

# Bayesian Nonparametric Survival Regression for Optimizing Precision Dosing of Intravenous Busulfan in Allogeneic Stem Cell Transplantation

Yanxun Xu<sup>†</sup>

*Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, USA*

Peter F. Thall

*Department of Biostatistics, U.T. M.D. Anderson Cancer Center, Houston, USA.*

William Hua

*Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, USA*

Borje S. Andersson

*Department of Stem Cell Transplantation and Cellular Therapy, U.T. M.D. Anderson Cancer Center, Houston, USA.*

**Abstract.** Allogeneic stem cell transplantation (allo-SCT) is now part of standard of care for acute leukemia (AL). To reduce toxicity of the pre-transplant conditioning regimen, intravenous busulfan is usually used as a preparative regimen for AL patients undergoing allo-SCT. Systemic busulfan exposure, characterized by the area under the plasma concentration versus time curve (AUC), is strongly associated with clinical outcome. An AUC that is too high is associated with severe toxicities, while an AUC that is too low carries increased risks of disease recurrence and failure to engraft. Consequently, an optimal AUC interval needs to be determined for therapeutic use. To address the possibility that busulfan pharmacokinetics and pharmacodynamics vary significantly with patient characteristics, we propose a tailored approach to determine optimal covariate-specific AUC intervals. To estimate these personalized AUC intervals, we apply a flexible Bayesian nonparametric regression model based on a dependent Dirichlet process and Gaussian process, DDP-GP. Our analyses of a dataset of 151 patients identified optimal therapeutic intervals for AUC that varied substantively with age and whether the patient was in complete remission or had active disease at transplant. Extensive simulations to evaluate the DDP-GP model in similar settings showed that its performance compares favorably to alternative methods. We provide an R package, `DDPGPSurv`, that implements the DDP-GP model for a broad range of survival regression analyses.

*Keywords:* Allogeneic stem cell transplantation; Bayesian nonparametrics; Personalized medicine; Survival regression.

<sup>†</sup>*Address for correspondence:* Yanxun Xu, Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA 21042. Email: yanxun.xu@jhu.edu.

## 1. Introduction

Allogeneic stem cell transplantation (allo-SCT) is an established treatment for various hematologic diseases, including acute myelogenous and lymphocytic leukemia and non-Hodgkins lymphoma. Intravenous (IV) busulfan has been established as a desirable component of the preparative regimen for allo-SCT, due to its absolute bioavailability and dosing accuracy, leading to improved patient survival (Andersson et al., 2002; Wachowiak et al., 2011; Nagler et al., 2013; Copelan et al., 2013; Bredeson et al., 2013). The patient’s busulfan systemic exposure (Bu-SE) represented by the area under the plasma concentration versus time curve, AUC, is crucial, as serious adverse events are associated with an AUC that is either very high or too low. Higher AUC values are associated with neurologic toxicity (grand mal seizures), hepatic veno-occlusive disease, mucositis, and/or gastro-intestinal toxicity (Dix et al., 1996; Ljungman et al., 1997; Kontoyiannis et al., 2001; Geddes et al., 2008). Lower AUC is associated with an increased likelihood of disease recurrence and thus shorter survival time (Slattery et al., 1997; McCune et al., 2002; Bartelink et al., 2009; Russell et al., 2013; Andersson et al., 2016).

Consequently, it is important to define an optimal AUC interval of busulfan exposure that maximizes survival while minimizing risk. Studies of fixed-dose oral busulfan regimens suggest that inter-individual variations in Bu-SE exposure may be as high as 10- to 20-fold. In contrast, IV delivery of busulfan is more consistent and reliable for controlling delivered dose (Andersson et al., 2000), and thus is better suited for obtaining optimal busulfan AUC intervals. Andersson et al. (2002) showed that an optimal interval of IV Bu-SE had AUC values approximately 950 to 1520  $\mu\text{Mol}\cdot\text{min}$  from one representative dose in a typical 16-dose treatment course, or a total course AUC of 15,200 – 24,400  $\mu\text{Mol}\cdot\text{min}$ , yielding longer survival times and lower toxicity rates compared with values outside this interval. More recently, Bartelink et al. (2016) reported that, in children and young adults, the optimal daily AUC range in a prototype 4-day Bu-based regimen was 78-101  $\text{mg}\cdot\text{h}/\text{L}$  (corresponding to a total course AUC of about 19,100 – 21,200  $\mu\text{Mol}\cdot\text{min}$ ), regardless of the type or stage of underlying disease and whether the patients were in complete remission (CR) if they had an underlying malignancy.

In this paper, we account for patient heterogeneity to assess the joint impact of patient age, CR status, and AUC on patient survival, with the goal to determine covariate-specific optimal daily AUC intervals. To evaluate possible interactions between covariates and Bu-SE, and their association with treatment outcome, we analyzed a dataset of 151 patients who underwent allo-SCT for acute myelogenous leukemia (AML) and myelodysplastic syndrome (MDS). It has been demonstrated that many different comorbidity conditions may affect the patient’s risk for developing complications with these procedures (Sorrer et al., 2005, 2014). Additionally, there commonly is a correlation between the severity of comorbidities and patient age. Therefore, we analyzed the outcome of our patients using age as a continuous covariate. We also included the indicator of whether the patient was in CR or had active disease at time of allo-SCT, since patients transplanted in CR have, on average, more favorable outcomes (De Lima et al., 2004; Kanakry et al., 2014). Our goal was to find patient-specific optimal AUC ranges that maximize expected survival time given the patient’s age and CR status. The results of this analysis may provide specific guidelines for so-called “personalized” or

”precision” medicine in clinical practice.

Andersson et al. (2002) estimated the optimal AUC range by fitting a Cox proportional hazards regression model for overall survival (OS) time and smoothing a martingale residual plot, which showed that the hazard of death was approximately a quadratic function of  $\log(\text{AUC})$ . Bartelink et al. (2016) used a fourth-order polynomial model to estimate the association between AUC and OS. Both of these methods assumed specific parametric distributions for survival time, and the latter analysis assumed a specific polynomial function for the relationship between AUC and OS. For our data set, Figure 1 shows histograms and estimated density plots of the patients’ OS times in weeks, with and without a log transformation of OS. The figure clearly presents a long-tailed distribution that might result from a mixture of several unknown distributions. Alternatively, the long-tailed distribution might be due to the fact patients who have survived at least four years from transplant are at risk of death from natural causes, rather than leukemia or transplant related causes. Consequently, the specific models and strong parametric assumptions made by Andersson et al. (2002) and Bartelink et al. (2016) may not be suitable to fit the current data set well. In particular, the proportional hazards assumption underlying the Cox model may not be valid. Even given a survival distribution that fits the data reasonably well, an additional problem is determining functional relationships between AUC, prognostic covariates, and the risk of death.

We present a flexible Bayesian nonparametric (BNP) survival regression model to estimate the relationship between survival time, AUC, and baseline covariates. Based on our analysis of the allo-SCT dataset, we determined personalized optimal AUC ranges based on patient’ age and CR status. An important advantage of BNP models is that they often fit complicated data structures better than parametric model-based methods because BNP models can accurately approximate essentially any distribution or function, a property known as “full support.” Another important advantage of BNP models is that they often identify unexpected structures in a dataset that cannot be seen using conventional statistical models and methods. BNP models have been used widely for survival analysis. Hanson and Johnson (2002) proposed a mixture of Polya tree priors in semiparametric accelerated failure time (AFT) models, while Gelfand and Kottas (2003) developed the corresponding Dirichlet process (DP) mixture approach. Zhou and Hanson (2017) presented a unified approach for modeling survival data by exploiting and extending the three most commonly-used semiparametric models: proportional hazards, proportional odds, and accelerated failure time. Despite the flexibility of these approaches for modeling baseline survival distributions, they are restricted in the way that covariates may affect the baseline distribution. Fully nonparametric tree-based survival models have been developed, such as the use of random forests (Ishwaran et al., 2008) and Bayesian additive regression trees (Sparapani et al., 2016). De Iorio et al. (2009) proposed an unconstrained survival regression model with a dependent Dirichlet process (DDP) (MacEachern, 1999) prior in order to incorporate covariates in a naturally interpretable way. Xu et al. (2016) developed a DDP prior with Gaussian process as the base measure, the DDP-GP model, to evaluate overall survival (OS) times of complex dynamic treatment regimes including multiple transition times. However, a non-trivial limitation of the DDP-GP model in Xu et al. (2016) is that it gives the same fixed weight to all covariates, regardless of their numerical domains, when

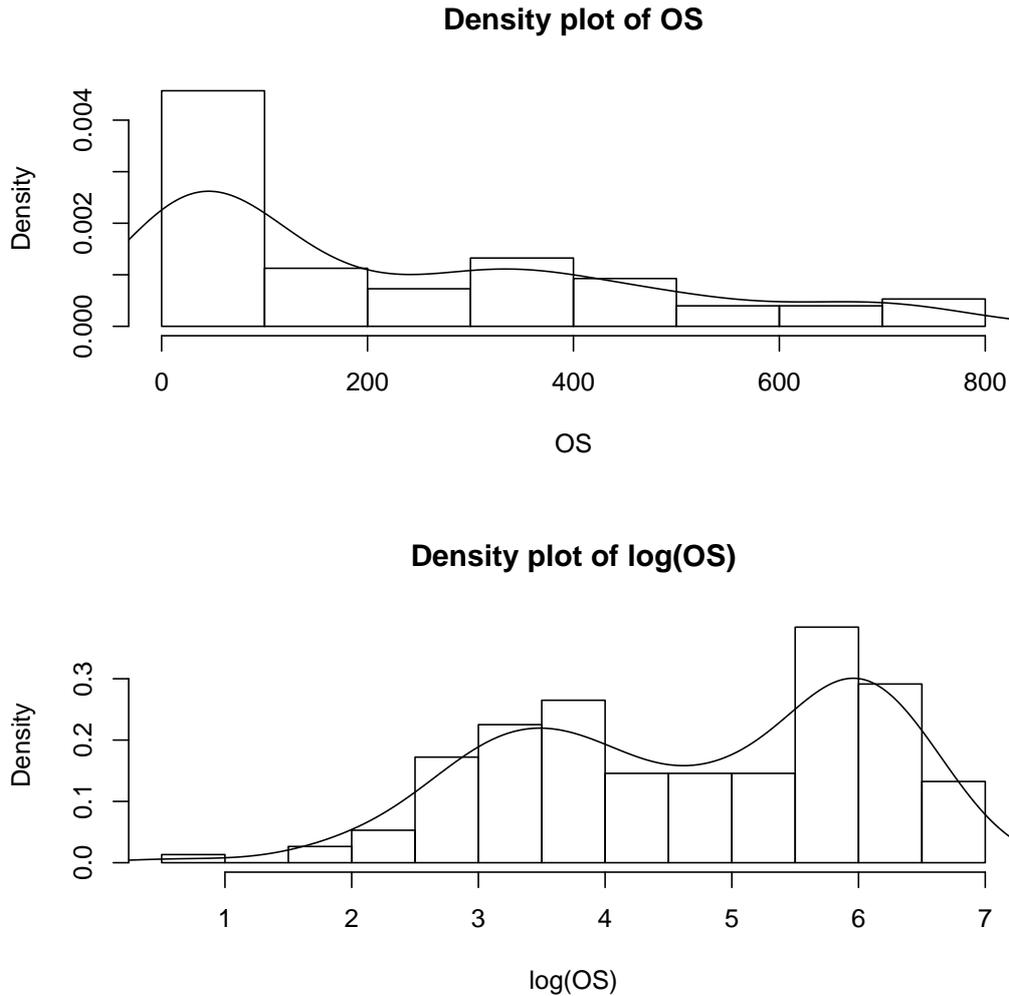


Figure 1: Histograms of overall survival time in weeks (top) and log overall survival time (bottom), with nonparametric density estimates.

quantifying dependence between patients. Such a restriction may cause posterior inferences to be inconsistent (Tokdar and Ghosh, 2007).

Building on the work in Xu et al. (2016), in this paper we propose a flexible survival regression framework by formulating a DDP with a more general covariance function for the GP prior that includes an individual scale parameter for each covariate and additional hyperparameters for model flexibility and robustness. The proposed model provides easy-to-implement posterior inferences in settings where the proportional hazards assumption, specific parametric models such as AFT models, or semi-parametric models may not fit the data well. Currently, the R package `survival`, which is limited to such models, remains a standard tool for statistical and medical researchers. One of the main contributions of our paper is to provide a new, easy-to-use R package, `DDPGPSurv`, that implements the proposed DDP-GP survival regression model for a broad range of survival analyses. A major goal is that `DDPGPSurv` will become a new

standard computational tool for implementing this generalized DDP-GP to conduct survival analysis in medical research.

The rest of the paper is organized as follows. In Section 2 we review the motivating dataset. We present the DDP-GP survival regression model in Section 3. Section 4 gives a brief introduction to the R package DDPGPSurv. Extensive simulation studies with comparison to alternative methods are conducted in Section 5. We analyze our dataset in Section 6, and conclude with a brief discussion in Section 7.

## 2. Motivating Study

When total body irradiation was replaced with high-dose oral busulfan (Santos et al., 1983; Tutschka et al., 1987), it quickly became clear that unpredictable, often lethal, toxicities limited the use of a busulfan-based conditioning program. Several retrospective studies indicated an association between systemic drug exposure and clinical treatment outcome (Dix et al., 1996; Slattery et al., 1997). This spawned an interest in exploring pharmacokinetic dose guidance, but the erratic bioavailability of oral busulfan prevented its successful implementation in a prospective fashion. The advent of IV Busulfan, which guarantees complete bioavailability with absolute assurance for systemic dose delivery has changed this. Routine application of therapeutic dose guidance for IV busulfan in pre-transplant conditioning therapy now makes it possible to accurately deliver a predetermined systemic exposure dose in terms of AUC, thereby optimizing treatment. This is important, since in (myeloid) leukemia the cytotoxic drug dose and accurate dose delivery are associated with clinical treatment outcome (Andersson et al., 2002; De Lima et al., 2004; Russell et al., 2013; Bartelink et al., 2016; Andersson et al., 2016). A question that has not yet been resolved satisfactorily is what optimal systemic exposure dose to target in an individual patient. To address this decisively, we have retrieved data in The University of Texas MD Anderson Cancer Center from 151 AML/MDS patients who received a standardized 4-day fludarabine-IV busulfan combination, with both agents administered based on body surface area. Pharmacological studies of busulfan were performed as an optional procedure, but the information was not used for busulfan dose-adjustments. The dataset includes overall survival (OS) times and the covariates age, CR status, and AUC. Table 1 summarizes the characteristics of the study population at baseline.

## 3. Probability Model

### 3.1. *Dependent Dirichlet process-Gaussian process prior*

Denote the log time to death by  $Y$  and censoring time on the log(time) domain by  $C$ , with  $T = Y \wedge C$  the observed log time of the event or censoring, and  $\delta = I(Y \leq C)$ . Indexing patients by  $i = 1, \dots, n$ , the observed outcome data for patient  $i$  are  $(T_i, \delta_i)$ , and we let  $\mathbf{x}_i$  denote the baseline covariate vector, including age, CR status, and AUC.

We construct a Bayesian nonparametric (BNP) survival regression model for  $F(Y | \mathbf{x})$ , the distribution of  $[Y | \mathbf{x}]$ , as follows. We start with a model for a discrete random distribution  $G(\cdot)$ , then use a Gaussian kernel to extend this to a prior for a continuous random distribution,

		Patients( $n = 151$ )
Age(years)		
	$\leq 25$	12 (8%)
	26 – 35	22 (15%)
	36 – 45	32 (21%)
	46 – 55	50 (33%)
	$\geq 56$	35 (23%)
Sex		
	Male	77 (51%)
	Female	74 (49%)
In CR at transplantation		
	Yes	80 (53%)
	No	71 (47%)
AUC quantile		
	10%	3,928
	25%	4,328
	50%	5,077
	75%	5,754
	90%	6,371

Table 1: Patient characteristics at time of transplantation.

and finally we replace the kernel means by a regression structure to define the desired prior on  $\{F(Y | \mathbf{x}), \mathbf{x} \in X\}$ . The constructions of  $G(\cdot)$  and  $F(\cdot)$  are elaborated below, by way of a brief review of BNP models. See, for example, Müller and Mitra (2013) and Müller and Rodriguez (2013) for more extensive reviews.

First proposed by Ferguson et al. (1973), the Dirichlet process (DP) prior has been used widely in Bayesian analyses as a prior model for random unknown probability distributions. A  $DP(\alpha_0, G_0)$  involves a positive scaling parameter  $\alpha_0$  and a base probability measure  $G_0$ . A constructive definition of a DP is provided by Sethuraman (1994), the so-called “stick-breaking” construction, given by  $G = \sum_{h=1}^{\infty} w_h \delta_{\theta_h}$ , where  $\theta_h \sim G_0$ , and  $w_h = v_h \prod_{l < h} (1 - v_l)$  with  $v_h \sim \text{be}(1, \alpha)$ . Here,  $\delta_{\theta_h}(\cdot)$  denotes the Dirac delta function, which is equal to 1 at  $\theta_h$  and is equal to 0 everywhere else. In many applications, the discrete nature of  $G$  is not appropriate. To deal with this, a DP mixture model extends the DP model by replacing each point mass  $\delta(\theta_h)$  with a continuous kernel, such as a DP mixture of normals  $G = \sum_h w_h N(\theta_h, \sigma^2)$ , where  $N(\theta_h, \sigma^2)$  denotes a normal distribution with mean  $\theta_h$  and standard deviation  $\sigma$ .

To include regression on covariates, MacEachern (1999), extended the DP mixture model by replacing each mean parameter  $\theta_h$  in the sum with a function  $\theta_h(\mathbf{x})$  of covariates  $\mathbf{x}$ . This is called a dependent Dirichlet process (DDP), obtained by assuming the regression model

$$F(y | \mathbf{x}) = \sum_{h=1}^{\infty} w_h N(y; \theta_h(\mathbf{x}), \sigma^2),$$

where one can specify a stochastic process prior for  $\{\theta_h(\mathbf{x})\}$ . As a default assumption MacEach-

ern (1999) proposed a Gaussian process (GP) prior. Here, the GP is indexed by  $\mathbf{x}$ . Temporarily suppressing the subindex  $h$ , a GP prior is characterized by the marginal distribution for any  $n$ -tuple  $(\theta(\mathbf{x}_1), \dots, \theta(\mathbf{x}_n))$  being a multivariate normal distribution with mean vector  $(\mu(\mathbf{x}_1), \dots, \mu(\mathbf{x}_n))$  and  $(n \times n)$  covariance matrix with  $(i, j)$  element  $C(\mathbf{x}_i, \mathbf{x}_j)$ , for any set of  $n \geq 1$  covariate vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . We denote this model by  $\theta(\mathbf{x}) \sim \text{GP}(\mu, C)$ . Extensive reviews of the GP are given by MacKay (1999) and Rasmussen and Williams (2006).

In the context of modeling each patient’s transition times between successive disease states in a dataset arising from multi-stage chemotherapy of acute leukemia, Xu et al. (2016) modeled  $\{\theta_h(\mathbf{x})\} \sim \text{GP}(\mu_h(\cdot), C(\cdot, \cdot))$ ,  $h = 1, 2, \dots$  with  $\mu_h(\mathbf{x}_i; \beta_h) = \mathbf{x}_i \beta_h$  and  $C(\mathbf{x}_i, \mathbf{x}_\ell) = \exp\{-\sum_{d=1}^D (x_{id} - x_{\ell d})^2\} + \delta_{i\ell} J^2$ . Here,  $D$  is the dimension of the covariate vector, with  $\delta_{i\ell} = I(i = \ell) = 1$  if  $i = \ell$  and 0 otherwise. The term  $J^2$  is jitter added to provide numerical stability by avoiding singular covariance matrices, with a small value such as  $J = 0.1$  typically used.

A non-trivial limitation of this covariance function is that it gives the same weight to all covariates, regardless of their numerical domains, when quantifying dependence between patients. This implies that different covariates  $x_d$  and  $x_{d'}$  contribute the same to the correlation between patients  $i$  and  $l$  as long as  $(x_{id} - x_{\ell d})^2$  and  $(x_{id'} - x_{\ell d'})^2$  are the same. Furthermore, without including hyperparameters with prior distributions in the covariance function, the posterior inference using a Gaussian process prior may not be consistent. Technical proofs can be found in Tokdar and Ghosh (2007). To avoid these limitations, we extend the DDP-GP model by including an additional scale parameter,  $\lambda_d$ , for each covariate  $x_d$  and also an overall multiplicative scale parameter  $\sigma_0^2$  in the covariance function:

$$C(\mathbf{x}_i, \mathbf{x}_\ell) = \sigma_0^2 \exp\left\{-\sum_{d=1}^D \frac{(x_{id} - x_{\ell d})^2}{\lambda_d^2}\right\} + \delta_{i\ell} J^2. \quad (3.1)$$

The multiplicative scale parameter  $\sigma_0^2$  accounts for variability in the data that is not accommodated by the variance  $\sigma^2$  of the normal component distributions.

The model can be summarized as

$$p(y_i | \mathbf{x}_i, F) = F_{\mathbf{x}_i}(y_i) \\ \{F_{\mathbf{x}}\} \sim \text{DDP-GP} \left\{ \{\mu_h\}, C; \alpha, \{\beta_h\}, \{\lambda_d^2\}, \sigma_0^2, \sigma^2 \right\}. \quad (3.2)$$

We use the acronym DDP-GP to refer to the proposed model with the DDP mixture of normals with this particular GP prior on the mean of the normal kernel. Thus,

$$F_{\mathbf{x}}(y) = \sum_{h=0}^{\infty} w_h \mathbb{N}(y; \theta_h(\mathbf{x}), \sigma^2) \text{ with } \{\theta_h(\mathbf{x})\} \sim \text{GP}(\mu_h(\cdot), C(\cdot, \cdot)), \quad (3.3)$$

where  $h = 1, 2, \dots$ ,  $\mu_h(\mathbf{x}_i) = \mathbf{x}_i \beta_h$ , and  $C(\cdot, \cdot)$  is defined in (3.1).

### 3.2. DDP-GP survival regression model

Denote the vector of all model parameters by  $\Theta$  and the data by  $\mathcal{D}_n = \{T_i, \delta_i, \mathbf{x}_i\}_{i=1}^n$ . The likelihood function is the usual form

$$L(\Theta | \mathcal{D}_n) = \prod_{i=1}^n f_{\mathbf{x}_i}(t_i | \Theta)^{\delta_i} \{1 - F_{\mathbf{x}_i}(t_i | \Theta)\}^{1-\delta_i},$$

where  $f_{\mathbf{x}}(\cdot)$  and  $F_{\mathbf{x}}(\cdot)$  denote the density and cumulative distribution function of  $Y$  for an individual with covariates  $\mathbf{x}$ . Given the assumed DDP-GP prior on  $F_{\mathbf{x}}(\cdot)$ , shown in (3.2), we complete the model by assuming the priors  $\beta_h \sim N(\beta_0, \Sigma_0)$ ,  $1/\sigma^2 \sim \text{Gamma}(a_1, b_1)$ , the precision parameter  $\alpha \sim \text{Gamma}(a_2, b_2)$ ,  $\sigma_0 \sim N(0, \tau_\sigma^2)$ , and the covariate scale parameters  $\lambda_d \sim \text{iid } N(0, \tau^2)$ ,  $d = 1, \dots, D$ . Thus, the DDP-GP's hyperparameters are  $\theta^* = (\beta_0, \Sigma_0, a_1, b_1, a_2, b_2, \tau_\sigma^2, \tau^2)$ .

To implement the DDP-GP model, one first must determine numerical values for the hyperparameters  $\theta^*$ . We introduce default choices for fixing  $\theta^*$  in our `DDPGPSurv` package below, although users can define their own preferred values, if desired. We suggest using an empirical Bayes method to obtain  $\beta_0$  by fitting a normal distribution for patient response on the log scale,  $\log(Y) | \mathbf{x} \sim N(\mathbf{x}\beta_0, \hat{\sigma}^2)$  and assuming  $\Sigma_0$  to be a diagonal matrix with all diagonal values 10. Once an empirical estimate  $\hat{\sigma}^2$  of  $\sigma^2$  is obtained, one can tune  $(a_1, a_2)$  so that the prior mean of  $\sigma^2$  matches the empirical estimate and the variance equals 10 or a suitably large value to ensure a vague prior. The total mass parameter  $\alpha$  in the stick-breaking construction determines the number of unique clusters in the underlying DP Polya urn scheme. Usually, the DP yields many small clusters, therefore changing the prior of  $\alpha$  does not significantly alter posterior predictive inference, which we will use for estimating the survival function and optimal AUC ranges. We assume  $a_2 = b_2 = 1$  to ensure a vague prior on  $\alpha$ . Lastly, we assume  $\tau = \tau_\sigma = 10$  so that the ranges of  $\lambda_d$ 's and  $\sigma_0$  in the covariance function are large enough to cover variability in the data.

To obtain posterior inference for a DDP-GP survival regression model, we first marginalize (3.2) analytically with respect to the random probability measures  $F_{\mathbf{x}}(\cdot)$ . To do this, we first rewrite (3.3) equivalently as a hierarchical model with a set of new latent indicator variables  $\gamma_i$  as

$$(Y_i | \gamma_i = h, \mathbf{x}_i) \sim N(\theta_h(\mathbf{x}_i), \sigma^2) \quad \text{and} \quad p(\gamma_i = h) = w_h, \quad (3.4)$$

for  $i = 1, \dots, n$ . If clusters of patients are defined as  $S_h = \{i : \tilde{\theta}_i = \theta_h\}$ , then the  $\gamma_i$ 's are interpreted as cluster membership indicators. Posterior simulation makes use of these indicators and the vectors  $\theta_h = (\theta_h(\mathbf{x}_1), \dots, \theta_h(\mathbf{x}_n))$ . After marginalization with respect to  $F_{\mathbf{x}}$ , we are left with the marginal model for  $\{\gamma_i, \theta_h(\mathbf{x}_i); i = 1, \dots, n, h = 1, \dots\}$ . We implement posterior sampling based on the collapsed Gibbs sampler (Escobar and West, 1995) in the R package `DDPGPSurv`. Details of the MCMC computations are provided in the supplement Section 1.

### 3.3. Personalized optimal AUC range estimation

Let  $\rho_n = (S_1, \dots, S_H)$  denote the partition of the  $n$  patients, determined by the clusters induced by the  $\gamma_i$ 's. A key advantage of the proposed BNP model is that we can easily write down the posterior predictive distribution of the outcome  $Y_{n+1}$  for a future patient with covariate vector  $\mathbf{x}_{n+1}$ , given by

$$p(Y_{n+1} | \mathbf{x}_{n+1}, \mathcal{D}_n) = \sum_{\rho_n} p(\rho_n | \mathcal{D}_n) \int p(\Theta | \rho_n, \mathcal{D}_n) \times \left\{ \sum_{h=1}^{H+1} p(Y_{n+1} | n+1 \in S_h, \theta_h(\mathbf{x}_{n+1}), \Theta) p(n+1 \in S_h | \mathbf{x}_{n+1}, \mathcal{D}_n, \rho_n, \Theta) \right\} d\Theta. \quad (3.5)$$

The innermost sum averages with respect to the cluster membership for the  $(n+1)^{st}$  patient during the MCMC. The term  $h = H+1$  corresponds to the case that this new patient may form his/her own singleton cluster. The posterior average with respect to  $p(\rho_n | \mathcal{D})$  and  $p(\Theta | \rho_n, \mathcal{D}_n)$  is evaluated as an average over the MCMC sample.

For the IV busulfan allo-SCT data,  $\mathbf{x}$  includes the key treatment variable AUC, which quantifies the patient's delivered dose of IV busulfan and thus may be targeted by the treating physician. From (3.5), based on our analysis of the IV busulfan data using the DDP-GP, we can use the predictive distribution to compute the optimal AUC for the future patient  $n+1$  as that which maximizes expected log survival time,

$$\widehat{\text{AUC}}_{n+1} = \operatorname{argmax}_{\text{AUC}} E(Y_{n+1} | \mathbf{x}_{n+1}, \mathcal{D}_n), \quad (3.6)$$

where  $\mathbf{x}_{n+1}$  includes patients' age, CR status, and AUC. The laboratory-based method for determining the median specific daily Bu-SE has about a 3% error when sampling is carried over 12-14 hours (or about 3-4 drug half-lives). However, if sampling is restricted to 4-6 hours (1.0 - 1.5 drug half-lives), as is done with many PK evaluation methods, the error increases to at least 6%. Based on these considerations, we decided to use optimal AUC +/- 10% as a reasonable interval for targeting, since it is not possible to detect any difference in covariate impact on outcome between patients with AUC values falling within this narrow Bu-SE interval. Therefore, we define the optimal AUC interval for future patient  $n+1$  as

$$[ 0.9 \times \widehat{\text{AUC}}_{n+1}, 1.1 \times \widehat{\text{AUC}}_{n+1} ],$$

bearing in mind that  $\widehat{\text{AUC}}_{n+1}$  depends on the patient's covariates  $\mathbf{x}_{n+1}$ .

## 4. R package: DDPGPSurv

One of the main contributions of this paper is that we have developed an R package, DDPGPSurv, that implements the proposed DDP-GP model as a general tool for survival analysis. The functions in the package perform inference via MCMC simulations from the posterior distributions based on a DDP-GP prior using a collapsed Gibbs sampler (`mcmc_DDPGP`). The

outputs from `mcmc_DDPGP` are then used as the inputs for the other functions to evaluate and plot the estimated posterior predicted density, survival, and hazard functions for new observations/patients. The package also includes a function for evaluating posterior mean survival for specified values of the covariate vector. For example, this allows the user to determine the optimal value of a specific covariate (with the other covariates fixed) that maximizes posterior expected survival time. The R package `DDPGPSurv` can be downloaded from <https://cran.r-project.org/web/packages/DDPGPSurv/index.html>.

Current standard methods for survival analysis typically involve the Kaplan-Meier (KM) estimator for unadjusted survival times with independent right censoring, accelerated failure time (AFT) models, or the Cox proportional hazards (PH) model. The KM estimator is a non-parametric statistic and is constructed using a finite number of conditional probabilities of survival at successive time intervals. To analyze the effects of specific covariates using the KM estimator, the most common approach is simply to compute the KM for particular patient subsets that may be defined from  $\mathbf{x}$ , which reduces reliability. AFT regression models are fully parametric, which may be problematic if the baseline hazard function does not fit the specified AFT distribution. Comparisons between the DDP-GP and AFT models via simulations show that the DDP-GP is more robust, with much more accurate predictions across a range of various distributions (Weibull, lognormal, exponential). That is, if the distribution selected for the AFT model does not match the truth, the predictions will be inaccurate. The Cox model, which is semi-parametric, relies on the PH assumption, which states that the each covariate has a constant effect on the hazard function that does not vary over time. This assumption may not always be true, and it is not required by the DDP-GP model. Additionally, as with any BNP model, the DDP-GP accommodates irregularly shaped survival distributions, for example having multiple modes. Thus, the DDP-GP, implemented by the `DDPGPSurv` package, provides many advantages over these conventional methods, including robustness and accuracy across a wide range of possible distributions.

## 5. Simulation Studies

We conducted simulation studies to evaluate the DDP-GP model in terms of estimation of survival densities and optimal personalized AUC ranges, with the data simulated to mimic the structure of the allo-SCT dataset. We generated  $T$ =survival time, the covariates  $x_1$  = age,  $x_2$  = AUC, and  $x_3$  = CR status for each patient, as follows. Let  $LN(m, s)$  denote a log normal distribution with location and scale parameters  $m$  and  $s$ , and let  $\mathbf{x}_i = (x_{i,1}, x_{i,2}, x_{i,3})$  denote the covariates for patient  $i$ . Patients' ages and AUC values were sampled with replacement from the actual ages and AUC values in the allo-SCT dataset. We generated patients' CR statuses as independently and identically distributed (i.i.d.) binary variables from a Bernoulli(0.5). We simulated the  $Y_i$ 's from a lognormal distribution,  $Y_i | \mathbf{x}_i \sim LN(\mu(\mathbf{x}_i), \sigma_0^2)$ , where the location parameter is the following function of  $\mathbf{x}_i$ ,

$$\mu(\mathbf{x}_i) = 4 - 0.1x_{i,1} + 0.7x_{i,2} + 0.3x_{i,3} - 0.07x_{i,2}^2 - 0.1x_{i,1}x_{i,2} + 0.2x_{i,2}x_{i,3} - 0.18x_{i,1}x_{i,2}x_{i,3},$$

for  $i = 1, \dots, n$ , and  $\sigma_0 = 0.4$ . We deliberately designed the form of  $\mu(\mathbf{x}_i)$  based on clinical knowledge, including a quadratic term for AUC to reflect the fact that a Bu-SE that is either too high or too low is associated with shorter survival time. We also included interaction terms between AUC and covariates so that the relationship between survival and AUC may vary depending on each patient’s age and CR status.

We considered two scenarios, one with  $n = 200$  observations without censoring and the other with  $n = 200$  and 25% censoring. For each scenario,  $B = 100$  trials were simulated, and the proposed DDP-GP survival regression model was fit to each simulated dataset. The MCMC sampler was implemented for posterior inference and run for 5,000 iterations with an initial burn-in of 2,000 iterations, thinned by 10. We used the R package *coda* to check convergence, and both traceplots and autocorrelation plots (not shown) to check mixing of the Markov chain, which showed no convergence problems.

### 5.1. Survival density estimation

For simulated trials indexed by  $b = 1, \dots, B$ , let  $\bar{S}_b(t | \mathbf{x}) = p(Y_{n+1} \geq t | \mathbf{x}_{n+1} = \mathbf{x}, \text{data})$  denote the posterior predicted probability that a future patient  $n + 1$  with covariate  $\mathbf{x}$  in trial  $b$  survives beyond time  $t$ . To estimate  $\bar{S}_b(t | \mathbf{x})$  using our package *DDPGPSurv*, we first run the MCMC using the function *mcmc\_DDPGP*. Then, the output from the function *mcmc\_DDPGP* serves as the input to the function *DDPGP\_Surv*, which returns the mean and 95% credible intervals for the survival function across the saved MCMC posterior samples. Using the empirical covariate distribution  $\frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{x}_i}$  to marginalize w.r.t.  $\mathbf{x}_{n+1}$  and averaging across simulations, we get

$$\bar{S}(t) = \frac{1}{B} \sum_{b=1}^B \frac{1}{n} \sum_{i=1}^n \bar{S}_b(t | \mathbf{x}_i).$$

For comparators, we considered six alternative methods. First, we fit AFT regression models using either lognormal or Weibull distributions by assuming

$$\log(Y_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,2}^2 + \beta_5 x_{i,1} x_{i,2} + \beta_6 x_{i,2} x_{i,3} + \beta_7 x_{i,1} x_{i,3} + \sigma \epsilon_i. \quad (5.1)$$

Here, assuming a normal distribution on  $\epsilon_i$  implies that  $Y_i$  follows a lognormal distribution, while the extreme value distribution assumption on  $\epsilon_i$  implies that  $Y_i$  follows a Weibull distribution. We also considered two flexible semiparametric survival methods that model the baseline survival using a Polya Trees (PT) prior (Hanson and Johnson, 2002) or a transformed Bernstein polynomials (TBP) prior (Zhou and Hanson, 2017), respectively. Both models were implemented in the R package *spBayesSurv*. We assumed the AFT regression model as the frailty model in both the PT and TBP methods with the same setup as in (5.1). Lastly, we compared the proposed DDP-GP model to two fully nonparametric survival models using random forests (RF) (Ishwaran et al., 2008) and Bayesian additive regression trees (BART) (Sparapani et al., 2016). We used the R packages *randomForestSRC* and *BART* to implement the RF method and the BART method, respectively.

Figure 2 compares  $\bar{S}(\cdot)$  estimated under the DDP-GP model to the simulation truth,

$$S_0(t) = \frac{1}{B} \sum_{b=1}^B \frac{1}{n} \sum_{i=1}^n S_0(t | \mathbf{x}_i),$$

the maximum likelihood estimates (MLEs) obtained under each of the two AFT models, and the estimated survival curves under the PT, TBP, BF, and BART methods. In each scenario, the true curve is given as a solid black solid line and the posterior mean survival function under the DDP-GP model as a solid red line with point-wise 95% posterior credible bands as two dotted red lines. In both scenarios, the DDP-GP model based estimate, as well as the RF

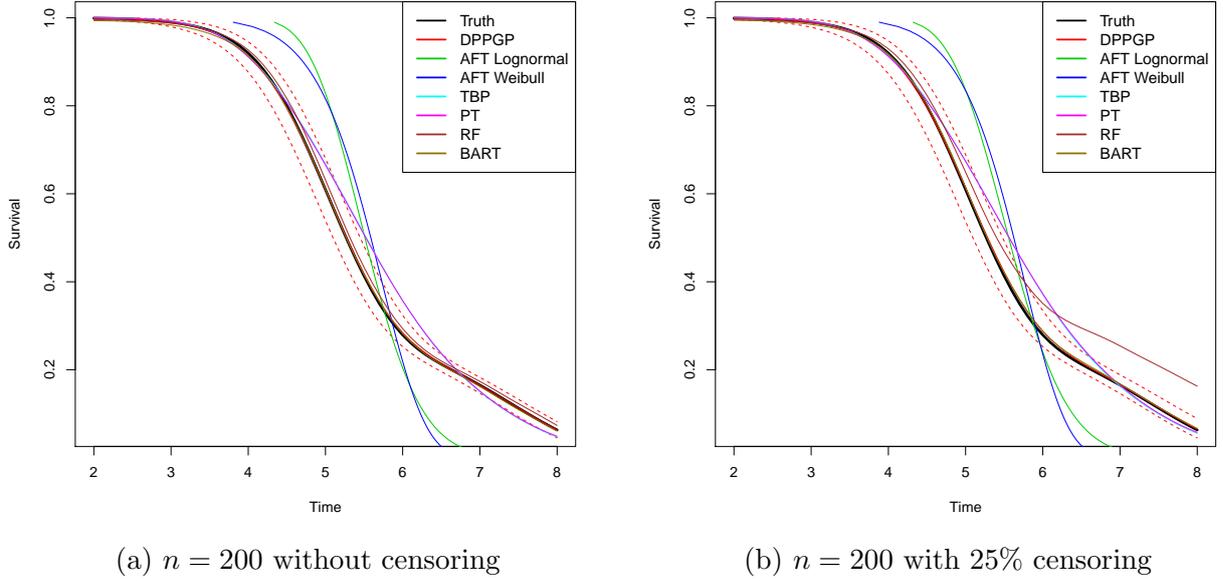


Figure 2: Survival function estimates for the simulated data, with survival time on the log scale. True survival functions are in black, and estimated posterior mean survival functions under the DDP-GP model are in red with point-wise 95% credible bands as two dotted red lines, for  $n = 200$  (left) and  $n = 200$  with 25% censoring (right). For comparators, we also show the survival function estimates under AFT regression models using the lognormal and Weibull distributions, TBP, PT, RF, and BART.

and BART methods, reliably recovered the shape of the true survival function, while the four other methods (AFT Lognormal, AFT Weibull, TBP, and PT) showed substantial bias.

## 5.2. Personalized optimal AUC estimation

For the simulated data, we next evaluated the ability of the DDP-GP survival regression model to estimate optimal personalized AUC ranges, computed by (3.6). This estimation can be performed by the function `DDPGP_meansurvival` in our R package `DDPGPSurv`. This function takes the output from `mcmc_DDPGP` and calculates the posterior mean survival times and 95% credible intervals for patients of interest. Figure 3 compares the simulated true optimal AUC and the optimal AUC range estimates for a 30-year old patient with two different CR statuses,

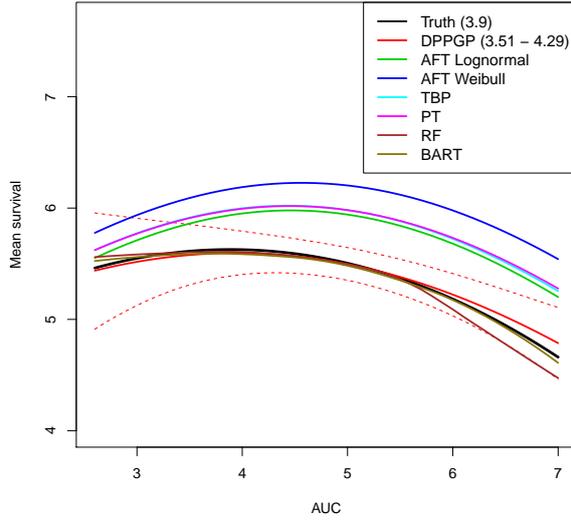
under each of seven models: the DDP-GP model, and the lognormal or Weibull AFT models in (5.1), TBP, PT, RF, and BART. Since the RF and BART methods do not have closed-forms for mean survival times and only provide the estimated survival probabilities at the time points observed in the original data, we estimated the mean survival time as the area under the survival curve in the interval  $(0, t_{\max})$ , where  $t_{\max}$  is the largest observed time point in the data. In Figure 3, the numbers in parentheses in the legend represent the simulated true optimal AUC and the optimal AUC range estimated by the DDP-GP survival regression model. The figure shows that the DDP-GP model accurately estimates the mean survival function and identifies the optimal AUC, with the simulated true AUC being in the estimated optimal AUC range. In contrast, the mean survival functions and the optimal AUC estimates given by the AFT models, PT, and TBP are considerably different from the simulation truth. For instance, when CR=No with 25% censoring, the AFT models with lognormal or Weibull distributions estimate the optimal AUC to be 4.4 and 4.5, respectively, while the true AUC is 3.9. While the RF and BART methods are able to accurately estimate the survival function, the estimates of mean survival are biased, especially when the data are censored.

In summary, the DDP-GP is more robust than alternative methods in the sense that it can better fit the survival functions and more accurately estimate personalized optimal AUC ranges, even while only including the main effects  $(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3})$  in the mean of the Gaussian process prior. In contrast, the alternative parametric and semiparametric models which do not perform as well as the DDP-GP, include not only main effects but also quadratic terms and interactions between covariates as in (5.1). This illustrates an important advantage of the DDP-GP model. It allows one to include covariates as simple linear combinations, but still is able to identify quite general interactions that are not limited to conventional multiplicative interaction terms, such as  $\beta_1 AUC \times age + \beta_2 AUC \times CR$ , that typically are included in the linear components of conventional Cox or AFT models. Such a construction is extremely useful especially when the covariates are high-dimensional, in which case including all the interactions among covariates in the regression model is infeasible. We also present one additional simulation study with ten covariates and a more complex mean structure in the supplement Section 2, demonstrating that the DDP-GP model can accurately recover the simulation truth under more complicated scenarios and compare favorably to alternative methods.

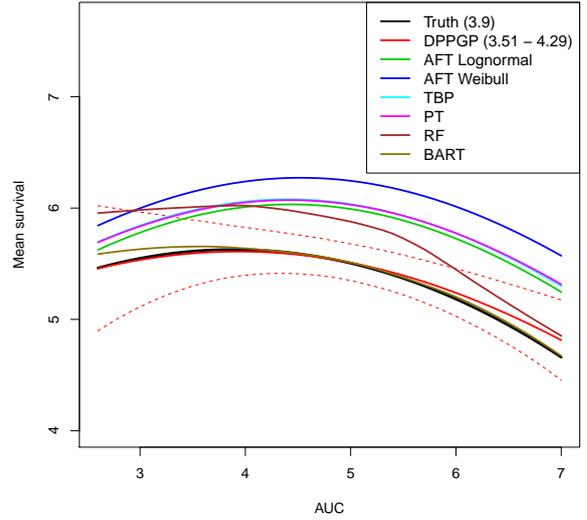
## 6. IV busulfan Data Analysis

While an optimal AUC interval has been determined previously for use in all patients (Andersson et al., 2002), the underlying statistical analyses motivating this assume homogeneity, and thus do not allow the possibility that the optimal interval may vary non-trivially with patient characteristics. Here, we approach the problem differently by estimating mean survival time as a function of (age, CR status, AUC), allowing the possibility that the effect of AUC on survival may vary with age and CR.

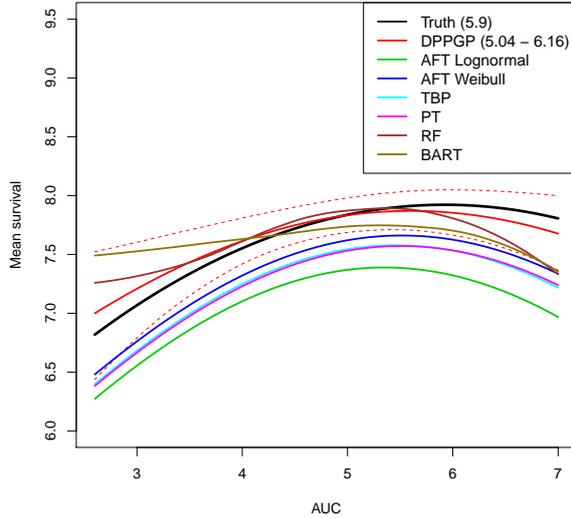
The AUC values in our analysis are in units of thousands of mean daily  $\mu M$ /\*min. An initial analysis of the IV Bu dataset using Kaplan-Meier estimates is given in Figure 4. We divided



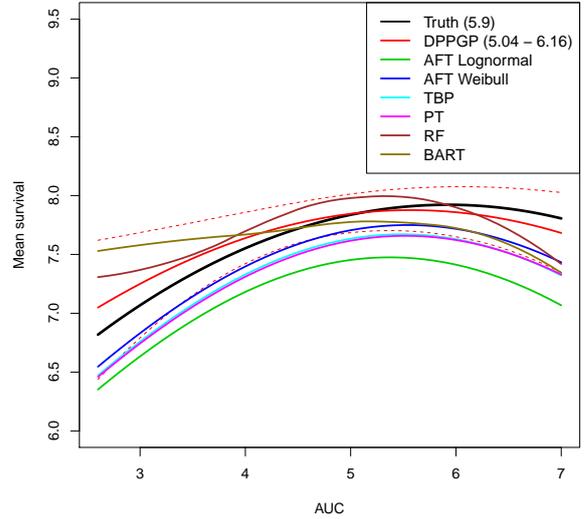
(a)  $n = 200$  without censoring, CR=No



(b)  $n = 200$  with 25% censoring, CR=No



(c)  $n = 200$  without censoring, CR=Yes



(d)  $n = 200$  with 25% censoring, CR=Yes

Figure 3: Optimal AUC estimation for the simulated data, with both survival time and AUC on the log scale. True mean survival functions versus AUC are in black and estimated mean survival functions under the DDP-GP model are in red with point-wise 95% credible bands as two dotted red lines for  $n = 200$  (left plots) and  $n = 200$  with 25% censoring (right plots). For the comparators, we also show the mean survival function estimates under AFT regression models using the lognormal and Weibull distributions, TBP, PT, RF, and BART. The numbers in parentheses in the legend are the true and estimated optimal AUC values.

patients into four groups based on CR status and age, dichotomized as being above or below the median age of 49, and plotted their survival probabilities. Figure 4 illustrates the well-known

fact that being in CR at transplant yields higher survival probabilities. Similarly, younger patients are also expected to have higher survival probabilities. The  $p$ -value obtained from the log rank test comparing the survival distributions between the four groups is significant, indicating that CR and age are important covariates for any survival regression model. The cut-off 49 for dichotomizing age was chosen for convenience, however, as is commonly done in survival analyses. In addition to loss of information about the joint effect of age and CR status on survival caused by dichotomizing age, the reliability of each Kaplan-Meier estimate is reduced because it is based on a subsample.

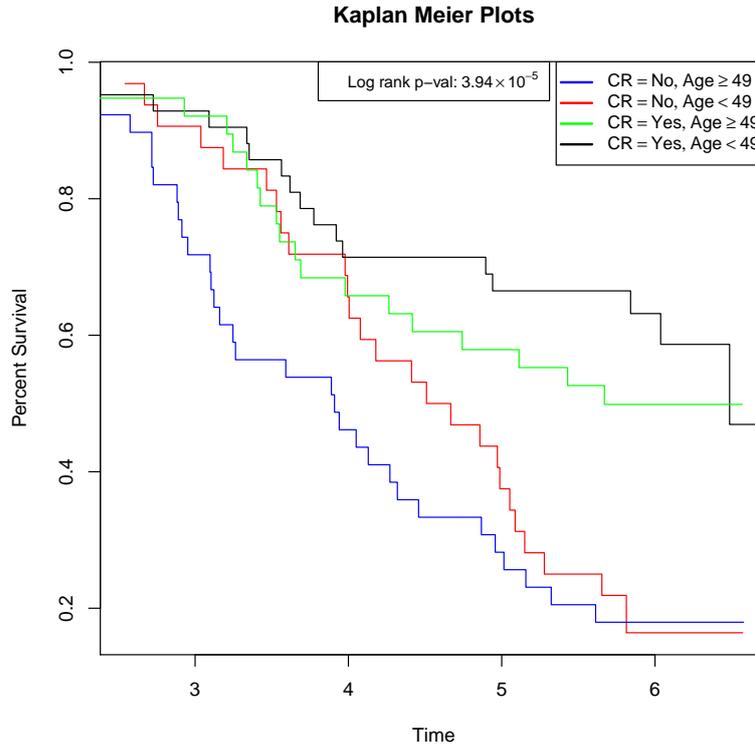


Figure 4: Kaplan Meier Plots. The time in weeks (log scale) versus probability of survival for four different groups are plotted. The  $p$ -value from the log rank test for comparison between the survival distributions between the four groups is given at the top of the figure.

We fit the DDP-GP survival regression model to the allo-SCT dataset with 10,000 Gibbs sampler iterations and a burn-in of 5,000 iterations. The estimated posterior survival distributions with 95% credible intervals under the DDP-GP for patients with different CR statuses and ages 30, 40, 50, or 60, given AUC=5, are shown in Figure 5, respectively. For each (CR status, age) combination, the optimal AUC range is defined as the AUC value that maximizes estimated posterior mean survival,  $\pm 10\%$ . Given CR status and AUC, Figure 5 shows that the estimated posterior mean survival function decreases for older patients, agreeing with what was seen in the preliminary Kaplan-Meier estimates.

For the eight combinations of CR status and Age, we calculated predicted posterior mean survival time as a function of AUC, to address the primary goal of the analyses. These plots

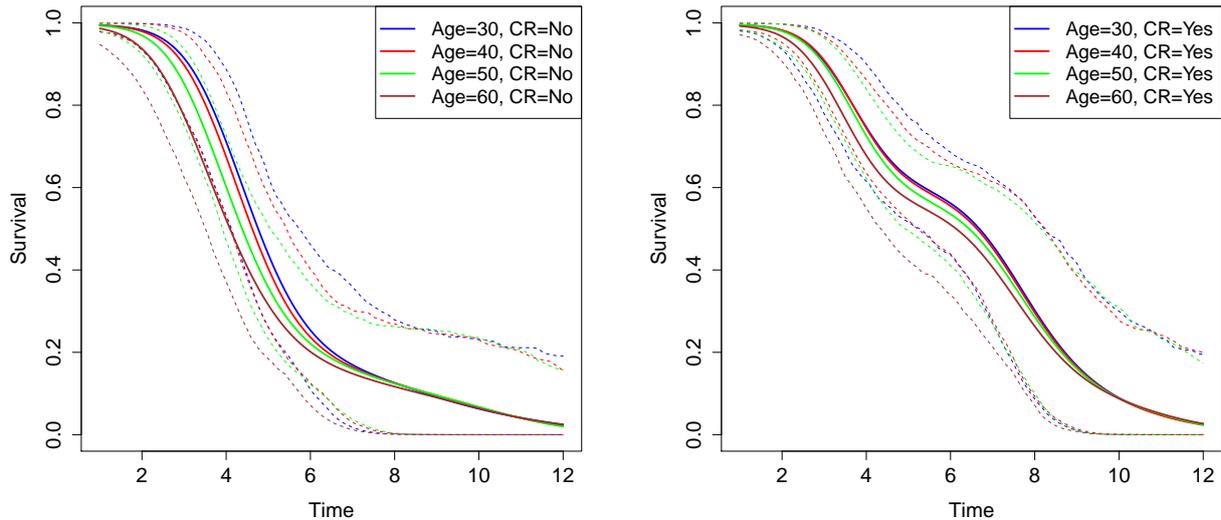


Figure 5: Estimated survival functions under the DDP-GP survival regression model for patients with different CR status (Yes or No) and ages (30, 40, 50, 60). The patients are assigned  $AUC=5$ . The dashed lines represent the point-wise 95% credible intervals for each survival curve.

are given in Figure 6. Our analyses confirm the existence, for each combination of CR status and Age, of an optimal AUC range that yields higher expected survival times compared to an AUC that is either below or above the optimal range. A very important inference is that these optimal AUC ranges differ substantially between many of the (CR status, Age) combinations. This has extremely important therapeutic implications when choosing an individual patient’s targeted AUC. For example, the optimal AUC interval for a patient not in CR with Age=50 is  $4.7 \pm 0.47 = [4.23, 5.17]$  compared with the optimal interval  $5.8 \pm 0.58 = [5.22, 6.38]$  for a patient in CR with Age=40. Since these intervals are disjoint, they suggest that these two patients should have very different targeted AUC values to maximize their expected survival times. The estimated mean survival times versus AUC under the alternative methods, PT, TBP, RF, and BART, are included in the supplement Section 3. There are no meaningful patterns we can observe in these figures.

In contrast with our inferences, (Bartelink et al., 2016) concluded that CR status has a negligible effect on the optimal AUC. However, the results reported by Bartelink et al. (2016) were based on data from a large number of different medical centers, many different pretransplant conditioning regimens were used, the PK-data were obtained from different laboratories, with a very heterogeneous pediatric patient population having a large number of different diagnostic categories, including patients with malignant and non-malignant genetic disorders. In contrast, our analyses are based on a much more homogeneous dataset. Our results indicate that CR status is an important covariate, and that the optimal dose of AUC is higher for patients who are in CR at transplant. Furthermore, the increased optimal AUC for patients in CR at transplant versus patients not in CR is much larger in older patients, whereas these

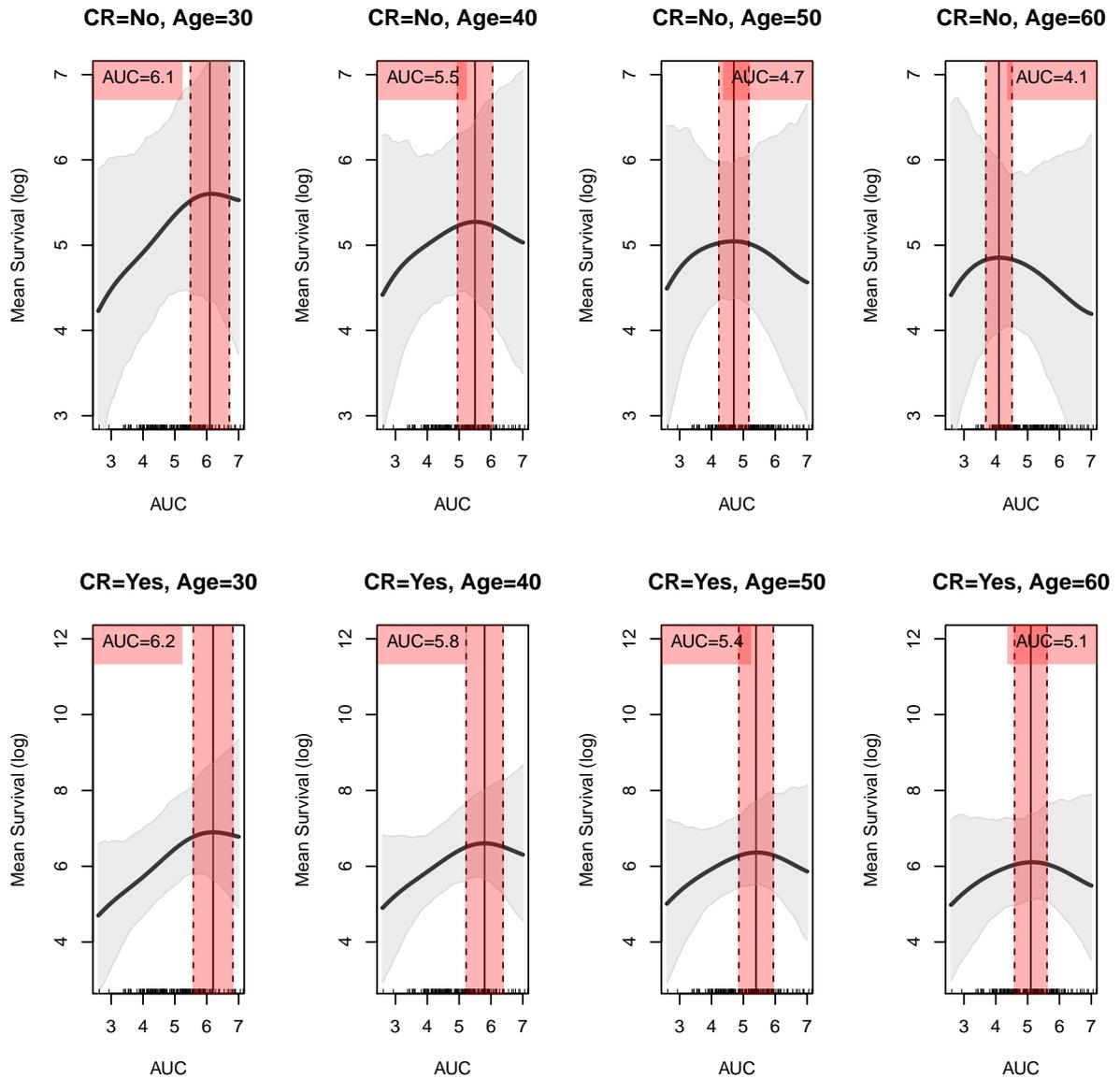


Figure 6: Mean log survival time estimates under the DDP-GP model, as a function of AUC, for each of eight (CR status, Age) combinations. The gray area in each plot represents the 95% credible interval for estimated mean survival, and the tick marks on the horizontal axis (rug plot) indicate the AUC values for patients in the data set. The red area represents the optimal AUC range, defined as the estimated mean  $\pm 10\%$ .

differences appear negligible in adolescents or young adults, similar to what was reported by Bartelink, et al. (2016). Our results also demonstrate that, across all ages, mean survival time for patients in CR is larger compared with those not CR.

To further illustrate how the optimal AUC ranges change with both CR and Age, we plotted the optimal AUC ranges as Age is varied continuously, for CR=Yes and CR=No, in Figure 7. The negative association between optimal AUC and Age is clearly shown by this figure. It also shows that, while CR status has virtually no effect on the optimal AUC interval for very young patients with Age  $\leq 28$ , the optimal AUC for patients in CR at transplant is increasingly higher as Age increases, with the optimal intervals for CR = Yes versus CR = No becoming completely disjoint for patients above 55 years of age. Thus, the lower portions of the curves in Figure 7 for Age  $\leq 28$ , agree with the conclusion of (Bartelink et al., 2016) for pediatric and adolescent patients, while the higher portions for Age  $> 28$ , provide news insights. Again, this demonstrates the importance of considering both CR status and Age when planning a targeted AUC for a patient with a diagnosis of AML or MDS.

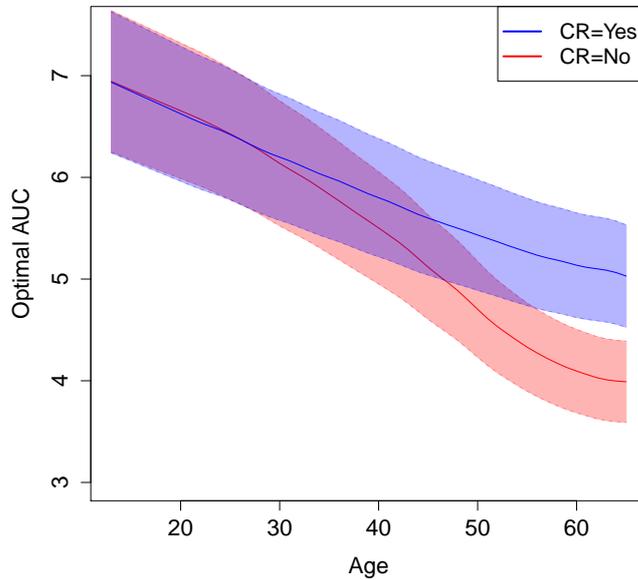


Figure 7: Optimal AUC ranges versus age given CR status. The blue and red lines represent the optimal AUC for CR=Yes and No, respectively. The optimal AUC ranges are represented by the shaded regions above and below the optimal AUC.

## 7. Conclusions

We have proposed an extended Bayesian nonparametric DDP-GP model for survival regression having a generalized covariance structure, studied it by simulation, and applied it to estimate personalized optimal dose intervals for IV busulfan in allo-SCT for AML/MDS. Our simulations, constructed to mimic the dataset, show that the DDP-GP model provides more

accurate survival function estimates and optimal AUC range estimates compared with conventional parametric to AFT models. Our analyses of the IV busulfan allo-SCT dataset identified optimal AUC intervals, varying with the patient’s CR status and Age, that previously have not been known for this treatment. Our results may have profound therapeutic implications, since they provide a basis for personalized medicine by enabling physicians to prospectively assign an optimized therapeutic target interval for each patient based on his/her CR status and age.

More generally, we have developed an R package, `DDPGPSurv`, that implements the DDP-GP model for a broad range of survival regression analyses. While the DDP-GP is more complex than conventional survival regression models, its robustness and broad applicability make it an attractive methodology for survival analysis. The DDP-GP based data analysis reported here, while important in its own right, identified a nonlinear three-way interaction between age, CR status, and AUC in their joint effect on survival time, as shown by Figures 6 and 7. This pattern was identified despite the fact, noted above, that only the main effects were included in the mean of the Gaussian process prior via the linear term  $\beta_0 + \beta_1 Age + \beta_2 CR + \beta_3 AUC$ . This is because the DDP-GP is essentially a mixture model, hence it can identify complex patterns in the data that may be missed by conventional models. For the allo-SCT IV busulfan data, this may be related to the multi-modality of the survival time distribution, seen in Figure 1. This illustrates the practical advantage that, when applying the DDP-GP, one need not guess or search for complex patterns in the linear term of the covariates, as is done routinely when applying conventional survival regression models.

## Acknowledgements

Peter Thall’s research was supported by NCI grant 5-R01-CA083932.

## References

- Andersson, B., Thall, P., Valdez, B., Milton, D., Al-Atrash, G., Chen, J., Gulbis, A., Chu, D., Martinez, C., Parmar, S. et al. (2016) Fludarabine with pharmacokinetically guided IV busulfan is superior to fixed-dose delivery in pretransplant conditioning of AML/MDS patients. *Bone Marrow Transplantation*.
- Andersson, B. S., Madden, T., Tran, H. T., Hu, W. W., Blume, K. G., Chow, D. S.-L., Champlin, R. E. and Vaughan, W. P. (2000) Acute safety and pharmacokinetics of intravenous busulfan when used with oral busulfan and cyclophosphamide as pretransplantation conditioning therapy: a phase i study. *Biology of Blood and Marrow Transplantation*, **6**, 548–554.
- Andersson, B. S., Thall, P. F., Madden, T., Couriel, D., Wang, X., Tran, H. T., Anderlini, P., De Lima, M., Gajewski, J. and Champlin, R. E. (2002) Busulfan systemic exposure relative to regimen-related toxicity and acute graft-versus-host disease: defining a therapeutic window for iv bucy2 in chronic myelogenous leukemia. *Biology of Blood and Marrow Transplantation*, **8**, 477–485.

- Bartelink, I. H., Bredius, R. G., Belitser, S. V., Suttorp, M. M., Bierings, M., Knibbe, C. A., Egeler, M., Lankester, A. C., Egberts, A. C., Zwaveling, J. et al. (2009) Association between busulfan exposure and outcome in children receiving intravenous busulfan before hematologic stem cell transplantation. *Biology of Blood and Marrow Transplantation*, **15**, 231–241.
- Bartelink, I. H., Lalmohamed, A., van Reij, E. M., Dvorak, C. C., Savic, R. M., Zwaveling, J., Bredius, R. G., Egberts, A. C., Bierings, M., Kletzel, M. et al. (2016) Association of busulfan exposure with survival and toxicity after haemopoietic cell transplantation in children and young adults: a multicentre, retrospective cohort analysis. *The Lancet Haematology*, **3**, e526–e536.
- Bredeson, C., LeRademacher, J., Kato, K., DiPersio, J. F., Agura, E., Devine, S. M., Appelbaum, F. R., Tomblyn, M. R., Laport, G. G., Zhu, X. et al. (2013) Prospective cohort study comparing intravenous busulfan to total body irradiation in hematopoietic cell transplantation. *Blood*, **122**, 3871–3878.
- Chipman, H. A., George, E. I. and McCulloch, R. E. (2012) BART: Bayesian additive regression trees. *Annals of Applied Statistics*, **6**, 266–298.
- Copelan, E. A., Hamilton, B. K., Avalos, B., Ahn, K. W., Bolwell, B. J., Zhu, X., Aljurf, M., Van Besien, K., Bredeson, C. N., Cahn, J.-Y. et al. (2013) Better leukemia-free and overall survival in aml in first remission following cyclophosphamide in combination with busulfan compared to tbi. *Blood*, **122**, 3863–3870.
- De Iorio, M., Johnson, W. O., Müller, P. and Rosner, G. L. (2009) Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics*, **65**, 762–771.
- De Lima, M., Couriel, D., Thall, P. F., Wang, X., Madden, T., Jones, R., Shpall, E. J., Shahjahan, M., Pierre, B., Giralt, S. et al. (2004) Once-daily intravenous busulfan and fludarabine: clinical and pharmacokinetic results of a myeloablative, reduced-toxicity conditioning regimen for allogeneic stem cell transplantation in aml and mds. *Blood*, **104**, 857–864.
- Dix, S., Wingard, J., Mullins, R., Jerkunica, I., Davidson, T., Gilmore, C., York, R., Lin, L., Devine, S., Geller, R. et al. (1996) Association of busulfan area under the curve with veno-occlusive disease following bmt. *Bone marrow transplantation*, **17**, 225–230.
- Escobar, M. D. and West, M. (1995) Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577–588.
- Ferguson, T. S. et al. (1973) A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209–230.
- Geddes, M., Kangarloo, S. B., Naveed, F., Quinlan, D., Chaudhry, M. A., Stewart, D., Savoie, M. L., Bahlis, N. J., Brown, C., Storek, J. et al. (2008) High busulfan exposure is associated with worse outcomes in a daily iv busulfan and fludarabine allogeneic transplant regimen. *Biology of Blood and Marrow Transplantation*, **14**, 220–228.

- 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.
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H. and Lauer, M. S. (2008) Random survival forests. *Annals of Applied Statistics*, **2**, 841–860.
- Kanakry, C. G., O'Donnell, P. V., Furlong, T., De Lima, M. J., Wei, W., Medeot, M., Mielcarek, M., Champlin, R. E., Jones, R. J., Thall, P. F. et al. (2014) Multi-institutional study of post-transplantation cyclophosphamide as single-agent graft-versus-host disease prophylaxis after allogeneic bone marrow transplantation using myeloablative busulfan and fludarabine conditioning. *Journal of Clinical Oncology*, **32**, 3497–3505.
- Kontoyiannis, D. P., Andersson, B. S., Lewis, R. E. and Raad, I. I. (2001) Progressive disseminated aspergillosis in a bone marrow transplant recipient: response with a high-dose lipid formulation of amphotericin b. *Clinical Infectious Diseases*, **32**, e94–e96.
- Ljungman, P., Hassan, M., Bekassy, A., Ringden, O. and Öberg, G. (1997) High busulfan concentrations are associated with increased transplant-related mortality in allogeneic bone marrow transplant patients. *Bone Marrow Transplantation*, **20**, 909–913.
- MacEachern, S. N. (1999) Dependent nonparametric processes. In *ASA proceedings of the Section on Bayesian Statistical Science*, 50–55. American Statistical Association, pp. 50–55, Alexandria, VA.
- MacKay, D. (1999) Introduction to Gaussian processes. *Tech. rep.*, Cambridge University, <http://wol.ra.phy.cam.ac.uk/mackay/GP/>.
- McCune, J., Gooley, T., Gibbs, J., Sanders, J., Petersdorf, E., Appelbaum, F., Anasetti, C., Risler, L., Sultan, D. and Slattery, J. (2002) Busulfan concentration and graft rejection in pediatric patients undergoing hematopoietic stem cell transplantation. *Bone Marrow Transplantation*, **30**, 167.
- Müller, P. and Mitra, R. (2013) Bayesian nonparametric inference – Why and how. *Bayesian Analysis*, **8**, 269–302.
- Müller, P. and Rodriguez, A. (2013) Nonparametric Bayesian inference. *IMS-CBMS Lecture Notes. IMS*, **270**.
- Nagler, A., Rocha, V., Labopin, M., Unal, A., Ben Othman, T., Campos, A., Volin, L., Poire, X., Aljurf, M., Masszi, T. et al. (2013) Allogeneic hematopoietic stem-cell transplantation for acute myeloid leukemia in remission: comparison of intravenous busulfan plus cyclophosphamide (Cy) versus total-body irradiation plus Cy as conditioning regimen – a report from the acute leukemia working party of the European group for blood and marrow transplantation. *Journal of Clinical Oncology*, **31**, 3549–3556.

- Rasmussen, C. and Williams, C. (2006) *Gaussian Processes for Machine Learning*. ISBN 0-262-18253-X. MIT Press.
- Russell, J. A., Kangaroo, S. B., Williamson, T., Chaudhry, M. A., Savoie, M. L., Turner, A. R., Larratt, L., Storek, J., Bahlis, N. J., Shafey, M. et al. (2013) Establishing a target exposure for once-daily intravenous busulfan given with fludarabine and thymoglobulin before allogeneic transplantation. *Biology of Blood and Marrow Transplantation*, **19**, 1381–1386.
- Santos, G. W., Tutschka, P. J., Brookmeyer, R., Saral, R., Beschoner, W. E., Bias, W. B., Braine, H. G., Burns, W. H., Eifenbein, G. J., Kaizer, H. et al. (1983) Marrow transplantation for acute nonlymphocytic leukemia after treatment with busulfan and cyclophosphamide. *New England Journal of Medicine*, **309**, 1347–1353.
- Sethuraman, J. (1994) A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**, 639–650.
- Slattery, J., Clift, R., Buckner, C., Radich, J., Storer, B., Bensinger, W., Soll, E., Anasetti, C., Bowden, R., Bryant, E. et al. (1997) Marrow transplantation for chronic myeloid leukemia: the influence of plasma busulfan levels on the outcome of transplantation. *Blood*, **89**, 3055–3060.
- Sorrer, M. L., Maris, M. B., Storb, R., Baron, F., Sandmaier, B. M., Maloney, D. G. and Storer, B. (2005) Hematopoietic cell transplantation (hct)-specific comorbidity index: a new tool for risk assessment before allogeneic hct. *Blood*, **106**, 2912–2919.
- Sorrer, M. L., Storb, R. F., Sandmaier, B. M., Maziarz, R. T., Pulsipher, M. A., Maris, M. B., Bhatia, S., Ostronoff, F., Deeg, H. J., Syrjala, K. L. et al. (2014) Comorbidity-age index: a clinical measure of biologic age before allogeneic hematopoietic cell transplantation. *Journal of Clinical Oncology*, **32**, 3249–3256.
- 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.
- Tokdar, S. T. and Ghosh, J. K. (2007) Posterior consistency of logistic Gaussian process priors in density estimation. *Journal of Statistical Planning and Inference*, **137**, 34–42.
- Tutschka, P. J., Copelan, E. A. and Klein, J. P. (1987) Bone marrow transplantation for leukemia following a new busulfan and cyclophosphamide regimen. *Blood*, **70**, 1382–1388.
- Wachowiak, J., Sykora, K., Cornish, J., Chybicka, A., Kowalczyk, J., Gorczyńska, E., Choma, M., Grund, G. and Peters, C. (2011) Treosulfan-based preparative regimens for allo-hsct in childhood hematological malignancies: a retrospective study on behalf of the ebmt pediatric diseases working party. *Bone marrow transplantation*, **46**.
- 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.

Zhou, H. and Hanson, T. (2017) A unified framework for fitting bayesian semiparametric models to arbitrarily censored survival data, including spatially-referenced data. *Journal of the American Statistical Association*, **In press**.