Flexible Bayesian survival modeling with nonparametric 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 readily available software to perform a fully Bayesian analysis of time-to-event data wherein the hazard is estimated nonparametrically. 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 nonparametric alternative to the piecewise exponential model which assumes continuity of the hazard, and to develop easily modifiable software thereby allowing the use of these methods in a variety of settings. To accomplish this aim, we construct a piecewise linear log-hazard model using a low-rank thin plate spline formulation for the log-hazard function. We discuss extensions that facilitate nonparametric adjustment for covariates with time-dependent effects and nonlinear time-independent effects possibly subject to shape restrictions. We investigate the properties of these modeling choices via simulation. We then apply our methods to colorectal cancer data from a clinical trial comparing the effectiveness of two novel treatment regimes relative to the standard of care with respect to overall survival. In particular, we characterize the hazard ratio as a function of time between each novel regime and 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.

Keywords: Bayesian hierarchical model, colorectal cancer, hazard regression, nonparametric methods, penalized splines, piecewise exponential model, shape restrictions, survival analysis, time-dependent effects

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 baseline hazard are inadequate for characterizing the curvature of non-unimodal functions. In fact, the Weibull model precludes the possibility of a non-monotone baseline hazard function. 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 nonparametric models for the hazard are useful for analysis of actual clinical
Flexible Bayesian survival modeling

data only in the presence of a framework that accommodates a diverse class of covariates for characterizing patient and intervention heterogeneity.

Moreover, deviations between the treatment-specific hazard functions may be insufficiently characterized by a single time-invariant scalar over the entire follow-up duration, as assumed in a proportional hazards model. Adjustments for continuous covariates using standard linear regressors may also fail to account for heterogeneity over the covariate domain derived from underlying biological mechanisms. There are currently few options with readily available software for conducting a fully Bayesian analysis of time-to-event data that facilitate nonparametric estimates of the baseline hazard and effects of covariates. A presently popular choice is the piecewise exponential model, which assumes that the hazard is piecewise constant over time (c.f. Ibrahim et al. 2001, Section 3.1). The strengths of this approach are its tractability, but a weakness is the discontinuous piecewise constant approximation to the hazard. In practice the hazard is typically believed to be continuous. Furthermore, estimation of the piecewise exponential model is sensitive to prespecification of the number and location of hazard function discontinuities along the time axis.

Perhaps a more robust Bayesian modeling option is discussed by Fahrmeir and Hennerfeind (2003), wherein the log-hazard model is constructed in an additive fashion with the baseline hazard and covariate effects being modeled nonparametrically as penalized B-splines. A similar approach is taken by Cai et al. (2002), though from a frequentist perspective, wherein the log-hazard is modeled using B-splines in a mixed model-based framework. 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 basis functions to be estimated, and facilitate mixtures with parametric hazard forms. All the previous additive log-hazard approaches require the evaluation of an intractable interval, owing to the formulation of B-splines. As a result, they rely on computationally efficient numerical integration implemented in specialized software (e.g., BayesX and the splinesurv package in R). These specialized software packages cannot easily be modified to handle shape-restricted effects or allow completely general prior distributional family choices.

The primary aim of this paper is to provide a unified framework and modifiable software that enables investigators to conduct fully Bayesian analyses of time-to-event data which is highly flexible with respect to hazard and covariate effect estimation. To do so, we construct a piecewise linear log-hazard model using low-rank thin plate splines that results in a tractable likelihood expression. We then extend this construct using an additive framework that facilitates a nonparametric estimate for time-dependent effects and shape-restricted proportional hazards effects. In doing so, we broaden the existing methodology to combine functional and time-dependent covariate estimation in the presence of a nonparametric continuous hazard, yielding a framework that facilitates probabilistic interpretations and seamless incorporation of prior information.

The motivating application involves data from a colorectal cancer clinical trial by Goldberg et al. (2004) that assessed the efficacy of three treatment regimes on patients with previously untreated metastatic colorectal cancer. The drug combinations consid-
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. The primary goal is to characterize the comparative effectiveness of each regime in terms of overall survival over the eight years following treatment initiation. A secondary goal is to jointly characterize the effect of aspartate transaminase (AST), a liver failure biomarker measured at baseline, that is thought to have a nondecreasing effect on the hazard for death over the domain of AST.

The remainder of our paper evolves as follows. In Section 2 we provide the details of likelihood construction and prior specification for the piecewise exponential model and the proposed piecewise linear log-hazard model in the absence of covariates. In Section 3 we extend the aforementioned models to accommodate covariates with various types of effects on the hazard, including time-dependent and shape-restricted effects. Section 4 provides the details and results of simulation studies designed to compare the proposed piecewise linear log-hazard model with other common models. In Section 5 we illustrate the proposed methods with an analysis of the colorectal cancer dataset. In this analysis we allow effects of the novel treatments relative to the standard of care to vary over time while adjusting for the effect of AST that is subject to a non-decreasing shape restriction. Finally, Section 6 closes with a discussion of our findings and directions for future work.

2 Hazard models

In this section we construct the likelihood for the piecewise exponential (PE) model and the proposed piecewise linear log-hazard (PL) model in the absence of covariates. Prior specification for each of these models is discussed later. We assume the data consist of $N$ independent observations, wherein $t_i$ denotes the observed time and $\delta_i$ denotes whether $t_i$ is an event ($\delta_i = 1$) or right-censored observation ($\delta_i = 0$). Thus, the data are $D = (t, \delta)$, where $t = (t_1, \ldots, t_N)$ and $\delta = (\delta_1, \ldots, \delta_N)$. Without loss of generality, we assume $t \in (0, 1]$.

The likelihood of the pair $(t, \delta)$ can be expressed as

$$L(t, \delta) = h(t)^\delta \exp \left\{ - \int_0^t h(u) du \right\} = h(t)^\delta \exp \{ -H(t) \},$$

where $h(t) > 0$ is the hazard function and $H(t) = \int_0^t h(u) du$ is the cumulative hazard function. The full-data likelihood is then $\prod_{i=1}^N L(t_i, \delta_i)$. The survival distribution is $S(t) = \exp \{ -H(t) \}$. Hence, an analysis of time-to-event data can be carried out entirely through a model for the hazard function, $h(t)$, or log-hazard function, $\log \{ h(t) \}$.
Flexible Bayesian survival modeling

2.1 Piecewise exponential model

The PE model improves upon parametric alternatives by introducing more parameters to accommodate diverse, possibly non-unimodal shapes of the hazard function (Ibrahim et al. 2001, Section 3.1). This model is constructed by partitioning the time axis into $K$ intervals $(0 = \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_{K-1} < \tilde{t}_K = 1)$, and taking

$$h(t; \lambda) = \lambda_k \quad \text{for} \quad t \in I_k, \quad k = 1, \ldots, K.$$  \hspace{1cm} (2)

Thus, the hazard function in (2) is assumed to be a discontinuous piecewise constant function, where $\lambda_k > 0$ is the value of the hazard function in the $k$th interval of the time axis partition. Under (2), the cumulative hazard function arises as

$$H(t; \lambda) = \sum_{k=1}^{K} \left[ \min \left( \max \left( t, \tilde{t}_{k-1} \right), \tilde{t}_k \right) - \tilde{t}_{k-1} \right] \lambda_k.$$  \hspace{1cm} (3)

The likelihood using this model is constructed by plugging (2) and (3) into (1).

The PE model depends upon the choice of $K$, which needs to be large enough to ensure the model is sufficiently flexible to capture the true features of the hazard function (Wand 2000). After choosing $K$, the placement of the $\tilde{t}_k$'s over the domain of $t$ can be done in any manner, say at the quantiles of the observed event times; however, 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. Another approach is to allow $K$ and $\tilde{t}$ to be unknown parameters, though the additional computational burden typically does not justify the marginal gains in approximation accuracy over a reasonable prespecified partition (Sharef et al. 2010). In practice, we suggest considering an increasing set of $K$ values and selecting the smallest $K$ such that the estimated curve does not change substantially for larger values. For a thorough investigation and discussion of selecting a partition for penalized splines, see Ruppert (2002).

Numerous prior specifications for $\lambda$ in (2) 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 $\log(\lambda_k)$’s. This choice is parsimonious and allows strength to be borrowed across adjacent intervals, thereby improving efficiency and resisting overfitting the data. Formally, the random walk prior process is specified as

$$\log(\lambda_1) \sim \mathcal{N}(0, 10^4), \quad \text{and} \quad \log(\lambda_k) \mid \log(\lambda_{k-1}), \sigma_\lambda \sim \mathcal{N}\left( \log(\lambda_{k-1}), \sigma_\lambda^2 \right) \quad \text{for} \quad k = 2, \ldots, K,$$  \hspace{1cm} (4)

where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean $\mu$ and variance $\sigma^2$. 

2.2 Piecewise linear log-hazard model

The PE model described in Section 2.1 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. A hazard function is necessarily non-negative, so we avoid the complexities of imposing positivity constraints by modeling the unrestricted log-hazard. Another complication arises from the fact that the likelihood expression in (1) requires the evaluation of the definite integral \( \int_0^t h(s)ds \). Since quadratic and higher order polynomial models for log\( \{h(t)\} \) do not produce analytical forms, we focus on first-order polynomial log-hazard models. Penalized splines are a computationally simple, yet highly flexible option that facilitates using a piecewise linear model. Crainiceanu et al. (2005) demonstrate that low-rank thin plate splines exhibit fast Markov chain Monte Carlo (MCMC) convergence with low posterior autocorrelation relative to truncated basis splines, and they offer a tractable alternative to B-splines. A partition of the time axis into \( K \) intervals \((0 = \tilde{t}_0 < \tilde{t}_1 < \ldots < \tilde{t}_{K-1} < \tilde{t}_K = 1)\) is again required to model the log-hazard using a low-rank thin plate spline. As with the PE model, we select \( K \) to be large enough to capture the interesting features in the hazard and employ equally spaced partitions. The resulting model imposes a piecewise constant derivative upon the hazard, thereby facilitating a model that is more robust to misspecification than the PE model.

Given the selected partition, we take

\[
\log \{h(t; \alpha)\} = \alpha_0 + \alpha_1 t + \sum_{k=2}^{K} \alpha_k (|t - \tilde{t}_{k-1}| - |\tilde{t}_{k-1}|),
\]

where \( \alpha = (\alpha_0, \alpha_1, \ldots, \alpha_K)^\top \). The model defined in (5) replaces the typical basis terms, \(|t - \tilde{t}_{k-1}|\), with modified basis terms, \(|t - \tilde{t}_{k-1}| - |\tilde{t}_{k-1}|\), selected so that \( \log \{h(0; \alpha)\} = \alpha_0 \), thereby easing prior elicitation for the intercept \( \alpha_0 \) and improving MCMC convergence.

Under (5), the cumulative hazard arises as

\[
H(t; \alpha) = \sum_{k=1}^{K} h(s_k; \alpha) \left\{ 1 - e^{-\left( s_k - \tilde{t}_{k-1} \right) \left( u_{k,K}^\top \alpha_{(-1)} \right)} \right\},
\]

where \( s_k = \max\{\text{min}\{t, \tilde{t}_k\}, \tilde{t}_{k-1}\} \), \( \alpha_{(-1)} = (\alpha_1, \ldots, \alpha_K)^\top \), \( u_{k,K}^\top = (1_{k,1}^\top, -1_{k,K-k}^\top) \), for \( k = 1, \ldots, K \), and \( 1_k^\top \) denotes a \( k \)-dimensional row vector of ones.

Following the work of Crainiceanu et al. (2005), we implement a series of transformations, including a one-to-one transformation from \( \alpha \) to \( a = (a_0, \ldots, a_K)^\top \). The details of this procedure are provided in the Appendix. We then use the standard prior specification for a low-rank thin plate spline, wherein

\[
a_0 \sim \mathcal{N}(0, \sigma_0^2), \quad a_1 \sim \mathcal{N}(0, \sigma_1^2), \quad \text{and} \quad a_k | \sigma_a \overset{iid}{\sim} \mathcal{N}(0, \sigma_a^2) \quad \text{for} \quad k = 2, \ldots, K.
\]
This prior specification penalizes the coefficients pertaining to the modified radial basis functions in (5), thereby smoothing the resulting estimator and resisting overfitting the data (Ruppert et al. 2003). The hyperparameters $c_0$ and $c_1$ should be specified so that the prior distributions for $a_0$ and $a_1$ are sufficiently vague for the setting.

3 Covariate Adjustment

Often data derive from a heterogeneous population with measured covariates. Robust estimation methods using flexible hazard models are useful in practice only when incorporated into a modeling framework that accommodates a diverse class of covariates. In this section, we discuss extensions to the PE and PL hazard models developed in Section 2 to incorporate covariates. A diverse class of covariates and their effects can be encountered in the context of time-to-event data. Time-invariant or baseline covariates assume a fixed value throughout the time period of interest (e.g., gender, race, a biomarker measured at baseline, etc.). While not discussed in detail in this paper, 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.). Note that with either type of covariate, the effect on the hazard may vary over time or remain constant. We refer to the former as a time-dependent effect, and the latter as a time-independent or proportional hazards effect.

3.1 Time-dependent effect

Assume the data now include information about a time-invariant covariate $z$ that is assumed to have a time-dependent effect on the hazard function. For example, $z$ may indicate assignment to a novel treatment that is thought to have a diminishing benefit relative to the standard treatment over time. Cox’s proportional hazards model does not facilitate estimation of time-dependent effects, rather it requires a time-independent effects (i.e. proportional hazards) assumption (Cox 1975). Extensions to Cox’s proportional hazards model that facilitate estimation of time-independent effects do not estimate the baseline hazard function, so jointly characterizing the survival distributions for the two (or more) treatment groups is not feasible. By contrast, the PE and PL models can be extended to provide a uniform approach for modeling the conditional hazard $h(t|z)$, thereby facilitating investigation of the time-dependent effect of treatment, as well as the hazard functions and survival distributions in each treatment group.

Following the work of Gamerman (1991), the PE model defined in (2) can be extended to accommodate $z$ so that the conditional hazard is

$$h(t|z; \lambda, \gamma) = \lambda_k \exp(\gamma_k z) \quad \text{for} \quad t \in I_k, \quad \text{where} \quad I_k = (\tilde{t}_{k-1}, \tilde{t}_k], \quad k = 1, \ldots, K.$$ (8)

The $\gamma = (\gamma_1, \ldots, \gamma_K)$ are allowed to realize distinct values in each of the $K$ intervals, thereby facilitating time-dependent effects with a piecewise constant structure using the same time axis partition. The cumulative conditional hazard $H(t|z; \lambda, \gamma)$ arises by
replacing $\lambda_k$ in (5) with $\lambda_k \exp(\gamma_k z)$. Typically, $\lambda$ and $\gamma$ are assumed to be independent a priori, whence the prior specification for $\lambda$ still follows (4). The prior specification for $\gamma$ to induce temporal correlation can also follow the structure of (4), wherein $\gamma_k|\gamma_{k-1}, \sigma_\gamma^2 \sim \mathcal{N}(\gamma_{k-1}, \sigma_\gamma^2)$ for $k = 2, \ldots, K$.

To extend the PL model defined in (5), we model

$$
\log \{h(t|z; \alpha)\} = (\alpha_{0,0} + \alpha_{1,0}z) + (\alpha_{0,1} + \alpha_{1,1}z)t + \sum_{k=2}^K (\alpha_{0,k} + \alpha_{1,k}z) \left( |t - \tilde{t}_k - 1| - |\tilde{t}_k - 1| \right),
$$

where $\alpha = (\alpha_{0,1})$ and $\alpha_q = (\alpha_{q,0}, \ldots, \alpha_{q,K})^\top$, for $q = 0, 1$. As in Section 2.2, we apply a series of one-to-one transformations that result in a parameterization of (9) in terms of $a_0$ and $a_1$. The details for this procedure are also provided in the Appendix. We then assume that $a_0$ and $a_1$ are independent a priori, a standard assumption in the context of an additive model, and use prior specifications for $a_0$ and $a_1$ analogous to (7).

Under (9), the log-hazard ratio for an individual having $z = 1$ relative to $z = 0$ is a piecewise linear function given by a low-rank thin plate linear spline with the same set of basis functions as before. By contrast, the extended PE model in (8) assumes that the log-hazard ratio is piecewise constant over time. Thus, when the time-dependent effect of $z$ is of interest, the extended PL model offers a continuous alternative to the extended PE model. Nevertheless, either approach facilitates a nonparametric 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.

The PE model can easily be extended to handle $p$ time-invariant covariates with time-dependent effects by assuming the conditional hazard in the $k$th interval is $\lambda_k \exp(z^\top \gamma_k)$, where $z^\top = (z_1, \ldots, z_p)$ and $\gamma_k = (\gamma_{k,1}, \ldots, \gamma_{k,p})^\top$, $k = 1, \ldots, K$. Similarly, the PL model can be extended by replacing the basis coefficients, $(\alpha_{0,k} + \alpha_{1,k}z)$, in (9) with $z^\top \alpha_k$, where $z^\top = (1, z_1, \ldots, z_p)$ and $\alpha_k = (\alpha_{0,k}, \ldots, \alpha_{q,k})$, for $k = 0, \ldots, K$. Extensions of either model to handle a time-varying covariate $z(t)$ with a time-dependent effect are also straightforward, requiring a minor adjustment in the calculation of the cumulative hazard to account for the time-varying nature of $z(t)$.

### 3.2 Nonparametric time-independent effects

Assume the data also include information about a time-invariant continuous covariate $x$ that is assumed to have a time-independent effect on the hazard function. For example, $x$ may denote age at baseline which is assumed to have a constant effect on the hazard over the course of follow up. In this case, a proportional hazards approach to adjusting for $x$ is sensible; thus, $h(t|x, z) = h^*(t|z) \exp \{f(x)\}$, wherein $h^*(t|z)$ denotes the conditional baseline hazard. Notice that $\exp \{f(x)\}$ factors out in the calculations of the cumulative conditional hazard so that $H(t|x, z) = H^*(t|z) \exp \{f(x)\}$, thereby affording the use of either (8) or (9) as the model for $h^*(t|z)$. Likewise, in the absence of any covariate $z$ with
Flexible Bayesian survival modeling

a time-dependent effect, \( h(t|z) = h(t) \), allowing the use of either (4) or (5) as a model for \( h(t) \). Hence, the developments of the previous subsection can be combined with these methods to allow models that adjust for a set covariates with a mix of time-varying and time-invariant effects.

Often a standard linear regressor (i.e. \( f(x) = \beta x \)) will be a sufficiently flexible choice for the effect of \( x \). However, when \( x \) is a continuous covariate, a more general assumption for the relationship between \( x \) and the hazard may be necessary. For example, a nonparametric model may be warranted when the effect of \( x \) is believed to be a smooth nonlinear function over the domain of \( x \). Alternatively, a shape-restricted model may be more appropriate when there is a scientific basis to assume that the effect of \( x \) is monotonically increasing.

**Smooth nonlinear effect**

When \( f(x) \) is assumed to be a smooth function over the domain of \( x \), say continuously differentiable, cubic splines are a natural choice (Ruppert et al. 2003). In the context of a Bayesian analysis, we prefer low-rank thin plate cubic splines for their simple construction and tendency to exhibit fast MCMC convergence. The construction of a low-rank thin plate cubic spline for \( f(x) \) is similar to the approach of Section 2.2. Without loss of generality, we assume \( x \in [0, 1] \) and specify a partition with \( J \) pieces \((0 = \tilde{x}_0 < \tilde{x}_1 < \ldots < \tilde{x}_J = 1)\) over this domain. After selecting a partition, we model

\[
\begin{align*}
  f(x; \beta) &= \beta_1 x + \sum_{j=2}^{J} \beta_j (|x - \tilde{x}_{j-1}|^3 - |\tilde{x}_{j-1}|^3).
\end{align*}
\]

(10)

Note that the model in (10) fixes the intercept at zero, ensuring identifiability since \( h(t|z, x = 0) = h^*(t|z) \). Thus, the conditional baseline hazard represents the hazard of an individual having \( x = 0 \) and arbitrary \( z \).

In practice, we again implement a series of one-to-one transformations leading to (10) being parametrized in terms of \( b = (b_1, \ldots, b_J) \). The details of this procedure are provided in the Appendix. Next, we assume that \( b \) is independent of \( a \) a priori, specify a vague prior for \( b_1 \), and take \( b_j | \sigma_b \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_b^2) \) for \( j = 2, \ldots, J \).

**Monotone nonlinear effect**

Sometimes, understanding of an underlying biological mechanism may justify assuming that \( f(x) \) is of some particular shape. Many types of shape restrictions can be imposed tractably through the prior when \( f(x) \) is modeled as a truncated quadratic spline (c.f. Shively et al. 2011). A truncated quadratic spline is formulated as

\[
\begin{align*}
  f(x; \psi) &= \psi_1 x + \psi_2 x^2 + \sum_{j=2}^{J} \psi_{j+1} (x - \tilde{x}_{j-1})^2.
\end{align*}
\]

(11)
where \((x)^2 = \max(x, 0)^2\). Note that the model in (11), like the model in (10), fixes the intercept at zero. Thus, \(h^*(t|z)\) again represents the hazard for an individual having \(x = 0\) and arbitrary \(z\), and \(e^{f(x; \psi)}\) again measures the hazard ratio of an individual having arbitrary \(x\) compared to one having \(x = 0\).

Using (11), monotonicity can be imposed by forcing the first derivative, \(f'(x; \psi)\), to be non-negative for all \(x \in [0, 1]\). Since \(f'(x; \psi) = \psi_1 + \sum_{j=1}^{J} 2\psi_{j+1}(x - \tilde{x}_{j-1})_+\) is a piecewise linear function, the local minima in each interval \((\tilde{x}_{j-1}, \tilde{x}_j)\) will be realized at boundaries (i.e. \(\tilde{x}_{j-1}\) or \(\tilde{x}_j\)). Therefore, the necessary constraints are identified by evaluating \(f'(x; \psi)\) at each interval boundary (i.e. \(\tilde{x}_0, \ldots, \tilde{x}_J\)), and requiring that each of the resulting \(J + 1\) expressions be non-negative. Doing so, the constraints arise as

\[
\psi_1 \geq 0 \text{ and } \psi_1 + \sum_{k=1}^{j} 2(\tilde{x}_j - \tilde{x}_{k-1})\psi_{j+1} \geq 0, \quad j = 1, \ldots, J.
\]

Notice that the constraints in (12) are linear combinations of \(\psi\). Applying the transformation \(p = L\psi\), where \(L\) is a \((J + 1)\) dimensional lower triangular matrix with \(j\)th row up to the diagonal given by \((1, 2(\tilde{x}_j - \tilde{x}_0), \ldots, 2(\tilde{x}_j - \tilde{x}_{j-1}))\), the monotonicity constraints in (12) are simply \(p_j \geq 0, j = 1, \ldots, J + 1\).

The monotonicity constraints can be imposed by specifying prior distributions for each \(p_j\) that have nonnegative support \cite{Shively2009}, e.g.,

\[
\pi(p_j) = \nu_j N(p_j|0, 10^4)I_{[0, \infty)}(p_j) + (1 - \nu_j)I_0(p_j), \quad \text{for } j = 1, \ldots, J + 1,
\]

where \(N(p_j|0, 10^4)I_{[0, \infty)}(p_j)\) denotes a truncated normal distribution with positive support, \(I_0(p_j)\) denotes the indicator function, and the \(\nu_j\) are prespecified mixture weights. Since \(f'(\tilde{x}_j; \psi) = p_j\), the \(\nu_j\) control the prior probability that \(f'(x; \psi)\) is positive at \(\tilde{x}_j\). Thus, \(\nu_j\) is the prior probability that \(f(x; \psi)\) is unchanging at \(\tilde{x}_{j-1}\), allowing control over the probability of invariant intervals in \(f(x; \psi)\) along the domain of \(x\). The prior specification in (13) deviates slightly from that of \cite{Shively2009}, who use a mixture of a \((J + 1)\) dimensional multivariate normal distribution truncated to have positive support in each dimension of \(\psi\), and probability masses corresponding to each boundary wherein a subset of \(\psi\) is exactly zero. The univariate mixture structure we propose in (13) is substantially easier to construct in practice, and the resulting posterior estimates are similar.

Extensions to handle an arbitrarily large set of covariates containing a mix of effect types are straightforward. Further, the time-independent effects may include a mix of smooth nonlinear, shape-restricted, or linear effects. Using an additive framework, we can model \(h(t|x, z) = h^*(t|z)\exp\left\{\sum_{\ell=1}^{p} f(x_\ell)\right\}\) where \(\log\{h^*(t|z)\}\) has the generalized piecewise constant or piecewise linear structure discussed in Section 3.1 and the \(f(x_\ell)\) are modeled as discussed in Section 3.2. Lastly, we could in theory consider a time-dependent nonlinear effect, however the amount of data and computation time needed
Flexible Bayesian survival modeling
to reliably estimate such a complex effect may be prohibitive in practice.

4 Simulation Investigations

This section consists of three simulation studies that investigate the performance of the PL model relative to the PE model and their extensions. First, we compare their performance for estimation of a hazard function with a complex shape and resultant survival distribution in a homogeneous population. Second, we compare the extended PL and PE models developed in Section 3.1 for estimation of a time-dependent effect. Third, we compare the nonparametric time-independent effect estimation approaches developed in Section 3.2 for estimation of a monotone effect.

4.1 Hazard Function Estimation

Here we evaluate how the performance of the PL and PE models are affected by the shape of the true hazard function in a homogeneous population. We consider estimation of three true hazard functions using two sample sizes, \( N = 50 \) and \( 200 \), and consider equally-spaced partitions of the time axis. The piecewise constant assumption requires a finer partition than the piecewise linear assumption to reasonably approximate a complex hazard function, so we consider varying partition sizes of \( K = 5, 10, \) and \( 20 \) for the PE model, and \( K = 3, 8, \) and \( 15 \) for the PL model.

We evaluate the performance of each model over \( R \) simulation runs, wherein we generate pairs \( (t_i, \delta_i) \) of independent, possibly right censored observation times, \( i = 1, \ldots, N \). In each run, the \( N \) independent observations are drawn from a survival distribution with a prespecified underlying hazard function \( h(t) \). To draw an observation from a survival distribution with an arbitrary \( h(t) \), we follow the inverse cumulative density function method of Bender et al. (2005).

Formally, we specify a target function \( h(t) \) such that \( H(t) \) is available analytically. We 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 \( y_i = H^{-1}\{-\log(u_i)\} \). Since \( H(t) \) is available analytically, we can solve the latter identity numerically, thereby affording diverse classes of hazard function shapes from which to choose. We also draw a corresponding censoring time \( c_i \) from an independent distribution calibrated so that the resulting data exhibit approximately 15% censoring. Finally, we calculate the observed time \( t_i = \min(y_i, c_i) \) and event indicator \( \delta_i = I(t_i \leq c_i) \).

Throughout this subsection we specify vague \( N(0, 10^4) \) priors for the unpenalized parameters (i.e. \( \log(\gamma_1), a_0 \) and \( a_1 \)), and follow Gelman (2006) by specifying vague \( \mathcal{U}(0.01, 100) \) priors for the standard deviation hyperparameters (i.e. \( \sigma_\lambda, \sigma_a \)), where \( \mathcal{U}(a,b) \) denotes a uniform distribution on \((a,b)\).

In each simulation run, we fit the PE and PL model for each partition size \( K \) under consideration to the generated data. We then save the posterior mean hazard function and survival distribution estimates from each model at \( M = 100 \) equally spaced time
Figure 2: Each row corresponds to an investigation under a different true hazard (broken line). The two leftmost columns contain results pertaining to hazard estimation and the two rightmost columns pertain to survival distribution estimation. The first and third columns from the left correspond to the piecewise exponential model with $K = 20$, whereas the second and fourth column correspond to the piecewise linear log-hazard model with $K = 8$. All results are based on $R = 200$ simulation runs each with a sample size of $N = 200$.

We compare the pointwise posterior mean hazard estimates, $\tilde{h}(s_m) = \frac{1}{R} \sum_{r=1}^{R} \hat{h}^{(r)}(s_m)$, as well as the empirical pointwise 2.5%, 10%, 90% and 97.5% quantiles. We also calculate the root-integrated squared error of the pointwise mean hazard curve,

$$\text{RISE} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left[ \tilde{h}(s_m) - h(s_m) \right]^2}. \quad (14)$$

The evaluation criteria for the survival distribution $S(t)$ are defined analogously.
Flexible Bayesian survival modeling

The first and third columns of Figure 2 depict the results of the PE model with $K = 20$, whereas the second and fourth columns depict the results of the PL model with $K = 8$. The three true hazard functions are depicted by the broken lines in two leftmost columns, whereas their corresponding survival distributions are depicted in two rightmost columns. We omit the analytical expressions, but note that these hazard functions were selected to cover an array of shapes with $S(1) \approx 0.1$.

The pointwise mean estimate $\hat{h}(t)$ for the PE model (first column) is a discontinuous piecewise constant function, whereas our proposed PL model (second column) results in a continuous estimate. As evidenced by the results depicted in the first and third row, both approaches struggle to detect a shift in hazard beyond 80% follow-up (t-axis) where the data are quite sparse. The PL model exhibits slightly greater variability than the PE model in this data-sparse region. In the top row, RISE is slightly larger for the PL model than the PE model, however, the PL model provides a markedly smaller RISE than the PE model when only integrating over the initial 80% of follow-up (0.061 versus 0.132). 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, the PL model (fourth column) results in RISE values that are slightly smaller in magnitude than the PE model (third column) across all three scenarios. The differences are slight; however, the proposed PL model better captures the dips in the true survival distribution. Results for other partition sizes and for $N = 50$ are discussed in the Appendix.

4.2 Time-dependent Effect Estimation

For this investigation we introduce a binary treatment indicator $z$ with $Pr(z = 1) = Pr(z = 0) = 0.5$. We consider a scenario where the effect of $z$ is time-dependent, and another where the effect of $z$ is time-independent (i.e. satisfies the proportional hazards assumption). Here we evaluate the performance of the extended PL model defined in (9) compared to the similarly extended PE model defined in (8). For the time-independent scenario the extended models provide more flexibility than is needed, so to investigate their possible loss of efficiency, we also fit proportional hazards (PH) versions of the PE and PL models. That is, we assume $h(t|z) = h^*(t) \exp \{z \beta \}$ where the baseline hazard, $h^*(t)$, is defined by (2) for the PE-PH model and by (5) for the PL-PH model. Lastly, we also fit Cox’s PH model using the `coxph()` function from the `survival` package available in R.

To compare the methods, we generate $N = 200$ survival observations $(t_i, \delta_i, z_i)$ from a distribution with a prespecified conditional hazard $h(t|z)$ using the inverse cumulative density function method discussed in (4.1). We fit each model to these data, and save the posterior mean estimates at 100 equally spaced grid points over $t \in [0, 1]$ for the hazard ratio curve, $HR(t) = h(t|z = 1)/h(t|z = 0)$, the control hazard, $h(t|z = 0)$, and the treatment hazard, $h(t|z = 1)$. We note that hazard ratio for the three PH models is a constant, and that Cox’s PH model provides no estimate of $h(t|z = 0)$ or $h(t|z = 1)$.

We iterate this data generation and model fitting process over $R = 200$ simulation
Figure 4: Results for the scenario where $z$ has a time-dependent effect are depicted in the two leftmost columns, whereas results for the scenario where $z$ has a time-independent (i.e., proportional hazards) effect are in the two rightmost columns. From top to bottom the rows correspond to the hazard ratio curve, control group hazard curve and treatment group hazard curve. The first and third columns correspond to the extended piecewise exponential model with $K = 20$, and the second and fourth columns to the extended piecewise linear log-hazard model with $K = 8$. The top row evaluates each method for hazard ratio estimation, whereas the middle and bottom row pertain to estimation of the conditional hazard where $z = 0$ and $z = 1$, respectively. All results are based on $R = 200$ simulation runs each with a sample size of $N = 200$.

runs, after which we calculate the pointwise mean hazard ratio curve and conditional hazard function curves from each approach, as well as the empirical 2.5th, 10th, 90th and 97.5th pointwise percentiles. Using these quantities we again compare the behavior of the approaches visually. For a concise quantitative comparison, we calculate RISE for the hazard ratio curve and the conditional hazards in a similar fashion to (14).

For the time-dependent scenario in Figure 4, the extended PE model (first column) results in discontinuous estimates of each curve, whereas the extended PL model (second column) results in continuous estimates of the log-hazard. For all three quantities, the extended PL model exhibits slightly greater variability in the right tail where the data become sparse. However, the extended PL model still results in a smaller RISE than the extended PE model for all three curves. Though not depicted, the PH models result in constant estimates and exhibit RISE values two to three times larger in magnitude than the extended PL model. Cox’s PH model only provides estimates for the hazard ratio curve with a large RISE of 1.094.

Turning to the time-independent scenario in Figure 4, the extended PL model (fourth
Flexible Bayesian survival modeling

Table 1: Distributional summaries for $RISE^{(r)}$ values of the individual hazard ratio curve estimates.

<table>
<thead>
<tr>
<th>Model</th>
<th>True Time-dependent Effect</th>
<th>Mean (SD)</th>
<th>2.5%</th>
<th>10%</th>
<th>50%</th>
<th>90%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended PL</td>
<td></td>
<td>0.58 (0.20)</td>
<td>0.25</td>
<td>0.35</td>
<td>0.57</td>
<td>0.80</td>
<td>1.05</td>
</tr>
<tr>
<td>Extended PE</td>
<td></td>
<td>0.63 (0.15)</td>
<td>0.38</td>
<td>0.44</td>
<td>0.63</td>
<td>0.83</td>
<td>0.90</td>
</tr>
<tr>
<td>PL-PH</td>
<td></td>
<td>1.13 (0.09)</td>
<td>1.07</td>
<td>1.07</td>
<td>1.09</td>
<td>1.22</td>
<td>1.36</td>
</tr>
<tr>
<td>PE-PH</td>
<td></td>
<td>1.13 (0.10)</td>
<td>1.07</td>
<td>1.07</td>
<td>1.09</td>
<td>1.23</td>
<td>1.40</td>
</tr>
<tr>
<td>Cox’s PH</td>
<td></td>
<td>1.13 (0.09)</td>
<td>1.07</td>
<td>1.07</td>
<td>1.09</td>
<td>1.23</td>
<td>1.36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>True Time-independent Effect</th>
<th>Mean (SD)</th>
<th>2.5%</th>
<th>10%</th>
<th>50%</th>
<th>90%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended PL</td>
<td>0.48 (0.30)</td>
<td>0.14</td>
<td>0.21</td>
<td>0.41</td>
<td>0.74</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>Extended PE</td>
<td>0.33 (0.19)</td>
<td>0.09</td>
<td>0.12</td>
<td>0.29</td>
<td>0.55</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>PL-PH</td>
<td>0.18 (0.15)</td>
<td>0.01</td>
<td>0.02</td>
<td>0.14</td>
<td>0.41</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>PE-PH</td>
<td>0.18 (0.15)</td>
<td>0.00</td>
<td>0.02</td>
<td>0.14</td>
<td>0.40</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>Cox’s PH</td>
<td>0.18 (0.15)</td>
<td>0.01</td>
<td>0.02</td>
<td>0.14</td>
<td>0.40</td>
<td>0.52</td>
<td></td>
</tr>
</tbody>
</table>

The previous evaluation criteria assess the performance of the average posterior mean estimates, but we may additionally compare distributional summaries of the individual $RISE^{(r)}$ values for the posterior mean hazard ratio curve estimates amongst the $R = 200$ simulation runs. Letting $\hat{HR}^{(r)}(t)$ denote an estimate for the hazard ratio curve in the $r$th simulation run, we define

$$RISE^{(r)} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left( \hat{HR}^{(r)}(s_m) - HR(s_m) \right)^2}, \text{ for } r = 1, \ldots, R.$$ 

We then compare the mean, standard deviation, and selected empirical quantiles of the $RISE^{(r)}$’s for each model.

The results are summarized in Table 1. The extended PL model, owing to its flexibility, exhibits the greatest variation of all the methods, as evidenced by its largest standard deviation (SD). However, when the effect of $z$ is truly time-dependent, the extended PL model outperforms all the other models considered. The performances of the three PH models are nearly indistinguishable with respect to the hazard ratio curve for both the time-dependent and time-independent scenarios. When the effect
of \(z\) is time-independent, as these model correctly assume, the PH models outperform the highly parametrized extended PL and PE models. However, the \(RISE^{(r)}\) extended model distributions overlap substantially with those of the PH models.

### 4.3 Time-independent Monotone Effect Estimation

For our last investigation we introduce a continuous covariate \(x\), and consider a scenario where the effect of \(x\) is a monotone nonlinear function, and another where it is linear. Here we compare the performance of the nonparametric model defined in (11) that imposes monotonicity shape restrictions with that of the unrestricted nonparametric model defined in (10), using the PL model defined in (5) as the model for the baseline hazard. In the linear scenario, the proposed models facilitate more flexibility than is needed, so we also fit a model that correctly assumes linearity in the effect of \(x\) with the PL model for the baseline hazard. We also consider Cox’s PH model with an AIC-optimal smoothing spline estimate for the effect of \(x\). We estimate the latter model using the \texttt{coxph()} function with a \texttt{pspline(x,df=0)} term.

For evaluation, we generate \(N = 200\) survival observations \((t_i, \delta_i, x_i)\) from a survival distribution with a prespecified conditional hazard \(h(t|x) = h^*(t) \exp\{f(x)\}\) 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 for each run we are only generating new, possibly right censored observation times that are conditional on the same set of \(x_i\)’s. We then fit each model to these data, and save the posterior mean estimates for the log-hazard ratio curve at the \(N\) equally spaced \(x_i \in [0, 1]\) values. The log-hazard ratio curve can be calculated relative to any \(x\) value; we choose to calculate this curve relative to the central value, \(x = .5\). Hence, the estimated log-hazard ratio curve denotes the log-hazard ratio for an individual having arbitrary \(x\) versus an individual having \(x = 0.5\); therefore, the value of the estimated curve at \(x = 0.5\) is always zero. Iterating this process over \(R = 200\) simulation runs, for each model we calculate the pointwise mean log-hazard ratio curve, corresponding \(RISE\) values, and pointwise percentiles as we did in Section 4.1.

For the scenario with a true monotone nonlinear log-hazard ratio curve, Figure 6 shows the model with the standard linear regressor (top left panel) results in a suboptimal linear fit with the worst \(RISE\) value. Cox’s PH model with an AIC-optimal smoothing spline estimate of \(f(x)\) (top right panel) and the proposed unrestricted model (bottom left panel) perform similarly, though the former tends to exhibit greater variability as indicated by the wider empirical percentile regions. By contrast, the proposed shape-restricted model (bottom right panel) exhibits the best average behaviour as indicated by the lowest \(RISE\) value. In addition, the shaded percentile regions for the shape-restricted model confirm that the resulting estimate always satisfies \(f(x < .5) \leq f(x = .5) = 0 \leq f(x > .5)\).

Turning to the monotone linear scenario, Figure 6 shows that the performance of Cox’s PH model (top right panel) and the proposed unrestricted model (bottom left panel) do not deteriorate much relative to the “correct” model that uses a standard
Flexible Bayesian survival modeling

Figure 6: Results for the scenario with a monotone nonlinear effect of $x$ are displayed in the four leftmost panels, whereas the results for the monotone linear scenario are displayed in the four rightmost panels. On each respective side, the upper left panel corresponds to the model with a standard linear regressor, the upper right panel corresponds to Cox’s proportional hazards model with an AIC optimal smoothing spline estimator of $f(x)$, and the bottom left and right panels correspond to the proposed nonparametric methods without and with monotone shape restrictions, respectively. All results are based on $R = 200$ simulation runs each with a sample size of $N = 200$.

linear regressor (top left panel). Cox’s PH model again exhibits greater variability than the proposed unrestricted method as indicated by the wider and undulating percentile regions. By contrast, the proposed shape-restricted model (bottom right panel) seems to deteriorate slightly, perhaps owing to the relatively large prior weight given to the probability masses on zero (i.e. $Pr(p_j = 0) = .5$) 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 the methods proposed 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) in terms of overall survival, while jointly characterizing 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 allow time-dependent effects for each regime as described in Section 3.1. Furthermore, AST is thought to have a non-decreasing time-independent effect, so we consider jointly modeling the effect of AST with and without a monotonicity shape restriction, as described
Figure 8: Hazard ratio curves over the course of follow-up for each treatment comparison (first column), and over the domain of aspartate transaminase (AST) with and without shape restrictions (second column). From top to bottom, the hazard ratio curves on the left compare IROX to IFL, FOLFOX to IFL, and FOLFOX to IROX. Whereas, the hazard ratio curves on the right are the unrestricted estimate followed by the monotone shape-restricted estimate each calculated relative to an individual having an AST of 40 U/L. The light grey tick marks denote the observed event times for the relevant regimes in the left column, and the observed AST values in the right column, whereas the thick dark grey tick marks denote the quartiles.

To analyze these data, we standardize the observation times \((t_i)\) and AST values \((x_i)\) so that \(t_i \in (0, 1)\) and \(x_i \in [0, 1]\), \(i = 1, \ldots, N\). After preliminary investigation, we selected an equally spaced time-axis partition of size \(K = 20\), and equally spaced partitions of the AST covariate range of size \(J = 20\) for the unrestricted model and \(J = 10\) for the shape-restricted model. For the model where the effect of AST was restricted to be monotone, we chose to use prior mixture weights of \(v_j = .5, j = 1, \ldots, J + 1\). This choice indicates a moderate prior belief that the log-hazard ratio curve is constant for some interval along the domain of AST values. For estimation, we used 2 chains with 2,000 iterations of burn-in, followed by 20,000 iterations for posterior estimation.

The results of our analysis are displayed in Figure 8. The first column displays posterior summaries of the 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 monotonicity shape restrictions. We note that these results are nearly indistinguishable from the model that adjusts for AST nonparametrically without constraints, 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 amount of 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. Naturally, the amount of evidence decreases as the number of patients at risk shrinks as evidenced by the increasing width of the credible regions as follow-up accrues. The benefits of either IROX of 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 indeed be acceptable for these data.

The second column of Figure 8 compares the results from the unrestricted (top right panel) and shape-restricted (bottom right panel) analysis for the hazard ratio over the domain of AST. Both of these curves depict the hazard ratio for an individual having arbitrary AST units per liter (U/L) versus an individual having an AST of 40 U/L, which is the standard upper threshold for the normal range of AST. Hence, the hazard ratio must be one at an AST of 40 U/L. The unrestricted estimate suggests an increasing hazard of death up to 150 U/L, but this hazard diminishes for higher AST values. Since larger AST values are indicative of complications, 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. By contrast, the monotone estimate is static in the lower domain of AST values (≤30 U/L), then increases sharply until about 60 U/L, at which point it remains constant, suggesting that the hazard ratio of death remains constant for increasingly high levels of AST. Hence, the resulting estimate provides evidence that patients with AST at or above 60 U/L appear to have a hazard for death nearly 1.5 times higher than those with an AST of 40 U/L, whereas persons with AST below 30 U/L have a hazard about 0.75 times lower than those with an AST of 40 U/L.

6 Discussion

This article presented a highly flexible framework for conducting a fully Bayesian analysis of survival data that can adjust for covariates using nonparametric time-dependent effects and shape-restricted time-independent effects. 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 exhibit fast MCMC convergence, thereby making the estimation of these models computationally feasible.
The simulation investigations in Section 4 show that our proposed methods are competitive with similar existing approaches, and offer estimated curves that better align with prior beliefs about the smoothness of the true mechanism than similar versions of the presently popular piecewise exponential model. The simulations also showed that monotone shape restrictions are beneficial in the context of a strongly nonlinear log-hazard ratio curve, though in the context of a linear log-hazard ratio curve these restrictions can exhibit worse RISE properties 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. Further exploration into the benefits and disadvantages of shape-restricted proportional hazards models is a worthy topic for future research.

Regarding the estimation of a hazard function and time-dependent effects, the proposed methods suffer from more variability in upper domain, where data are often sparse, than a similar piecewise exponential model. A possible remedy for this somewhat undesirable behavior would be to constrain the piecewise linear log-hazard model to be constant beyond some boundary point along the time-axis. Regardless, the piecewise linear model for the log-hazard or baseline log-hazard function provides improvements over the piecewise exponential model for complex shapes that are not well approximated by a discontinuous function. In practice, we suggest using the piecewise linear log-hazard model and its extension that facilitates time-dependent effect estimation when there is reason to suspect moderate time-dependencies in the hazard function or the effect of a covariate.

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, our conclusions were unaffected using our more flexible models, and considering the possibility that the novel treatments have a time-dependent effect is good practice and often illuminating. The proposed methods afford great flexibility and robustness to rigid assumptions like time-independence and linearity. Lastly, these methods make it easy for an investigator to explore important questions about modeling assumptions, which are difficult to address using currently available software.

References


Flexible Bayesian survival modeling


**Acknowledgments**

This work was supported in part by NCI grant 1-R01-CA157458-01A1.

**Appendix**

**Low-Rank Thin Plate Spline Implementation**

To implement the low-rank thin plate spline model of Section 2.2 we construct a \((K + 1) \times (K + 1)\) design matrix \(T_K\) with the \(i\)th row given by  
\[
\begin{bmatrix} 1, t_i, |t_i - 1| & \cdots & |t_i - K| & -|\tilde{t}_K - 1| 
\end{bmatrix},
\]
so that \((5)\) can be rewritten succinctly as  
\[
h(t_i; \alpha) = \exp \left(t_i^T \alpha \right). 
\]
We then construct a \((K + 1) \times (K + 1)\) transformation matrix  
\[
D_{\alpha} = \begin{pmatrix} I_2 & 0 \\ \Omega_{\alpha}^{1/2} \\ 0 \end{pmatrix}
\]
where the \((\ell,k)\)th entry of the penalty matrix \(\Omega_{\alpha}\) is defined as  
\[
|\tilde{t}_\ell - \tilde{t}_k|,
\]
for \(\ell, k = 1, \ldots, K - 1\). We apply the transformations  
\[
a = D_{\alpha} \alpha, 
\]
\[
T = T_K D_{\alpha}^{-1},
\]
and  
\[
U = U_K D_{\alpha,-1}^{-1}
\]
where \(U_K\) is a \(K \times K\) matrix with \(k\)th row given by  
\[
u^T_k, 
\]
and \(D_{\alpha,-1}\) is the \(K \times (K + 1)\) matrix obtained by omitting the first row of \(D_{\alpha}^{-1}\).

Next, we rewrite the hazard defined in \((5)\) and cumulative hazard defined in \((2)\) as  
\[
h(t_i; \alpha) \equiv h(t_i; a) = \exp \left(t_i^T a \right)
\]
and  
\[
H(t_i; \alpha) \equiv H(t_i; a) = \sum_{k=1}^{K} \frac{h(s_{i,k}; a) \left\{ 1 - e^{-(s_{i,k} - \tilde{t}_{K-1}) u^T_k a} \right\}}{u^T_k a},
\]
where  
\[
t_i^T \text{ is the } i\text{th row of } T, \text{ and } u^T_k \text{ is the } k\text{th row of } U.
\]
The likelihood for the pair  
\[
(t_i, \delta_i)
\]
arises by plugging the hazard and cumulative hazard definitions from \((15)\) 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 \((9)\) for the \(i\)th observation succinctly as  
\[
h(t_i|z_i; \alpha) = \exp \left\{ t_i^T (\alpha_0 + \alpha_1 z_i) \right\}
\]
where  
\[
t_{i,K}^T \text{ is the } i\text{th row of } T.\]
We then use the aforementioned transformations, taking  
\[
a = D_{\alpha} \alpha, 
\]
\[
T = T_K D_{\alpha}^{-1}, \text{ and } U = U_K D_{\alpha,-1}^{-1},
\]
and rewrite  
\[
h(t|z; \alpha) \text{ and } H(t|z; \alpha)
\]
completely in terms of  
\[
a, T, \text{ and } U.
\]
We omit further details, since the resulting expressions mirror
Flexible Bayesian survival modeling

Figure 10: Each row corresponds to an investigation under a different true hazard (broken line). The two leftmost columns contain results pertaining to hazard estimation and the two rightmost columns pertain to survival distribution estimation. The first and third columns from the left correspond to the piecewise exponential model with $K = 20$, whereas the second and fourth column correspond to the piecewise linear log-hazard model with $K = 8$. All results are based on $R = 200$ simulation runs each with a sample size of $N = 50$.

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 proportional hazards non-parametric model defined in (10) using the $J \times J$ transformation matrix $D_\beta = \left( \begin{array}{cc} 1 & 0 \\ 0 & \Omega_\beta^{1/2} \end{array} \right)$, wherein the $(j,k)$th entry of $\Omega_\beta$ is defined as $|\tilde{x}_j - \tilde{x}_k|^3$, for $j,k = 1, \ldots, J - 1$. We let $x^i_j = (x_i, |x_i - \tilde{x}_1|^3, |\tilde{x}_1|^3, \ldots, |x_i - \tilde{x}_{J-1}|^3, |\tilde{x}_{J-1}|^3)$ denote the $i$th row of the $N \times J$ matrix $X_J$, and define $X = X_J D_\beta^{-1}$ and $b = D_\beta \beta$. Doing so, we have that $f(x_i; \beta) \equiv f(x_i; b) = x^i_1 b$ where $x^i_1$ denotes the $i$th row of $X$.

Additional Simulation Results
Results of the simulation investigation described in Section 4.1 for $N = 50$ are depicted in Figure 10. In this scenario, both models exhibit greater variation owing to the smaller sample size; however, the same trends hold, the PL model exhibits better RISE properties both in terms of hazard function and survival distribution estimation. For either $N = 50$ or $N = 200$, both the PL and PE models exhibit markedly worse properties for the smaller $K$ values considered, with which there are too few parameters to reasonably capture the features of the true hazard function.