
	Ave. CI length	0.060	<b>0.200</b>
	Coverage	<b>60.3%</b>	82.6%
<b>R-DP</b> (Recurrent events only with DP)	Mean	0.750	
	Bias	-0.050	
	Ave. CI length	<b>0.059</b>	
	Coverage	9.8%	
<b>S-DP</b> (Survival only with DP)	Mean		0.790
	Bias		0.090
	Ave. CI length		0.449
	Coverage		<b>90.6%</b>

(b) Parameters in Baseline Functions ( $\xi$  and  $\sigma$ )

Table 12: [Simulation 8] Simulation results for covariate effect estimation in (a) and estimation of parameters in baseline functions in (b). The values are based on 1000 simulated datasets, each fit using each model. Survival times are censored with probability 0.75. The four criteria are the mean of the point estimates (Mean), average difference between point estimates and the truth (bias), average length of 95% credible intervals (Ave. CI length), and the proportion of interval estimates capturing the truth (Coverage). The numbers for the best cases are in boldface.

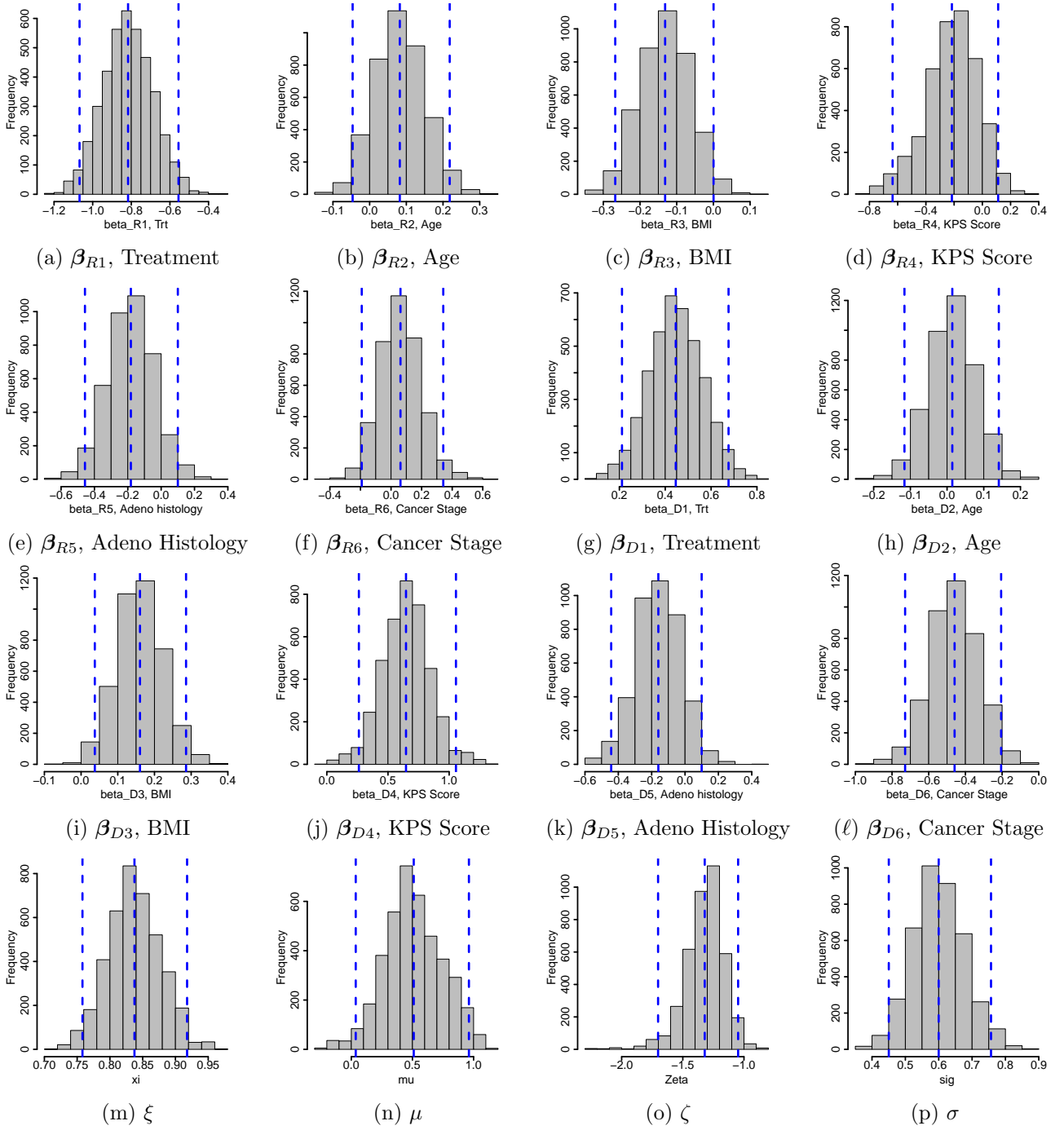


Figure 3: [Esophageal Cancer Data] Posterior distributions of the parameters, (a)–(f) for  $\beta_R$ , (g)–(l) for  $\beta_D$  and (m)–(p) for the baseline function parameters under the proposed joint model with DP. The blue dashed lines are at the posterior means and the lower and upper limits of 95% posterior credible intervals, respectively.