Biostatistics (2020) **00**, 00, *pp.* 1–16
doi:10.1093/biostatistics/kxaa008

A Bayesian nonparametric approach for evaluating the causal effect of treatment in randomized trials with semi-competing risks

YANXUN XU*

Department of Applied Mathematics and Statistics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

yanxun.xu@jhu.edu

DANIEL SCHARFSTEIN

Department of Biostatistics, Johns Hopkins University, 615 N Wolfe St, Baltimore, MD 21205, USA

PETER MÜLLER

Department of Mathematics, The University of Texas at Austin, 2515 Speedway, RLM 8.100, Austin, TX 78712, USA

MICHAEL DANIELS

Department of Statistics, University of Florida, Union Rd, Gainesville, FL 32603, USA

Summary

We develop a Bayesian nonparametric (BNP) approach to evaluate the causal effect of treatment in a randomized trial where a nonterminal event may be censored by a terminal event, but not vice versa (i.e., semi-competing risks). Based on the idea of principal stratification, we define a novel estimand for the causal effect of treatment on the nonterminal event. We introduce identification assumptions, indexed by a sensitivity parameter, and show how to draw inference using our BNP approach. We conduct simulation studies and illustrate our methodology using data from a brain cancer trial. The R code implementing our model and algorithm is available for download at https://github.com/YanxunXu/BaySemiCompeting.

Keywords: Bayesian nonparametrics; Brain cancer trial; Causal inference; Identification assumptions; Principal stratification; Sensitivity analysis.

1. Introduction

Semi-competing risks (Fine *and others*, 2001) occur in studies where observation of a nonterminal event (e.g., progression) may be pre-empted by a terminal event (e.g., death), but not vice versa. In randomized clinical trials to evaluate treatments of life-threatening diseases, patients are often observed for specific types of disease progression and survival. Often, the primary outcome is patient survival, resulting in data analyses focusing on the terminal event using standard survival analysis tools (Ibrahim *and others*, 2005). However, there may also be interest in understanding the causal effect of treatment on nonterminal

*To whom correspondence should be addressed.

outcomes such as progression, readmission, etc. An example is a randomized trial for the treatment of malignant brain tumors, where one of the important progression endpoints is based on deterioration of the cerebellum. An important feature of this progression endpoint is that it is biologically plausible that a patient could die without cerebellar deterioration. Thus, analyzing the effect of treatment on progression needs to account for the fact that progression is not well-defined after death.

Varadhan *and others* (2014) review models that have been proposed for analyzing semi-competing data. These models can be classified into two broad categories: models for the distribution of the observable data, e.g., cause-specific hazards, subdistribution functions (Fix and Neyman, 1951; Hougaard, 1999; Xu *and others*, 2010; Lee *and others*, 2015), and models for the distribution of the latent failure times (Robins, 1995a,b; Lin *and others*, 1996; Wang, 2003; Peng and Fine, 2007; Ding *and others*, 2009; Peng and Fine, 2012; Chen, 2012; Hsieh and Huang, 2012; Comment *and others*, 2019). Xu *and others* (2010) argued against the use of latent failure time models because the marginal distribution of the nonterminal event is hypothetical. This is because the joint distribution of the nonterminal event (Y_P) and terminal event (Y_D) is only identified on a wedge of R^2. Rather, they argued that "semi-competing risks data are better modeled using an illness-death compartment model," where "a subject can either transit directly to the terminal event or first to the nonterminal event and then to the terminal event." They proposed a Markov shared frailty model for the transition rates. Lee *and others* (2015) proposed a Bayesian semi-parametric extension, which focused on estimation of regression parameters, characterization of dependence between event times and prediction of event times for specific covariate profiles. The latent failure approaches of Fine *and others* (2001), Wang (2003), and Peng and Fine (2007) have focused on estimating regression parameters and estimating dependence between nonterminal and terminal event times using copula models. Robins (1995a,b) focused solely on estimating regression parameters and discusses causal interpretability. Recently, Comment *and others* (2019) proposed a casual estimand similar to the one we discuss here, but uses different models (i.e., parametric frailty models) and different causal assumptions (i.e., latent ignorability).

In this article, we are interested in estimating the causal effect of treatment on the nonterminal endpoint from a randomized trial generating semi-competing risk data. Using the potential outcomes framework (Rubin, 1974), we propose a principal stratification estimand (Frangakis and Rubin, 2002) to quantify the causal effect. Our estimand is a time-varying version of the survival average causal effect (see, e.g., Zhang and Rubin, 2003; Tchetgen Tchetgen, 2014), quantified on a relative risk scale. We introduce assumptions that utilize baseline covariates to identify this estimand from the distribution of the observable data and propose a Bayesian nonparametric (BNP) approach for modeling this distribution. An important feature of BNP models is their large support. For example, a Dirichlet process (Ferguson *and others*, 1973) location-scale mixture of normals (one of the most widely used BNP models), has full support on the space of absolutely continuous distributions (Lo, 1984). To handle covariates, our approach is based on the dependent Dirichlet process (DDP) prior introduced by MacEachern (1999).

The article is outlined as follows: Section 2 introduces the motivating brain tumor study. The formal definition of the causal estimand is introduced in Section 3. We introduce the BNP model in Section 4. A simulation study is summarized in Section 5. We analyze the brain tumor data in Section 6, and conclude with brief discussion in Section 7.

2. MOTIVATING BRAIN TUMOR STUDY

The methodology is motivated by a randomized and placebo-controlled Phase II trial for 222 recurrent gliomas patients, who were scheduled for tumor resection with recurrent malignant brain tumors (Brem *and others*, 1995). Eligible patients had a single focus of tumor in the cerebrum, had a Karnofsky score greater than 60, had completed radiation therapy, had not taken nitrosoureas within 6 weeks of enrollment, and had not had systematic chemotherapy within 4 weeks of enrollment. The data include 11 baseline

prognostic measures and a baseline evaluation of cerebellar function. The former includes age, race, Karnofsky performance score, local vs. whole brain radiation, percent of tumor resection, previous use of nitrosoureas, and tumor histology (glioblastoma, anapestic astrocytoma, oligodendrolioma, or other) at implantation. Patients were randomized to receive surgically implanted biodegradable polymer discs with or without 3.85% of carmustine. The follow-up duration was 1 year. Of the 219 patients with complete baseline measures, 204 were observed to die and 100 were observed to progress prior to death. Of the 15 patients who did not die, 4 were observed to have cerebellar progression. Our goal is to estimate the causal effect of treatment on time to cerebellar progression.

3. Causal estimand and identification assumptions

3.1. *Potential outcomes and causal estimand*

Let Y_P^z, Y_D^z, and C^z denote progression time, death time, and censoring time, under treatment z. Here, $z = 0, 1$ represents control and treatment group, respectively. All event times are log-transformed. Fundamental to our setting is that $Y_P^z \not> Y_D^z$ (i.e., progression cannot happen after death).

The causal estimand of interest is the function

$$\tau(u) = \frac{Pr[Y_P^1 < u \mid Y_D^0 \geq u, Y_D^1 \geq u]}{Pr[Y_P^0 < u \mid Y_D^0 \geq u, Y_D^1 \geq u]}, \tag{3.1}$$

where $\tau(\cdot)$ is a smooth function of u. Among patients who survive to time u under both treatments, this estimand contrasts the risk of progression prior to time u for treatment 1 relative to treatment 0, which is a causal effect in a subgroup defined by potential outcomes. This estimand is an example of a principal stratum causal effect (Frangakis and Rubin, 2002).

3.2. *Observed data*

Let Z denote treatment assignment and $\boldsymbol{X}$ denote a vector of the baseline covariates. Let $Y_P = Y_P^Z, Y_D = Y_D^Z$, and $C = C^Z$. Let $T_1 = Y_P \wedge Y_D \wedge C$, $\delta = I(Y_P < Y_D \wedge C)$, $T_2 = Y_D \wedge C$, and $\xi = I(Y_D < C)$ denote the observed event times and event indicators. The observed data for each patient are $\boldsymbol{O} = (T_1, T_2, \delta, \xi, Z, \boldsymbol{X})$. We assume that we observe n i.i.d. copies of $\boldsymbol{O}$. Throughout, variables subscripted by i will denote data specific to patient i.

3.3. *Identification assumptions*

We introduce the following four assumptions that are sufficient for identifying our causal estimand.

Assumption 1: Treatment is randomized, i.e.,

$$Z \perp (Y_P^z, Y_D^z, C^z, \boldsymbol{X}); \;\; z = 0, 1,$$

and $0 < Pr[Z = 1] < 1$.

This obviously holds by design in randomized trials as considered here.

Assumption 2: Censoring is noninformative in the sense that

$$C^z \perp (Y_P^z, Y_D^z) \,|\, \boldsymbol{X} = \boldsymbol{x}; \;\; z = 0, 1,$$

and $Pr[C^z > Y_P^z, C^z > Y_D^z | \boldsymbol{X} = \boldsymbol{x}] > 0$ for all $\boldsymbol{x}$.

Let $\lambda_{\boldsymbol{x}}^z$ and $G_{\boldsymbol{x}}^z$ denote the conditional hazard function and conditional distribution function of Y_D^z given $\boldsymbol{X} = \boldsymbol{x}$, respectively. Under Assumptions 1 and 2, $\lambda_{\boldsymbol{x}}^z$ and $G_{\boldsymbol{x}}^z$ are identified via the following formulae:

$$\lambda_{\boldsymbol{x}}^z(t) = \lim_{dt \to 0} \left\{ \frac{Pr[t \leq T_2 < t + \mathrm{d}t, \xi = 1 \mid T_2 \geq t, \boldsymbol{X} = \boldsymbol{x}, Z = z]}{\mathrm{d}t} \right\}$$

and

$$G_{\boldsymbol{x}}^z(t) = 1 - \exp\left\{ - \int_0^t \lambda_{\boldsymbol{x}}^z(s) \mathrm{d}s \right\}. \tag{3.2}$$

Furthermore, the conditional subdistribution function of Y_P^z given Y_D^z and $\boldsymbol{X} = \boldsymbol{x}$, $V_{\boldsymbol{x}}^z$, is identified via the following formula:

$$V_{\boldsymbol{x}}^z(s|t) = Pr[T_1 \leq s, \delta = 1 \mid T_2 = t, \xi = 1, \boldsymbol{X} = \boldsymbol{x}, Z = z], \tag{3.3}$$

where $s \leq t$. Together $G_{\boldsymbol{x}}^z(t)$ and $V_{\boldsymbol{x}}^z(s|t)$ identify the joint subdistribution $V_{\boldsymbol{x}}^z(s,t)$ for (Y_P^z, Y_D^z) given $\boldsymbol{X} = \boldsymbol{x}$.

Assumption 3: The conditional joint distribution function of (Y_D^0, Y_D^1) given $\boldsymbol{X} = \boldsymbol{x}$, $G_{\boldsymbol{x}}$, follows a Gaussian copula model, i.e.,

$$G_{\boldsymbol{x}}(v, w; \rho) = \Phi_{2,\rho}[\Phi^{-1}\{G_{\boldsymbol{x}}^0(v)\}, \Phi^{-1}\{G_{\boldsymbol{x}}^1(w)\}], \tag{3.4}$$

where Φ is a standard normal c.d.f. and $\Phi_{2,\rho}$ is a bivariate normal c.d.f. with mean 0, marginal variances 1, and correlation ρ. For fixed ρ, $G_{\boldsymbol{x}}$ is identified since $G_{\boldsymbol{x}}^0$ and $G_{\boldsymbol{x}}^1$ are identified. Similar assumptions have been used in the causal mediation literature (Daniels *and others*, 2012).

Assumption 4: Progression time under treatment z is conditionally independent of death time under treatment $1 - z$ given death time under treatment z and covariates $\boldsymbol{X} = \boldsymbol{x}$, i.e.,

$$Y_P^z \perp Y_D^{1-z} \mid Y_D^z, \boldsymbol{X} = \boldsymbol{x}; \;\; z = 0, 1.$$

Under Assumptions 1–4, $\tau(\cdot)$ is identified from the distribution of the observed data as follows:

$$\tau(u) = \frac{\int_{\boldsymbol{x}} \int_{s<u} \int_{v \geq u} \int_{t \geq u} \mathrm{d}V_{\boldsymbol{x}}^1(s|t) \mathrm{d}G_{\boldsymbol{x}}(v,t) \mathrm{d}K(\boldsymbol{x})}{\int_{\boldsymbol{x}} \int_{s<u} \int_{v \geq u} \int_{t \geq u} \mathrm{d}V_{\boldsymbol{x}}^0(s|t) \mathrm{d}G_{\boldsymbol{x}}(v,t) \mathrm{d}K(\boldsymbol{x})}, \tag{3.5}$$

where $K(\boldsymbol{x})$ is the empirical distribution of $\boldsymbol{X}$.

4. BAYESIAN REGRESSION MODEL

In this section, we propose a BNP survival regression model on the unknown conditional (on $\boldsymbol{X} = \boldsymbol{x}$) distribution of (Y_P^z, Y_D^z). However, any alternative Bayesian survival regression models could be implemented (Hanson and Johnson, 2002; Gelfand and Kottas, 2003; Zhou and Hanson, 2018; Sparapani *and others*, 2016; Xu *and others*, 2019); however, the first three are restrictive in how covariates are entered and the fourth one is semi-parametric.

4.1. *Dependent Dirichlet process—Gaussian process prior*

We start with a review of the Dirichlet process (DP) as a prior for an unknown distribution and step by step extend it to the dependent Dirichlet process—Gaussian process prior.

The DP prior has been widely used as a prior model for a random unknown probability distribution. We write $H \sim \text{DP}(\alpha, H_0)$ if a random distribution H of a J-dimensional random vector $\boldsymbol{V}$ follows a DP prior, where α is known as the total mass parameter and H_0 is known as the base measure. Sethuraman (1994) provides a constructive definition of a DP, where $dH(\boldsymbol{v}) = \sum_{h=1}^{\infty} w_h \delta_{\boldsymbol{\theta}_h}(\boldsymbol{v})$, $w_h = \nu_h \prod_{l<h}(1-\nu_l)$, $\boldsymbol{\theta}_h = \{\theta_{h1}, \ldots, \theta_{hJ}\} \overset{i.i.d.}{\sim} H_0$ and $\nu_h \overset{i.i.d.}{\sim} \text{Be}(1, \alpha)$. In many applications, the discrete nature of H is not appropriate. A DP mixture model extends the DP model by replacing each point mass $\delta_{\boldsymbol{\theta}_h}(\cdot)$ with a continuous kernel. For example, a DP mixture of normals takes the form: $\text{d}H(\boldsymbol{v}) = \sum_h w_h \phi(\boldsymbol{v}; \boldsymbol{\theta}_h, \boldsymbol{\Sigma})\, \text{d}\boldsymbol{v}$, where $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{S})$ is the density function of a multivariate normal random vector with mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $\boldsymbol{S}$.

To introduce a prior on the conditional (on covariates $\boldsymbol{X} = \boldsymbol{x}$) distribution ($H_{\boldsymbol{x}}$) of $\boldsymbol{V}$, the DP mixture model has been extended to a dependent DP (DDP) by replacing $\boldsymbol{\theta}_h$ in each term with $\boldsymbol{\theta}_h(\boldsymbol{x}) = \{\theta_{h1}(\boldsymbol{x}), \ldots, \theta_{hJ}(\boldsymbol{x})\}$, which is a multivariate stochastic process indexed by $\boldsymbol{x}$. A DDP mixture of normals takes the form:

$$\text{d}H_{\boldsymbol{x}}(\boldsymbol{v}) = \sum_h w_h \phi(\boldsymbol{v}; \boldsymbol{\theta}_h(\boldsymbol{x}), \boldsymbol{\Sigma})\, \text{d}\boldsymbol{v}. \tag{4.6}$$

To complete the prior specification, we need to posit a stochastic process prior for $\{\boldsymbol{\theta}_h(\boldsymbol{x}) : \boldsymbol{x}\}$. A common specification are independent Gaussian process (GP) priors (MacEachern, 1999; Xu *and others*, 2016) on $\{\theta_{hj}(\boldsymbol{x}) : \boldsymbol{x}\}$. A GP prior is specified such that for all $L \geq 1$ and $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L)$, the distribution of $(\theta_{hj}(\boldsymbol{x}_1), \ldots, \theta_{hj}(\boldsymbol{x}_L))$ follows a multivariate normal distribution with mean vector $(\mu_{hj}(\boldsymbol{x}_1), \ldots, \mu_{hj}(\boldsymbol{x}_L))$ and $(L \times L)$ covariance matrix where the (l, l') entry is $R_j(\boldsymbol{x}_l, \boldsymbol{x}_{l'})$. We write $\{\theta_{hj}(\boldsymbol{x}) : \boldsymbol{x}\} \sim \text{GP}(\mu_{hj}(\cdot), R_j(\cdot, \cdot))$. For an extensive review of the GP priors, see Rasmussen and Williams (2006) and MacKay (1999). We model the mean function $\mu_{hj}(\cdot)$ as a linear regression on covariates $\mu_{hj}(\boldsymbol{x}_l; \boldsymbol{\beta}_{hj}) = \boldsymbol{x}_l \boldsymbol{\beta}_{hj}$, with covariance process specified as

$$R_j(\boldsymbol{x}_l, \boldsymbol{x}_{l'}) = \exp\left\{-\sum_{d=1}^{D}(x_{ld} - x_{l'd})^2\right\} + \delta_{ll'}\epsilon^2, \tag{4.7}$$

where D is the dimension of the covariate vector, $\delta_{ll'} = I(l = l')$ and ϵ is a small constant (e.g., $\epsilon = 0.1$) used to ensure that the covariance function is positive definite. To ensure a reasonable covariance structure, continuous covariates should be standardized to have mean 0 and variance 1. More flexible covariance functions can be considered if desired. Additional priors are introduced on the $\boldsymbol{\beta}_j$'s and $\boldsymbol{\Sigma}$, the details of which are discussed in Appendix A.1. We write $\{H_{\boldsymbol{x}}\} \sim \text{DDP-GP}(\alpha, \Sigma, \text{GP}(\mu_j(\cdot), R_j(\cdot, \cdot)), j = 1, \ldots, J)$.

4.2. *Application to semi-competing risks data*

Separately for each treatment group z, we posit independent DDP-GP's on the unknown conditional (on $\boldsymbol{X} = \boldsymbol{x}$) probability measure ($H^z_{\boldsymbol{x}}$) of (Y^z_P, Y^z_D). Since $V^z_{\boldsymbol{x}}(s|t) = H^z_{\boldsymbol{x}}(Y^z_P \leq s, Y^z_P \leq Y^z_D | Y^z_D = t)$ $(s \leq t)$ and $G^z_{\boldsymbol{x}}(t) = H^z_{\boldsymbol{x}}(Y^z_D \leq t)$, the prior on $H^z_{\boldsymbol{x}}$ induces priors on $V^z_{\boldsymbol{x}}(s|t)$ and $G^z_{\boldsymbol{x}}(t)$ (identified under Assumptions 1 and 2) and together with the Gaussian copula for $G_{\boldsymbol{x}}$ implies a prior on the estimand $\tau(\cdot)$. The prior on $H^z_{\boldsymbol{x}}$ also induces priors on non-identified quantities which have no impact on our analysis. More specifics about our prior are presented in Appendix A.1.

Before transitioning to the posterior sampling algorithm, note that the relevant portion of the observed data likelihood for individual i, with data $O_i = (T_{1i}, T_{2i}, \delta_i, \xi_i, Z_i, \boldsymbol{X}_i)$ is

$$\begin{aligned} L(O_i) &= \left\{ \mathrm{d}V_{\boldsymbol{X}_i}^{Z_i}(T_{1i}|T_{2i}) \mathrm{d}G_{\boldsymbol{X}_i}^{Z_i}(T_{2i}) \right\}^{\delta_i \xi_i} \left\{ \int_{t>T_{2i}} \mathrm{d}V_{\boldsymbol{X}_i}^{Z_i}(T_{1i}|t) \mathrm{d}G_{\boldsymbol{X}_i}^{Z_i}(t) \right\}^{\delta_i(1-\xi_i)} \times \\ &\quad \left\{ \left(1 - V_{\boldsymbol{X}_i}^{Z_i}(T_{2i}|T_{2i})\right) \mathrm{d}G_{\boldsymbol{X}_i}^{Z_i}(T_{2i}) \right\}^{(1-\delta_i)\xi_i} \left\{ \int_{t>T_{2i}} \left(1 - V_{\boldsymbol{X}_i}^{Z_i}(T_{2i}|t)\right) \mathrm{d}G_{\boldsymbol{X}_i}^{Z_i}(t) \right\}^{(1-\delta_i)(1-\xi_i)} \\ &= \left\{ \mathrm{d}H_{\boldsymbol{X}_i}^{Z_i}(T_{1i}, T_{2i}) \right\}^{\delta_i \xi_i} \left\{ \int_{t>T_{2i}} \mathrm{d}H_{\boldsymbol{X}_i}^{Z_i}(T_{1i}, t) \right\}^{\delta_i(1-\xi_i)} \left\{ \int_{s>T_{1i}} \mathrm{d}H_{\boldsymbol{X}_i}^{Z_i}(s, T_{2i}) \right\}^{(1-\delta_i)\xi_i} \times \\ &\quad \left\{ \int_{t>T_{2i}} \int_{s>T_{2i}} \mathrm{d}H_{\boldsymbol{X}_i}^{Z_i}(s, t) \right\}^{(1-\delta_i)(1-\xi_i)}. \end{aligned}$$

We include the second equality because it allows us to see that, using data augmentation to replace the integrals, the joint full data likelihood is $\prod_{i=1}^{n} \mathrm{d}H_{\boldsymbol{X}_i}^{Z_i}(Y_{Pi}, Y_{Di})$. This will allow us to use existing posterior simulation techniques for DDP-GP models.

4.3. *Posterior simulation*

The details of the Markov chain Monte Carlo (MCMC) algorithm are presented in Appendix A.2. Here, we focus on individuals assigned to treatment z and suppress the dependence of the notation on z. As noted above, the MCMC implementation is based on the full data likelihood. While $\mathrm{d}H_{\boldsymbol{X}_i}$ is an infinite mixture of normals, we approximate it by a finite mixture with $K < \infty$ components. This finite mixture model for (Y_{Pi}, Y_{Di}) can be replaced by a hierarchical model where (1) γ_i is a latent variable that selects mixture component h ($h = 1, \ldots, K$) with probability w_h (properly normalized to handle the finite number of mixture components) and (2) given γ_i, the pair (Y_{Pi}, Y_{Di}) follows a multivariate normal distribution with mean $\boldsymbol{\theta}_{\gamma_i}(\boldsymbol{X}_i)$ and variance $\boldsymbol{\Sigma}$.

Posterior simulation is based on this hierarchical model characterization. Importantly, all of the full conditionals in the MCMC algorithm have a closed form representation. Details of the MCMC posterior simulation can be found in Appendix A.2.

5. SIMULATION STUDIES

5.1. *Simulation setup*

We considered three simulation scenarios to evaluate the performance of our proposed approach with 500 repeated simulations for each scenario. We generated $Z \sim \mathrm{Bern}(0.5)$. Independently of Z, we generated two independent covariates X_1 and X_2, where X_1 followed a truncated normal distribution with mean 4.5, variance 1, and truncation interval $(2, 7.5)$, and $X_2 \sim \mathit{Bern}(0.4)$. For the first two simulation scenarios, we simulated progression time and death time on the log scale as follows:

$$\begin{aligned} Y_P^z &= 1.5z + 0.6X_1 + 2X_2 + \epsilon, \\ Y_D^z &= 4z + 0.3X_1 + X_2 + \nu. \end{aligned}$$

In Scenario 1, we assumed (ϵ, ν) followed a bivariate normal distribution with mean $(0, 1.5)'$, marginal variances $S_{11} = S_{22} = 1$, and correlation $S_{12} = 0.75$. In Scenario 2, we assumed (ϵ, ν) to be a scaled multivariate t distribution with degree of freedom $\nu = 3$, mean $(0, 1.5)'$, marginal variance

Table 1. *(a) For each scenario, mean and standard deviation of RMSE across 500 simulated data sets under the proposed BNP method, the naive Bayesian method (Naive), and the LinearDDP method. Bold values indicate that the proposed BNP yields the smallest mean RMSE when* $\rho = 0.5$. *(b) Means and standard deviations of RMSE for estimating* $\hat{\tau}(u)$ *across 500 simulations in three scenarios under the proposed BNP approach, the naive Bayesian method (Naive), and the LinearDDP method, respectively*

Scenario	$Z = 0$			$Z = 1$		
	BNP	Naive	LinearDDP	BNP	Naive	LinearDDP
1	0.012 (0.007)	0.013 (0.007)	0.014 (0.007)	0.012 (0.006)	0.013 (0.007)	0.015 (0.008)
2	0.042 (0.022)	0.088 (0.032)	0.063 (0.020)	0.019 (0.007)	0.073 (0.035)	0.058 (0.023)
3	0.012 (0.006)	0.013 (0.007)	0.014 (0.007)	0.012 (0.007)	0.014 (0.007)	0.016 (0.008)

(a)

Scenario	$\rho = 0.2$			$\rho = 0.5$		
	BNP	Naive	LinearDDP	BNP	Naive	LinearDDP
1	0.286 (0.087)	0.328 (0.126)	0.332 (0.087)	**0.059 (0.035)**	0.073 (0.051)	0.091 (0.016)
2	0.277 (0.128)	0.493 (0.250)	0.449 (0.189)	**0.090 (0.062)**	0.199 (0.169)	0.179 (0.123)
3	0.106 (0.032)	0.105 (0.038)	0.115 (0.043)	**0.033 (0.016)**	0.035 (0.021)	0.043 (0.027)

Scenario	$\rho = 0.8$		
	BNP	Naive	LinearDDP
1	0.185 (0.037)	0.207 (0.047)	0.181 (0.042)
2	0.261 (0.070)	0.243 (0.111)	0.203 (0.084)
3	0.086 (0.028)	0.097 (0.034)	0.086 (0.035)

(b)

$S_{11} = S_{22} = 1$, and correlation $S_{12} = 0.75$. Scenario 3 explored performance under a nonlinear covariate effect specification on progression and death times. We generated $Y_P^z = 1.5z + 0.6X_i + 2X_2 + \epsilon$ and $Y_D^z = 4z + 0.3X_1 + X_2 + 0.5\sqrt{X_1} + \nu$, with (ϵ, ν) following the same distribution as in Scenario 1.

In all scenarios, the censoring time C^z on the log scale was generated independently according to a Unif(8, 10) distribution. In Scenario 1, 56.6% of the patients' deaths and progressions were both observed ($\delta = \xi = 1$), 2% of the patients' deaths and progressions were both censored ($\delta = \xi = 0$), 36.4% of the patients' deaths were observed and progressions were censored ($\delta = 0, \xi = 1$), 5% of the patients' deaths were censored and progressions were observed ($\delta = 1, \xi = 0$). In Scenario 2, 55.8% of the patients' deaths and progressions were both observed, 4.8% of the patients' deaths and progressions were both censored, 33.6% of the patients' deaths were observed and progressions were censored, 5.8% of the patients' deaths were censored and progressions were observed. In Scenario 3, these percentages were 69.4%, 3.4%, 10.6%, and 16.6%, respectively. For the joint distribution of Y_D^0 and Y_D^1 in (3.4), we set $\rho = \rho^o = 0.5$ in the Gaussian copula as the truth. We generated $(Z_i, X_{1i}, X_{2i}, Y_{Pi}, Y_{Di}, C_i)$ for $n = 500$ independent patients and then coarsened to $(Z_i, X_{1i}, X_{2i}, T_{i1}, T_{i2}, \delta_i, \xi_i)$.

To explore sensitivity of $\tau(u)$ with respect to ρ, we conducted inference for $\tau(u)$ under several values of $\rho = 0.2, 0.5, 0.8$. For all three scenarios, we specified hyperparameters as described in Appendix A.1.

For comparative purposes, we implemented two alternative models. The first one is a naive Bayesian (Naive) model by assuming that the conditional probability measure $(H_{\boldsymbol{x}}^z)$ of (Y_P^z, Y_D^z) follows a multivariate normal distribution with mean $(\boldsymbol{x}'\boldsymbol{\beta}_P^z, \boldsymbol{x}'\boldsymbol{\beta}_D^z)$ and variance–covariance matrix $\boldsymbol{S}^z$, with conjugate multivariate normal priors on $\boldsymbol{\beta}_P^z$ and $\boldsymbol{\beta}_D^z$ and an inverse Wishart prior on $\boldsymbol{S}^z$ (i.e., $\boldsymbol{\beta}_P^z \sim \text{MN}(\boldsymbol{0}, \tau_P^2 \boldsymbol{I})$, $\boldsymbol{\beta}_P^z \sim \text{MN}(\boldsymbol{0}, \tau_D^2 \boldsymbol{I})$, and $\boldsymbol{S}^z \sim$ Inverse Wishart(ν, Ψ), $z = 0, 1$). The second one is the linear dependent DDP (LinearDDP) model proposed in De Iorio *and others* (2009), which simplifies the proposed BNP model by assuming that $\boldsymbol{\theta}_h(\boldsymbol{x})$ in (4.6) is a linear regression on $\boldsymbol{x}$, instead of a Gaussian process prior on $\boldsymbol{\theta}_h(\boldsymbol{x})$ used in the proposed BNP model.

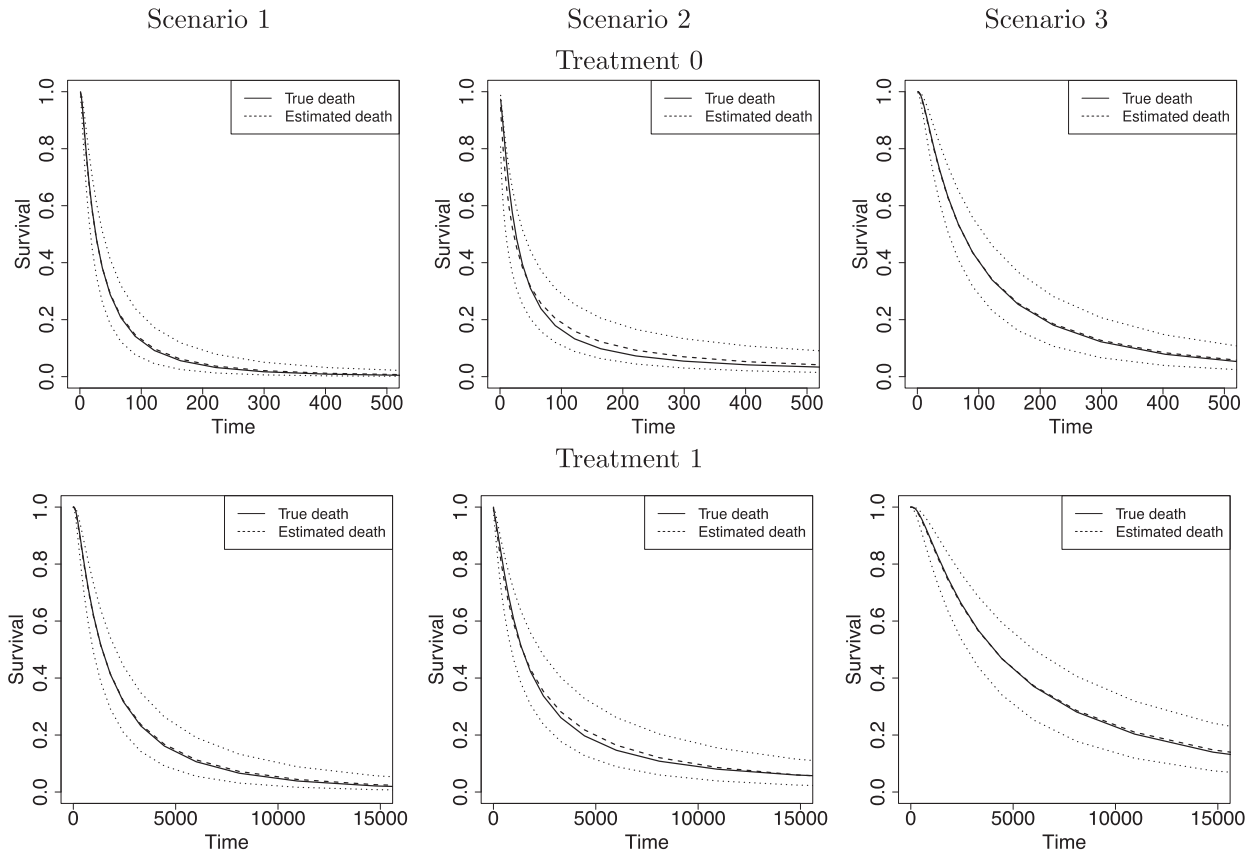

Fig. 1. For each simulation scenario and by treatment group (first and second rows refer to treatments 0 and 1, respectively), the true survival functions (solid line), the posterior mean survival functions averaged over simulated data sets (dashed line), and 95% point-wise credible intervals (computed using quantiles) averaged over simulated data sets (dotted lines). Survival times are on the original scale (days).

For each analysis, we ran 5000 MCMC iterations with an initial burn-in of 2000 iterations and a thinning factor of 10. The convergence diagnostics using the R package `coda` show no evidence of practical convergence problems.

5.2. *Simulation results*

We first report on the performance in terms of recovering the true treatment-specific marginal survival functions for time to death. For the BNP approach, Figure 1 shows, for each of the three simulation scenarios and by treatment group (first and second rows refer to treatments 0 and 1, respectively), the true survival functions (solid line), the posterior mean survival functions averaged over simulated data sets (dashed line), and 95% point-wise credible intervals (computed using quantiles) averaged over simulated data sets (dotted lines) on the original time scale (days). As another metric of performance, we computed, for each simulated data set, the root mean squared error (RMSE) taken as the square root of the average of the squared errors at 34 equally spaced grid points in log-scaled time interval $(0, 10)$. For each scenario, Table 1(a) summarizes the mean and standard deviation of RMSE across the 500 simulated data sets. Both Figure 1 and Table 1(a) show that our proposed BNP procedure performs well, for each of the three scenarios, in terms of recovering the true survival function.

Table 1(a) also shows the mean and standard deviation of RMSE for the Naive and the LinearDDP models. In Scenario 1, the two models match the true simulation model, thereby yielding comparable

results as the proposed BNP model. In contrast, the Naive and the LinearDDP models perform worse than the BNP model in Scenario 2 when the fitted model does not match the simulation truth. In Scenario 3, the BNP model performs slightly better than the Naive and the LinearDDP models. Overall, the proposed BNP model is more robust compared to the Naive and the LinearDDP models.

Evaluation of $\tau(\cdot)$ requires evaluation of $G^z_{\boldsymbol{x}}$ as the second marginal under $H^z_{\boldsymbol{x}}$. Expression (3.5) allows us now to estimate $\tau(u)$. Both the numerator and denominator can be evaluated as functionals of the currently imputed random probability measure $H^z_{\boldsymbol{x}}(Y^z_P, Y^z_D)$ of time to log progression Y^z_P and time to log death Y^z_D under treatment z, marginalizing with respect to the empirical distribution of covariates $\boldsymbol{x}$'s. Each iteration of the posterior MCMC simulation evaluates a point-wise estimate and we estimate the posterior mean of $\tau(u)$ as $\hat{\tau}(u) = \text{Mean}\{\tau(u) \mid \textit{data}\}$ across iterations. We also report the mean RMSE in estimating the $\hat{\tau}(u)$ by averaging over 500 repeated simulations under the proposed BNP, the Naive, and the LinearDDP models. Table 1(b) summarizes the results.

Figure 2 shows $\hat{\tau}(u)$ versus u in the three scenarios, respectively, using $\rho = 0.2, 0.5$ and 0.8. As shown in Figure 2, in all three scenarios, when $\rho = \rho^o = 0.5$, the estimates under the proposed BNP model reliably recover the simulated true $\tau(u)$ and avoid the excessive bias seen with other ρ values. This agrees with the results reported in Table 1(b) that $\rho = 0.5$ always yields the smallest mean RMSE in all three scenarios. Furthermore, when $\rho = 0.5$, the proposed BNP model has smaller mean RMSE compared to the Naive and the LinearDDP models. When $\rho = 0.2$ or 0.8, the BNP model performs better or comparable to the Naive model in terms of providing smaller mean RMSE and variability of RMSE across simulations.

6. Brain tumor data analysis

An initial analysis of the brain tumor death outcome using Kaplan–Meier is given in Figure 3, indicating that the treatment group has higher estimated survival probabilities. The estimated difference at 365 days is 2.6% (95% CI −8.1% to 13.3%). Figure 3 plots the estimated posterior survival curves for treatment and control groups marginalized over the distribution of covariate with 95% credible intervals; panels (a), (b), and (c) display the results for the BNP, Naive, and LinearDDP approaches, respectively. Using the BNP approach, the estimated posterior difference in survival at 365 days is 6.2% (95% CI −1.2% to 13.3%). For the Naive approach, the estimated posterior difference in survival at 365 days is 8.4% (95% CI 0.2% to 17.9%). The LinearDDP approach estimated the posterior difference in survival at 365 days to be 9.9% (95% CI 0.9% to 20.8%). The BNP approach produces comparable or higher treatment-specific estimates of survival and greater treatment differences than Kaplan–Meier. In contrast, the Naive and LinearDDP approaches produce comparable or lower (higher) estimate of survival for the control (treatment) group than Kaplan–Meier. Comparatively speaking, the Naive and LinearDDP approaches produce lower treatment-specific posterior estimates of survival than the BNP approach. When we compare the fit to the observed survival data of the three approaches using the log-pseudo marginal likelihood (LPML; (Geisser and Eddy, 1979), a leave-one-out cross-validation statistic, we see the BNP performs better. Specifically, the LPML for the treatment arm is −144, −161, −147 for the BNP, Naive, and LinearDDP approaches, respectively. The corresponding numbers for the control arm are −137, −174, and −139.

For the BNP (panel (a)), Naive (panel (b)), and LinearDDP (panel (c)) approaches, Figure 4 plots the posterior estimates (along with point-wise 95% credible intervals) of the causal estimand $\tau(u)$ versus u for three choices of ρ, 0.2, 0.5, and 0.8. Except near $u = 0$, there are no appreciable differences between the two approaches. In addition, the results are insensitive to choice of ρ. Overall, this analysis shows that there is a lower estimated risk of progression for treatment versus of control at all time points, except near zero. However, there is appreciable uncertainty, characterized by wide posterior credible intervals, that precludes more definitive conclusions about the difference between treatment groups with regards to progression. When we compare the fit to the observed survival *and* progression data of the BNP and Naive

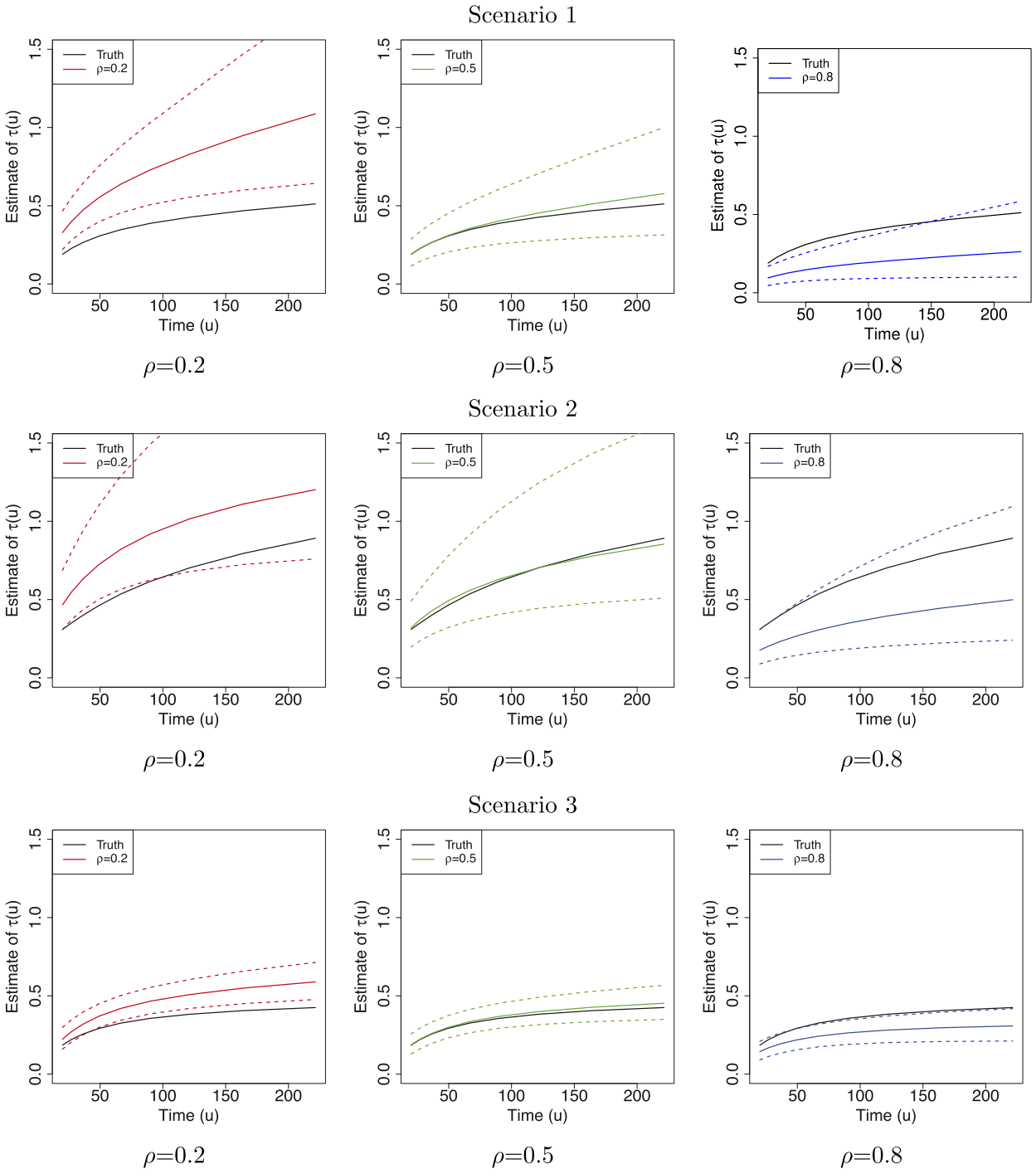

Fig. 2. The posterior estimates (dashed lines) of $\tau(u)$ versus u on the original scale (days) for the three scenarios using $\rho = 0.2, 0.5, 0.8$, respectively. The solid lines represent the simulation truth using $\rho^o = 0.5$. The dotted lines represent 95% point-wise credible intervals (computed using quantiles) averaged over simulated datasets.

approaches using LPML, we see that the approaches perform comparably. Specifically, the LPML for the treatment arm is -227, -232, and -235 for the BNP, Naive, and LinearDDP approaches, respectively. The corresponding numbers for the control arm are -215, -214, and -219.

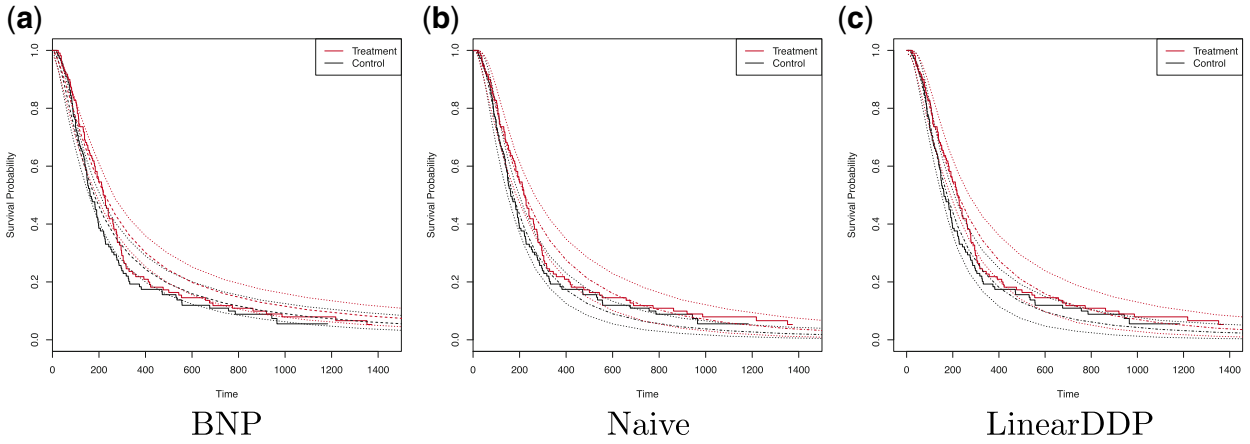

Fig. 3. The dashed lines in (a) represent the estimated posterior mean survival curves for the proposed BNP method. The dotdash lines in (b) and (c) represent the estimated posterior mean survival curves for the Naive method and LinearDDP method, respectively. In all figures, the solid lines represent the Kaplan–Meier curves of the observed survival data in control and treatment groups, and the dotted lines represent 95% point-wise credible intervals of the posterior estimated survival curves. Survival times are on the original scale (days).

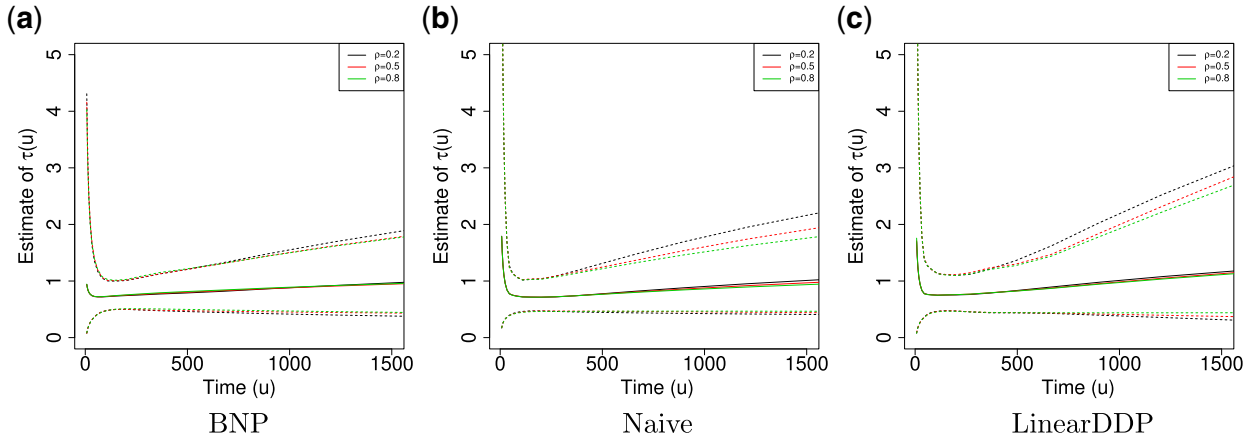

Fig. 4. Posterior estimated $\tau(u)$ versus u on the original scale (days) in brain tumor data analysis for different ρ's under the proposed BNP method, the Naive method, and the LinearDDP method, respectively. The solid lines represent the posterior estimated $\tau(u)$, and the dashed lines represent 95% point-wise credible intervals. (a) BNP, (b) Naive, and (c) LinearDDP.

7. Discussion

In this article, we proposed a causal estimand for characterizing the effect of treatment on progression in a randomized trials with a semi-competing risks data structure. We introduced a set of identification assumptions, indexed by a non-identifiable sensitivity parameter that quantifies the correlation between survival under treatment and survival under control. Selecting a range of the sensitivity parameter ρ in a specific trial will depend on clinical considerations. For example, in trial of a biomarker targeted therapy, one might expect weaker correlation, since survival under control might be primarily determined by co-morbidities and the survival under treatment might be more determined by the presence of the targeted molecular aberration. For example, a recent FDA-approved drug LOXO-101 (Hyman *and others*, 2017) targeting NTRK fusion has an overall response rate of 78% in the treatment group, while only 10% in the

control group. In contrast, for some chemotherapies, the same factors that impact survival under control may equally impact survival under treatment, e.g., co-morbidities, social support (Kaufman *and others*, 2015). Then we would suggest a medium or high ρ, say $\rho = 0.8$. Fortunately, the sensitivity parameter is bounded between -1 and 1 and, in most settings, should be positive; a range should be selected in close collaboration with subject matter experts.

We proposed a flexible BNP approach for modeling the distribution of the observed data. Since the causal estimand is a functional of the distribution of the observed data and ρ, we draw inference about it using posterior summarization. Our procedure can easily be extended to accommodate a prior distribution on ρ, which will allow for integrated inference. Our procedure also allows for posterior inferences about other identified causal contrasts such as the distribution of survival under treatment versus under control. The procedure can also be used for predictive inference for patients with specific covariate profiles.

ACKNOWLEDGMENTS

The authors would like to thank Drs Henry Brem and Steven Piantadosi for providing access to data from the brain cancer trial. *Conflict of Interest:* None declared.

FUNDING

This research is supported by National Institute Health NIH CA183854 and NIH GM 112327, and National Science Foundation NSF1918854.

APPENDIX A

A.1 Determining prior hyperparameters

As priors for $\boldsymbol{\beta}^z_{hj}$ in the GP mean function, we assume $\boldsymbol{\beta}^z_{hj} \sim N(\boldsymbol{\beta}^z_{0j}, \boldsymbol{\Lambda}^z_{0j})$. We assume $\boldsymbol{\Lambda}^z_{0j} \sim$ Inverse-Wishart$(\lambda^z_0, \boldsymbol{\Psi}^z)$, where $E[\boldsymbol{\Lambda}^z_{0j}] = \frac{\boldsymbol{\Psi}^z}{\lambda^z_0 - 3}$. The precision parameter α in the DDP is assumed to be distributed Ga(λ_1, λ_2).

In applications of Bayesian inference with small to moderate sample sizes, a critical step is to fix values for all hyperparameters $\omega = \{\boldsymbol{\beta}^z_{0j}, \boldsymbol{\Sigma}^z_{0j}, \lambda^z_0, \Phi^z, z = 0, 1, j = 0, 1, \lambda_1, \lambda_2\}$. Inappropriate information could be introduced by improper numerical values, leading to inaccurate posterior inference. We use an empirical Bayes method to obtain $\boldsymbol{\beta}^z_0 = (\boldsymbol{\beta}^z_{01}, \boldsymbol{\beta}^z_{02})$ by fitting a bivariate normal distribution for responses of patients under treatment z, $\boldsymbol{Y} \mid Z = z \sim N(\boldsymbol{x}\boldsymbol{\beta}^z_0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_0})$. For $\boldsymbol{\Sigma}^z_{0j}$, we assume a diagonal matrix with the diagonal values being 10. After an empirical estimate of $\hat{\Sigma}^z$ is computed, we tune λ^z_0 and Φ^z so that the prior mean of $\boldsymbol{\Sigma}^z$ matches the empirical estimate, $\lambda^z_0 = 4$ and $\Phi^z = \hat{\boldsymbol{\Sigma}}^z$. Finally, we assume $\lambda_1 = \lambda_2 = 1$.

A.2 MCMC computational details

Unless required for clarity, we suppress dependence of the notation on treatment z. Here, j is used to denote endpoint ($j = 1$ for progression and $j = 2$ for death). We define

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}.$$

Let $A_h = \{i : \gamma_i = h\}$ and $n_h = |A_h|$, $\boldsymbol{Y}_i = (Y_{Pi}, Y_{Di})$, $\boldsymbol{\theta}^{\star}{}_{hj} = (\theta_{hj}(\boldsymbol{X}_1), \ldots, \theta_{hj}(\boldsymbol{X}_n))$ $(h = 1, \ldots, K)$, $\boldsymbol{X}$ is an $n \times D$ matrix where the ith row contains the D-dimensional covariate vector $\boldsymbol{X}_i$ for the ith patient, $\boldsymbol{R}_j$ is an $n \times n$ matrix where the (i, i') entry is $R_j(\boldsymbol{X}_i, \boldsymbol{X}_{i'})$, $\boldsymbol{U}_h$ is an $n_h \times n$ matrix where the kth row refers to

the kth patient in A_h, the ith column refers to patient i and (k, i) element is the indicator that the patient in kth row is the same as the patient in the ith column, $\boldsymbol{I}_n$ is an $n \times n$ identity matrix, $\tilde{\boldsymbol{Y}}_{hj} = \{\tilde{Y}_{ji} : \gamma_i = h\}$, where $\tilde{Y}_{1i} = Y_{Pi} - \frac{\sigma_{12}}{\sigma_2^2}\left(Y_{Di} - \theta_{\gamma_i 2}(\boldsymbol{X}_i)\right)$ and $\tilde{Y}_{2i} = Y_{Di} - \frac{\sigma_{21}}{\sigma_1^2}\left(Y_{Pi} - \theta_{\gamma_i 1}(\boldsymbol{X}_i)\right)$, $\tilde{\sigma}_1^2 = \sigma_1^2 - \frac{\sigma_{12}\sigma_{21}}{\sigma_2^2}$, and $\tilde{\sigma}_2^2 = \sigma_2^2 - \frac{\sigma_{12}\sigma_{21}}{\sigma_1^2}$.

For $z = 0, 1$, we iterate through the following six updating steps:

(1) Update w_h

$$v_h \sim Beta(1 + n_h, \alpha + \sum_{j>h} n_j),$$

where $n_h = \sum_{i=1}^{n_k} I(\gamma_i = h)$ is the number of observations such that $\gamma_i = h$. Then, $w_1 = v_1$ and $w_h = v_h \prod_{j<h}(1 - v_j)$.

(2) Update α
Assuming that $\alpha \sim \text{Ga}(\lambda_1, \lambda_2)$,

$$\alpha \sim \text{Ga}(\lambda_1 + H - 1, \lambda_2 - \sum_{h=1}^{H-1} \log(1 - v_h)),$$

where v_h is generated from Step 1.

(3) Update $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} \mid \cdot \sim \text{Inverse Wishart}\left(\lambda_0 + n, \boldsymbol{\Psi} + \sum_{h=1}^{K} \sum_{i:\gamma_i=h} (\boldsymbol{Y}_i - \boldsymbol{\theta}_h(\boldsymbol{X}_i))(\boldsymbol{Y}_i - \boldsymbol{\theta}_h(\boldsymbol{X}_i))'\right)$$

(4) Update $\boldsymbol{\theta}^{\star}_{hj}, j = 1, 2$

$$\begin{aligned} p(\boldsymbol{\theta}^{\star}_{hj} \mid \cdot) &\propto p(\boldsymbol{\theta}^{\star}_{hj}) \prod_{i:\gamma_i=h} p(\boldsymbol{Y}_i \mid \boldsymbol{\theta}_h(\boldsymbol{X}_i)) \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^{\star}_{hj} - \boldsymbol{X}\boldsymbol{\beta}_{hj})' \boldsymbol{R}_j^{-1} (\boldsymbol{\theta}^{\star}_{hj} - \boldsymbol{X}\boldsymbol{\beta}_{hj})\right\} \\ &\quad \times \exp\left\{-\frac{1}{2} \sum_{i:\gamma_i=h} (\boldsymbol{Y}_i - \boldsymbol{\theta}_h(\boldsymbol{X}_i))' \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y}_i - \boldsymbol{\theta}_h(\boldsymbol{X}_i))\right\} \\ &\sim \text{N}\left(\left\{\boldsymbol{R}_j^{-1} + \frac{\boldsymbol{U}_h' \boldsymbol{U}_h}{\tilde{\sigma}_j^2} \boldsymbol{I}_n\right\}^{-1} \left\{\boldsymbol{U}_h' \frac{\tilde{\boldsymbol{Y}}_{hj}}{\tilde{\sigma}_j^2} + \boldsymbol{R}_j^{-1} \boldsymbol{X}\boldsymbol{\beta}_{hj}\right\}, \left\{\boldsymbol{R}_j^{-1} + \frac{\boldsymbol{U}_h' \boldsymbol{U}_h}{\tilde{\sigma}_j^2} \boldsymbol{I}_n\right\}^{-1}\right) \end{aligned}$$

(5) Update $\boldsymbol{\beta}_{hj}$

$$\begin{aligned} p(\boldsymbol{\beta}_{hj} \mid \cdot) &\propto p(\boldsymbol{\beta}_{hj}) \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^{\star}_{hj} - \boldsymbol{X}\boldsymbol{\beta}_{hj})' \boldsymbol{R}_j^{-1} (\boldsymbol{\theta}^{\star}_{hj} - \boldsymbol{X}\boldsymbol{\beta}_{hj})\right\} \\ &\sim \text{N}(\boldsymbol{\Lambda}_{hj}\left(\boldsymbol{X}' \boldsymbol{R}_j^{-1} \boldsymbol{\theta}^{\star}{}_{hj} + \boldsymbol{\Lambda}_{0j}^{-1} \boldsymbol{\beta}_{0j}\right), \boldsymbol{\Lambda}_{hj}), \end{aligned}$$

where $\boldsymbol{\Lambda}_{hj} = (\boldsymbol{X}'\boldsymbol{R}_j^{-1}\boldsymbol{X} + \boldsymbol{\Lambda}_{0j}^{-1})^{-1}$.

(6) Update $(\gamma_i, \boldsymbol{Y}_i)$, where $\boldsymbol{Y}_i = (Y_{Pi}, Y_{Di})$.
We write $p(\gamma_i, \boldsymbol{Y}_i \mid \cdot)$ as $p(\boldsymbol{Y}_i \mid \gamma_i, \cdot)p(\gamma_i \mid \cdot)$ where $\cdot$ includes $\boldsymbol{O}_i$.

- If $\delta_i = \xi_i = 1$

$$Pr(\gamma_i = h \mid \cdot) \propto w_h \phi(T_{1,i}, T_{2,i}; \boldsymbol{\theta}_h(\boldsymbol{X}_i), \boldsymbol{\Sigma});$$

$p(\boldsymbol{Y}_i \mid \gamma_i, \cdot)$ is point mass at $\boldsymbol{Y}_i = (T_{1i}, T_{2i})$.

- If $\delta_i = \xi_i = 0$ (i.e., $Y_{Pi} > T_{2i}, Y_{Di} > T_{2i}$),

$$Pr(\gamma_i = h \mid \cdot) \propto \int_{t>T_{2i}} \int_{s>T_{2i}} w_h \phi(s, t; \boldsymbol{\theta}_h(\boldsymbol{X}_i), \boldsymbol{\Sigma}) \mathrm{d}s\mathrm{d}t.$$

$$p(\boldsymbol{Y}_i \mid \gamma_i, \cdot) = \frac{\phi(\boldsymbol{Y}_i; \boldsymbol{\theta}_{\gamma_i}(\boldsymbol{X}_i), \boldsymbol{\Sigma})}{\int_{t>T_{2i}} \int_{s>T_{2i}} \phi(s, t; \boldsymbol{\theta}_{\gamma_i}(\boldsymbol{X}_i), \boldsymbol{\Sigma}) \mathrm{d}s\mathrm{d}t},$$

where $Y_{Pi} > T_{2i}, Y_{Di} > T_{2i}$.

- If $\delta_i = 1$ and $\xi_i = 0$

$$Pr(\gamma_i = h \mid \cdot) \propto \int_{t>T_{2i}} w_h \phi(T_{1i}, t; \boldsymbol{\theta}_h(\boldsymbol{X}_i), \boldsymbol{\Sigma}) \mathrm{d}t;$$

$$p(\boldsymbol{Y}_i \mid \gamma_i, \cdot) = \frac{\phi(\boldsymbol{Y}_i; \boldsymbol{\theta}_{\gamma_i}(\boldsymbol{X}_i), \boldsymbol{\Sigma})}{\int_{t>T_{2i}} \phi(T_{1i}, t; \boldsymbol{\theta}_{\gamma_i}(\boldsymbol{X}_i), \boldsymbol{\Sigma}) \mathrm{d}t},$$

where $Y_{Pi} = T_{1i}, Y_{Di} > T_{2i}$.

- If $\delta_i = 0$ and $\xi_i = 1$

$$Pr(\gamma_i = h \mid \cdot) \propto \int_{s>T_{1i}} w_h \phi(s, T_{2i}; \boldsymbol{\theta}_h(\boldsymbol{X}_i), \boldsymbol{\Sigma}) \mathrm{d}s;$$

$$p(\boldsymbol{Y}_i \mid \gamma_i, \cdot) = \frac{\phi(\boldsymbol{Y}_i; \boldsymbol{\theta}_{\gamma_i}(\boldsymbol{X}_i), \boldsymbol{\Sigma})}{\int_{s>T_{1i}} \phi(s, T_{2i}; \boldsymbol{\theta}_{\gamma_i}(\boldsymbol{X}_i), \boldsymbol{\Sigma}) \mathrm{d}s},$$

where $Y_{Pi} > T_{1i}, Y_{Di} = T_{2i}$.

REFERENCES

BREM, H., PIANTADOSI, S., BURGER, P. C., WALKER, M., SELKER, R., VICK, N. A., BLACK, K., SISTI, M., BREM, S., MOHR, G. *and others*. (1995). Placebo-controlled trial of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy for recurrent gliomas. *The Lancet* **345**, 1008–1012.

CHEN, Y.-H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime Data Analysis* **18**, 36–57.

COMMENT, L., MEALLI, F., HANEUSE, S. AND ZIGLER, C. (2019). Survivor average causal effects for continuous time: a principal stratification approach to causal inference with semicompeting risks. *arXiv preprint arXiv:1902.09304*.

DANIELS, M. J., ROY, J. A., KIM, C., HOGAN, J. W. AND PERRI, M. G. (2012). Bayesian inference for the causal effect of mediation. *Biometrics* **68**, 1028–1036.

DE IORIO, M., JOHNSON, W. O., MÜLLER, P. AND ROSNER, G. L. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics* **65**, 762–771.

DING, A., SHI, G., WANG, W. AND HSIEH, J.-J. (2009). Marginal regression analysis for semi-competing risks data under dependent censoring. *Scandinavian Journal of Statistics* **36**, 481–500.

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

FINE, J. P., JIANG, H. AND CHAPPELL, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907–919.

FIX, E. AND NEYMAN, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*, 205–241.

FRANGAKIS, C. E. AND RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

GEISSER, S. AND EDDY, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74**, 153–160.

GELFAND, A. E. AND KOTTAS, A. (2003). Bayesian semiparametric regression for median residual life. *Scandinavian Journal of Statistics* **30**, 651–665.

HANSON, T. AND JOHNSON, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* **97**, 1020–1033.

HOUGAARD, P. (1999). Multi-state models: a review. *Lifetime Data Analysis* **5**, 239–264.

HSIEH, J.-J. AND HUANG, Y.-T. (2012). Regression analysis based on conditional likelihood approach under semi-competing risks data. *Lifetime Data Analysis* **18**, 302–320.

HYMAN, D. M., LAETSCH, T. W., KUMMAR, S., DUBOIS, S. G., FARAGO, A. F., PAPPO, A. S., DEMETRI, G. D., EL-DEIRY, W. S., LASSEN, U. N., DOWLATI, A. *and others*. (2017). The efficacy of larotrectinib (LOXO-101), a selective tropomyosin receptor kinase (TRK) inhibitor, in adult and pediatric TRK fusion cancers. *Journal of Clinical Oncology* **18_suppl**, LBA2501.

IBRAHIM, J. G., CHEN, M.-H. AND SINHA, D. (2005). *Bayesian Survival Analysis*. Hoboken, NJ: Wiley Online Library.

KAUFMAN, P. A, AWADA, A., TWELVES, C., YELLE, L., PEREZ, E. A., VELIKOVA, G., OLIVO, M. S., HE, Y., DUTCUS, C. E. AND CORTES, J. (2015). Phase III open-label randomized study of eribulin mesylate versus capecitabine in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline and a taxane. *Journal of Clinical Oncology* **33**, 594.

LEE, K. H., HANEUSE, S., SCHRAG, D. AND DOMINICI, F. (2015). Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64**, 253–273.

LIN, D. Y., ROBINS, J. M. AND WEI, L. J. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* **83**, 381–393.

LO, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics* **12**, 351–357.

MACEACHERN, S. N. (1999). Dependent nonparametric processes. In: *ASA Proceedings of the Section on Bayesian Statistical Science*. Alexandria, VA: American Statistical Association, pp. 50–55.

MACKAY, D. (1999). Introduction to Gaussian processes. *Technical Report*. Cambridge University. http://wol.ra.phy.cam.ac.uk/mackay/GP/.ter.

PENG, L. AND FINE, J. P. (2007). Regression modeling of semicompeting risks data. *Biometrics* **63**, 96–108.

PENG, L. AND FINE, J. P. (2012). Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American Statistical Association*.

RASMUSSEN, C. E. AND WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

ROBINS, J. M. (1995a). An analytic method for randomized trials with informative censoring: Part II. *Lifetime Data Analysis* **1**, 417–434.

ROBINS, J. M. (1995b). An analytic method for randomized trials with informative censoring: Part 1. *Lifetime Data Analysis* **1**, 241–254.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688.

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

SPARAPANI, R. A., LOGAN, B. R., MCCULLOCH, R. E. AND LAUD, P. W. (2016). Nonparametric survival analysis using Bayesian Additive Regression Trees (BART). *Statistics in Medicine* **35**, 2741–2753.

TCHETGEN TCHETGEN, E. J. (2014). Identification and estimation of survivor average causal effects. *Statistics in Medicine* **33**, 3601–3628.

VARADHAN, R., XUE, Q.-L. AND BANDEEN-ROCHE, K. (2014). Semicompeting risks in aging research: methods, issues and needs. *Lifetime Data Analysis* **20**, 538–562.

WANG, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**, 257–273.

XU, J., KALBFLEISCH, J. D. AND TAI, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics* **66**, 716–725.

XU, Y., MÜLLER, P., WAHED, A. S. AND THALL, P. F. (2016). Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *Journal of the American Statistical Association* **111**, 921–950.

XU, Y., THALL, P. F., HUA, W. AND ANDERSSON, B. S. (2019). Bayesian non-parametric survival regression for optimizing precision dosing of intravenous busulfan in allogeneic stem cell transplantation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **68**, 809–828.

ZHANG, J. L. AND RUBIN, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics* **28**, 353–368.

ZHOU, H. AND HANSON, T. (2018). A unified framework for fitting Bayesian semiparametric models to arbitrarily censored survival data, including spatially-referenced data. *Journal of the American Statistical Association* **113**, 571–581.

[*Received July 21, 2019; revised January 30, 2020; accepted for publication February 3, 2020*]