



A dependent Dirichlet process model for survival data with competing risks

Yushu Shi¹ · Purushottam Laud² · Joan Neuner²

Received: 5 February 2019 / Accepted: 12 September 2020 / Published online: 12 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

In this paper, we first propose a dependent Dirichlet process (DDP) model using a mixture of Weibull models with each mixture component resembling a Cox model for survival data. We then build a Dirichlet process mixture model for competing risks data without regression covariates. Next we extend this model to a DDP model for competing risks regression data by using a multiplicative covariate effect on subdistribution hazards in the mixture components. Though built on proportional hazards (or subdistribution hazards) models, the proposed nonparametric Bayesian regression models do not require the assumption of constant hazard (or subdistribution hazard) ratio. An external time-dependent covariate is also considered in the survival model. After describing the model, we discuss how both cause-specific and subdistribution hazard ratios can be estimated from the same nonparametric Bayesian model for competing risks regression. For use with the regression models proposed, we introduce an omnibus prior that is suitable when little external information is available about covariate effects. Finally we compare the models' performance with existing methods through simulations. We also illustrate the proposed competing risks regression model with data from a breast cancer study. An R package “DPWeibull” implementing all of the proposed methods is available at CRAN.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10985-020-09506-0>) contains supplementary material, which is available to authorized users.

✉ Yushu Shi
ys8wp@umsystem.edu
Purushottam Laud
laud@mcw.edu
Joan Neuner
jneuner@mcw.edu

¹ University of Missouri, Columbia, Middlebush Hall, Columbia, MO 65201, USA

² Medical College of Wisconsin, CAPS, 8701 Watertown Plank Rd, Milwaukee, WI 53226, USA

Keywords Survival analysis · Competing risks · Nonparametric Bayesian model · Time-dependent covariate

1 Introduction

Among survival regression models, Cox model is used most frequently (Cox 1972). Taking the multiplicative covariate effect assumption from the Cox model, Fine and Gray (1999) proposed a model based on the subdistribution hazards for competing risks data. As both models are valid only when the proportional hazards (subdistribution hazards) assumption is not strongly violated, it is desirable to have a model without such an assumption that provides robust yet interpretable results.

To provide such a flexible model, De Iorio et al. (2004) proposed a nonparametric Bayesian ANOVA method employing a dependent Dirichlet process (DDP). They then adapted their model for continuous covariates in De Iorio et al. (2009). Inspired by the latter DDP model based on mixture of normal distributions and Kottas (2006)'s mixture of Weibull distributions model, we propose here a dependent Dirichlet process mixture model that combines both. Our model inherits many good properties from the Weibull kernel such as a positive domain for observations and explicit likelihood expressions for censored data, while maintaining flexibility in hazard and density functions of the observed data.

Nonparametric Bayesian methods have been proposed for competing risks data, including those based on frailty models (Naskar et al. 2005; Zhang et al. 2014; de Castro et al. 2015) or on the pseudo-likelihood in the Fine and Gray model (Ge and Chen 2012; Lee et al. 2016). The DDP competing risks model we propose in this paper is also based on the subdistribution hazard of Fine and Gray, but we utilize the full likelihood with the data generating distribution taken as a mixture of Weibull-based kernels.

The paper is arranged in the following order. Section 2 shows our model for survival regression data without competing risks. Section 3 enriches the survival model by adding an external time-dependent covariate. Section 4 describes our competing risks model without regression covariates. Section 5 extends the competing risks model to the regression case and discusses how (time varying) cause-specific hazard ratios can also be estimated from posterior samples. Section 6 recommends a prior specification when little external information is available. Section 7 compares results from our model with those from traditional non-Bayesian methods through simulation studies. Section 8 implements the proposed competing risks regression model in a breast cancer dataset. Section 9 concludes the paper with some future plans.

2 Survival model with covariates

Our proposed model combines two models designed specifically for time-to-event data, the Cox proportional hazards model (Cox 1972) and the Dirichlet process mixture (DPM) of Weibull distributions model (Kottas 2006). Although it uses the multiplica-

tive covariate effect assumption on the hazards for each mixture component, this assumption is not inherited by the proposed mixture model.

To incorporate covariates in the Dirichlet process mixtures, we adopt the “fixed-p” dependent Dirichlet process proposed by MacEachern (1999), whose implementation can be achieved using the methods developed for posterior sampling of Dirichlet process mixture models. To write down the model, let t_i be the event time for the i th patient. While this time may or many not be exactly observed due to censoring, we take it to be drawn from a Weibull distribution as

$$t_i | \alpha_i, \lambda_i \stackrel{ind.}{\sim} \text{Weib}(\alpha_i, \lambda_i) \quad i = 1, 2, \dots, n.$$

We use the shape and rate parametrization, i.e., $\text{Weib}(\alpha, \lambda)$ with shape parameter $\alpha > 0$ and rate parameter $\lambda > 0$ has density $f_{\text{Weib}}(t|\alpha, \lambda) = \alpha \lambda t^{\alpha-1} \exp(-\lambda t^\alpha)$, $t > 0$. Denoting $\theta_i, (\alpha_i, \lambda_i) \triangleq \theta_i$, we take

$$\theta_i | \mathbf{Z}_i \sim G_{\mathbf{Z}_i},$$

where \mathbf{Z}_i is the covariate vector of the i th patient. Under the dependent Dirichlet process (DDP) framework, the discrete distribution $G_{\mathbf{Z}_i}$, is one instance from a collection of random distributions $\{G_{\mathbf{Z}}, \mathbf{Z} \in \mathcal{Z}\}$, where \mathcal{Z} is the covariate space. For each $\mathbf{Z} \in \mathcal{Z}$, $G_{\mathbf{Z}}$ can be represented through the stick-breaking method for the Dirichlet process, i.e.,

$$G_{\mathbf{Z}} = \sum_{s=1}^{\infty} \rho_s I(\theta_{\mathbf{Z},s}),$$

where weights ρ_s 's are controlled by the concentration parameter of this Dirichlet process, ν (Sethuraman 1994). The connection with the covariate space is achieved by linking $\theta_{\mathbf{Z},s} = (\alpha_{\mathbf{Z},s}, \lambda_{\mathbf{Z},s})$ with covariate \mathbf{Z} via

$$\alpha_{\mathbf{Z},s} = \alpha_s, \quad \lambda_{\mathbf{Z},s} = \lambda_s \exp(\mathbf{Z}^T \boldsymbol{\beta}_s).$$

We assume the covariates only affect the rate parameter of the Weibull distribution. The covariate coefficient for component s is $\boldsymbol{\beta}_s$. Given covariate \mathbf{Z} , where $\mathbf{Z} \in \mathcal{Z}$, the distribution of the event time is a mixture of Weibull distributions:

$$t | \mathbf{Z} \sim \sum_{s=1}^{\infty} \rho_s \text{Weib}(t | \alpha_s, \lambda_s \exp(\mathbf{Z}^T \boldsymbol{\beta}_s)).$$

Observations having the same parameter values are considered as belonging to the same cluster. For each component, the hazard function with respect to time t is

$$h_s(t) = \alpha_s \lambda_s \exp(\mathbf{Z}^T \boldsymbol{\beta}_s) t^{\alpha_s-1} = \exp(\mathbf{Z}^T \boldsymbol{\beta}_s) h_{0s}(t).$$

Componentwise, covariates have a multiplicative effect on hazard, similar to the Cox model. It is interesting to note that the multiplicative covariate effect on the hazard function is equivalent to an inverse multiplicative effect on the mean survival time and the median survival time. This follows from the expression for the mean and the median of the Weibull component as $\Gamma(1 + \frac{1}{\alpha_s}) / [\lambda_s \exp(\mathbf{Z}^T \boldsymbol{\beta}_s)]$ and $\log(2)^{\frac{1}{\alpha_s}} / [\lambda_s \exp(\mathbf{Z}^T \boldsymbol{\beta}_s)]$, respectively. Changing the sign of the regression coefficient, this means that the model is also componentwise multiplicative in the mean as well as the median.

The parameters in each cluster, i.e., α_s, λ_s and $\boldsymbol{\beta}_s$ are modeled as arising from the base distribution of the Dirichlet process, denoted by G_0 . With the hyperparameter specification of Shi et al. (2019) for G_0 , we complete the model construction as

$$G_0 = \text{Ga}(\lambda|\alpha_0, \lambda_0) \text{I}_{(g(\lambda), \infty)}(\alpha) \text{Ga}(\alpha|\alpha_\alpha, \lambda_\alpha) \text{C}(\boldsymbol{\beta});$$

$$\lambda_0 \sim \text{Ga}(\alpha_{00}, \lambda_{00}); \quad \nu \sim \text{Ga}(a, b). \tag{1}$$

To clarify, $\text{Ga}(\omega, \zeta)$ denotes a Gamma distribution with shape parameter $\omega > 0$ and rate parameter $\zeta > 0$. Its density function is $f_{\text{Ga}}(x|\omega, \zeta) = \frac{\zeta^\omega}{\Gamma(\omega)} x^{\omega-1} \exp(-\zeta x)$. Choice of prior parameters $\alpha_\alpha, \lambda_\alpha, \alpha_{00}, \lambda_{00}, a, b$ as well as the distribution $\text{C}(\boldsymbol{\beta})$ are discussed in Sect. 6.

To sample the posterior, we implemented Neal (2000)'s method, which is based on the Pólya urn scheme representation of the Dirichlet process. Denoting the cluster indicator of the samples as $\mathbf{r} = (r_1, r_2, \dots, r_n)$, the sampling processes are essentially the iterations between two steps.

1. Assign observations sequentially to mixture clusters. For the i th observation, we generate m new possible clusters besides the existing k^- clusters, (If the i th observation is the only one in its cluster, then k^- is the current number of clusters -1 , otherwise k^- is the current number of clusters.) and denote the total number of candidate clusters as V . The probability of being assigned to cluster r is

$$P(r_i = r | r_{-i}, t_i, \mathbf{Z}_i, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_V) \propto \begin{cases} \frac{n_{-i,r}}{n-1+\nu} \text{L}(t_i | \mathbf{Z}_i, \boldsymbol{\theta}_r) & \text{for } 1 \leq r \leq k^-; \\ \frac{\nu/m}{n-1+\nu} \text{L}(t_i | \mathbf{Z}_i, \boldsymbol{\theta}_r) & \text{for } k^- < r \leq V. \end{cases}$$

Here r_{-i} is the cluster assignment for all the observations except for the i th observation. $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_V$ are the parameters in each candidate cluster. For clusters from $1 \leq r \leq k^-$, $\boldsymbol{\theta}_r$ is the posterior given the observations assigned to the r th cluster. For the rest m clusters, $\boldsymbol{\theta}_r$'s are sampled directly from the base distribution. L is the likelihood of the i th observation.

2. Update the θ in each cluster, which contains $\alpha, \lambda, \boldsymbol{\beta}$ for this case.

One advantage of using the Weibull distribution as mixing kernel for time-to-event data is the closed mathematical form of the likelihood for censored observations. If patient i is right censored at time t , the likelihood is $\exp(-\lambda_i t^\alpha)$. If a patient is interval censored and the event time lies between $[t_1, t_2]$, the corresponding likelihood is $\exp(-\lambda_i t_1^{\alpha_i}) - \exp(-\lambda_i t_2^{\alpha_i})$.

It is worth noticing that although the covariate effect is constant in time for each component, the mixture model is not restricted to constant covariate effects. This can

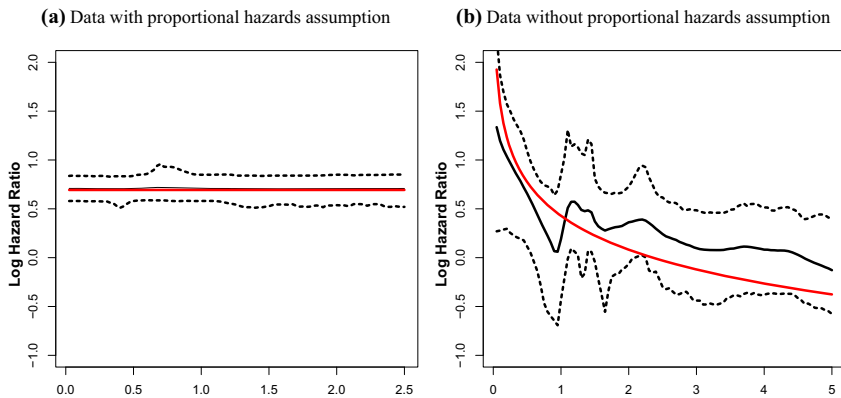


Fig. 1 Visual inspection of proportional hazard assumption. Red lines represent true true log hazard ratio; black solid lines are estimates from the DDP model, dashed lines show 95% pointwise credible intervals (Color figure online)

be seen by writing the hazard ratio given covariate \mathbf{Z} to that given $\mathbf{Z} = \mathbf{0}$ as

$$\frac{h(t|\mathbf{Z})}{h(t|\mathbf{0})} = \frac{\sum_{s=1}^{\infty} \rho_s d_{\text{Weib}}(t|\alpha_s, \lambda_s \exp(\mathbf{Z}^T \boldsymbol{\beta}_s)) / \sum_{s=1}^{\infty} \rho_s S_{\text{Weib}}(t|\alpha_s, \lambda_s \exp(\mathbf{Z}^T \boldsymbol{\beta}_s))}{\sum_{s=1}^{\infty} \rho_s d_{\text{Weib}}(t|\alpha_s, \lambda_s) / \sum_{s=1}^{\infty} \rho_s S_{\text{Weib}}(t|\alpha_s, \lambda_s)}$$

Clearly, this is time dependent and the model is more flexible than allowing only constant hazard ratios. Also notice that because covariates are linked through the scale parameter of the Weibull kernel, the shape of the hazard function is fixed in each cluster, however, the shape of the hazard function can be different across the covariate space through the mixture. The log of the hazard ratio provides an interpretable effect size in a familiar form and is a useful tool for inspecting time varying covariate effects. To illustrate, we generated 1000 observations with the proportional hazard assumption: $x \sim \text{Bernulli}(0.5)$, $y \sim \begin{cases} \text{Weib}(2, 1) & \text{if } x = 0 \\ \text{Weib}(2, 2) & \text{if } x = 1. \end{cases}$ The resulting estimates are shown in Fig. 1a.

Also, 1000 observations without the proportional hazard assumption were generated as: $x \sim \text{Bernulli}(0.5)$, $y \sim \begin{cases} \text{Weib}(1.3, 0.19) & \text{if } x = 0 \\ \text{Weib}(0.8, 0.48) & \text{if } x = 1. \end{cases}$ Estimates are shown in Fig. 1b. The red lines represent the true log hazard ratio from the data generating distributions. The black solid lines represent the estimates from the DDP model, while the dashed lines represent 95% pointwise credible intervals. As can be seen readily, the log hazard ratio estimate remains constant with the proportional hazards data and varies with time for the nonproportional hazards data.

3 Survival model with a time-dependent covariate

We address a situation that often appears in clinical studies, namely, patients are allowed to switch from the initially assigned treatment to an alternative treatment. This results in an external and binary time-dependent covariate. Notationally, we

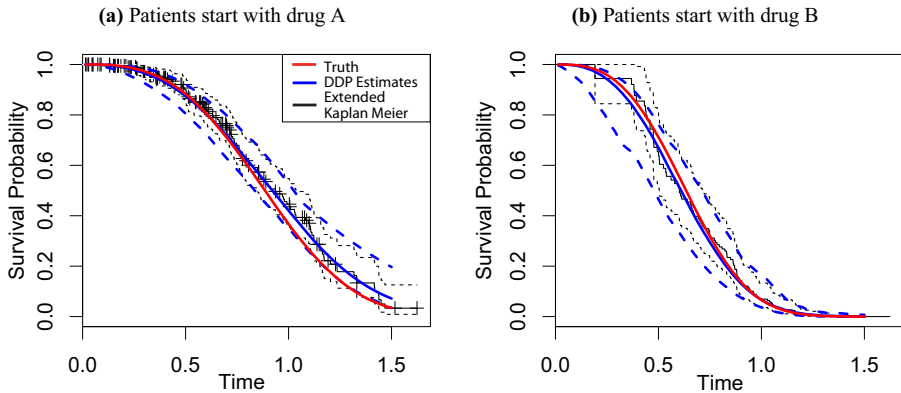


Fig. 2 Estimated survival for dataset with time-dependent covariate. Solid lines are estimates, dashed lines show 95% pointwise credible (in blue) and confidence (in black) intervals (Color figure online)

define the covariate as $Z_d(t) = \begin{cases} 0, & t < t_d \\ 1, & t \geq t_d \end{cases}$, where t_d is the treatment switching time.

If we assume the event time for the baseline is from a Weib(α, λ) distribution and the effect on hazard is multiplicative $h(t) = h_0(t) \exp(Z(t)\beta)$, where β is the regression coefficient for the time-varying covariate, the survival and density functions can be written as

$$S(t) = \begin{cases} \exp(-\lambda t^\alpha), & t < t_d \\ \exp(-\lambda(t_d)^\alpha + \lambda \exp(\beta)(t_d)^\alpha - \lambda \exp(\beta)t^\alpha), & t \geq t_d \end{cases}$$

$$f(t) = \begin{cases} \lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha), & t < t_d \\ \lambda \exp(\beta) \alpha t^{\alpha-1} \exp(-\lambda(t_d)^\alpha + \lambda \exp(\beta)(t_d)^\alpha - \lambda \exp(\beta)t^\alpha), & t \geq t_d. \end{cases}$$

While a likelihood resulting from this density function can be employed in posterior calculations, we find that a technique similar to the extended Kaplan–Meier method proposed by Snapinn et al. (2012) simplifies expressions and provides numerical stability. This consists of splitting the likelihood contribution of one patient into the product of the likelihoods of two “pseudo” patients. For a patient with time-varying covariate changing from 0 to 1 at time t_d and experiencing an event at time t , the likelihood contribution is equivalent to the product of: (i) the likelihood of one pseudo patient with time-varying covariate being the constant 0 and censored at time t_d , equaling $L_1(t) = \begin{cases} \lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha) & t < t_d \\ \exp(-\lambda(t_d)^\alpha) & t \geq t_d \end{cases}$, and (ii) the likelihood of another patient with time-varying covariate being the constant 1 left-truncated at time t_d , equaling $L_2(t) = \lambda \exp(\beta) \alpha t^{\alpha-1} \exp(-\lambda \exp(\beta)t^\alpha + \lambda \exp(\beta)t_d^\alpha)$, $t \geq t_d$. We have employed this method in the reported calculations.

To illustrate, we simulated a dataset with 200 observations all starting with drug A. During the study, some patients may switch to drug B, and the switching times in this simulation are uniformly sampled on the interval $[0, 2]$. For patients using drug A, the hazard is t^2 , whereas for patients using drug B, the hazard is $2.718t^2$. In Fig. 2, the black and gray lines are the extended Kaplan Meier estimates provided

by the R package “survival” (Therneau 2015). The red lines are the “true” survival for patients who start with drug A and B separately, while the blue lines are the estimates and pointwise credible intervals provided by our DDP model. The latter are consistent with the estimates provided by the “survival” package and the credible intervals contain the true survival distributions. When the time-varying covariate does not have a simple binary form, one can still derive the survival and the density function from the relationship $h(t) = h_0(t) \exp(Z(t)\beta)$ and calculate the likelihood, though the likelihood may not have a closed form that is easy to compute and would not bear the analogy with the extended Kaplan Meier method.

4 Competing risks model without covariates

In competing risks data, multiple causes of events can lead to the final outcome while we only observe the event time due to the first cause of the event. This is because the occurrence of one type of event precludes us from observing events of other types. Time until the first of any type of event T and event type k are recorded, $k = 1, \dots, K$.

Following Fine and Gray (1999)’s suggestion of combining multiple causes of secondary interests, we assume there are only two potential causes of events. For the i th observation, $c_i = 1$ indicates event from cause 1 (primary), $c_i = 2$ indicates event from cause 2 (secondary), and $c_i = 0$ indicates right censoring. $F_k(t)$ is the cumulative incidence function (CIF), $F_k(t) = P(T \leq t, c_i = k)$, $k = 1, 2$. The $\lim_{t \rightarrow \infty} F_1(t) = P(c = 1)$ is denoted as p . The cause specific density for the k th cause

is $f_k(t) = \lim_{\Delta \rightarrow 0} \frac{P(t \leq T < t + \Delta, c = k)}{\Delta}$, and $F_k(t) = \int_0^t f_k(s)ds$. The overall survival function has a relationship with the two CIFs that $S(t) = P(T > t) = 1 - \sum_{k=1}^2 F_k(t)$. The subdistribution hazard for cause k is defined as $\eta_k(t) = -d \log\{1 - F_k(t)\}/dt$. For individual i the likelihood can be written as

$$L(t_i) = \{f_1(t_i)\}^{I(c_i=1)} \{f_2(t_i)\}^{I(c_i=2)} \{1 - F_1(t_i) - F_2(t_i)\}^{I(c_i=0)}. \tag{2}$$

By implementing Fan (2008)’s idea of normalizing cause specific density functions into legitimate density functions $d_1(t) = f_1(t)/p$, $d_2(t) = f_2(t)/(1 - p)$ and cumulative incidence functions into legitimate cumulative distribution functions $D_1(t) = F_1(t)/p$ and $D_2(t) = F_2(t)/(1 - p)$, where the normalizing parameter p is set to be $p = F_1(\infty)$, the likelihood can be written as

$$L(t_i) = \{pd_1(t_i)\}^{I(c_i=1)} \{(1 - p)d_2(t_i)\}^{I(c_i=2)} \{1 - pD_1(t_i) - (1 - p)D_2(t_i)\}^{I(c_i=0)}.$$

Considering $d_1(t)$ and $d_2(t)$ as Weibull densities within each component, we obtain a DPM model as follows:

$$t_i | \alpha_{i1}, \lambda_{i1}, \alpha_{i2}, \lambda_{i2}, p_i, c_i \stackrel{ind.}{\sim} L(t_i | \alpha_i, \lambda_i, p_i, c_i) \quad \text{for } i = 1, 2, \dots, n;$$

$$\begin{aligned}
 (\alpha_{i1}, \lambda_{i1}, \alpha_{i2}, \lambda_{i2}, p_i) | G &\stackrel{ind.}{\sim} G; \\
 G &\sim \text{DP}(G_0, \nu); \\
 G_0 &= [\text{Ga}(\lambda|\alpha_0, \lambda_0)\mathbf{I}_{(g(\lambda), \infty)}(\alpha)\text{Ga}(\alpha|\alpha_\alpha, \lambda_\alpha)]^2 \text{Unif}_{[0,1]}(p).
 \end{aligned}
 \tag{3}$$

We use a $\text{Unif}(0, 1)$ as the base distribution for the scaling factor p (Fan 2008). Compared with various frequentists’ strategies for censoring (Fine and Gray 1999; Chen et al. 2012), our model takes advantage of the simple form of Weibull survival function and handles censoring without extra modeling or imputation.

5 Competing risks data with covariates

Fine and Gray (1999) incorporated the core idea of the Cox model with the subdistribution hazard and proposed a model assuming that covariates modify the subdistribution hazard of the primary interest, $\eta_1(t)$, in a proportional manner. In particular,

$$\eta_1(t|\mathbf{Z}) = \eta_{01}(t) \exp(\mathbf{Z}^T \boldsymbol{\beta}_1).$$

where $\eta_{01}(t)$ is the baseline subdistribution hazard function for cause 1. Equivalently, one can write the Fine and Gray model in terms of cumulative incidence function as

$$F_1(t|\mathbf{Z}, \boldsymbol{\beta}_1) = 1 - (1 - F_{01}(t))^{\exp(\mathbf{Z}^T \boldsymbol{\beta}_1)}.
 \tag{4}$$

Here $F_{01}(t)$ is the cumulative incidence function of a patient with baseline covariate ($\mathbf{Z} = 0$) for cause 1. Similarly, we define the cumulative incidence function of a patient with baseline covariate for cause 2 as $F_{02}(t)$, and the corresponding baseline cause specific density functions as $f_{01}(t)$ and $f_{02}(t)$. If we use the normalizing constant $p = F_{01}(\infty)$ to normalize $F_{01}(t)$, $F_{02}(t)$, $f_{01}(t)$ and $f_{02}(t)$, the normalized baseline cumulative incidence functions are denoted as $D_{01}(t) = F_{01}(t)/p$, $D_{02}(t) = F_{02}(t)/(1 - p)$, and the normalized cause specific densities are denoted as $d_{01}(t) = f_{01}(t)/p$ and $d_{02}(t) = f_{02}(t)/(1 - p)$. Equation 4 can be rewritten as:

$$F_1(t|\mathbf{Z}, \boldsymbol{\beta}_1) = 1 - (1 - pD_{01}(t))^{\exp(\mathbf{Z}^T \boldsymbol{\beta}_1)}.
 \tag{5}$$

Differentiating Eq. 5, we obtain the cause specific density for cause 1:

$$f_1(t|\mathbf{Z}, \boldsymbol{\beta}_1, p) = \exp(\mathbf{Z}^T \boldsymbol{\beta}_1)[1 - pD_{01}(t)]^{\exp(\mathbf{Z}^T \boldsymbol{\beta}_1)-1} p d_{01}(t).
 \tag{6}$$

Because of the constraint $F_1(\infty|\mathbf{Z}) + F_2(\infty|\mathbf{Z}) = 1$, we can not apply the same regression model on both causes at the same time. With the information that $F_1(\infty|\mathbf{Z}, \boldsymbol{\beta}_1) = 1 - (1 - p)^{\exp(\mathbf{Z}^T \boldsymbol{\beta}_1)}$, we use the method proposed by Fan (2008) to model cause 2 as

$$F_2(t|\mathbf{Z}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, p) = (1 - p)^{\exp(\mathbf{Z}^T \boldsymbol{\beta}_1)}(1 - (1 - D_{02}(t))^{\exp(\mathbf{Z}^T \boldsymbol{\beta}_2)}).
 \tag{7}$$

Differentiating 7 can lead to the cause specific density function for cause 2,

$$f_2(t|\mathbf{Z}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, p) = (1-p)^{\exp(\mathbf{Z}^T \boldsymbol{\beta}_1)} (1-D_{02}(t))^{\exp(\mathbf{Z}^T \boldsymbol{\beta}_2)-1} \exp(\mathbf{Z}^T \boldsymbol{\beta}_2) d_{02}(t). \tag{8}$$

Plugging Eqs. 5, 6, 7 and 8 into the likelihood described in Eq. 2, the DDP model with competing risks can be written as

$$\begin{aligned}
 t_i | \alpha_{i1}, \lambda_{i1}, \alpha_{i2}, \lambda_{i2}, p_i, \boldsymbol{\beta}_{1i}, \boldsymbol{\beta}_{2i}, c_i, \mathbf{Z}_i &\stackrel{ind.}{\sim} L(t_i | \alpha_{i1}, \lambda_{i1}, \alpha_{i2}, \lambda_{i2}, p_i, \boldsymbol{\beta}_{1i}, \boldsymbol{\beta}_{2i}, c_i, \mathbf{Z}_i) \tag{9} \\
 &\text{for } i = 1, 2, \dots, n; \\
 (\alpha_{i1}, \lambda_{i1}, \alpha_{i2}, \lambda_{i2}, p_i, \boldsymbol{\beta}_{1i}, \boldsymbol{\beta}_{2i}) | G &\stackrel{ind.}{\sim} G; \\
 G &\sim \text{DP}(G_0, \nu); \\
 G_0 &= [\text{Ga}(\lambda | \alpha_0, \lambda_0) \text{I}_{(g(\lambda), \infty)}(\alpha) \text{Ga}(\alpha | \alpha_\alpha, \lambda_\alpha) \text{C}(\boldsymbol{\beta})]^2 \text{Unif}_{[0,1]}(p).
 \end{aligned}$$

Remark In traditional approaches to competing risks data, there has been a healthy debate about whether to model the subdistribution hazards or the cause-specific hazards as time-constant multiplicative functions of covariates. The two approaches clearly lead to different interpretations of the regression coefficients, as is also well-emphasized (Dignam et al. 2012; Allison 2018). In our model in Eq. 9, although the mixture employs component models that are formulated in terms of multiplicative subdistribution hazards, cause-specific (as well as subdistribution) hazard ratios can be recovered from posterior samples, as may be convenient for interpretation. (As noted before, both types of hazard ratios are allowed to be functions of time in the mixture model.) In particular, since the cause specific hazard for cause 1 at time t is $f_1(t|\mathbf{Z})/S_1(t|\mathbf{Z})$, where $S_1(t|\mathbf{Z}) = 1 - F_1(t|\mathbf{Z})/F_1(\infty|\mathbf{Z})$ is the marginal survival function, the cause specific hazard ratio can be written as:

$$\begin{aligned}
 h_1(t|\mathbf{Z})/h_{01}(t) &= \frac{f_1(t|\mathbf{Z})/S_1(t|\mathbf{Z})}{f_{01}(t)/S_{01}(t)} \\
 &= \frac{f_1(t|\mathbf{Z})/(1 - F_1(t|\mathbf{Z})/F_1(\infty|\mathbf{Z}))}{f_{01}(t)/(1 - F_{01}(t)/F_{01}(\infty))},
 \end{aligned}$$

where $f_1(t|\mathbf{Z})$, $f_{01}(t)$, $F_1(t|\mathbf{Z})$, $F_{01}(t)$, $F_1(\infty|\mathbf{Z})$ and $F_{01}(\infty)$ can be estimated from the MCMC posteriors of the DDP mixture of proportional subdistribution hazards model.

6 Prior for DDP model

For the DPM model without covariates and with a Weibull kernel, Shi et al. (2019) have proposed and studied a low-information omnibus (LIO) prior. For the regression model of Sect. 2 (Eq. 1) above, we use and recommend this prior in the absence of external information. Specifically, in Eq. 1 we take $\alpha_\alpha = 0.2$, $\lambda_\alpha = 0.1$, $\alpha_0 = 0.035$, $\alpha_{00} = 1.354$ and $\lambda_{00} = 7.181$. The lower limit for α is defined as a function of λ , $g(\lambda) = \max(0, \frac{\log(3/\lambda)}{3.22})$. This specification is intended to avoid near-zero values for

both shape parameter α and rate parameter λ , as such values correspond to distributions that have an infinite spike at 0 yet assign substantial probabilities to large values. The concentration parameter of the DP, ν , is given a Gamma prior with $a = 1$ and $b = 1$ (Escobar and West 1995).

For competing risks data without covariates (Sect. 4, Eq. 3), we recommend the same prior for the two normalized cumulative incidence functions $D_1(t)$ and $D_2(t)$. The main idea is to ensure the availability of a suitably wide variety of mixture components in the DPM. If p has a $\text{Unif}(0, 1)$ distribution, the desirable components from distribution $pD_1(t) + (1 - p)D_2(t)$ are the same components needed for $D_1(t)$ or $D_2(t)$, thus our previous specification is portable to the competing risks data.

We now address the prior for regression coefficients, namely $C(\beta)$ in Eqs. 1 and 9 by mainly following and adapting to models for time-to-event data the recommendation of Gelman et al. (2008). The first step is to standardize covariates as follows:

- Binary covariates are coded to have a mean of 0 and a difference of 1 between their lower and upper conditions. For example, if 10% of the study cohort uses drug A, and 90% uses drug B, define the centered “Drug Assignment” covariate to take 0.9 for drug A users and -0.1 for drug B users. Covariates with m categories are first converted to $m - 1$ binary covariates;
- Continuous covariates are shifted to have a mean of 0 and scaled to have a standard deviation of 0.5.

To specify priors for regression coefficients in the case of time-to-event data, we follow reasoning similar to that in Gelman et al. (2008). Consider a trial with only two patients, A and B, and only one binary covariate Z . The covariate value for patient B is 0.5, while for patient A is -0.5 . The time that A dies is a random variable T_A , while the time that B dies is a random variable T_B . Denote the corresponding density functions for event times by $f_A(t)$ and $f_B(t)$, the survival functions by $S_A(t)$ and $S_B(t)$, and the hazard functions by $h_A(t)$ and $h_B(t)$. If we assume that the proportional hazards assumption is valid, then $f_B(t) = h_B(t)S_B(t) = \exp(\beta)h_A(t)S_B(t)$. We thus have

$$P(T_A < T_B) = \int_0^\infty S_B(t)f_A(t)dt = \int_0^\infty S_B(t)h_A(t)S_A(t)dt$$

and

$$P(T_A > T_B) = \int_0^\infty S_A(t)f_B(t)dt = \int_0^\infty \exp(\beta)S_B(t)h_A(t)S_A(t)dt.$$

Given the sum of the two probabilities is 1, one can compute the probability of A dying first as $1/(1 + \exp(\beta))$, while the probability of B dying first is $\exp(\beta)/(1 + \exp(\beta))$. By assuming each of these equals 1/2, the likelihood of β is $\exp(\beta)^{1/2}/(1 + \exp(\beta))$. As shown in Fig. 3, we borrow Gelman et al. (2008)’s approximation to the likelihood with a Cauchy(0, 2.5). When combined with the standardization, this prior implies that the absolute difference in log hazard ratio should be less than 5 (equivalent to 148 fold change in terms of hazard) when moving from one standard deviation below the mean, to one standard deviation above the mean for any covariate, which is adequate for most data situations in practice.

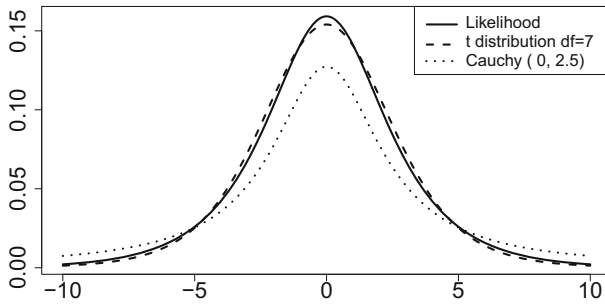


Fig. 3 Regression coefficient prior selection

7 Simulation study

We examine three models with repeated data simulations: survival regression model; competing risks without covariates model; competing risks with covariates model. For each, we first lay out the simulation settings and then discuss the results. All simulations use the LIO prior (Shi et al. 2019) as well as the prior for all regression coefficients of centered and scaled covariates as described in Sect. 6.

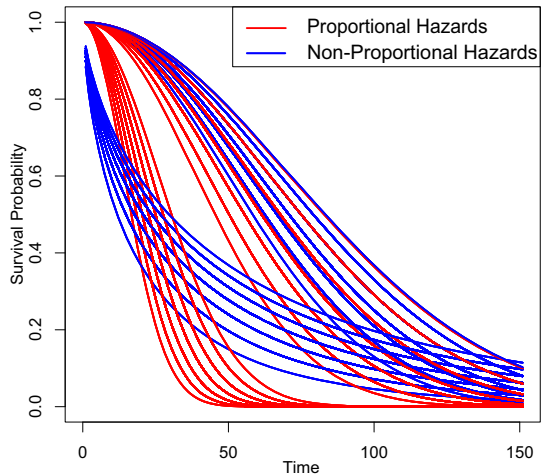
7.1 Survival data with covariates

We compare posterior estimates obtained by Bayesian analysis of our model with two popular methods, the Cox model estimates provided by the R package “survival” (Therneau 2015), and random survival forest estimates from the R package “randomForestSRC” (Ishwaran and Kogalur 2018). Because of restrictions of these two packages, we only simulate right censored data. Our DDP model and its posterior sampling available in the DPWeibull package have the flexibility to handle interval censored data also. We include both categorical and continuous covariates.

7.1.1 Categorical covariates

We borrow the binary covariate simulation settings from Sparapani et al. (2016), which cover both the proportional hazards scenario as well as the non-proportional hazards scenario. In each scenario, we generate 200 datasets, each of which contains 400 observations with 20% censoring and 9 covariates that are equally likely to be 0 or 1. Censoring is achieved through an exponential distribution. For observations from proportional hazards setting, the event times without censoring come from the following distribution: $y|\alpha, \lambda \sim \text{Weib}(\alpha, \lambda)$ with $\alpha = 2$ and $\lambda = \exp(-6 - 0.2(x_1 + x_2 + x_3 + x_4 + x_5 + x_6) - 2x_7)$. For non-proportional hazards setting, the data generating distribution is $y|\alpha, \lambda \sim \text{Weib}(\alpha, \lambda)$ with $\alpha = 0.7 + 1.3x_7$, $\lambda = 1/[20 + 5(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + 10x_7)]^{0.7+1.3x_7}$. Here Covariates x_8 and x_9 are used as noises. All $2^9 = 512$ data generating distributions are shown in Fig. 4, with red lines representing all possible proportional hazard data generating distributions and the blue ones representing all possible non-proportional data generating distributions. The

Fig. 4 Simulation settings for survival data with categorical covariates



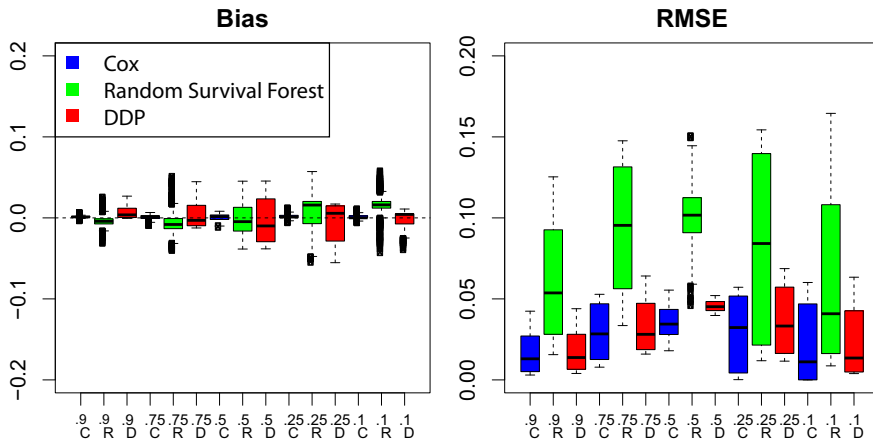
obvious crossings in the non-proportional data generating indicate clear violations of the hazard proportionality assumption.

Though the flexible DDP-of-Weibulls model can provide a wide variety of inferences, we focus on estimation of the survival function at certain times for comparison purposes. These time points are chosen such that the average of all 512 possible data-generating survival functions at these points reaches 0.90, 0.75, 0.50, 0.25 and 0.10. Figure 5 and the supplemental Figure 1 summarize the performance of three estimation methods. Due to a limitation of the package “randomForestSRC”, we cannot provide confidence intervals for random survival forest. The x-axes mark the 5 time points of interest by the average survival. Each box-plot is based on the predicted values of all 512 possible combinations of covariates. In Supplement Figures 1–5, CI stands for confidence interval for frequentist methods and credible interval for Bayesian methods; RMSE stands for root mean squared error. In Fig. 5a, we see that for proportional hazards data, our method does not lose much even compared with Cox model estimates and greatly outperforms random survival forest model in terms of RMSE. For the non-proportional hazards setting, where Cox model is inappropriate, our model gives reasonable RMSE and credible intervals. However, in both cases the Bayesian method shows slightly more bias than RSF; and greater bias than Cox model estimates for proportional hazards. The proposed Bayesian method provides higher CI coverage in most cases.

7.1.2 Continuous covariates

Here, we use Friedman’s function (also employed by Sparapani et al. (2016) in their simulations). There are 10 covariates x_1, \dots, x_{10} , each of which is generated from a $\text{Unif}(0, 1)$, with 5 of these covariates actually affecting data generation. The data are drawn from a Weibull distributions $y|\alpha, \lambda \sim \text{Weib}(\alpha, \lambda)$ with $\alpha = 2$ and $\lambda = \exp\{-6 - 5 \sin(\pi x_1 x_2) - 2(x_3 - 0.5)^2 - x_4 - 0.5x_5\}$.

(a) Proportional hazards data



(b) Nonproportional hazards data

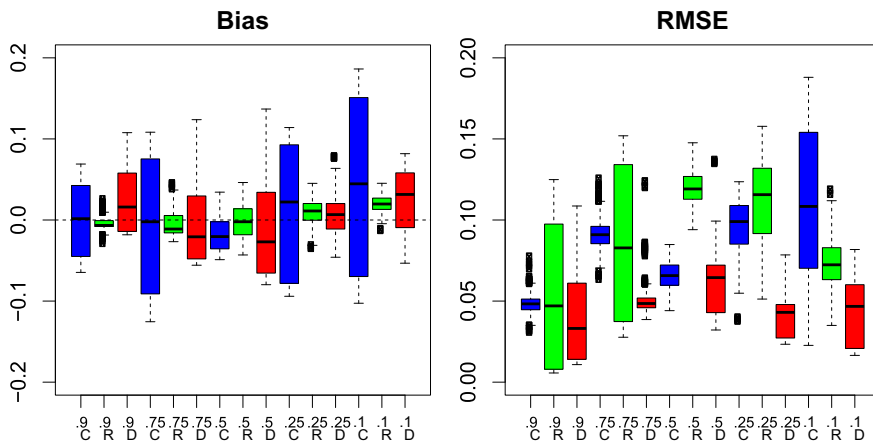


Fig. 5 Comparison of DDP posterior estimates (D) with frequentist methods (Cox model (C) and random survival forest (R)) at 5 time-points where the average survival equals 0.9, 0.75, 0.5, 0.25 and 0.1

In this simulation, we first randomly generate 10,000 covariate combinations and calculate the time when the averaged survival function reaches 9 deciles (denoted as t_{base}). Conditioning on N Weibull observations generated at randomly sampled covariate values, we generate posterior MCMC samples. We then use another K randomly generated covariate combinations for out-of-sample prediction, comparing true survivals with predicted survival probabilities from posterior samples for pre-fixed t_{base} .

Figure 6 plots the predicted survival versus the true survival for $K = 100$ when $N = 400$ and 4000. The dots in the “Zeppelin plot” lie around the 45-degree line, with a high coefficient of determination and a low median absolute error. These together

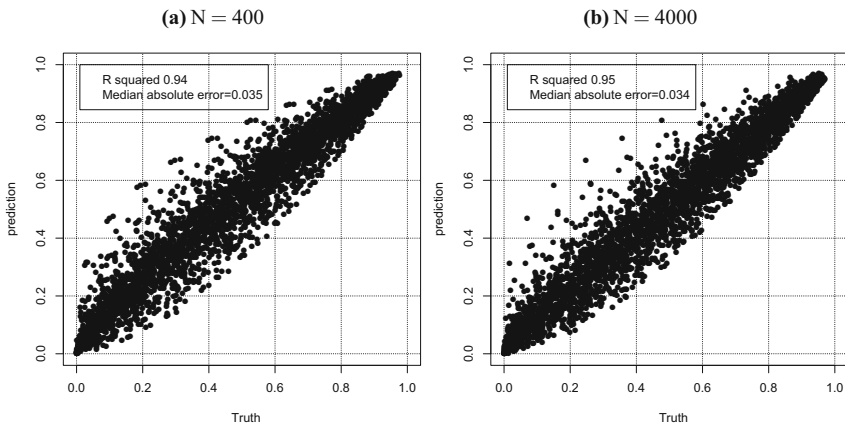


Fig. 6 True survival vs DDP estimates for survival data with continuous covariates

suggest that our method performs well even for a complicated data structure. For comparison purpose, we also test the same data on Cox model and the results are presented in supplemental Figure 6. For this complicated data structure, Cox model does worse in terms of coefficient of determination (0.76 for $N = 400$ and 0.79 for $N = 4000$) and median absolute error (0.087 for $N = 400$ and 0.085 for $N = 4000$). As shown in supplemental Figure 6, the zeppelin shape for Cox model is much wider than the zeppelin for the DDP survival model.

7.2 Competing risks data without covariates

We compare our DPM method for competing risks data with empirical estimates with 200 datasets, each of which contains 200 observations. The two causes have identical hazard before $t = 0.5$. The cause specific hazard function for cause 1 is piecewise constant, $\eta_1(t) = \begin{cases} 1/3 & \text{before } t = 0.5 \\ 1 & \text{after } t = 0.5 \end{cases}$, while the cause specific hazard for cause 2 is set to be a constant $\eta_2(t) = 1/3$. The true cumulative incidence functions can be calculated as

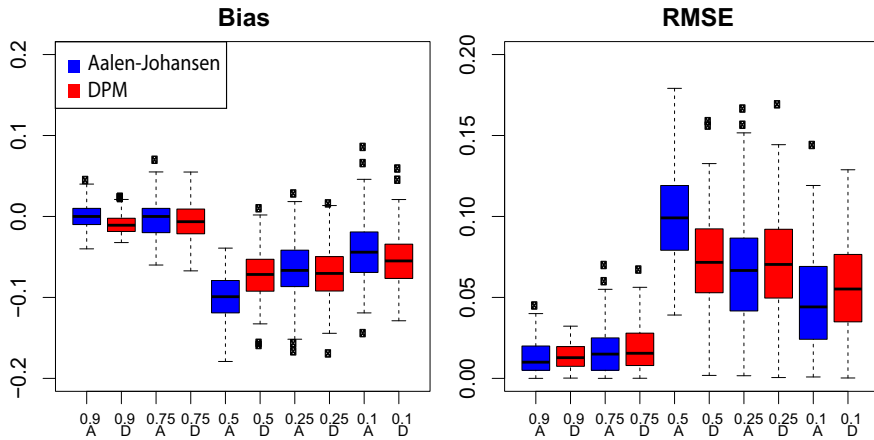
$$F_1(t) = \begin{cases} \frac{1}{2} - \frac{1}{2} \exp(-\frac{2}{3}t) & t \leq 0.5 \\ \frac{1}{2} + \frac{1}{4} \exp(-\frac{1}{3}) - \frac{3}{4} \exp(\frac{1}{3} - \frac{4}{3}t) & t > 0.5 \end{cases},$$

and

$$F_2(t) = \begin{cases} \frac{1}{2} - \frac{1}{2} \exp(-\frac{2}{3}t) & t \leq 0.5 \\ \frac{1}{2} - \frac{1}{4} \exp(-\frac{1}{3}) - \frac{1}{4} \exp(\frac{1}{3} - \frac{4}{3}t) & t > 0.5 \end{cases}.$$

The censoring distribution is set to be Unif(0, 2.5), which results in 35.91% events from cause 1, 23.83% events from cause 2, with 40.27% right censoring.

(a) Cause 1



(b) Cause 2

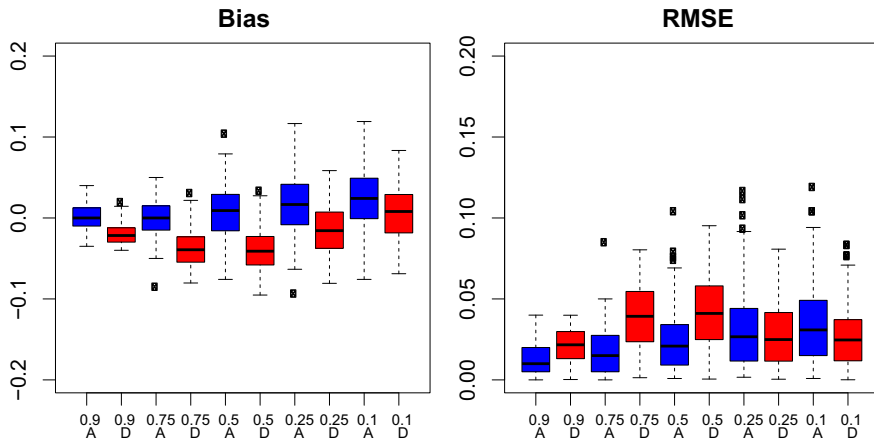


Fig. 7 Comparison of the DPM competing risks posterior estimates (D) with Aalen–Johansen empirical estimates (A) at 5 time-points where the survival equals 0.9, 0.75, 0.5, 0.25 and 0.1

Figure 7 shows the comparison of our estimates with Aalen–Johansen’s estimates in terms of bias and RMSE. Comparisons of confidence/credible interval (CI) coverage and length are included in the supplemental materials, together with additional simulation results (supplemental Figure 2–4). The CIs are compared at the times when the overall survival reaches 0.9, 0.75, 0.5, 0.25 and 0.1. For cause 1 with piecewise constant hazard function, our estimate provides some protection against extremely low CI coverage (see supplemental Figure 2) and from strong bias (at 50th percentile of the overall survival in Fig. 7a), whereas for cause 2 with constant hazard, it produces larger bias and RMSE. However, the performance of DPM estimates is still acceptable as almost all biases are less than 0.1 and all RMSEs are well below 0.15. As shown

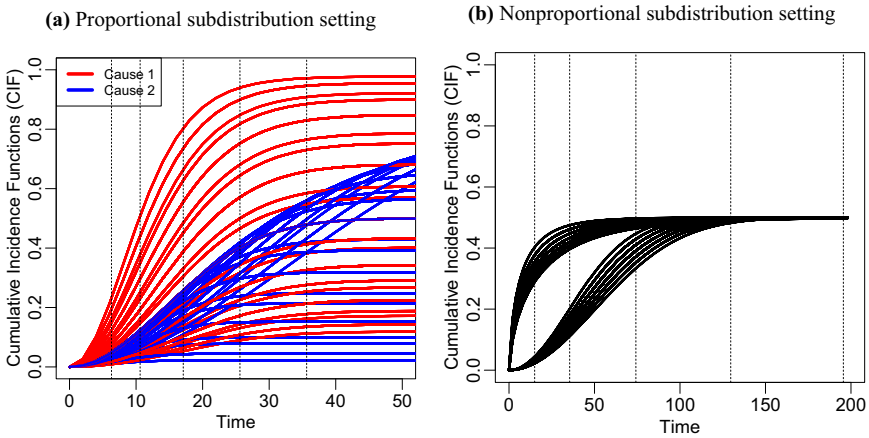


Fig. 8 Simulation settings for competing risks regression data

in supplemental Figure 2, the coverage is better for Scrucca et al. (2007)’s confidence intervals before the median overall survival time, while the credible intervals from DPM catch up later in their coverage.

7.3 Competing risks data with covariates

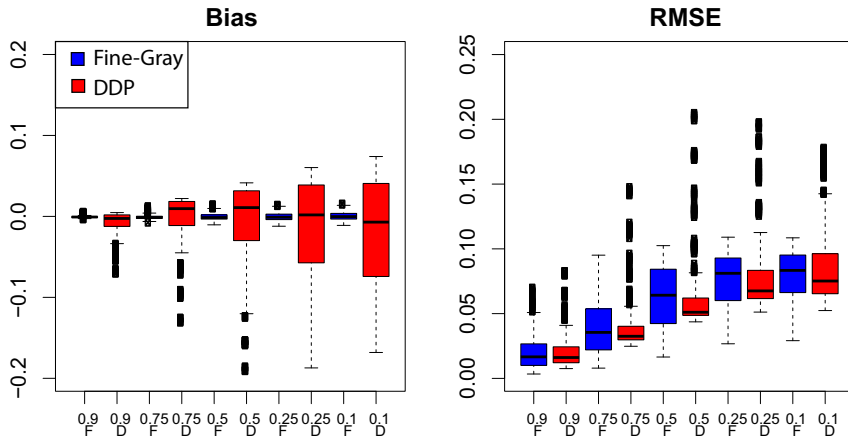
The comparisons are made first under a scenario where the subdistribution hazard has a multiplicative relationship with the covariates, i.e., Fine and Gray model assumption holds; then under a scenario where the subdistribution hazard does not have a multiplicative relationship with the covariates. There are 100 datasets for each scenario, each of which has 400 observations with 9 binary covariates $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9)$, and each covariate has 50% chance of taking the value 1. All inferences shown are for cause 1.

In the multiplicative hazard scenario, the regression coefficients for the two causes are set equal: $\beta_1 = \beta_2 = (-0.2, 0.2, 0.5, -0.5, 0, -1, 1, 0, 0)$. The normalized cumulative incidence functions G_{01} and G_{02} are also identical Weibull distributions with shape parameters $\alpha_1 = \alpha_2 = 2$ and rate parameters $\lambda_1 = \lambda_2 = \exp(-6)$. The censoring times come from a $\text{Unif}(0, 200)$ distribution, while $F_{01}(\infty) = p = 0.5$. It is worth mentioning that though the normalized baseline distributions and the regression coefficients are the same for the two causes, the CIFs do not overlap.

In the non-multiplicative hazard scenario, the two competing risks are set to have the same CIF and there is no censoring. The data are generated by taking the smaller of two values generated from a Weibull distribution with shape parameter $\alpha(\mathbf{x}) = 0.7 + 0.5x_7$ and rate parameter $\lambda(\mathbf{x}) = (20 + 5(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + 10x_7))^{-\alpha(\mathbf{x})}$. This corresponds to the common CIF $t|\mathbf{x} = (1 - S_{\text{Weib}}^2(t|\alpha(\mathbf{x}), \lambda(\mathbf{x}))/2$ for both causes. Figure 8 shows the true CIFs in the two simulation scenarios.

We focus on CIF estimates when the averaged overall survival over all possible 512 configurations of the 9 binary covariates reaches 0.9, 0.75, 0.5, 0.25 and 0.1.

(a) With multiplicative subdistribution hazards assumption



(b) Without multiplicative subdistribution hazards assumption

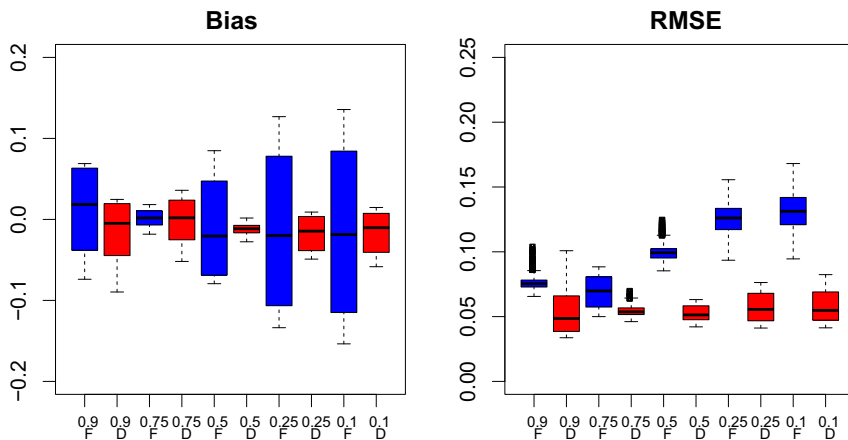


Fig. 9 Comparison of DDP competing risks regression model posterior estimates (D) with Fine and Gray estimates (F) of cause 1 CIF at time points where the average overall survival reaches 0.9, 0.75, 0.5, 0.25 and 0.1

The boxplots in Fig. 9 are made from the 512 estimates corresponding to all covariate configurations, where each estimate is an average over 100 simulated datasets. We compare our method with the Fine and Gray estimates from R package “timereg” in terms of bias, RMSE, CI coverage and CI length. The Fine and Gray estimates are in red and our DDP estimates are in blue.

As shown in Fig. 9, when the proportional subdistribution hazards assumption holds (Fig. 9a), the proposed DDP model gives larger bias, with the vast majority of the bias confined in the interval $[-0.1, 0.1]$ and just a few beyond this range. However, the proposed method shows comparable performance in terms of RMSE with a few deviants.

When the proportional subdistribution hazards assumption is violated (Fig. 9b), proposed DDP method substantially outperforms the Fine and Gray estimates for most time points in terms of bias and RMSE. As shown in the supplemental Figure 5, the credible intervals provided by the DDP method are wider in both scenarios, yet provide more coverage than the confidence intervals from the Fine and Gray method.

8 A breast cancer study

In this study, we focus on comparing the risk of bone fracture under two hormone therapy drugs for breast cancer survivors: Tamoxifen, which is thought to be protective for bone fracture, but carries risk for endometrial cancer and venous thromboembolism; and aromatase inhibitors (AIs), which are well tolerated by most patients (Murthy et al. 2004), and are suggested to reduce breast cancer mortality, but may increase the risk of fracture as shown by a previous clinical trial (Greep et al. 2003).

Our study is based on the Medicare Part D and SEER records of 20,119 postmenopausal breast cancer survivors aged 65 and over from across the U.S. The event of interest is the occurrence of first fracture, and death without having a fracture is considered as a competing risk in this study. Except for 2375 fractures (cause 1) and 2657 deaths (cause 2), this dataset contains 15,087 censored observations which indicate that either a patient has no event over the observation period or is lost to follow-up via a change in Medicare coverage to a type that does not provide patient-level billing data. Fracture was observed in 1870 out of 15,605 patients who started with AIs, and 505 had a fracture out of 4514 patients who began with Tamoxifen. In addition to hormonal therapy drug assignment, fracture and death information, the dataset has ten categorical covariates listed below. Five of these have more than one category, while the others are binary.

- Covariates with multiple categories:

Age groups: 65–69, 70–74, 75–79, 80–84, 85+

Race group: White, Black, Hispanic, Other

Income Per capita income at zipcode level: < 16,483, 16,483–19,612, 19,612–25,541, > 25,541, Missing

Education Percentage of population with less than high school education level: > 0.09, 0.06–0.09, 0.03–0.06, 0–0.03, Missing

Comorbidities Index (Klabunde et al. 2007): 0, 0–0.72, > 0.72

- Binary Covariates: Prior Fracture, Regular Anticonvulsants, Pregabalin/Gabapentin, IV Bisphosphonate, Low Income Subsidies.

We apply our method to explore whether AIs will give breast cancer survivors a higher chance of having a fracture. To be brief, we present only two plots: predicted cumulative incidence functions at the average values of other covariates (Fig. 10a) and log subdistribution hazard ratio of AIs versus Tamoxifen over time (Fig. 10b) from a rich variety of inference that the DDP method can make. The blue and red solid lines in Fig. 10a represent predicted CIFs for patients who assigned AIs and Tamoxifen throughout the 60 months period. The shaded areas bordered by the dashed lines show

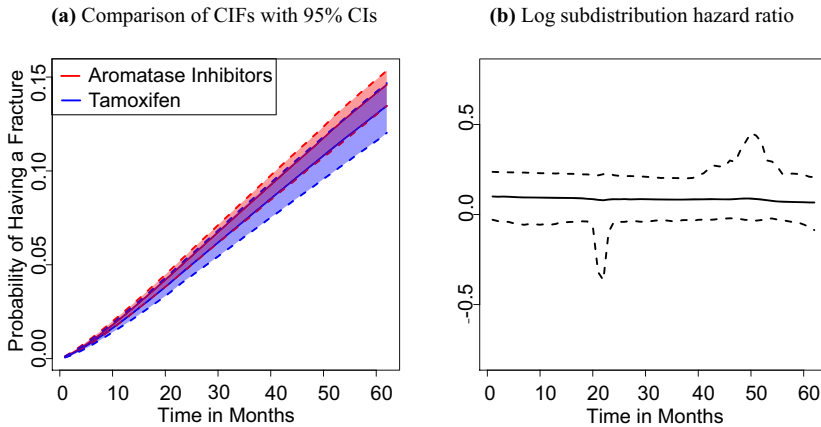


Fig. 10 Effect of AIs versus Tamoxifen

the 95% pointwise CIs of our estimates. Results from the DDP model do not support the hypothesis that AIs increase fracture risk. This is consistent with the conclusion reached using the Fine and Gray model, where the subdistribution hazard ratio for AI versus Tamoxifen is estimated as 1.09, 95% CI [0.99, 1.20] with a p value of 0.09. Results of the original study, using traditional methods and including propensity adjustments, are reported in Neuner et al. (2018).

9 Discussion

In this paper, we present nonparametric Bayesian regression methods for survival and competing risks data, applying this method to a breast cancer study. The R package “DPWeibull” containing an implementation of the proposed methods with functions for result visualization is available at CRAN (Shi 2019). Its data structure and syntax is similar to the “survival” package (Therneau 2015). The “DPWeibull” package provides reasonable computing speed with moderate sample size and number of covariates. On a computer with an Intel Core i7 2.9GHz processor and 16GB memory, it takes 163 s for 10,000 MCMC iterations with 2 covariates and 400 observations for the Weibull DDP competing risks model, yet the time is linear with respect to the sample size and the number of covariates.

The breast cancer dataset in Sect. 8 motivates us to extend our model in several directions in the future. First, drug assignment is treated as a binary covariate for now, using only the initial drug choice. In reality, long-term AIs takers may have higher chance of having a fracture than short-term AIs takers, which requires a model to accommodate cumulative effects. Second, in our example, event of secondary interest is defined as death without having a fracture. However, death will preclude fracture, but not vice versa. It would be meaningful to adapt our model to a semi-competing risks model to account for this fact. Third, some adverse events of breast cancer drugs, such as peripheral neuropathy can be precursors of fracture. It would be ideal to extend

the current model to a multistate model which addresses transfer probabilities between various states over time.

Acknowledgements The study was funded by the American Cancer Society (RSG-11-098-01-CPHPS to JMN).

References

- Allison P (2018) For causal analysis of competing risks, don't use fine and Gray's subdistribution method. <https://statisticalhorizons.com/for-causal-analysis-of-competing-risks>, statistical Horizons
- Chen DGD, Sun J, Peace KE (2012) Interval-censored time-to-event data: methods and applications. CRC Press, Boca Raton
- Cox DR (1972) Regression models and life-tables. *J R Stat Soc Ser B* 34(2):187–220
- de Castro M, Chen M, Zhang Y (2015) Bayesian path specific frailty models for multi-state survival data with applications. *Biometrics* 71(3):760–771
- De Iorio M, Müller P, Rosner GL, MacEachern SN (2004) An ANOVA model for dependent random measures. *J Am Stat Assoc* 99(465):205–215
- De Iorio M, Johnson WO, Müller P, Rosner GL (2009) Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics* 65(3):762–771
- Dignam JJ, Zhang Q, Kocherginsky M (2012) The use and interpretation of competing risks regression models. *Clin Cancer Res Off J Am Assoc Cancer Res* 18(8):2301–2308. <https://doi.org/10.1158/1078-0432.CCR-11-2097> [doi]
- Escobar MD, West M (1995) Bayesian density estimation and inference using mixtures. *J Am Stat Assoc* 90(430):577–588
- Fan X (2008) Bayesian nonparametric inference for competing risks data. The Medical College of Wisconsin, Ph.D. dissertation
- Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 94(446):496–509
- Ge M, Chen MH (2012) Bayesian inference of the fully specified subdistribution model for survival data with competing risks. *Lifetime Data Anal* 18(3):339–363
- Gelman A, Jakulin A, Pittau MG, Su YS (2008) A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2(4):1360–1383
- Greep NC, Giuliano AE, Hansen NM, Taketani T, Wang HJ, Singer FR (2003) The effects of adjuvant chemotherapy on bone density in postmenopausal women with early breast cancer. *Am J Med* 114(8):653–659
- Ishwaran H, Kogalur U (2018) Random forests for survival, regression, and classification (RF-SRC)
- Klabunde CN, Legler JM, Warren JL, Baldwin LM, Schrag D (2007) A refined comorbidity measurement algorithm for claims-based studies of breast, prostate, colorectal, and lung cancer patients. *Ann Epidemiol* 17(8):584–590
- Kottas A (2006) Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *J Stat Plan Inference* 136(3):578–596
- Lee KH, Dominici F, Schrag D, Haneuse S (2016) Hierarchical models for semicompeting risks data with application to quality of end-of-life care for pancreatic cancer. *J Am Stat Assoc* 111(515):1075–1095
- MacEachern SN (1999) Dependent nonparametric processes. In: *ASA proceedings of the section on Bayesian statistical science*, pp 50–55
- Murthy VH, Krumholz HM, Gross CP (2004) Participation in cancer clinical trials: race-, sex-, and age-based disparities. *J Am Med Assoc* 291(22):2720–2726
- Naskar M, Das K, Ibrahim JG (2005) A semiparametric mixture model for analyzing clustered competing risks data. *Biometrics* 61(3):729–737
- Neal RM (2000) Markov chain sampling methods for Dirichlet process mixture models. *J Comput Graph Stat* 9(2):249–265
- Neuner JM, Shi Y, Kong AL, Kamaraju S, Smith EC, Smallwood AJ, Laud PW, Charlson JA (2018) Fractures in a nationwide population-based cohort of users of breast cancer hormonal therapy. *J Cancer Surviv* 12(2):268–275. <https://doi.org/10.1007/s11764-017-0666-4>

- Scrucca L, Santucci A, Aversa F (2007) Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transpl* 40(4):381–387
- Sethuraman J (1994) A constructive definition of Dirichlet priors. *Stat Sin* 4:639–650
- Shi Y (2019) DPWeibull: Dirichlet process Weibull mixture model for survival data. <https://CRAN.R-project.org/package=DPWeibull>, r package version 1.4
- Shi Y, Martens M, Banerjee A, Laud P (2019) Low information omnibus (lio) priors for dirichlet process mixture models. *Bayesian Anal* 14(3):677–702. <https://doi.org/10.1214/18-BA1119>
- Snappin SM, Jiang Q, Iglewicz B (2012) Illustrating the impact of a time-varying covariate with an extended Kaplan–Meier estimator. *Am Stat* 59(4):301–307
- Sparapani RA, Logan BR, McCulloch RE, Laud PW (2016) Nonparametric survival analysis using Bayesian additive regression trees (BART). *Stat Med* 35(16):2741–2753
- Therneau TM (2015) A package for survival analysis in S. <https://CRAN.R-project.org/package=survival>, version 2.38
- Zhang Y, Chen MH, Ibrahim JG, Zeng D, Chen Q, Pan Z, Xue X (2014) Bayesian gamma frailty models for survival data with semi-competing risks and treatment switching. *Lifetime Data Anal* 20(1):76–105

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.