

Flexible Bayesian Survival Modeling with Semiparametric Time-Dependent and Shape-Restricted Covariate Effects

Thomas A. Murray^{*}, Brian P. Hobbs[†], Daniel J. Sargent[‡], and Bradley P. Carlin[§]

Abstract. Presently, there are few options with available software to perform a fully Bayesian analysis of time-to-event data wherein the hazard is estimated semi- or non-parametrically. One option is the piecewise exponential model, which requires an often unrealistic assumption that the hazard is piecewise constant over time. The primary aim of this paper is to construct a tractable semiparametric alternative to the piecewise exponential model that assumes the hazard is continuous, and to provide modifiable, user-friendly software that allows the use of these methods in a variety of settings. To accomplish this aim, we use a novel model formulation for the log-hazard based on a low-rank thin plate linear spline that readily facilitates adjustment for covariates with time-dependent and proportional hazards effects, possibly subject to shape restrictions. We investigate the performance of our model choices via simulation. We then analyze colorectal cancer data from a clinical trial comparing the effectiveness of two novel treatment regimes relative to the standard of care for overall survival. We estimate a time-dependent hazard ratio for each novel regime relative to the standard of care while adjusting for the effect of aspartate transaminase, a biomarker of liver function, that is subject to a non-decreasing shape restriction.

MSC 2010 subject classifications: Primary 62F15, 62N86; secondary 62F30.

Keywords: Bayesian methods, survival analysis, semiparametric methods, penalized splines, shape-restricted effects, time-dependent effects, colorectal cancer.

1 Introduction

Confirmatory tests of novel medical interventions measure evidence of effectiveness through comparative evaluations of time-to-event endpoints. Yet, treatment comparisons in confirmatory studies routinely rely on statistical models that suffer from several limiting assumptions. Parametric models for the hazard are often inadequate for characterizing the curvature of non-unimodal functions. For example, the Weibull model precludes the possibility of a non-monotone hazard over time. Despite its popularity,

^{*}Department of Biostatistics – Unit 1411, The University of Texas MD Anderson Cancer Center, P.O. Box 301402, Houston, TX 77230-1402, TAMurray@MDAnderson.org

[†]Department of Biostatistics – Unit 1411, The University of Texas MD Anderson Cancer Center, P.O. Box 301402, Houston, TX 77230-1402, BPHobbs@MDAnderson.org

[‡]Mayo Clinic Cancer Center, 200 First Street SW, Rochester, MN 55905, Sargent.Daniel@Mayo.edu

[§]Division of Biostatistics, Mayo Mail Code 303, School of Public Health, Minneapolis, MN 55455-0392, Carli002@UMN.edu; url: <http://www.biostat.umn.edu/~brad>

Cox's proportional hazards model only facilitates estimation of covariate effects, which is limiting when the hazard is of interest or the proportional hazards assumption is violated (Cox, 1975). Recent developments in nonparametric time-to-event modeling have predominately focused on imparting flexibility for estimation of a single feature in isolation, or concern nonparametric methods for continuous characterization of the hazard function (Müller and Mitra, 2013). Yet, flexible models for the hazard are useful for analysis of actual clinical data only in the presence of a framework that accommodates a diverse class of covariate effects for characterizing patient and intervention heterogeneity.

The primary aim of this paper is to provide a unified framework for highly flexible, fully Bayesian analyses of time-to-event data, along with user-friendly software that enables investigators to use our methods in a variety of settings. There are currently few options with available software for conducting a fully Bayesian analysis of time-to-event data that accommodate flexible hazard regression. A presently popular choice is the semiparametric piecewise exponential model, which assumes that the hazard is piecewise constant over time (see Ibrahim et al., 2001, Section 3.1). This approach is tractable and readily facilitates time-dependent and proportional hazards covariate effects, but the discontinuous piecewise constant approximation for the hazard and time-dependent effects, when present, is unrealistic and makes posterior inference sensitive to prespecification of the number and location of hazard function discontinuities along the time axis. We propose a more realistic piecewise linear log-hazard formulation that is still tractable and facilitates time-dependent and proportional hazards covariate effects.

Other flexible survival models have been proposed in the literature. Fahrmeir and Hennerfeind (2003) and Cai et al. (2002) model the log-hazard additively using B-splines. Henschel et al. (2009) extend this additive framework to include random effects in an effort to handle clustering in the data; Hennerfeind et al. (2006) also include structured spatial effects. Sharef et al. (2010) further allow the set of B-spline basis functions to be estimated, and facilitate mixtures with parametric baseline hazard forms. These additive log-hazard models formulated using B-splines all require numerical integration to evaluate the likelihood. Existing software packages that fit B-spline log-hazard models (e.g., `BayesX` and the `splinesurv` package in R) are limited by the fact that they do not facilitate shape-restricted covariate effects and they restrict prior specification to a particular distributional family. We propose a novel log-hazard model formulation using a low-rank thin plate (LRTP) linear spline that results in a closed-form likelihood, thereby avoiding numerical integration when evaluating the likelihood and even making estimation possible in popular Gibbs sampling software, i.e., `JAGS` (Plummer, 2003) and `BUGS` (Lunn et al., 2009), which are not well-suited to numerically-evaluated likelihoods. Our LRTP linear spline formulation of the log-hazard tends to result in fast posterior convergence relative to a similarly tractable truncated basis spline formulation (Crainiceanu et al., 2005), and we provide modifiable software for implementation.

Lin et al. (2014) considered a similar setting to ours for interval-censored survival data and developed a tractable sampling method. However, their approach uses a monotone spline to model the *cumulative* baseline hazard, which is monotone by definition, and their method was limited to proportional hazards covariate effects; see also Gelfand

and Mallick (1995). By way of contrast, we model the log-hazard and thereby avoid shape restrictions on the hazard domain, and introduce monotone splines to model proportional hazards *covariate effects* that are subject to shape restrictions.

The motivating application involves data from a colorectal cancer clinical trial by Goldberg et al. (2004) that assessed the efficacy of three treatment regimes for overall survival during the eight years following treatment initiation on patients with previously untreated metastatic colorectal cancer. The drug combinations considered were irinotecan and bolus fluorouracil plus leucovorin (IFL), and two novel regimes: oxaliplatin and infused fluorouracil plus leucovorin (FOLFOX), and irinotecan and oxaliplatin (IROX). The trial enrolled 795 patients, and randomly allocated 264 to each of the IFL and IROX regimes, and 267 to the FOLFOX regime. A secondary goal is to jointly characterize the effect of aspartate transaminase (AST), a liver prognostic biomarker measured at baseline, that is thought to have a non-decreasing effect on the hazard for death over the domain of AST.

The remainder of this paper evolves as follows. In Section 2, we develop the foundation of our flexible survival model using LRTP linear splines to formulate the log-hazard without covariates. In Section 3, we discuss adjustment for covariates with various types of effects on the hazard, including time-dependent and proportional hazards, possibly subject to shape-restrictions. In Section 4, we compare the proposed model with other common models using simulation. In Section 5, we analyze the colorectal cancer data, using time-dependent adjustment for the effect of each novel treatment relative to the standard of care, while adjusting for AST using a proportional hazards effect that is subject to a non-decreasing shape restriction. We close in Section 6 with an overview of our findings and directions for future work.

2 Log-Hazard Model Formulation

In this section, we develop the foundation for our flexible survival modeling approach without covariates using LRTP linear splines. We assume the data consist of N independent, possibly right-censored observations. We let $\mathbf{t} = (t_1, \dots, t_N)$, assuming without loss of generality that $t_i \in (0, 1]$, and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_N)$, where t_i denotes the i th observed time, and δ_i denotes whether t_i is an event (defining $\delta_i = 1$) or a right-censored observation (defining $\delta_i = 0$), for $i = 1, \dots, N$. The likelihood of these data is

$$L(\mathbf{t}, \boldsymbol{\delta}) = \prod_{i=1}^N h(t_i)^{\delta_i} \exp \left\{ - \int_0^{t_i} h(s) ds \right\} = \prod_{i=1}^N h(t_i)^{\delta_i} \exp \{ -H(t_i) \}, \quad (1)$$

where $h(t)$ is the hazard function and $H(t)$ is the cumulative hazard function (Klein and Moeschberger, 2003, Section 3.5). Since $h(t) \geq 0$, $H(t)$ is non-decreasing by definition. The survival distribution, i.e., $Pr(T > t)$, is $S(t) = \exp\{-H(t)\}$. Using (1), an analysis of time-to-event data can be conducted through a model for the hazard function. For example, the parametric Weibull model defines $h(t) = \psi\phi^{-\psi}t^{\psi-1}$, where $\phi, \psi > 0$.

2.1 Piecewise Exponential Model

The piecewise exponential (PE) model suggested by Ibrahim et al. (2001, Section 3.1) improves upon parametric alternatives by introducing more parameters to accommodate diverse, possibly non-unimodal shapes of the hazard function. We provide some details about the PE model because we will use it as a comparator. The PE model is constructed by partitioning the time axis into K intervals ($0 = \tilde{t}_0 < \tilde{t}_1 < \dots < \tilde{t}_{K-1} < \tilde{t}_K = \infty$), and modeling

$$h(t; \gamma) = \exp(\gamma_k) \quad \text{for } t \in [\tilde{t}_{k-1}, \tilde{t}_k), \quad k = 1, \dots, K. \quad (2)$$

The PE model assumes the hazard is a discontinuous piecewise constant function, where γ_k is the value of the log-hazard function in the k th interval of the prespecified time axis partition. We discuss the choice of a time-axis partition later. Using (2), the cumulative hazard function is

$$H(t; \gamma) = \sum_{k=1}^K [\{\min(\max(t, \tilde{t}_{k-1}), \tilde{t}_k) - \tilde{t}_{k-1}\} \exp\{\gamma_k\}]. \quad (3)$$

The likelihood for the PE model is a tractable expression that arises by plugging (2) and (3) into (1).

To complete the Bayesian specification of the PE model, a prior is assigned to $\gamma = (\gamma_1, \dots, \gamma_K)$. Numerous prior specifications have been proposed; for a discussion established options, see Ibrahim et al. (2001, Section 3). We focus on the random walk prior process proposed by Fahrmeir and Lang (2001) that assumes a smoothing marginal dependence structure on the γ_k 's. This prior process specifies

$$\gamma_1 \sim \mathcal{N}(0, 10^4), \quad \text{and} \quad \gamma_k | \gamma_{k-1}, \sigma_\gamma \sim \mathcal{N}(\gamma_{k-1}, \sigma_\gamma^2), \quad \text{for } k = 2, \dots, K, \quad (4)$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 . This choice is parsimonious and allows strength to be borrowed across successive intervals, thereby improving efficiency. Lastly, motivated by the work of Gelman (2006), we specify $\sigma_\gamma \sim \mathcal{U}(0.01, 100)$, where $\mathcal{U}(a, b)$ denotes a uniform distribution with positive support on the interval $[a, b]$.

2.2 Piecewise Linear Log-Hazard Model

The PE model defined in (2) is popular because it is tractable, yet flexible; however, it imposes unrealistic discontinuities in the resulting hazard estimate. Inference derived from a model that uses a higher-order approximation to the hazard remedies the discontinuity limitation of the PE model; however, since the likelihood expression in (1) requires the evaluation of the definite integral $H(t) = \int_0^t h(s) ds$, mathematical tractability is usually sacrificed. Using a first-order polynomial to model the log-hazard, we can retain tractability while gaining continuity in the resulting hazard estimate.

Penalized splines are a simple and flexible option for a first-order polynomial log-hazard model. Ruppert et al. (2003) discuss various spline constructions, including

B-splines, truncated basis splines and radial basis splines. Crainiceanu et al. (2005) demonstrate that low-rank thin-plate (LRTP) splines, a type of radial basis spline, tend to exhibit fast Markov chain Monte Carlo (MCMC) convergence relative to truncated basis splines. In our experience, B-splines, which require an intractable recursive algorithm to define the basis functions, exhibit MCMC convergence properties and resulting inferences that are similar to LRTP splines. The relative intractability of B-splines compared to LRTP splines has little practical consequence in the context of semiparametric regression, so we find either construction to be an attractive Bayesian modeling choice. However, in the present context of time-to-event outcomes, we prefer a LRTP spline construction because it results in a tractable likelihood expression, whereas a B-spline construction would require numerical integration to evaluate (1).

To formulate our model, we prespecify a partition of the time axis ($0 = \tilde{t}_0 < \tilde{t}_1 < \dots < \tilde{t}_{K-1} < \tilde{t}_K = \infty$) and define

$$\log \{h(t; \boldsymbol{\alpha}^*)\} = \alpha_0^* + \alpha_1^* t + \sum_{k=2}^K \alpha_k^* (|t - \tilde{t}_{k-1}| - |\tilde{t}_{k-1}|), \tag{5}$$

where $\boldsymbol{\alpha}^* = (\alpha_0^*, \alpha_1^*, \dots, \alpha_K^*)'$. In (5), we replace the usual radial basis terms, $|t - \tilde{t}_{k-1}|$, with modified terms, $|t - \tilde{t}_{k-1}| - |\tilde{t}_{k-1}|$, to ensure $\log \{h(0; \boldsymbol{\alpha}^*)\} = \alpha_0^*$. This modification improves MCMC convergence and may ease prior elicitation for α_0^* . Our proposed model in (5) assumes the log-hazard is a piecewise linear function, so we hereafter refer to this model as the piecewise linear (PL) model.

Under (5), the cumulative hazard arises as

$$H(t; \boldsymbol{\alpha}^*) = \sum_{k=1}^K \frac{h(s_k; \boldsymbol{\alpha}^*) \left[1 - e^{-(s_k - \tilde{t}_{k-1}) (\mathbf{u}'_{k,K} \boldsymbol{\alpha}^*_{(-1)})} \right]}{\mathbf{u}'_{k,K} \boldsymbol{\alpha}^*_{(-1)}}, \tag{6}$$

where $s_k = \max\{\min\{t, \tilde{t}_k\}, \tilde{t}_{k-1}\}$, $\boldsymbol{\alpha}^*_{(-1)} = (\alpha_1^*, \dots, \alpha_K^*)'$, $\mathbf{u}'_{k,K} = (\mathbf{1}'_k, -\mathbf{1}'_{K-k})$, for $k = 1, \dots, K$, and $\mathbf{1}'_k$ denotes a k -dimensional row vector of ones. The likelihood for the PL model is a tractable expression that arises by plugging (5) and (6) into (1).

Following the work of Crainiceanu et al. (2005), we implement a series of transformations, including a one-to-one transformation from $\boldsymbol{\alpha}^*$ to $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_K)'$, and then reformulate (5) and (6) in terms of $\boldsymbol{\alpha}$. The details of this procedure are provided in the Appendix. We specify priors for $\boldsymbol{\alpha}$ as follows

$$\alpha_0 \sim \mathcal{N}(0, 10^4), \quad \alpha_1 \sim \mathcal{N}(0, 10^4), \quad \text{and } \alpha_k | \sigma_{\boldsymbol{\alpha}} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{\boldsymbol{\alpha}}^2), \quad \text{for } k = 2, \dots, K. \tag{7}$$

We also assume $\sigma_{\boldsymbol{\alpha}} \sim \mathcal{U}(0.01, 100)$. The marginal prior distribution induced on α_k^* , $k = 2, \dots, K$, has the shape of a double-exponential distribution with mean zero, so the prior defined in (7) shrinks the modified radial basis coefficients in (5) toward zero, thereby smoothing the resulting estimator and resisting overfitting the data.

Partition Specification

The PE model defined in (2) and the proposed PL model defined in (5) each require a partition of the time domain. Specification of K and \tilde{t} can be avoided by treating them as unknown parameters and estimating them in a Bayesian framework, see Sharef et al. (2010). However, the additional computational burden typically does not justify the marginal gains in approximation accuracy over a reasonable prespecified partition. Ruppert (2002) shows that the most important issue is selecting K large enough so that the resulting model can adequately capture the hazard function curvature. Given K , the placement of the \tilde{t}_k 's over the domain of t can be done in any sensible manner, say at the quantiles of the observed event times. For time-to-event data, we prefer equally-spaced partitions because the hazard may still exhibit interesting features in an area where there is a dearth of event times. For example, the hazard may exhibit a sharp drop that results in event times being distributed away from this feature.

For the simulation, we use an equally-spaced partition with $K = 20$ intervals, thereby providing ample flexibility and consistency across scenarios, and computational feasibility. To select a partition for the analysis of the colorectal cancer data, we conduct a grid search using a modification of the deviance information criterion (DIC) (Spiegelhalter et al., 2002). Specifically, we select the partition that minimizes $\overline{D} + K$, instead of $DIC = \overline{D} + p_D$. Since the prior shrinks the effective number of parameters, p_D does not necessarily increase with K , despite our preference for smaller K owing to the added computational burden. Furthermore, the fit as measured by \overline{D} also does not improve by increasing K when K is already large. Therefore, the proposed criterion will identify a partition with small K that still provides a good fit, whereas DIC tends to have difficulties distinguishing among partitions with large K .

3 Covariate Adjustment

In this section, we incorporate covariate adjustment into the PE and PL hazard models developed in Section 2. In practice, data often derive from a heterogeneous population with measured covariates, so robust estimation methods using flexible hazard models are useful only when incorporated into a modeling framework that accommodates a diverse class of covariates. We focus on *baseline* covariates, which assume a fixed value throughout the time period of interest (e.g., gender, race, a biomarker measured at baseline, etc.). The effect a baseline covariate has on the hazard may be *time-dependent* (e.g., the beneficial effect of a novel treatment may diminish relative to the standard treatment over time) or satisfy the *proportional hazards* (i.e., time-independent) assumption. In contrast, *time-varying* covariates assume a value that may change over the course of follow-up (e.g., in-patient versus out-patient status, modifications in the course of treatment, measurements of surrogate markers, etc.), and they too may have a time-dependent or proportional hazards effect on the hazard. We first consider baseline covariates with time-dependent effects, and then address proportional hazards effect, including those subject to shape-restrictions.

3.1 Time-Dependent Effect

We assume the data now include information about a baseline covariate z that is assumed to have a time-dependent effect on the hazard. In the motivating data, the covariate z may indicate assignment to *FOLFOX* or *IROX*. The PE and PL models developed in Section 2 readily extend to accommodate z using an similar formulation of the conditional hazard $h(t|z)$. This extension facilitates investigation of the time-dependent effect of treatment, as well as the hazard functions and survival distributions in each treatment group.

Gamerman (1991) extends the PE model defined in (2) to accommodate z by assuming

$$h(t|z; \boldsymbol{\gamma}) = \exp(\gamma_{0,k} + \gamma_{1,k}z) \text{ for } t \in [\tilde{t}_{k-1}, \tilde{t}_k), \quad k = 1, \dots, K, \quad (8)$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0 \mid \boldsymbol{\gamma}_1)$ and $\boldsymbol{\gamma}_q = (\gamma_{q,1}, \dots, \gamma_{q,K})'$, for $q = 0, 1$. The entries of $\boldsymbol{\gamma}_q$ are allowed to realize distinct values, thereby facilitating time-dependent effects with a piecewise constant structure using the same time axis partition. The cumulative conditional hazard $H(t|z; \boldsymbol{\gamma})$ arises by replacing γ_k in (3) with $(\gamma_{0,k} + \gamma_{1,k}z)$. Typically, the $\boldsymbol{\gamma}_q$ are assumed to be independent *a priori*, whence the prior specification for each can follow analogous to (4).

In a similar fashion, we propose to extend the PL model defined in (5) by assuming

$$\begin{aligned} \log \{h(t|z; \boldsymbol{\alpha}^*)\} &= (\alpha_{0,0}^* + \alpha_{1,0}^* z) + (\alpha_{0,1}^* + \alpha_{1,1}^* z) t \\ &+ \sum_{k=2}^K (\alpha_{0,k}^* + \alpha_{1,k}^* z) (|t - \tilde{t}_{k-1}| - |\tilde{t}_{k-1}|), \end{aligned} \quad (9)$$

where $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_0^* \mid \boldsymbol{\alpha}_1^*)$ and $\boldsymbol{\alpha}_q^* = (\alpha_{q,0}^*, \dots, \alpha_{q,K}^*)'$, for $q = 0, 1$. Like in Section 2.2, we apply a series of one-to-one transformations that result in a parameterization of (9) in terms of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$. The details for this procedure are also provided in the Appendix. We then assume that $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ are independent *a priori*, a standard prior assumption in the context of an additive model, and use prior specifications for $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ analogous to (7).

Using (9), the log-hazard ratio for an individual having $z = 1$ relative to $z = 0$ is a piecewise linear function given by a LRTP linear spline with the previous set of basis functions. In contrast, the extended PE model in (8) assumes that the log-hazard ratio is piecewise constant over time. Therefore, if the time-dependent effect of z is of interest, our extended PL model offers a continuous alternative to the extended PE model. Nevertheless, either approach facilitates a flexible semiparametric estimate for the effect of z over the course of follow-up, whereas a proportional hazards model rigidly assumes that the effect of z is constant. Extending either (8) or (9) to handle an arbitrary number of baseline covariates with time-dependent effects is trivial. Furthermore, extending either model to accommodate a time-varying covariate $z(t)$ with a time-dependent effect requires only a minor adjustment in the calculation of the likelihood to account for the time-varying nature of $z(t)$.

3.2 Proportional Hazards Effect

Assume the data also include information for a baseline continuous covariate x that is thought to have a proportional hazards effect on the hazard function. In this case, we can model $h(t|x, z) = h_0(t|z) \exp\{f(x)\}$, wherein $\exp\{f(x)\}$ denotes the multiplicative effect of x on the conditional baseline hazard, $h_0(t|z)$. When x is continuous, a standard linear regressor (i.e., $f(x) \equiv \beta x$) does not always sufficiently characterize its effect on the conditional baseline hazard, and a more flexible model may thus be required. In this case, modeling the effect of x as a smooth nonlinear function is usually sensible; alternatively, a shape-restricted model may be more appropriate, say if there is a scientific basis to assume that the effect of x is monotonically increasing.

In this subsection, we extend the previously developed PE and PL models to accommodate semiparametric proportional hazards covariate adjustment, with and without shape restrictions. Before doing so, we note that the effect of x factors out in the calculations of the cumulative conditional hazard so that $H(t|x, z) = H_0(t|z) \exp\{f(x)\}$, thereby affording the use of either (8) or (9) as a model for $h_0(t|z)$. Likewise, in the absence of any covariate z with a time-dependent effect, $h_0(t|z) \equiv h(t)$, thereby allowing the use of either (2) or (5) as a model for $h(t)$. It follows that the previous models can be combined with the methods in this subsection, thereby providing a flexible approach that can accommodate the diverse sets of covariates actually encountered in practice.

Smooth Proportional Hazards Effect

If the proportional hazards effect of x is not adequately captured by a linear term, but its effect is smooth, then splines are a natural choice (Ruppert et al., 2003). We use a LRTP cubic spline without sacrificing tractability because proportional hazards effects factor conveniently out of the definite integral in (1). We again prefer LRTP splines for their simple construction and tendency to exhibit fast MCMC convergence, though a B-spline would be another sensible choice. The construction of our LRTP cubic spline model for the effect of x is similar to the approach of Section 2.2, so we briefly provide the details.

We assume without loss of generality that $x \in [0, 1]$ and specify a partition with J equally-spaced pieces ($0 = \tilde{x}_0 < \tilde{x}_1 < \dots < \tilde{x}_J = 1$) over this domain. We use a two dimensional grid search to select J jointly with K . Like K , using too small J may result in a model that does not adequately characterize the effect of x on the hazard, whereas using an unnecessarily large J will increase the computational burden while providing a similar fit as that of a model with a smaller, yet adequate J . Given this partition, we model

$$f(x; \beta^*) = \beta_1^*(x - \bar{x}) + \sum_{j=2}^J \beta_j^* (|x - \tilde{x}_{j-1}|^3 - |\bar{x} - \tilde{x}_{j-1}|^3), \quad (10)$$

where \bar{x} is the sample mean of the covariate. This model fixes $f(\bar{x}; \beta^*) = 0$, thereby ensuring interpretability of the baseline conditional hazard, since $h(t|z, x = \bar{x}) \equiv h_0(t|z)$, and improving MCMC convergence. Using (10), the conditional baseline hazard represents the hazard of an individual having $x = \bar{x}$ and arbitrary z . Recall, either (9) or (8) can be used to model $h_0(t|z)$ in this context.

In practice, we again implement a series of one-to-one transformations leading to (10) being parametrized in terms of $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)'$. The details of this procedure are provided in the Appendix. For prior specification, we assume *a priori* that $\boldsymbol{\beta}$ is mutually independent of the parameters characterizing the baseline conditional hazard (e.g., $\boldsymbol{\alpha}$ or $\boldsymbol{\gamma}$). We then specify the usual LRTP spline prior for $\boldsymbol{\beta}$, i.e., $\beta_1 \sim \mathcal{N}(0, 10^4)$ and $\beta_j | \sigma_{\boldsymbol{\beta}} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{\boldsymbol{\beta}}^2)$ for $j = 2, \dots, J$ (Crainiceanu et al., 2005). Lastly, we specify $\sigma_{\boldsymbol{\beta}} \sim \mathcal{U}(0.01, 100)$.

Shape-Restricted Proportional Hazards Effect

In some settings, understanding of an underlying biological mechanism may justify assuming that the effect of x is of some particular shape. For example, the effect of AST on overall survival is thought to be non-decreasing over its observed domain in our motivating colorectal cancer data. Shively et al. (2011) show that many types of shape restrictions can be imposed tractably through the prior when the effect of x is modeled as a truncated quadratic spline. In this context, motivated by tractability for the constraints, we model the effect of x using a “centered” truncated basis quadratic spline as

$$f(x; \boldsymbol{\psi}^*) = \psi_1^* x + \sum_{j=1}^J \psi_{j+1}^* (x - \tilde{x}_{j-1})_+^2, \tag{11}$$

where $(x)_+^2 = \max(x, 0)^2$. This model fixes $f(0; \boldsymbol{\psi}^*) = 0$, so $h_0(t|z)$ is the conditional hazard for an individual having $x = 0$ and arbitrary z . We considered a “centered” version of the shape-restricted model, i.e., fixing $f(\bar{x}; \boldsymbol{\psi}^*) = 0$, however, MCMC convergence did not improve relative the uncentered version in (11), and the presentation of the uncentered version is simpler.

Using (11), we can impose monotonicity by forcing the first derivative of $f(x; \boldsymbol{\psi}^*)$,

$$f'(x; \boldsymbol{\psi}^*) = \psi_1^* + 2 \sum_{j=1}^J \psi_{j+1}^* (x - \tilde{x}_{j-1})_+,$$

to be non-negative for all $x \in [0, 1]$. Since $f'(x; \boldsymbol{\psi}^*)$ is a piecewise linear function, the local minima in each interval $(\tilde{x}_{j-1}, \tilde{x}_j)$ will be realized at the boundaries, i.e., at $x = \tilde{x}_{j-1}$ or $x = \tilde{x}_j$. It follows that the necessary constraints are identified by evaluating $f'(x; \boldsymbol{\psi}^*)$ at each interval boundary (i.e., $\tilde{x}_0, \dots, \tilde{x}_J$), and requiring that each of the resulting $J + 1$ expressions be non-negative. Doing so, the constraints arise as

$$\psi_1^* \geq 0 \text{ and } \psi_1^* + 2 \sum_{k=1}^j \psi_{k+1}^* (\tilde{x}_j - \tilde{x}_{k-1}) \geq 0, \quad j = 1, \dots, J. \tag{12}$$

Following Shively and Sager (2009), these constraints can be tractably reformulated as $\psi_j \geq 0$, for $j = 1, \dots, J + 1$, after applying the linear transformation $\boldsymbol{\psi} = \mathbf{L}\boldsymbol{\psi}^*$, where \mathbf{L} is a $(J + 1)$ -dimensional lower triangular matrix with $(j + 1)$ th row given by $\{1, 2(\tilde{x}_j - \tilde{x}_0), \dots, 2(\tilde{x}_j - \tilde{x}_{j-1}), \mathbf{0}'\}$. Therefore, these constraints can be imposed by specifying a prior distribution for $\boldsymbol{\psi}$ that has non-negative support.

We assign a hierarchical prior for ψ with non-negative support using mixture distributions as follows

$$\begin{aligned} \psi_j | \iota_j &\sim [\mathcal{N}(0, 10^4)_{[0, \infty)}]^{\iota_j} \times [\delta_0]^{(1-\iota_j)}, \text{ and} \\ \iota_j &\sim \text{Bern}(p_0) \text{ for } j = 1, \dots, J + 1, \end{aligned} \quad (13)$$

where $\mathcal{N}(0, 10^4)_{[0, \infty)}$ denotes a truncated normal distribution with positive support, δ_0 denotes the Dirac delta function with infinite density at zero, and $p_0 \in [0, 1]$ is a prespecified hyperparameter. Since $f'(\tilde{x}_{j-1}; \psi^*) = \psi_j$, the value of p_0 represents the prior probability that $f(x; \psi^*)$ is increasing at \tilde{x}_{j-1} . Thus, p_0 provides control over the probability of invariant intervals in $f(x; \psi^*)$ along the domain of x ; we use a default value of $p_0 = 0.50$. Using a normal distribution truncated to have non-positive support in (13) would instead restrict the proposed effect to be non-increasing.

Our proposed prior specification in (13) deviates slightly from that of Shively and Sager (2009), who use a mixture of a $(J + 1)$ -dimensional multivariate normal distribution truncated to have positive support in each dimension of ψ , and probability masses corresponding to each boundary wherein a subset of ψ is exactly zero. The univariate mixture structure we propose in (13) is substantially easier to construct in practice, and in our experience the resulting posterior estimates are similar. Many other monotone model formulations have been suggested; see, e.g., Brezger and Steiner (2004) and Dunson (2005).

4 Simulations

In this section, we report three simulation studies comparing proposed PL model framework with the PE model framework. First, we compare the estimation of a hazard function for a variety of complex shapes and the resultant survival distribution in a homogeneous population. Second, we compare the estimation of a time-dependent treatment effect. Third, we compare the estimation of a monotone proportional hazards effect.

4.1 Log-Hazard and Survival Distribution Estimation

We evaluate the performance of the PL and PE models for various shapes of the true hazard function in a homogeneous population. We use R simulation runs to do so, wherein we generate pairs (t_i, δ_i) , $i = 1, \dots, N$, of independent, possibly right-censored observation times from a survival distribution with complex hazard $h(t)$. To draw an observation from a survival distribution with hazard $h(t)$, we follow the inverse cumulative density function method of Bender et al. (2005). We specify $h(t)$ such that $H(t)$ is available analytically, then generate an event time y_i from $S(t) = \exp\{-H(t)\}$ by drawing a $u_i \sim \mathcal{U}(0, 1)$, and defining $y_i = H^{-1}\{-\log(u_i)\}$. We solve the latter identity numerically, thereby affording diverse classes of hazard function shapes from which to choose. We also generate censoring times c_i from an independent uniform distribution, and set the observed time $t_i = \min(y_i, c_i)$ and event indicator $\delta_i = I(t_i \leq c_i)$. We use $N = 200$ and specify the survival and censoring distributions so that $S(1)$ is

about 0.10 and the resulting data exhibit about 15% censoring. These choices reflect the approximate characteristics of each treatment group in the colorectal cancer data.

For the r th run, we fit the PE and PL models defined in Sections 2.1 and 2.2 to the generated data and save the posterior mean parameter estimates, denoting these by $\hat{\gamma}^{(r)}$ and $\hat{\alpha}^{(r)}$. We conduct posterior estimation by calling JAGS in R via the R2jags package, and using snowfall for parallel computing. Following 2,000 iterations of burn-in, we used 10,000 iterations from 2 chains for posterior estimation. These choices reflect a preliminary assessment of MCMC convergence based on the potential scale reduction factor and the resulting effective sample size for each parameter, see Gelman et al. (2014, Chapter 11) R code to reproduce each simulation assessment in this paper is available on the fourth author's software webpage, www.biostat.umn.edu/~brad/software.html.

For visual comparison, we calculate the *average* pointwise log-hazard estimate, $\widehat{\log\{h(t)\}}$, and the empirical pointwise 2.5% and 97.5% quantiles at 10,000 equally spaced time points over the unit interval. Specifically, in each of the $R = 200$ runs, we saved the posterior mean log-hazard estimate. We then estimated $\widehat{\log\{h(t)\}}$ from these posterior mean log-hazard estimates. For quantitative comparison, we also calculate the root-integrated square error (RISE) of $\widehat{\log\{h(t)\}}$, defined by

$$\text{RISE} = \int_0^1 \left[\widehat{\log\{h(t)\}} - \log\{h(t)\} \right]^2 dt. \quad (14)$$

We use a Riemannian approximation for (14) based on a grid of 10,000 equally spaced points. For further visual comparison, we display the *best* individual estimate defined as having the smallest RISE among the $R = 200$ posterior mean estimates. The evaluation criteria for the survival distribution $S(t)$ are defined analogously.

The results with $K = 20$ are displayed in Figure 1, wherein each row corresponds to a different scenario. We provide the analytical definitions in the Appendix for the three $h(t)$ considered in this simulation. The log-hazard estimates for the PE model are discontinuous piecewise constant functions, whereas our proposed PL model results in continuous piecewise linear estimates. As evidenced by the results depicted in the first and third row, both approaches struggle to detect a shift in log-hazard beyond 80% follow-up (t-axis) where the data are quite sparse. In this data-sparse region, the PL model results in more variable hazard estimates than those of the PE model. This is because a linear function is more flexible than a constant, and, where the data are sparse, a linear model, e.g., the PL model, will exhibit greater variability than a constant model, e.g., the PE model. In the top row, RISE of the average estimate is slightly larger for the PL model than the PE model; however, the PL model provides a smaller RISE than the PE model when only integrating over the initial 80% of follow-up (0.06 versus 0.07). The PL model provides an improvement in RISE over the PE model for the scenarios depicted the middle and bottom rows. Turning to the survival distribution evaluation, from top to bottom row, the PL model results in a RISE of the average estimate that is 30%, 20% and 15% smaller in magnitude than the PE model. This suggest that the proposed PL model better captures the curvature of the true survival distribution on average than the PE model.

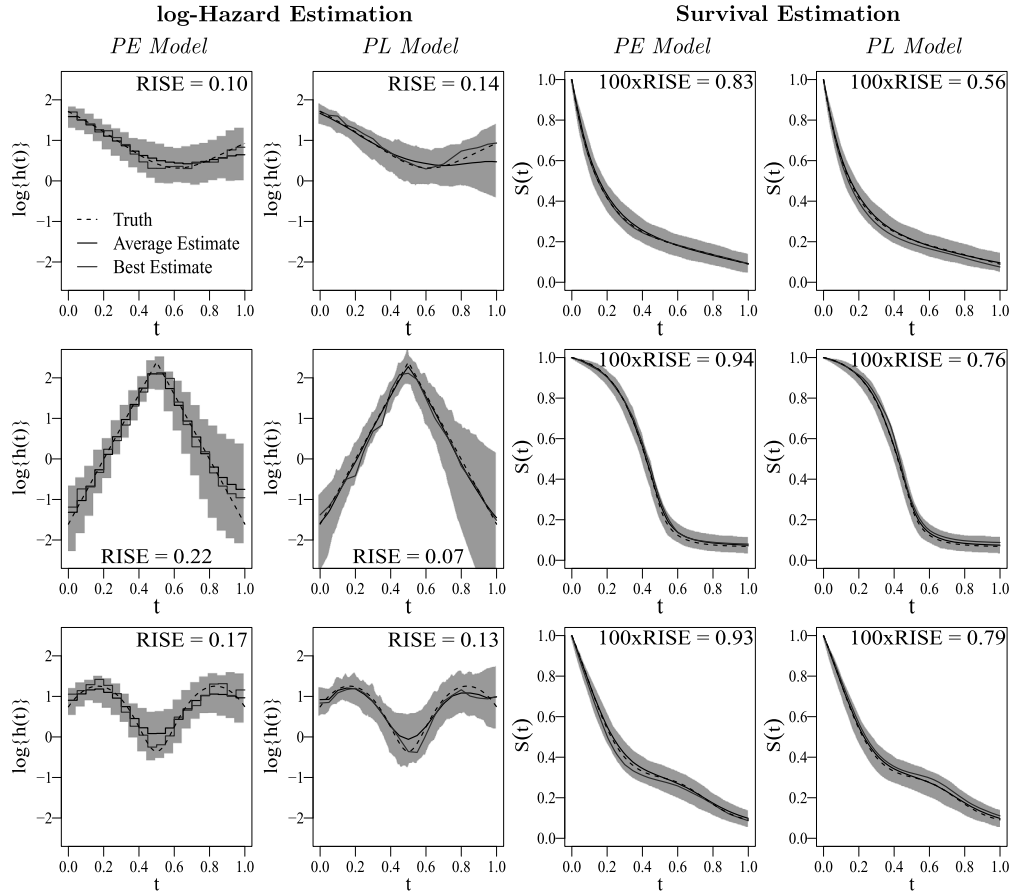


Figure 1: Log-hazard and survival distribution estimates for the PL and PE model with $K = 20$ and $N = 200$. All results are based on $R = 200$ runs and the reported RISE is for the average estimate.

4.2 Time-Dependent Effect Estimation

For this investigation we introduce a binary treatment indicator z with $Pr(z = 1) = Pr(z = 0) = 0.50$. We consider two scenarios, the first where z has a time-dependent effect and the second where it has a proportional hazards effect. We evaluate the performance of the extended PL model with a time-dependent effect defined in (9) compared to the similarly extended PE model defined in (8). We refer to these as the “PL-TD” and “PE-TD” models. For the proportional hazards scenario the PL-TD and PE-TD models provide more flexibility than is needed, so to investigate their possible loss of efficiency, we also fit proportional hazards (PH) versions of the PL and PE models. That is, we assume $h(t|z) = h_0(t) \exp\{\beta z\}$ where $h_0(t)$ is defined by (2) for the PE model and by (5) for the PL model. We refer to these as the “PL-PH” and “PE-PH” models.

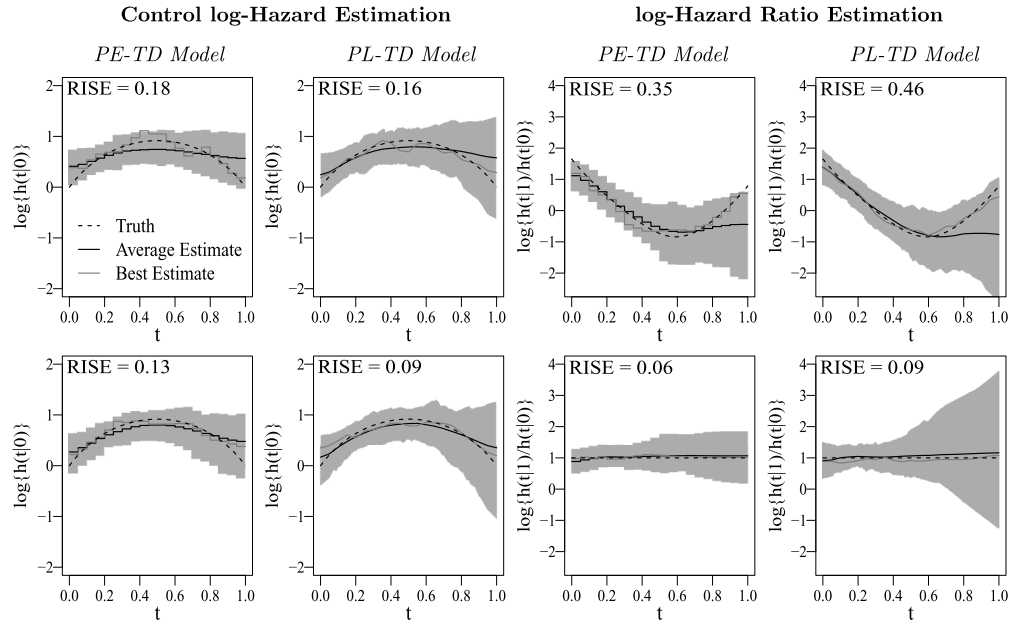


Figure 2: Time-dependent effect estimation comparison of PE-TD and PL-TD models with $N = 200$ and $K = 20$. All results are based on $R = 200$ simulation runs and the reported RISE is for the average estimate.

Lastly, we fit Cox’s PH model using the `coxph()` function from the `survival` package available in R.

To compare the methods, we again generate $N = 200$ survival observations (t_i, δ_i, z_i) from a distribution with a prespecified conditional hazard $h(t|z)$ using the inverse cumulative density function method discussed in (4.1). We then fit each model to these data and save the posterior mean parameter estimates, e.g., $\hat{\alpha}$, $\hat{\gamma}$, and $\hat{\beta}$. We iterate this process over $R = 200$ simulation runs. We compare the models visually using the average pointwise log-hazard estimate for the control group, $\log\{\widehat{h(t|z=0)}\}$, and log-hazard ratio estimate, $\log\{\widehat{h(t|z=1)}\} - \log\{\widehat{h(t|z=0)}\}$. We also calculate the 2.5% and 97.5% pointwise percentiles, RISE of the average estimates, and display the individual estimate with the smallest RISE over all $R = 200$ runs.

The results of this investigation using $K = 20$ are displayed in Figure 2. We provide the analytical definitions in the Appendix for the $h(t|z)$ considered in this simulation. The log-hazard ratio estimate for the three PH models is a constant, and Cox’s PH model provides no estimate log-hazard curve for the control group, so we have omitted these models from the display and will evaluate their performance below using only the RISE of the average estimate. For both the time-dependent (first row) and proportional hazards scenario (second row), the PL-TD model exhibits greater variability than the PE-TD model in the right tail where the data become sparse. For the time-dependent

scenario, the PL-TD model results in a smaller control log-hazard RISE than the PE-TD model. In contrast, the PL-TD model results in a larger log-hazard ratio RISE value than the PE-TD model; however, focusing on $t \in [0, 0.60]$, the PL-TD model result in a much smaller RISE than the PE-TD model (0.09 versus 0.17). The PE-PH and PL-PH models have control log-hazard RISE values of 0.34 and 0.38, which are much larger than those of the TD models. All three PH models exhibit a relatively large log-hazard ratio RISE value of 0.81. For the time-independent scenario (second row), the three PH models result much smaller log-hazard ratio RISE values than the highly parametrized models, with the PL-PH model exhibiting the smallest at 0.007. By contrast, the control log-hazard RISE values are only slightly larger for the highly parametrized models than their PH equivalents. For example, the PL-PH model has 0.08 versus 0.09 for the PL-TD model.

4.3 Monotone Proportional Hazards Effect Estimation

For our last investigation we introduce a continuous covariate x with a monotone proportional hazards effect. We compare the performance of the shape-restricted model defined in (11) with that of the unrestricted model defined in (10), using the PL model defined in (5) with $K = 20$ as the model for the baseline hazard. We consider two nonlinear monotone effect scenarios, wherein the first has static intervals and the second does not, and a third linear monotone effect scenario. For the linear effect scenario, the proposed models facilitate more flexibility than needed, so we also fit a model that correctly assumes linearity in the effect of x , again using the PL model defined in (5) with $K = 20$ for the baseline hazard.

For evaluation, we generate $N = 200$ survival observations (t_i, δ_i, x_i) from a survival distribution with a prespecified conditional hazard $h(t|x) = h_0(t) \exp\{f(x)\}$ using a straightforward extension of the inverse cumulative density function methods discussed in Section 4.1. We fix the x_i at N equally spaced points across $[0, 1]$ for all R simulation runs, so that each run uses the same set of x_i 's. We then fit each model to these data, and save the posterior mean parameter estimates relating to the proportional hazards effect, $\hat{\beta}$. We compare the models by characterizing the effect of x on the hazard as a log-hazard ratio curve defined relative to the sampling average, $\widehat{f(x)} - \widehat{f(\bar{x})}$. By doing so, the estimated effect denotes the log-hazard ratio for an individual having arbitrary x versus an individual having $x = 0.50$; therefore, the value of the estimated curve at $x = 0.50$ is zero by definition. Iterating this process over $R = 200$ simulation runs, we compare the models visually using the average pointwise estimate, along with 2.5% and 97.5% pointwise quantiles, and the best estimate defined by having the smallest RISE among all $R = 200$ simulation runs. We also compare the average estimates quantitatively using RISE as we did in Section 4.1.

The results are depicted in Figure 3. We provide the analytical definitions in the Appendix for the baseline hazard $h_0(t)$ and the three $f(x)$ considered in this simulation. For the two nonlinear scenarios (first and second row), the linear effect model results in a suboptimal linear fit with the largest RISE value, and the smooth effect model exhibits the best average behavior as indicated by the lowest RISE value. The

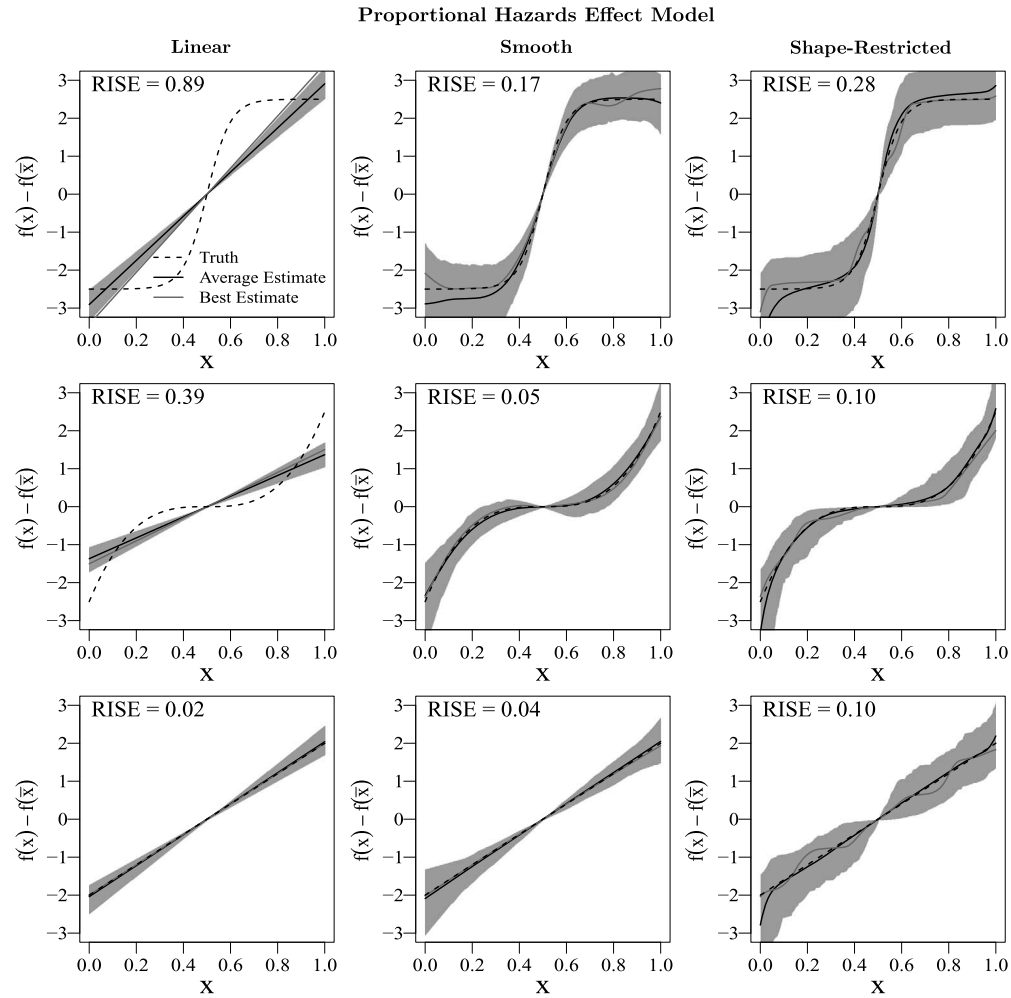


Figure 3: Monotone proportional hazards effect estimation with $N = 200$ and $J = 20$. All results are based on $R = 200$ simulation runs and the reported RISE is for the average estimate.

shape-restricted effect model (using $p_0 = 0.5$) exhibits an average estimate with a RISE value about twice as large as the smooth effect model, and slightly greater estimate variability as evidenced by wider shaded percentile regions. In addition, the shaded percentile regions for the shape-restricted model confirm that the resulting estimate always satisfies $f(x < 0.5) \leq f(x = 0.5) = 0 \leq f(x > 0.5)$, whereas the smooth effect model does not. Turning to the linear scenario (second row), the smooth effect model does not deteriorate much relative to the “correct” linear effect model, although it exhibits greater variability. In contrast, the shape-restricted effect model seems to deteriorate slightly, perhaps owing to the relatively large prior weight given

to the probability masses on zero (i.e., $p_0 = 0.50$) which indicates a moderate prior belief that f contains static intervals along the domain of x , which is not the case here.

5 Colorectal Cancer Clinical Trial Application

In this section, we combine our proposed methods in Sections 2 and 3 to evaluate the performance of the three regimes (i.e., IFL, IROX and FOLFOX) assigned in the clinical trial reported by Goldberg et al. (2004) for overall survival, while adjusting for the effect of AST. We do not assume that the effect of IROX and FOLFOX relative to IFL satisfy the proportional hazards assumption, but rather use the extended PL model we proposed in Section 3.1. We consider two such models that differ in the adjustment for the effect of AST corresponding to the smooth and shape-restricted models discussed in Section 3.2.

Before fitting the proposed models, we divided the observed times by the maximum observed time so that $t_i \in (0, 1]$, and we standardized the observed AST values x_i^* by defining $x_i = (x_i^* - x_{min}^*) / (x_{max}^* - x_{min}^*)$, where $x_{min}^* = \min\{x_i^* : i = 1, \dots, N\}$ and $x_{max}^* = \max\{x_i^* : i = 1, \dots, N\}$, so that $x_i \in [0, 1]$, $i = 1, \dots, N$. We fit each model by calling JAGS from R using the R2jags package. For posterior estimation, we ran two MCMC chains for 20,000 iterations, following 2,000 iterations of burn-in. Following Gelman et al. (2014, Section 11.4), we monitored MCMC convergence using the potential scale reduction factor. To jointly select the size of the time axis partition, K , and AST domain partition, J , we fit each AST effect model using combinations of K and J in the set $\{5, 10, 15, 20\}$ and selected the partitions that resulted in the smallest $\bar{D} + K + J$. For each AST effect model, these were the partitions with $K = 10$ and $J = 5$.

The results of our analysis are displayed in Figure 4. The first column displays posterior summaries of the log-hazard ratio curves from the three possible treatment comparisons over the initial five years of follow-up since treatment initiation, while adjusting for AST subject to a non-decreasing shape restriction. The results depicted in Figure 4 are nearly indistinguishable from the model that adjusts for AST using the smooth effect formulation discussed in Section 3.2, a lack of change apparently due to the even spread of AST demographics across the three treatment regimes. The top left panel shows that IROX appears to reduce the hazard for death relative to IFL by about 20% during the initial year and a half, but the evidence is not substantial and its advantage over IFL is less clear after that point in time. The middle panel on the left shows that there is substantial evidence that FOLFOX reduces the hazard for death relative to IFL by about 40% during the initial four years of follow-up. The lower left panel suggests FOLFOX even reduces the hazard for death relative to IROX by about 20% during the initial two and half years, though the evidence is less substantial as the 95% credible intervals contain one throughout much of this period. The amount of evidence decreases as the number of patients at risk shrinks, hence the increasing width of the credible regions as follow-up accrues. The benefits of either IROX or FOLFOX relative to IFL do appear to diminish slightly over the course of follow-up, though there is no substantial evidence for time dependency in either case, suggesting that a proportional hazards model may be acceptable for these data.

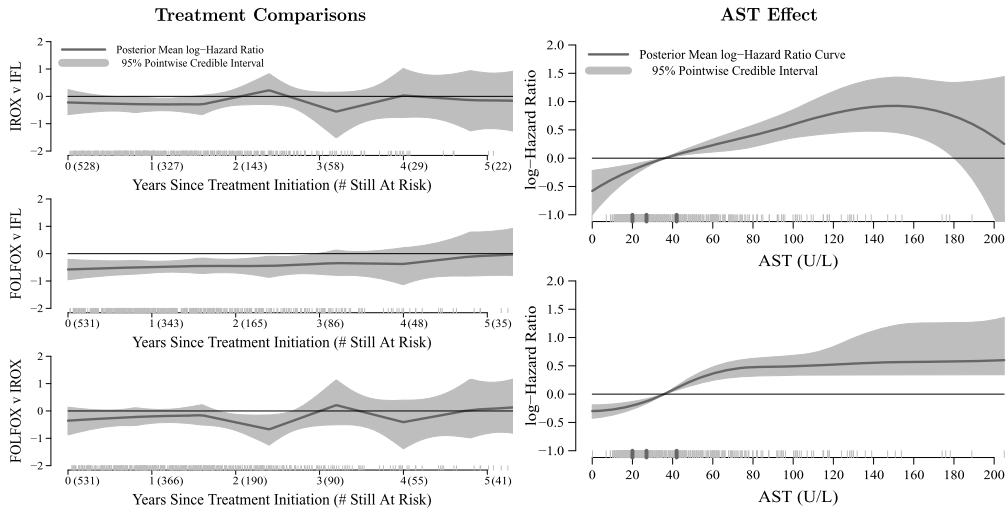


Figure 4: Treatment comparisons are based on the log-hazard ratio over time and the effect of AST is calculated as the log-hazard ratio relative to an individual having an AST of 40 U/L. The light grey ticks on the x -axis are the observed event times among individuals assigned the relevant regimes and the observed AST values for AST effect estimation. The dark grey ticks are the quantiles of the observed AST values.

The second column of Figure 4 compares the results from the model that uses a smooth proportional hazards effect (top right panel) and the model that uses a shape-restricted proportional hazards effect (bottom right panel) for AST. We report the effect of AST relative to an individual having an AST of 40 U/L, which is the standard upper threshold for the normal range of AST. The smooth effect model suggests an increasing hazard of death up to 150 U/L, but this hazard diminishes for higher AST values. Larger AST values are indicative of complications, so this is a non-intuitive signal and may purely be a result of sampling variation and sparse data for long follow-ups rather than a true effect. The shape-restricted effect model suggests the effect of AST is static in the lower domain of AST values (≤ 20 U/L), then increases sharply until about 60 U/L, at which point it remains constant. The resulting estimate provides evidence that patients with AST at or above 60 U/L have a hazard for death nearly 1.5 times higher than those with an AST of 40 U/L, whereas persons with AST below 20 U/L have a hazard about 0.75 times lower than those with an AST of 40 U/L. We also fit a “centered” version of the shape-restricted model, i.e., fixing $f(\bar{x}; \psi^*) = 0$, and posterior inference was not affected.

6 Discussion

This article presented a highly flexible framework for conducting a fully Bayesian analysis of survival data that can adjust for covariates using semiparametric time-dependent

effects and proportional hazards effects subject to shape-restrictions. These developments provide a unified framework to conduct a fully Bayesian analyses of complex survival data that we hope will encourage more comprehensive analyses, which currently often rely on some version of Cox’s proportional hazards model without further exploration. The modifiability of our approach eases investigations into prior sensitivity and assumptions about the relationship between covariates and the hazard. Furthermore, our choice to rely on low-rank thin plate splines ensures that the proposed methods attain fast MCMC convergence, thereby making the estimation of these models computationally feasible.

The simulations in Section 4 showed that where the data is non-sparse, estimates of the log-hazard and log-hazard ratio curves resulting from the proposed piecewise linear (PL) modeling approach are better than those of piecewise exponential (PE) modeling approach; however, where the data is sparse, estimates from PL models can exhibit greater variability and thereby diminish efficiency when compared to results obtained from PE models. To reduce variability in sparse regions, one possible remedy is a hybrid approach that constrains the PL formulation to be constant beyond some time-point. The simulations also showed that monotone shape restrictions can result in worse average behavior than unrestricted methods. The performance of the shape-restricted model in these contexts may be improved by careful selection of the mixture weights that control the prior probability for static regions in the hazard ratio curve. We considered estimating p_0 using a vague beta prior, but this did not seem to improve the average RISE properties. Further exploration into the usefulness of shape-restricted proportional hazards effect models is a worthy topic for future research.

The colorectal clinical trial data analysis in Section 5 illustrates the vast modifiability of the proposed methods, and verifies the conclusion of Goldberg et al. (2004) that FOLFOX is indeed the superior regime. Our choice to allow time-varying treatment effects was certainly reasonable, but the data did not provide substantial evidence that this flexibility was necessary. The comparative effectiveness of the three treatments turned out to be reasonably handled within the proportional hazards framework. However, the tools presented here help investigators to explore and infer time-dependent effects, which may promote better understanding of the biological mechanism at play that determine the relative long-term effectiveness of emerging therapies. The proposed methods provide a flexible, robust alternative to linear proportional hazards covariate effect assumptions. Lastly, these methods facilitate feasible exploration of critical modeling assumptions, which can be difficult to address using currently available software.

Appendix A

A.1 Low-Rank Thin Plate Spline Implementation

To implement the LRTP spline model of Section 2.2, we construct a $(K + 1) \times (K + 1)$ design matrix \mathbf{T}_K with the i th row given by $\mathbf{t}_{i,K} = (1, t_i, |t_i - \tilde{t}_1| - |\tilde{t}_1|, \dots, |t_i - \tilde{t}_{K-1}| - |\tilde{t}_{K-1}|)'$, so that (5) can be rewritten succinctly as $h(t_i; \boldsymbol{\alpha}^*) = \exp(\mathbf{t}'_{i,K} \boldsymbol{\alpha}^*)$. We then

construct a $(K + 1) \times (K + 1)$ transformation matrix

$$D_{\alpha} = \begin{bmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & \Omega_{\alpha}^{1/2} \end{bmatrix}$$

where the (ℓ, k) th entry of the penalty matrix Ω_{α} is defined as $|\tilde{t}_{\ell} - \tilde{t}_k|$, for $\ell, k = 1, \dots, K - 1$. We apply the transformations $\alpha = D_{\alpha}\alpha^*$, $T = T_K D_{\alpha}^{-1}$, and $U = U_K D_{\alpha,(-1)}^{-1}$ where U_K is a $K \times K$ matrix with k th row given by $\mathbf{u}'_{k,K}$, and $D_{\alpha,(-1)}^{-1}$ is the $K \times (K + 1)$ matrix obtained by omitting the first row of D_{α}^{-1} .

Next, we rewrite the hazard defined in (5) and cumulative hazard defined in (2) as

$$h(t_i; \alpha^*) \equiv h(t_i; \alpha) = \exp(\mathbf{t}'_i \alpha)$$

and $H(t_i; \alpha^*) \equiv H(t_i; \alpha) = \sum_{k=1}^K \frac{h(s_{i,k}; \alpha) \{1 - e^{-(s_{i,k} - \tilde{t}_{k-1}) \mathbf{u}'_k \alpha}\}}{\mathbf{u}'_k \alpha}, \tag{15}$

where \mathbf{t}'_i is the i th row of T , and \mathbf{u}'_k is the k th row of U . The likelihood arises by plugging these expressions into (1).

To implement the extended PL log-hazard model of Section 3.1, we also conduct a series of transformations. In the presence of a covariate z that is assumed to have a time-dependent effect, we can write the extended PL model defined in (8) for the i th observation succinctly as $h(t_i|z_i; \alpha^*) = \exp\{\mathbf{t}'_{i,K}(\alpha_0^* + \alpha_1^* z_i)\}$, where $\mathbf{t}'_{i,K}$ is the i th row of the matrix T_K . We then use the aforementioned transformations, taking $\alpha = D_{\alpha}\alpha^*$, $T = T_K D_{\alpha}^{-1}$, and $U = U_K D_{\alpha,(-1)}^{-1}$, and rewrite $h(t|z; \alpha^*)$ and $H(t|z; \alpha^*)$ completely in terms of α , T , and U . We omit further details, since the resulting expressions mirror (15) with $(\alpha_0 + \alpha_1 z)$ replacing α .

Finally, to implement the low-rank thin plate spline model for the effect of a continuous covariate x discussed in Section 3.2, we reformulate the model defined in (10) using the $J \times J$ transformation matrix

$$D_{\beta} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \Omega_{\beta}^{1/2} \end{bmatrix},$$

wherein the (j, k) th entry of Ω_{β} is defined as $|\tilde{x}_j - \tilde{x}_k|^3$, for $j, k = 1, \dots, J - 1$. We let $\mathbf{x}_{i,J} = (x_i, |x_i - \tilde{x}_1|^3 - |\tilde{x}_1|^3, \dots, |x_i - \tilde{x}_{J-1}|^3 - |\tilde{x}_{J-1}|^3)'$ denote the i th row of the $N \times J$ matrix \mathbf{X}_J , and define $\mathbf{X} = \mathbf{X}_J D_{\beta}^{-1}$ and $\beta = D_{\beta}\beta^*$. Doing so, we have that $f(x_i; \beta^*) \equiv f(x_i; \beta) = \mathbf{x}'_i \beta$ where \mathbf{x}'_i denotes the i th row of \mathbf{X} .

A.2 Simulation Function Definitions

Log-Hazard and Survival Distribution Estimation

For the simulation in Section 4.1, we used the following definitions of $h(t)$ for $t \in [0, 1]$:

Scenario 1 $h(t) = 2 \cos(\pi t + 0.5\pi) + 6/(t + 1) - 4 \log(1.125),$

Scenario 2 $h(t) = 0.2 \exp\left\{(8t)^{I(t < 0.5)}[-8(t - 1)]^{I(t \geq 0.5)}\right\},$

Scenario 3 $h(t) = 1.4 \sin(3\pi t) + 2.1.$

Time-Dependent Effect Estimation

For the simulation in Section 4.2, we used the following definitions of $h(t|z)$ for $t \in [0, 1]$:

All Scenarios $h(t|z = 0) = 1.5 \sin(\pi t) + 1,$

Scenario 1 $h(t|z = 1) = 2 \cos(\pi t + 0.5\pi) + 6/(t + 1) - 0.8,$

Scenario 2 $h(t|z = 1) = h(t|z = 0) \exp(1).$

Monotone Proportional Hazards Effect Estimation

For the simulation in Section 4.3, we used the following definitions of $h_0(t)$ for $t \in [0, 1]$ and $f(x)$ for $x \in [0, 1]$:

All Scenarios $h_0(t) = 4 [0.5 \cos(\pi t + 0.5\pi) + 1.5/(t + 1) - \log(1.125)],$

Scenario 1 $f(x) = 5 \exp\{20(x - 0.5)\} / [1 + \exp\{20(x - 0.5)\}] - 2.5,$

Scenario 2 $f(x) = 20(x - 0.5)^3,$

Scenario 3 $f(x) = 4(x - 0.5).$

References

- Bender, R., Augustin, T., and Blettner, M. (2005). "Generating Survival Times to Simulate Cox Proportional Hazards Models." *Statistics in Medicine*, 24(11): 1713–1723. MR2137646. doi: <http://dx.doi.org/10.1002/sim.2059>. 390
- Brezger, A. and Steiner, W. (2004). "Monotonic regression Bayesian P-splines." *Journal of Business and Economic Statistics*, 26(1): 90–104. MR2422064. doi: <http://dx.doi.org/10.1198/073500107000000223>. 390
- Cai, T., Hyndman, R., and Wand, M. (2002). "Mixed model-based hazard estimation." *Journal of Computational and Graphical Statistics*, 11(4): 784–798. MR1944263. doi: <http://dx.doi.org/10.1198/106186002321018786>. 382
- Cox, D. R. (1975). "Partial likelihood." *Biometrika*, 62(2): 269–276. MR0400509. 382
- Crainiceanu, C. M., Ruppert, D., and Wand, M. P. (2005). "Bayesian Analysis for Penalized Spline Regression Using WinBUGS." *Journal of Statistical Software*, 14(14): 1–24. 382, 385, 389

- Dunson, D. B. (2005). “Bayesian semiparametric isotonic regression for count data.” *Journal of the American Statistical Association*, 100(470): 618–627. MR2160564. doi: <http://dx.doi.org/10.1198/016214504000001457>. 390
- Fahrmeir, L. and Hennerfeind, A. (2003). “Nonparametric Bayesian hazard rate models based on penalized splines.” *Discussion paper // Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München* 361. 382
- Fahrmeir, L. and Lang, S. (2001). “Bayesian inference for generalized additive mixed models based on Markov random field priors.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(2): 201–220. MR1833273. doi: <http://dx.doi.org/10.1111/1467-9876.00229>. 384
- Gamerman, D. (1991). “Dynamic Bayesian Models for Survival Data.” *Applied Statistics*, 40(1): 63–79. 387
- Gelfand, A. E. and Mallick, B. K. (1995). “Bayesian analysis of proportional hazards models built from monotone functions.” *Biometrics*, 51: 843–852. 382
- Gelman, A. (2006). “Prior distributions for variance parameters in hierarchical models.” *Bayesian Analysis*, 1: 1–19. MR2221284. 384
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis, 3rd Edition*. Boca-Raton, FL: Chapman & Hall/CRC Press. MR3235677. 391, 396
- Goldberg, R. M., Sargent, D. J., Morton, R. F., Fuchs, C. S., Ramanathan, R. K., Williamson, S. K., Findlay, B. P., Pitot, H. C., and Alberts, S. R. (2004). “A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer.” *Journal of Clinical Oncology*, 22(1): 23–30. 383, 396, 398
- Hennerfeind, A., Brezger, A., and Fahrmeir, L. (2006). “Geoaddivitive Survival Models.” *Journal of the American Statistical Association*, 101(475): 1065–1075. MR2324146. doi: <http://dx.doi.org/10.1198/016214506000000348>. 382
- Henschel, V., Engel, J., Hölzel, D., and Mansmann, U. (2009). “A semiparametric Bayesian proportional hazards model for interval censored data and frailty effects.” *BMC Medical Research Methodology*, 9(1): 1–15. 382
- Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer. MR1876598. doi: <http://dx.doi.org/10.1007/978-1-4757-3447-8>. 382, 384
- Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. New York, NY: Springer. 383
- Lin, X., Cai, B., Wang, L., and Zhang, Z. (2014). “A Bayesian proportional hazards model for general interval-censored data.” *Lifetime Data Analysis*, 1–21. doi: <http://dx.doi.org/10.1007/s10985-014-9305-9> 382
- Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). “The BUGS project: Evolution, critique and future directions.” *Statistics in Medicine*, 28(25): 3049–3067. MR2750401. doi: <http://dx.doi.org/10.1002/sim.3680>. 382

- Müller, P. and Mitra, R. (2013). “Bayesian nonparametric inference: why and how.” *Bayesian Analysis*, 8(2): 269–302. MR3066939. doi: <http://dx.doi.org/10.1214/13-BA811>. 382
- Plummer, M. (2003). “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.” <http://mcmc-jags.sourceforge.net/>. 382
- Ruppert, D. (2002). “Selecting the number of knots for penalized splines.” *Journal of Computational and Graphical Statistics*, 11(1): 735–757. MR1944261. doi: <http://dx.doi.org/10.1198/106186002321018768>. 386
- Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric Regression*. New York: Cambridge University Press. MR1998720. doi: <http://dx.doi.org/10.1017/CB09780511755453>. 384, 388
- Sharef, E., Strawderman, R. L., Ruppert, D., Cowen, M., and Halasyamani, L. (2010). “Bayesian adaptive B-spline estimation in proportional hazards frailty models.” *Electronic Journal of Statistics*, 4: 606–642. MR2660535. doi: <http://dx.doi.org/10.1214/10-EJS566>. 382, 386
- Shively, T. S. and Sager, T. W. (2009). “A Bayesian approach to non-parametric monotone function estimation.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1): 159–175. MR2655528. doi: <http://dx.doi.org/10.1111/j.1467-9868.2008.00677.x>. 389, 390
- Shively, T. S., Walker, S. G., and Damian, P. (2011). “Nonparametric function estimation subject to monotonicity, convexity and other shape constraints.” *Journal of Econometrics*, 161: 166–181. MR2774935. doi: <http://dx.doi.org/10.1016/j.jeconom.2010.12.001>. 389
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). “Bayesian measures of model complexity and fit.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4): 583–639. MR1979380. doi: <http://dx.doi.org/10.1111/1467-9868.00353>. 386

Acknowledgments

Thanks to the Editor, Associate Editor and Referee for their helpful comments during the revision process. This work was supported in part by NCI grant 1-R01-CA157458-01A1 and CCSG grant P30-CA016672.