BAYESIAN SEMIPARAMETRIC JOINT REGRESSION ANALYSIS OF RECURRENT ADVERSE EVENTS AND SURVIVAL IN ESOPHAGEAL CANCER PATIENTS

BY JUHEE LEE[*,1], PETER F. THALL[†,2] AND STEVEN H. LIN[†]

*University of California, Santa Cruz[*] and
University of Texas MD Anderson Cancer Center[†]*

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 chemoradiation, including the times of recurrent effusions of fluid to the heart or lungs, survival time, prognostic covariates, and radiation therapy modality.

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 (percardial 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 nonfatal event, one for time to death following the nonfatal event, and one for time to death without the nonfatal 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 • at the time of death or ○ 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 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.
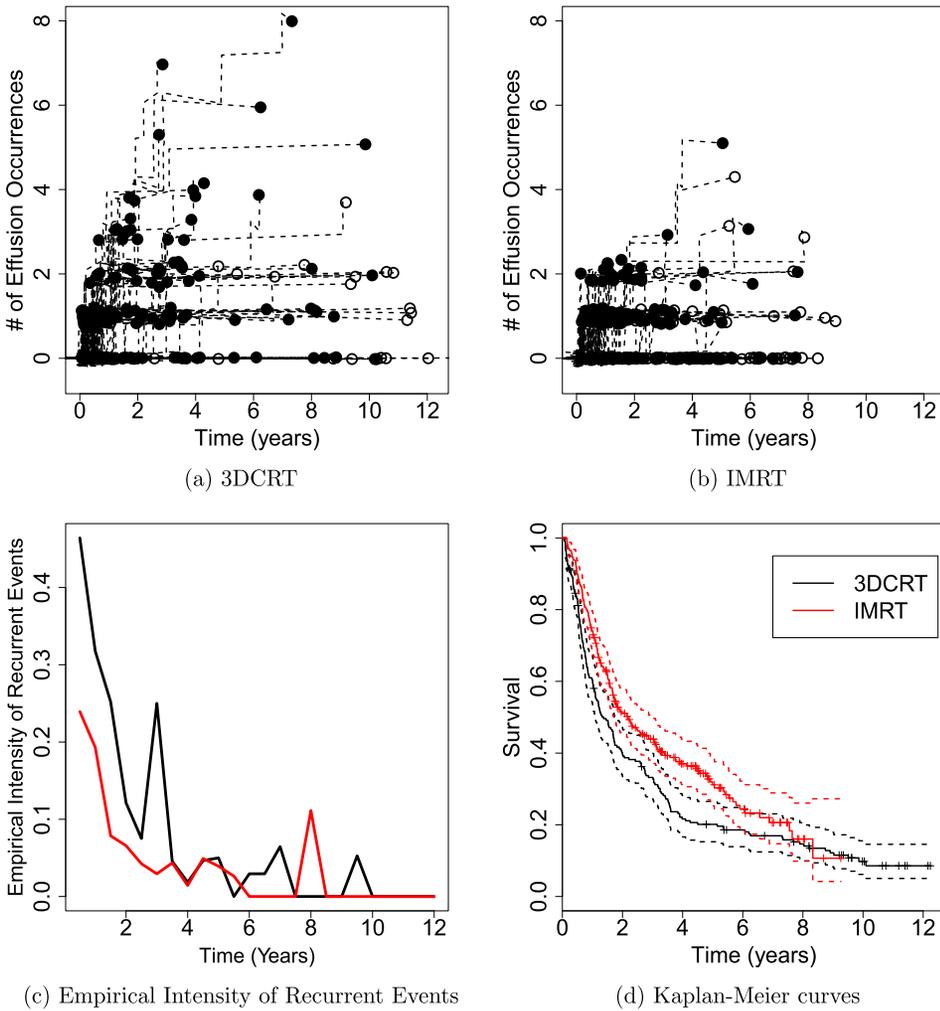
FIG. 1. *Esophageal cancer.* (a) *and* (b) *give plots of counts of recurrent effusion events over time with the two radiation therapy modalities for the esophageal cancer patients, with* ● *and* ○ *representing patients with observed and censored survival outcomes, respectively. For the two radiation therapy modalities,* 3*DCRT* (*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.*

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), Henderson, Diggle and Dobson (2000), Song, Davidian and Tsiatis (2002), Wulfsohn and Tsiatis (1997), Ye, Lin and Taylor (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, Boye and Carlin (2011), and Hatfield, Hodges and Carlin (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 i.i.d. subject-specific random effects (frailties) $\{\gamma_1, \ldots, \gamma_n\}$, with γ_i appearing in both the intensity function of the recurrent event process and the survival distribution of the ith subject. Conditional on γ_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), Kalbfleisch et al. (2013), Liu and Huang (2009), Liu, Wolfe and Huang (2004), Ouyang et al. (2013), Sinha et al. (2008), Wen et al. (2016) for examples. Cook and Lawless (2002) gave a thorough account of models and methods for nonfatal recurrent events. Liu, Wolfe and Huang (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 nonparametric 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 semiparametric approach to jointly model recurrent events and survival, accounting for baseline covariates. Like many existing methods, our modeling framework assumes i.i.d. 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 $i = 1, \ldots, n$, 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 $\mathbf{Z}_{Ri} = (Z_{Ri1}, \ldots, Z_{Rip})'$ and $\mathbf{Z}_{Di} = (Z_{Di1}, \ldots, Z_{Diq})'$ 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 jth occurrence for patient i. Note that $\{N_i(t), t > 0\}$ and $\{T_{i,1}, T_{i,2}, \ldots\}$ 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, Calabria and Pulcini (1989), Kuo and Yang (1996)]. Let $\gamma_1, \ldots, \gamma_n$ denote i.i.d. 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

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

The parameter ξ describes the time-varying pattern of the intensity function. If $\xi < 1 \, (> 1)$, the intensity decreases (increases) over time. The latent frailty γ_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 γ_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 identifiabilty issues. This is further discussed in Section 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 $\gamma_1, \ldots, \gamma_n$ and the truncation times $\tilde{\boldsymbol{D}}$ is

$$
p(\boldsymbol{N} \mid \boldsymbol{\psi}, \tilde{\boldsymbol{D}}) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{N_i(\tilde{D}_i)} \psi_R(T_{i,j}, \boldsymbol{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R) \right\}
$$

$$
(2) \qquad \times \exp \left\{ - \int_0^{\tilde{D}_i} \psi_R(t, \boldsymbol{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R) \, dt \right\}
$$

$$
= \prod_{i=1}^{n} \left\{ \prod_{j=1}^{N_i(\tilde{D}_i)} \gamma_i \xi T_{i,j}^{\xi-1} \exp(\boldsymbol{\beta}_R' \boldsymbol{Z}_{Ri}) \right\} \exp(-\gamma_i \tilde{D}_i^{\xi} e^{\boldsymbol{\beta}_R' \boldsymbol{Z}_{Ri}}).
$$

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

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

where the W_i's are i.i.d. following a distribution with c.d.f. F_W. Any tractable parametric or nonparametric model can be used for F_W, for example, 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 Section 2.3. With the choice of the extreme value distribution, the induced c.d.f. $F_{D,i}(t) = F_D(t \mid \boldsymbol{Z}_{Di}, \gamma_i, \mu, \zeta, \boldsymbol{\beta}_D)$ for $t > 0$ of D_i is the Weibull with shape parameter $1/\sigma$ and scale parameter $\exp\{-(\mu + \zeta \eta_i + \boldsymbol{\beta}_D' \boldsymbol{Z}_{Di})/\sigma\}$. This yields the survival hazard function

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

For each subject $i = 1, \ldots, n$, the random patient frailty, γ_i, is the key element that creates the Bayesian joint model from its two main components, the recurrent event process intensity function $\psi_R(t, \boldsymbol{Z}_{Ri}, \gamma_i, \boldsymbol{\beta}_R, \xi)$ in (1) and the survival process hazard function $\psi_D(t, \boldsymbol{Z}_{Di}, \gamma_i, \mu, \zeta, \boldsymbol{\beta}_D)$ in (4). The hazard function ψ_D has a common factor $\exp(-\mu/\sigma)$ and subject specific factors $\gamma_i^{-\zeta/\sigma}$. Compared to the intensity function ψ_R in (1) that is proportional to γ_i, the effect of γ_i on D_i is adjusted through the term $\gamma_i^{-\zeta/\sigma}$. This allows γ_i to have different effect sizes on the recurrent event intensity and the survival distribution hazard. Another way to

view γ_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 γ_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 γ_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) [Austin (2013), Robins, Hernan and Brumback (2000)] or doubly robust (augmented) IPTW [Bang and Robins (2005), Robins, Hernan and Brumback (2000)]. Instead, we account for possible bias through a flexible modeling strategy for the distribution of the γ_i's. To do this, we take a BNP approach by assuming a Dirichlet process (DP) [Ferguson (1973)] for the distribution of γ_i in equations (1) and (3). This adjusts for possible biases in the posterior estimates of the covariate effects, $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$, 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 $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$.

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 γ_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

$$(5) \qquad \eta_i \stackrel{\text{i.i.d.}}{\sim} G, \qquad G \sim \mathrm{DP}(\alpha G_0),$$

where α is the total mass parameter and G_0 is the baseline distribution of the DP. We assume $G_0 = \mathrm{N}(\bar{\eta}, v^2)$. G_0 determines the features of G such as shape, spread, and location, and α 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

idetifiability issue. A scale parameter is omitted in (1), and no fixed effect is paired with the random effect, while ζ and μ are common for all subjects in (3) [Li, Müller and Lin (2011)]. Following [Escobar and West (1995)], we assume that α is random with prior $\alpha \sim Ga(a_\alpha, b_\alpha)$ Using the stick-breaking representation [Sethuraman (1994)], G can be expressed as an infinite mixture,

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

where $v_k \mid \alpha \overset{\text{i.i.d.}}{\sim} Be(1, \alpha)$, $\eta_k^\star \overset{\text{i.i.d.}}{\sim} G_0$ and $\delta(\eta^\star)$ is a unit point mass at $\eta^\star$. As equation (6) shows, G is almost surely discrete and it allows ties in the η_i's with positive probabilities, resulting in clustering of the η_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 $\psi'_{Ri}(t) = \xi t^{\xi-1} \exp(\boldsymbol{\beta}'_R \mathbf{Z}_{Ri})$ and $\Psi'_{Ri}(t) = \int_0^t \psi'_{Ri}(u)\,du$. Similarly, let $\psi'_{Di}(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \exp(-\frac{\mu + \boldsymbol{\beta}'_D \mathbf{Z}_{Di}}{\sigma})$ and $\Psi'_{Di}(t) = \int_0^t \psi'_{Di}(u)\,du$. Then,

$$\psi_{Ri}(t \mid \mathcal{H}_i(t^-), D_i \geq t) \equiv \psi_{Ri}(t \mid N_i(t^-), D_i \geq t)$$

$$= \psi'_{Ri}(t) \frac{\mathrm{E}[\gamma_i^{N_i(t^-)+1} \exp\{-\gamma_i \Psi'_{Ri}(t) - \gamma_i^{-\zeta/\sigma} \Psi'_{Di}(t)\}]}{\mathrm{E}[\gamma_i^{N_i(t^-)} \exp\{-\gamma_i \Psi'_{Ri}(t) - \gamma_i^{-\zeta/\sigma} \Psi'_{Di}(t)\}]},$$

$$\psi_{Di}(t \mid \mathcal{H}_i(t^-)) \equiv \psi_{Di}(t \mid N_i(t^-))$$

$$= \psi'_{Di}(t) \frac{\mathrm{E}[\gamma_i^{N_i(t^-)-\zeta/\sigma} \exp\{-\gamma_i \Psi'_{Ri}(t) - \gamma_i^{-\zeta/\sigma} \Psi'_{Di}(t)\}]}{\mathrm{E}[\gamma_i^{N_i(t^-)} \exp\{-\gamma_i \Psi'_{Ri}(t) - \gamma_i^{-\zeta/\sigma} \Psi'_{Di}(t)\}]},$$

where the expectations are taken with respect to the distribution G of γ_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 σ, ζ 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 $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$.

For priors on $\boldsymbol{\beta}_R$, $\boldsymbol{\beta}_D$, ξ, μ, and ζ, we assume $\boldsymbol{\beta}_R \sim \mathrm{N}_p(\bar{\boldsymbol{\beta}}_R, \Sigma_R)$, $\boldsymbol{\beta}_D \sim \mathrm{N}_q(\bar{\boldsymbol{\beta}}_D, \Sigma_D)$, $\xi \sim \mathrm{Gamma}(a_\xi, b_\xi)$, $\zeta \sim \mathrm{N}(\bar{\zeta}, \omega^2)$ and $\mu \sim \mathrm{N}(\bar{\mu}, \tau^2)$. We complete the prior model specification by assuming $\sigma \sim \mathrm{IG}(a_\sigma, b_\sigma)$. Specification of numerical hyperparameter values is discussed in Section 3.

We use standard Markov chain Monte Carlo (MCMC) methods to implement posterior inference on the parameters $\boldsymbol{\beta}_R$, $\boldsymbol{\beta}_D$, $\{\eta_i\}$ (equivalently, $\{\gamma_i\}$), ξ, μ, ζ, and σ. 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 ζ and μ may greatly improve the mixing. For updating γ_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 [Lee, Thall and Lin (2019)]. 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 Section 4 simulation of 17,000 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/~juheelee/.

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, $\gamma_i \overset{\text{i.i.d.}}{\sim} \text{logNormal}(0, \tilde{u}^2)$ 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

(7) $$\log(D_i) = \mu_i^S + \boldsymbol{\beta}_D' \boldsymbol{Z}_{Di} + \sigma W_i.$$

That is, the "S" models include a component with both a subject-specific frailty and covariate effects, similar to (3). We assume either $\mu_i^S \overset{\text{i.i.d.}}{\sim} \text{N}(\bar{\mu}^S, (v^S)^2)$ with fixed $\bar{\mu}^S$ and $(v^S)^2$, or $\mu_i^S \overset{\text{i.i.d.}}{\sim} G_{\mu^S}$ and $G_{\mu^S} \sim \text{DP}(\alpha_{\mu^S} \cdot G_{\mu^S 0})$, where $\alpha_{\mu^S} \sim Ga(a_\alpha^S, b_\alpha^S)$ and $G_{\mu^S 0} = \text{N}(\bar{\mu}^S, (v^S)^2)$ with fixed $\bar{\mu}^S$ and $(v^S)^2$. We assume the same prior distributions for $\boldsymbol{\beta}_D$ and σ 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 γ_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 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 $\log(T_{i,j}) \propto -\boldsymbol{\beta}_R' \boldsymbol{Z}_{Ri}$ and $\log(D_i) \propto -\boldsymbol{\beta}_D' \boldsymbol{Z}_{Di}$, respectively. In particular, the intensity of the recurrent events takes the form $\lambda_i(t) = \gamma_i \lambda_0(t \exp(\boldsymbol{\beta}_R' \boldsymbol{Z}_{Ri})) \exp(\boldsymbol{\beta}_R' \boldsymbol{Z}_{Ri})$ and the hazard function of the survival time $h_i(t) = \gamma_i h_0(t \exp(\boldsymbol{\beta}_D' \boldsymbol{Z}_{Di})) \exp(\boldsymbol{\beta}_D' \boldsymbol{Z}_{Di})$ for patient i at time t. λ_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 $\exp(\boldsymbol{\beta}_R' \boldsymbol{Z}_{Ri})$, that is, for a patient with covariate $\boldsymbol{Z}_{Ri}$, the cumulative intensity is $\Lambda_i(t) = \int_0^t \lambda_i(u) \, du = \gamma_i \Lambda_0(t \exp(\boldsymbol{\beta}_R' \boldsymbol{Z}_{Ri}))$, 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 γ_i appears multiplicatively in $h_i(t)$ without any adjustment, and the distribution of γ_i is unspecified, with $\text{E}(\gamma \mid \boldsymbol{Z}, \boldsymbol{\beta}_D) = m_\gamma$. The parameter $\boldsymbol{\beta}_D$ in the JSCM has the opposite sign of the parameter $\boldsymbol{\beta}_D$ in (3) and (7), so to facilitate fair comparison, negative values of the JSCM-based estimates are used for the illustration in Sections 3 and 4. The JSCM computes parameter

estimates for the two processes sequentially, first estimating parameters for the re-
current 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 Section 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 dis-
tributions based on DIC values showed that the extreme value distribution for W
gave the best fit to the survival data in Section 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 includ-
ing this model selection procedure to choose a distribution for W_i. We examined
robustness of this extended approach empirically through simulation studies in
Section 2.3 of the Supplementary Material.

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 let-
ting $Z_{1i} \overset{\text{i.i.d.}}{\sim} 0.5N(-0.5, 0.25^2) + 0.5N(0.5, 0.25^2)$ and $Z_{2i} \overset{\text{i.i.d.}}{\sim} \text{Unif}(-\sqrt{3}, \sqrt{3})$.
We introduced bias in the treatment assignment by simulating the treatment indi-
cator $Z_{3i} \in \{0, 1\}$ 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 in-
creasing in Z_{1i}, patients with larger Z_{1i} are more likely to be given treatment

$Z_{3i} = 0$. We simulated patient frailties $\eta^{TR} = \log(\gamma^{TR})$ from a mixture distribution with $K = 5$ components, $p(\eta_i^{TR}) = \sum_{k=1}^{K} \pi_k \phi(\eta_i^{TR} \mid \eta_k^{\star TR}, 0.25)$, where $\phi(\cdot \mid \eta^{\star}, v^2)$ is the normal distribution with mean $\eta^{\star}$ and variance v^2. We fixed the mixture component means $(\eta_1^{\star TR}, \ldots, \eta_5^{\star TR}) = (-1.5, -0.4, 0.0, 0.4, 0.6)$ and simulated the component weights from $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_5) \sim \text{Dir}(3, 3, 3, 3, 3)$, where $\text{Dir}(\boldsymbol{a})$ denotes the Dirichlet distribution with parameter $\boldsymbol{a}$. We set the covariate effects $\boldsymbol{\beta}_R^{TR} = (0.30, 0.25, -0.30)'$, $\boldsymbol{\beta}_D^{TR} = (-0.30, -0.25, 0.30)'$, $\xi^{TR} = 0.8$, $\zeta^{TR} = -1.2$, $\mu^{TR} = 0.5$ and $\sigma^{TR} = 0.7$. 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 $\gamma_i^{TR} \xi^{TR} t^{\xi^{TR}-1} \exp(\boldsymbol{\beta}_R^{TR\prime} \boldsymbol{Z}_i)$, with $\gamma_i^{TR} = \exp(\eta_i^{TR})$ 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 $\bar{\boldsymbol{\beta}}_R = \bar{\boldsymbol{\beta}}_D = \bar{\eta} = \bar{\mu} = 0$, 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 α, 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 DP-package [Jara et al. (2018)]. We let $v^2 = 1$ for the base distribution G_0. For the priors of ξ and ζ, 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 η_i, we let each η_i have its own cluster and $\eta_k^{\star}$ be a random draw from G_0. We then implemented posterior inference using MCMC simulation over 10,000 iterations, discarding the first 4000 iterations as burn-in and choosing every other sample as thinning.

The simulation results are summarized for $\boldsymbol{\beta}$ in Table 1 and for ξ and σ 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 β'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 ξ and the shape parameter of the baseline hazards function σ.

For comparison, we fit each of the five models described in Section 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

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*

		Recurrent events			Survival		
	Parameters: Simulation truth:	β_{R1} 0.300	β_{R2} 0.250	β_{R3} −0.300	β_{D1} −0.300	β_{D2} −0.250	β_{D3} 0.300
Joint-DP	Mean	0.288	0.247	−0.312	−0.291	−0.251	0.311
(Joint with DP)	Bias	**−0.012**	**−0.003**	**−0.012**	**0.009**	**−0.001**	**0.011**
	Ave. CI length	**0.419**	**0.208**	0.464	**0.432**	**0.213**	0.483
	Coverage	93.1%	92.6%	93.2%	**94.1%**	93.5%	**94.3%**
Joint-logNormal	Mean	0.174	0.261	−0.622	−0.202	−0.255	0.516
(Joint with logNormal	Bias	−0.126	0.011	−0.322	0.098	−0.005	0.216
Frailty)	Ave. CI length	0.473	0.246	**0.392**	0.504	0.246	0.509
	Coverage	75.0%	**95.3%**	28.8%	85.3%	**94.4%**	60.6%
JSCM	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%
R-DP	Mean	0.260	0.223	−0.276			
(Recurrent events	Bias	−0.040	−0.027	0.024			
only with DP)	Ave. CI length	0.446	0.221	0.496			
	Coverage	**93.6%**	91.4%	**94.8%**			
R-logNormal	Mean	0.016	0.222	−0.856			
(Recurrent events	Bias	−0.284	−0.028	−0.556			
only with logNormal	Ave. CI length	0.490	0.260	**0.406**			
Frailty)	Coverage	38.9%	94.1%	3.3%			
S-DP	Mean				−0.276	−0.251	0.339
(Survival only	Bias				0.024	−0.001	0.039
with DP)	Ave. CI length				0.457	0.224	0.507
	Coverage				94.0%	93.7%	93.3%
S-logNormal	Mean				0.121	−0.253	1.270
(Survival only	Bias				0.421	−0.003	0.970
with logNormal	Ave. CI length				0.454	0.233	**0.425**
Frailty)	Coverage				16.6%	92.0%	0.7%

TABLE 2
Simulation results for parameters in baseline functions. Results are based on 1000 *simulated datasets, each fit with each model. The parameters* ξ *and* σ *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*

	Simulation truth:	Baseline intensity (ξ) 0.800	Baseline hazards (σ) 0.700
Joint-DP	Mean	0.796	0.702
(Joint with DP)	Bias	**−0.004**	**0.002**
	Ave. CI length	0.122	**0.202**
	Coverage	**96.1%**	97.1%
Joint-logNormal	Mean	0.848	0.617
(Joint with logNormal	Bias	0.048	−0.083
Frailty)	Ave. CI length	0.117	0.264
	Coverage	58.8%	74.9%
R-DP	Mean	0.686	
(Recurrent events	Bias	−0.114	
only with DP)	Ave. CI length	0	
	Coverage	3.3%	
R-logNormal	Mean	0.695	
(Recurrent events	Bias	−0.105	
only with logNormal	Ave. CI length	**0.099**	
Frailty)	Coverage	11.0%	
S-DP	Mean		0.736
(Survival only	Bias		0.036
with DP)	Ave. CI length		0.265
	Coverage		**95.4%**
S-logNormal	Mean		0.608
(Survival only	Bias		−0.092
with logNormal	Ave. CI length		0.291
Frailty)	Coverage		42.4%

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 β, especially for treatment effects β_{R3} and β_{D3}. For exam-

ple, (Bias, Ave. CI length, Coverage) are $(-0.322, 0.392, 28.8\%)$ for β_{R3} and $(0.216, 0.509, 60.6\%)$ for β_{D3} under the Joint-logNormal model. This comparatively poor performance in estimation of β_1 and β_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 ξ and σ compared to the Joint-DP model. The JSCM produces larger bias and larger confidence intervals for both $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$, although the estimates of β_{R1} and β_{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 $\boldsymbol{\beta}_D$ only are comparable to the simulation truth. The coverage probabilities of the confidence intervals under JSCM are above 99% for $\boldsymbol{\beta}_D$, possibly leading to the conclusion of no 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 $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$ 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 ξ and σ 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 $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$ 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 Section 2 of the Supplementary Material. 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 β_1 and β_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 $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$ 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 $\boldsymbol{\beta}$ 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 Section 2 of the Supplementary Material.

4. Analysis of the esophageal cancer data. In this section, we summarize our analyses of the esophageal cancer dataset described in Section 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 $p_R = p_D = 6$.

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 8000 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 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

TABLE 3

Fits of the esophageal cancer data under the joint models. Point estimates of β_R and β_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 frequenstist JSCM. Covariates having a statistically significant effect are given in boldface

Model	Covariates	β_R		β_D	
Joint-DP	IMRT	**−0.82**	**(−1.07, −0.56)**	**0.45**	**(0.21, 0.68)**
(Joint with DP)	Age	0.08	(−0.05, 0.22)	0.01	(−0.12, 0.14)
	BMI	−0.13	(−0.27, 0.00)	**0.16**	**(0.04, 0.29)**
	KPS Score	−0.22	(−0.64, 0.11)	**0.65**	**(0.26, 1.06)**
	Adeno Histology	−0.18	(−0.46, 0.10)	−0.16	(−0.44, 0.10)
	Cancer Stage	0.06	(−0.19, 0.34)	**−0.46**	**(−0.73, −0.21)**
Joint-logNormal	IMRT	**−0.88**	**(−1.15, −0.62)**	**0.45**	**(0.18, 0.66)**
(Joint with logNormal	Age	0.08	(−0.06, 0.22)	−0.01	(−0.15, 0.12)
Frailty)	BMI	−0.12	(−0.28, 0.03)	**0.15**	**(0.01, 0.28)**
	KPS Score	**−0.36**	**(−0.61, −0.10)**	**0.68**	**(0.38, 0.93)**
	Adeno Histology	−0.10	(−0.39, 0.20)	**−0.31**	**(−0.56, −0.02)**
	Cancer Stage	0.08	(−0.16, 0.33)	**−0.47**	**(−0.68, −0.22)**
JSCM	IMRT	**−0.85**	**(−1.45, −0.25)**	−0.22	(−0.98, 0.54)
	Age	**0.24**	**(0.06, 0.41)**	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)	**−0.84**	**(−1.64, 0.04)**

TABLE 4

Baseline function parameter estimates of the esophageal cancer data. Point estimates of ξ and σ are given with their 95% credible intervals in parentheses under Joint-DP, Joint-logNormal, R-DP, R-logNormal, S-DP, and S-logNormal

Model	Baseline intensity (ξ)	
Joint-DP	0.84	(0.76, 0.92)
Joint-logNormal	0.90	(0.83, 0.98)
R-DP	0.68	(0.62, 0.74)
R-logNormal	0.70	(0.64, 0.77)

Model	Baseline hazards (σ)	
Joint-DP	0.60	(0.45, 0.76)
Joint-logNormal	0.43	(0.30, 0.61)
S-DP	0.63	(0.49, 0.78)
S-logNormal	0.58	(0.44, 0.71)

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., Ji et al. (2016), Zhang et al. (2013)]. 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^{ξ} 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 Material illustrates the posterior distributions of all parameters. From panel (o) of the figure, the posterior mean of ζ 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 γ_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 γ_i. A very interesting result is that the posterior distribution of γ_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 γ_i's would miss this bimodality, this in turn would affect the posterior estimates of $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$.

Posterior inferences on $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_D$ 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 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 and Dey (1994), Gelfand, Dey and Chang (1992)]. DIC and LPML are commonly used for model comparison in the Bayesian paradigm. DIC measures posterior prediction error based on
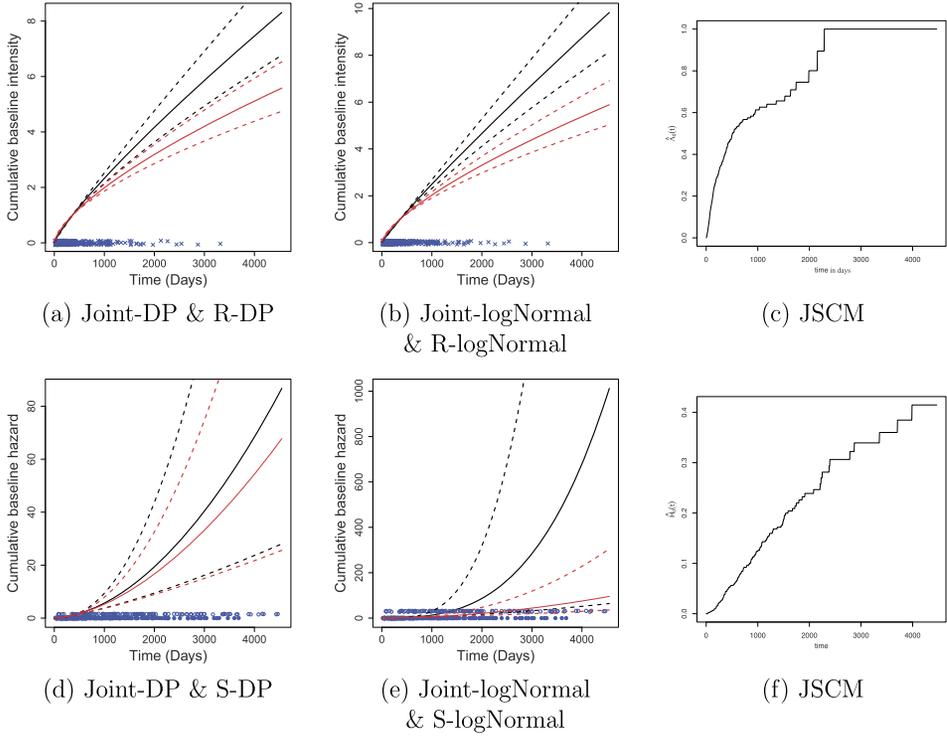
FIG. 2.   *Baseline functions for the esophageal cancer data. In panels* (a)–(c), *inferred baseline intensity functions under different models are illustrated. Panels* (d)–(f) *has plots of inferred baseline hazard functions. Solid lines and dotted lines are posterior means and* 95% *credible intervals. 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.*

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 [Lee et al. (2016), Millar (2009)]. Details of the DIC and LPML computation are discussed in Section 3 of the Supplementary Material. Table 6 provides DIC and LPML for the mode 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 $\boldsymbol{\beta}_D$.
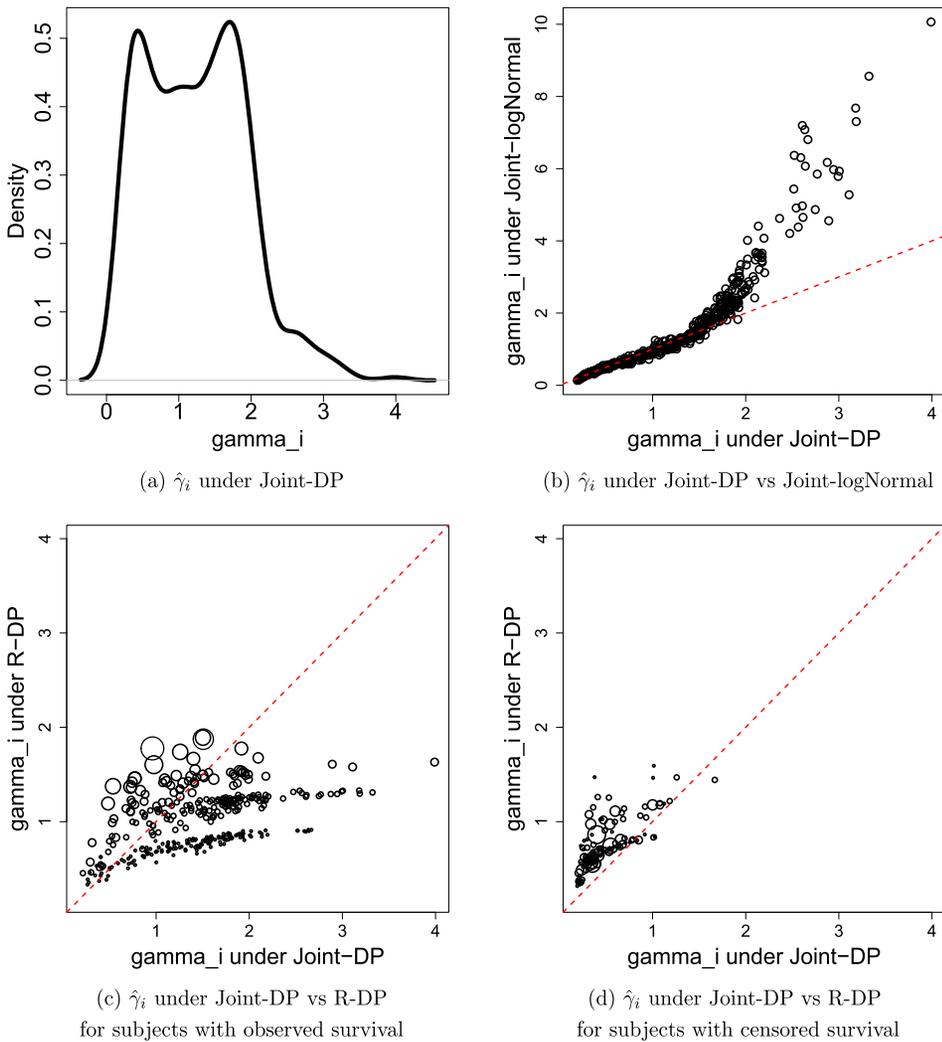
(a) $\hat{\gamma}_i$ under Joint-DP

(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

FIG. 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* δ_1 (*censored survival*), *respectively.*

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

TABLE 5

Fits of the esophageal cancer data under the single outcome models. Point estimates of β_R and β_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	Covariates	β_R		β_D	
R-DP	IMRT	**−0.73**	**(−0.98, −0.50)**		
(Recurrent events	Age	0.09	(−0.05, 0.23)		
only with DP)	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)		
R-logNormal	IMRT	**−0.85**	**(−1.14, −0.54)**		
(Recurrent events	Age	0.09	(−0.07, 0.24)		
only with logNormal	BMI	−0.10	(−0.27, 0.06)		
Frailty)	KPS Score	**−0.45**	**(−0.74, −0.13)**		
	Adeno Histology	−0.26	(−0.56, 0.04)		
	Cancer Stage	−0.09	(−0.37, 0.18)		
S-DP	IMRT			**0.49**	**(0.27, 0.74)**
(Survival only	Age			0.01	(−0.12, 0.15)
with DP)	BMI			0.13	(−0.01, 0.26)
	KPS Score			**0.74**	**(0.40, 1.07)**
	Adeno Histology			−0.13	(−0.43, 0.13)
	Cancer Stage			**−0.42**	**(−0.70, −0.13)**
S-logNormal	IMRT			**0.58**	**(0.38, 0.79)**
(Survival only	Age			0.03	(−0.10, 0.14)
with logNormal	BMI			0.09	(−0.03, 0.23)
Frailty)	KPS Score			**1.00**	**(0.75, 1.24)**
	Adeno Histology			−0.10	(−0.37, 0.11)
	Cancer Stage			−0.25	(−0.48, 0.01)

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 ξ 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 γ_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

TABLE 6
DIC and LPML under Bayesian models for the esophageal cancer data

Model	DIC	LPML
Joint-DP	5654.21	−4113.72
Joint-logNormal	14,206.75	−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

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, 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.

SUPPLEMENTARY MATERIAL

Supplement to "Bayesian semiparametric joint regression analysis of recurrent adverse events and survival in esophageal cancer patients" (DOI: 10.1214/18-AOAS1182SUPP; .pdf). Joint Bayesian semiparametric regression analysis of recurrent adverse events and survival in esophageal cancer patients are available under the paper information link at the Journal website.

REFERENCES

AUSTIN, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Stat. Med.* **32** 2837–2849. MR3069909

BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189

BROWN, E. R. and IBRAHIM, J. G. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* **59** 221–228. MR1987388

BUSH, C. A. and MACEACHERN, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83** 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. *Comput. Statist. Data Anal.* **112** 170–185. MR3645597

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. *Int. J. Radiat. Oncol. Biol. Phys.* **95** 488–497.

COOK, R. J. and LAWLESS, J. F. (2002). Analysis of repeated events. *Stat. Methods Med. Res.* **11** 141–166.

COX, D. R. (1955). Some statistical methods connected with series of events. *J. Roy. Statist. Soc. Ser. B* **17** 129–157; discussion, 157–164. MR0092301

DE GRUTTOLA, V. and TU, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50** 1003–1014.

ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

FAUCETT, C. L. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Stat. Med.* **15** 1663–1685.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. and MOLENBERGHS, G., eds. (2009). *Longitudinal Data Analysis*. CRC Press, Boca Raton, FL. MR1500110

GEISSER, S. (1993). *Predictive Inference*: *An Introduction*. *Monographs on Statistics and Applied Probability* **55**. Chapman & Hall, New York. MR1252174

GELFAND, A. E. and DEY, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *J. Roy. Statist. Soc. Ser. B* **56** 501–514. MR1278223

GELFAND, A. E., DEY, D. K. and CHANG, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. Technical report, Dept. Statistics, Stanford Univ., Stanford, CA.

GHOSH, S. K. and GHOSAL, S. (2006). Semiparametric accelerated failure time models for censored data. *Bayesian Stat. Appl.* **15** 213–229.

GHOSH, D. and LIN, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56** 554–562. MR1795021

GRANDELL, J. (1976). *Doubly Stochastic Poisson Processes*. *Lecture Notes in Mathematics* **529**. Springer, Berlin. MR0433591

GUIDA, M., CALABRIA, R. and PULCINI, G. (1989). Bayes inference for a non-homogeneous Poisson process with power intensity law [reliability]. *IEEE Trans. Reliab.* **38** 603–609.

HATFIELD, L. A., BOYE, M. E. and CARLIN, B. P. (2011). Joint modeling of multiple longitudinal patient-reported outcomes and survival. *J. Biopharm. Statist.* **21** 971–991. MR2823361

HATFIELD, L. A., HODGES, J. S. and CARLIN, B. P. (2014). Joint models: When are treatment estimates improved? *Stat. Interface* **7** 439–453. MR3302373

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. *Radiother. Oncol.* **121** 70–74.

HENDERSON, R., DIGGLE, P. and DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1** 465–480.

HUANG, C.-Y. and WANG, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc.* **99** 1153–1165. MR2109503

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** 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** 366–374. MR3071055

KUO, L. and YANG, T. Y. (1996). Bayesian computation for nonhomogeneous Poisson processes in software reliability. *J. Amer. Statist. Assoc.* **91** 763–773. MR1395743

LAWLESS, J. F. (1987). Regression methods for Poisson process data. *J. Amer. Statist. Assoc.* **82** 808–815. MR0909986

LEE, J., THALL, P. F. and LIN, S. H. (2019). Supplement to "Bayesian semiparametric joint regression analysis of recurrent adverse events and survival in esophageal cancer patients." DOI:10.1214/18-AOAS1182SUPP.

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. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **64** 253–273. MR3302299

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. *J. Amer. Statist. Assoc.* **111** 1075–1095. MR3561930

LI, Y., MÜLLER, P. and LIN, X. (2011). Center-adjusted inference for a nonparametric Bayesian random effect distribution. *Statist. Sinica* **21** 1201–1223. MR2827521

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. *Int. J. Radiat. Oncol. Biol. Phys.* **84** 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. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 65–81. MR2662234

LIU, L., WOLFE, R. A. and HUANG, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60** 747–756. MR2089451

MACEACHERN, S. N. and MÜLLER, P. (1998). Estimating mixture of Dirichlet process models. *J. Comput. Graph. Statist.* **7** 223–238.

MILLAR, R. B. (2009). Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. *Biometrics* **65** 962–969. MR2649870

MÜLLER, P. and RODRIGUEZ, A. (2013). *Nonparametric Bayesian Inference. NSF-CBMS Regional Conference Series in Probability and Statistics* **9**. IMS, Beachwood, OH. MR3113683

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. *Stat. Med.* **32** 2629–2642. MR3067412

ROBERT, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. Springer, New York. MR2723361

ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

SINHA, D., MAITI, T., IBRAHIM, J. G. and OUYANG, B. (2008). Current methods for recurrent events data with dependent termination: A Bayesian perspective. *J. Amer. Statist. Assoc.* **103** 866–878. MR2435473

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** 742–753. MR1945011

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 583–639. MR1979380

TORRE, L. A., BRAY, F., SIEGEL, R. L., FERLAY, J., LORTET-TIEULENT, J. and JEMAL, A. (2015). Global cancer statistics, 2012. *CA Cancer J. Clin.* **65** 87–108.

WALKER, S. and MALLICK, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics* **55** 477–483. MR1705102

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. *Stat. Med.* **35** 4794–4812. MR3554994

WULFSOHN, M. S. and TSIATIS, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53** 330–339. MR1450186

XU, Y., MÜLLER, P., WAHED, A. S. and THALL, P. F. (2016). Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *J. Amer. Statist. Assoc.* **111** 921–950. MR3561917

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. *J. Amer. Statist. Assoc.* **112** 794–805. MR3671771

YE, W., LIN, X. and TAYLOR, J. M. G. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data—a two-stage regression calibration approach. *Biometrics* **64** 1238–1246. MR2522273

ZHANG, S. S., YANG, H., LUO, K. J., HUANG, Q. Y., CHEN, J. Y., YANG, F., CAI, X. L., XIE, X., LIU, Q. W., BELLA, A. E. et al. (2013). The impact of body mass index on complication and survival in resected oesophageal cancer: A clinical-based cohort and meta-analysis. *Br. J. Cancer* **109** 2894–2903.

J. LEE
DEPARTMENT OF APPLIED MATHEMATICS
  AND STATISTICS
BASKIN SCHOOL OF ENGINEERING
UNIVERSITY OF CALIFORNIA, SANTA CRUZ
1156 HIGH STREET MAIL STOP SOE2
SANTA CRUZ, CALIFORNIA 95064
USA
E-MAIL: juheelee@soe.ucsc.edu

P. F. THALL
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF TEXAS
  MD ANDERSON CANCER CENTER
HOUSTON, TEXAS 77030
USA

S. H. LIN
DEPARTMENT OF RADIATION ONCOLOGY
UNIVERSITY OF TEXAS
  MD ANDERSON CANCER CENTER
HOUSTON, TEXAS 77030
USA