Biometrics 61, 403–412
June 2005

DOI: 10.1111/j.1541-0420.2005.00329.x

A General Class of Bayesian Survival Models with Zero and Nonzero Cure Fractions

Guosheng Yin

Department of Biostatistics and Applied Mathematics, M. D. Anderson Cancer Center, The University of Texas, Houston, Texas 77030, U.S.A.
email: gyin@odin.mdacc.tmc.edu

and

Joseph G. Ibrahim

Department of Biostatistics, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, U.S.A.

Summary. We propose a new class of survival models which naturally links a family of proper and improper population survival functions. The models resulting in improper survival functions are often referred to as cure rate models. This class of regression models is formulated through the Box–Cox transformation on the population hazard function and a proper density function. By adding an extra transformation parameter into the cure rate model, we are able to generate models with a zero cure rate, thus leading to a proper population survival function. A graphical illustration of the behavior and the influence of the transformation parameter on the regression model is provided. We consider a Bayesian approach which is motivated by the complexity of the model. Prior specification needs to accommodate parameter constraints due to the nonnegativity of the survival function. Moreover, the likelihood function involves a complicated integral on the survival function, which may not have an analytical closed form, and thus makes the implementation of Gibbs sampling more difficult. We propose an efficient Markov chain Monte Carlo computational scheme based on Gaussian quadrature. The proposed method is illustrated with an example involving a melanoma clinical trial.

Key words: Bayesian computation; Box–Cox transformation; Constrained parameter; Cure rate model; Gaussian quadrature; Gibbs sampling.

1. Introduction

In many clinical studies, especially in cancer research, there may be a certain percentage of patients who respond favorably to treatment and appear to be risk free or "cured" of the disease of interest after a sufficient period of follow-up. For such failure-time data, a proportion of subjects from the population are susceptible to, and others are not susceptible to, the target event. Empirical evidence to confirm this feature of the population is a long and stable plateau with heavy censoring at the tail of the Kaplan–Meier survival curve. With long-term survivors (i.e., survival data with a cure fraction), the usual Cox proportional hazards model (Cox, 1972) is not applicable since the proportional hazards assumption is violated. Cure rate models (Ibrahim, Chen, and Sinha, 2001a) are constructed specifically for modeling time-to-event data incorporating a cure fraction. These types of models are becoming increasingly useful in clinical trials, especially in oncology studies.

There are two major approaches to modeling survival data with a cure fraction. One is the standard mixture cure model (Berkson and Gage, 1952) where a proportion of the population, say $\theta (0 < \theta < 1)$, is cured and free of risk of the event, and $(1 - \theta)$ of the patients are not "cured" and follow a proper survival function. Let $S_{\text{pop}}(t)$ be the population survival function, and let $S(t)$ be the (proper) survival function for the "noncured" subjects. The standard mixture cure model is defined as

$$S_{\text{pop}}(t) = \theta + (1 - \theta)S(t). \tag{1}$$

Clearly, we see that (1) is improper since $S_{\text{pop}}(\infty) = \theta$. When covariates are included, we have a different θ_i for each subject, $i = 1, \ldots, n$. A logistic regression structure for θ_i is usually (Kuk and Chen, 1992) assumed, i.e.,

$$\theta_i = \frac{\exp(\boldsymbol{\beta}'\mathbf{Z}_i)}{1 + \exp(\boldsymbol{\beta}'\mathbf{Z}_i)},$$

where θ_i is a probability and cannot be zero. The standard mixture cure model has been extensively studied in the literature (Gray and Tsiatis, 1989; Kuk and Chen, 1992; Taylor, 1995; Maller and Zhou, 1996; Peng and Dear, 2000; Sy and Taylor, 2000; Betensky and Schoenfeld, 2001; among others). Although model (1) is intuitively attractive and

widely used, it does not have a proportional hazards structure for $S_{\text{pop}}(t)$, which is a desirable property in carrying out covariate analyses.

An alternative cure rate model, with a proportional hazards structure for the population, sometimes called the promotion time cure rate model (Yakovlev et al., 1993; Chen, Ibrahim, and Sinha, 1999), is given by

$$S_{\text{pop}}(t) = \exp\{-\theta F(t)\}, \tag{2}$$

where $F(t)$ is a proper cumulative distribution function (c.d.f.) and represents the promotion time, i.e., time to development of a detectable tumor mass (see Ibrahim et al., 2001a, Chapter 5). The corresponding population hazard function is $\lambda_{\text{pop}}(t) = \theta f(t)$, where $f(t) = dF(t)/dt$ is the density corresponding to the promotion time c.d.f. In this formulation, we typically model the covariates by letting $\theta_i = \exp(\boldsymbol{\beta}'\mathbf{Z}_i)$. The cure rate implied by (2) is $\exp(-\theta_i)$, and for subject i with covariate vector $\mathbf{Z}_i$, the population hazard can be written as

$$\lambda_{\text{pop}}(t \,|\, \mathbf{Z}_i) = f(t)\exp(\boldsymbol{\beta}'\mathbf{Z}_i). \tag{3}$$

Tsodikov (1998, 2002) among others investigated (2) in the frequentist framework, while Chen et al. (1999) and Ibrahim, Chen, and Sinha (2001b) formulated a Bayesian approach to (2). Model (2) is biologically motivated as follows. For a given patient, let N be the number of metastasis-competent tumor cells that remain active (capable of metastasizing) after treatment. Further, assume that N follows a Poisson distribution with mean θ, and let X_j be the time for the jth tumor cell in that individual to produce detectable metastatic disease. Given N, $(X_0, X_1, \ldots, X_N)$ are assumed to be independent and identically distributed (i.i.d.) with a common c.d.f. $F(x) = 1 - S(x)$ where $\Pr(X_0 = \infty) = 1$. Thus, the time to relapse of cancer is $T = \min(X_1, \ldots, X_N)$ which has a population survival function given by

$$\begin{aligned} S_{\text{pop}}(t) &= \Pr(\text{no metastatic cancer by time } t) \\ &= \Pr(N = 0) \\ &\quad + \Pr(X_1 > t, \ldots, X_N > t \,|\, N \geq 1)\Pr(N \geq 1) \\ &= \exp(-\theta) + \sum_{k=1}^{\infty} S(t)^k \frac{\theta^k \exp(-\theta)}{k!} \\ &= \exp\{-\theta F(t)\}. \end{aligned}$$

Chen et al. (1999) introduced the latent Poisson random variable N to facilitate the Markov chain Monte Carlo (MCMC) computation. They showed that under some mild conditions, a uniform improper prior for $\boldsymbol{\beta}$ in (3) still leads to a proper posterior, while it always yields an improper posterior for $\boldsymbol{\beta}$ if we use (1). This is a solid feature of the promotion time cure rate model, since it facilitates Bayesian inference with improper priors. Tsodikov, Ibrahim, and Yakovlev (2003) provide a comprehensive review of the recent developments of model (2). Both models (1) and (2) result in improper population survival functions, since they do not allow for a zero cure fraction, i.e., $S_{\text{pop}}(\infty) \neq 0$.

As opposed to a multiplicative structure in (2), analogous to the additive hazards model (Lin and Ying, 1994), we propose a new additive model,

$$\lambda_{\text{pop}}(t \,|\, \mathbf{Z}_i) = f(t) + \boldsymbol{\beta}'\mathbf{Z}_i, \tag{4}$$

where $f(t)$ is a proper density function instead of a baseline hazard function $\lambda_0(t)$, and "proper" means $\int_0^\infty f(t)\,dt = 1$. Note that (4) leads to a proper survival function since the population cumulative hazard function $\Lambda_{\text{pop}}(\infty \,|\, \mathbf{Z}_i) = \int_0^\infty \lambda_{\text{pop}}(t \,|\, \mathbf{Z}_i)\,dt$ is not bounded. In fact, both models (2) and (4) belong to a more general family of transformation models, as demonstrated in Section 2. In conventional linear models, the Box–Cox transformation (Box and Cox, 1964) is well studied and routinely used to transform the observed outcomes, which is defined as,

$$Y^{(\gamma)} = \begin{cases} (Y^\gamma - 1)/\gamma & \gamma \neq 0 \\ \log(Y) & \gamma = 0, \end{cases} \tag{5}$$

where γ is the transformation parameter and $\gamma \in R^1$ (the real line). This is a continuous transformation, since $\lim_{\gamma \to 0}(Y^\gamma - 1)/\gamma = \log(Y)$.

In this article, we unify the promotion time cure model (3) and the additive model (4) into a general class, by imposing the Box–Cox transformation on $\lambda_{\text{pop}}(t \,|\, \mathbf{Z}_i)$ and $f(t)$ in (4). Instead of modeling the baseline hazard, we build up this new family of survival models from the density function. This opens a whole new arena of modeling structures as alternatives to the Cox proportional hazards model. This class of transformation models is very general, which provides a natural link between proper and improper survival functions (i.e., zero and nonzero cure fractions), through a single Box–Cox transformation parameter γ. The transformation class has model (2) as a special case, which is the only one in this family that has a nonzero cure fraction, and it also includes a broad range of regression models yielding proper population survival functions (zero cure fractions). In the Bayesian framework, we derive the likelihood function based on a piecewise exponential assumption, and study the prior specification under a constrained parameter space. Because the full conditionals involve a complicated integral that does not have a closed form for some γ's, we propose an efficient Gaussian quadrature approximation to carry out Gibbs sampling.

The rest of this article is organized as follows. In Section 2, we introduce notation and a new class of transformation survival models. In Section 3, we formulate the likelihood function within the Bayesian paradigm, discuss the prior specification, and derive the full conditional distributions needed for Gibbs sampling. In Section 4, we propose two types of model assessment procedures based on the Conditional Predictive Ordinate (Geisser, 1993) and the Deviance Information Criterion (Spiegelhalter et al., 2002). We illustrate the proposed methods with a melanoma clinical trial in Section 5 and give concluding remarks in Section 6.

2. A Class of Transformation Survival Models

Let T_i be the failure time for the ith subject, let C_i be the censoring time, and $Y_i = \min(T_i, C_i)$ is the observed time for $i = 1, \ldots, n$. Correspondingly, the failure-time indicator is $\nu_i = I(T_i \leq C_i)$ where $I(\cdot)$ is the indicator function. Let $\mathbf{Z}_i$ be the $(p + 1) \times 1$ covariate vector where the first component of $\mathbf{Z}_i$ is a constant, 1, which corresponds to the intercept. We

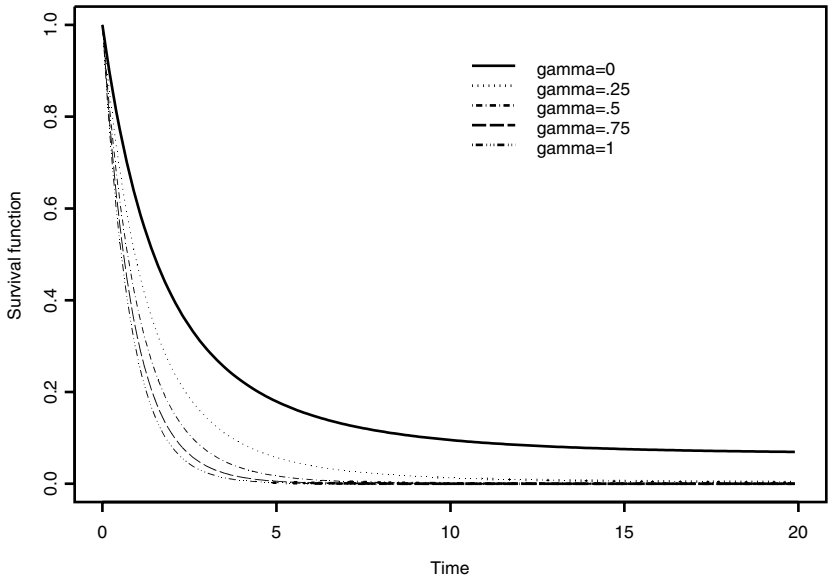

Figure 1. Simulated population survival curves for a set of γ's given $Z = 1$.

assume that T_i and C_i are independent conditional on $\mathbf{Z}_i$, and $\{(T_i, C_i, \mathbf{Z}_i), i = 1, \ldots, n\}$ are i.i.d.

By imposing the Box–Cox transformation on the population hazard function and the baseline density function, we propose a class of survival models as

$$\frac{\lambda_{\text{pop}}(t \mid \mathbf{Z}_i)^\gamma - 1}{\gamma} = \frac{f(t)^\gamma - 1}{\gamma} + \boldsymbol{\beta}'\mathbf{Z}_i, \quad \gamma \in [0, 1]. \tag{6}$$

It is apparent that we can rewrite (6) as

$$\lambda_{\text{pop}}(t \mid \mathbf{Z}_i) = \left\{ f(t)^\gamma + \gamma \boldsymbol{\beta}'\mathbf{Z}_i \right\}^{1/\gamma}. \tag{7}$$

As $\gamma \to 0$, model (6) becomes $\log\{\lambda_{\text{pop}}(t \mid \mathbf{Z}_i)\} = \log\{f(t)\} + \boldsymbol{\beta}'\mathbf{Z}_i$, and thus reduces to (3), and when $\gamma = 1$, model (6) reduces to (4) with an additive structure. Here, the utilization of the Box–Cox transformation is substantially different from (5), since it is applied to the unknown functions $\lambda_{\text{pop}}(t \mid \mathbf{Z}_i)$ and $f(t)$ rather than to observable quantities. We take γ as fixed throughout the article since treating γ as an unknown parameter leads to intractable posterior distributions as well as weak identifiability (more discussion on this issue is given in Section 6). Our primary interest for γ lies in [0, 1], since $\gamma = 0$ and $\gamma = 1$ represent the two extreme cases, i.e., the multiplicative and additive modeling schemes, respectively, and $0 < \gamma < 1$ results in an intermediate modeling structure. Although γ could mathematically take any value on the real line in (6), values of $\gamma \in [0, 1]$ are perhaps the most meaningful and useful. It is important to note that $\gamma = 0$ is the only case that results in an improper survival function, and that $\gamma \neq 0$ always leads to a proper survival function.

To gain some insight into the generality and flexibility of model (7), we present a numerical study here. We take $f(t)$ to be an exponential density with mean 5, i.e., $f(t) = \lambda \exp(-\lambda t)$ and $\lambda = .2$. The true values of the regression parameters are $\beta_0 = .1$ and $\beta_1 = 1$ and a single binary covariate Z takes a value of 0 or 1 with probability .5. We take $\gamma = (0, .25, .5, .75, 1)$. Figure 1 shows how the population survival functions for $Z = 1$ vary with respect to the different γ's. The solid line represents $\gamma = 0$, which is the only survival function with a nonzero cure fraction. The rest of the functions are proper survival functions, satisfying $S_{\text{pop}}(\infty) = 0$. Based on (7), the population cumulative hazard function for the ith subject is given by

$$\begin{aligned}\Lambda_{\text{pop}}(t \mid \mathbf{Z}_i) &= \int_0^t \left\{ f(u)^\gamma + \gamma \boldsymbol{\beta}'\mathbf{Z}_i \right\}^{1/\gamma} du \\ &= \int_0^t \left\{ \lambda^\gamma \exp(-\gamma\lambda u) + \gamma \boldsymbol{\beta}'\mathbf{Z}_i \right\}^{1/\gamma} du, \end{aligned} \tag{8}$$

and the population survival function is given by $S_{\text{pop}}(t \mid \mathbf{Z}_i) = \exp\{-\Lambda_{\text{pop}}(t \mid \mathbf{Z}_i)\}$. When $\gamma \neq 0$, clearly, $|\lambda^\gamma \exp(-\gamma\lambda u) + \gamma\boldsymbol{\beta}'\mathbf{Z}_i|^{1/\gamma} \geq |\gamma\boldsymbol{\beta}'\mathbf{Z}_i|^{1/\gamma}$, and thus $\Lambda_{\text{pop}}(\infty \mid \mathbf{Z}_i)$ is not bounded and $S_{\text{pop}}(\infty \mid \mathbf{Z}_i) = 0$, thus leading to a proper survival function. When $\gamma \to 0$,

$$\Lambda_{\text{pop}}(\infty \mid \mathbf{Z}_i) = \int_0^\infty f(t) \exp(\boldsymbol{\beta}'\mathbf{Z}_i)\, dt = \exp(\boldsymbol{\beta}'\mathbf{Z}_i),$$

which is bounded and the cure rate is $S_{\text{pop}}(\infty \mid \mathbf{Z}_i) = \exp\{-\exp(\boldsymbol{\beta}'\mathbf{Z}_i)\}$ resulting in an improper survival function. Specifically, we derive $\Lambda_{\text{pop}}(\infty \mid \mathbf{Z}_i)$ for several γ's under the exponential assumption for $f(t)$ to demonstrate how a variety of proper survival distributions can be generated. For example, when $\gamma = 1$,

$$\Lambda_{\text{pop}}(t \mid \mathbf{Z}_i) = 1 - \exp(-\lambda t) + \boldsymbol{\beta}'\mathbf{Z}_i t,$$

and when $\gamma = .5$,

$$\begin{aligned}\Lambda_{\text{pop}}(t \mid \mathbf{Z}_i) = 1 - \exp(-\lambda t) &+ (\boldsymbol{\beta}'\mathbf{Z}_i)^2 t/4 \\ &+ 2\boldsymbol{\beta}'\mathbf{Z}_i \lambda^{-1/2}\{1 - \exp(-\lambda t/2)\}.\end{aligned}$$

More generally, if $1/\gamma$ is an integer, we have a closed form for (8),

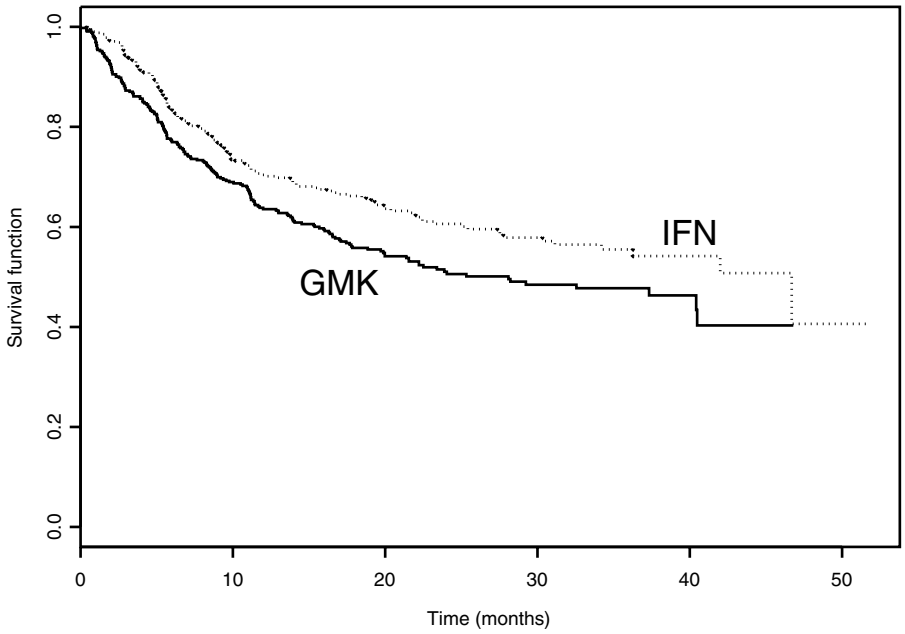

Figure 2. Kaplan–Meier curves of patient groups treated with IFN versus GMK vaccine in E1694.

$$\Lambda_{\text{pop}}(t \mid \mathbf{Z}_i) = \sum_{g=0}^{1/\gamma} \binom{1/\gamma}{g} \lambda^{\gamma g} (\gamma \boldsymbol{\beta}' \mathbf{Z}_i)^{1/\gamma - g} \int_0^t \exp(-\gamma g \lambda u)\, du, \tag{9}$$

where the integral in (9) equals $\{1 - \exp(-\gamma g \lambda t)\}/(\gamma g \lambda)$ if $g \neq 0$, and equals t if $g = 0$.

In practical applications, we can prespecify a set of γ's covering the interval [0, 1] in order to model a broad class of population hazard functions, then choose the best fitting model according to a suitable model selection criterion. Because the hazard function in (7) cannot be negative, the following parameter constraint is required when $\gamma \neq 0$:

$$f(t)^{\gamma} + \gamma \boldsymbol{\beta}' \mathbf{Z}_i \geq 0, \quad i = 1, \ldots, n. \tag{10}$$

Because $f(t)$ is a probability density function, $\lim_{t \to \infty} f(t) = 0$, and the constraint (10) reduces to $\boldsymbol{\beta}' \mathbf{Z}_i \geq 0$.

Due to the power parameter γ in (7), the regression parameters $\boldsymbol{\beta}$ are intertwined together with $f(t)$, thus making parameter estimation computationally challenging. To overcome the numerical difficulties, we propose to fit this transformation model using a Bayesian approach. A comprehensive discussion of Bayesian cure rate models is given in Chapter 5 of the book by Ibrahim et al. (2001a).

One interesting application of the proposed methodology is from a recent phase III melanoma cancer vaccine clinical trial conducted by the Eastern Cooperative Oncology Group (ECOG), the Southwest Oncology Group (SWOG), and the Cancer and Leukemia Group B, labeled E1694 (Kirkwood et al., 2001). Melanoma is a malignant form of skin cancer that affects melanin, the substance that pigments the skin. E1694 was a two-arm clinical trial comparing a new vaccine to high-dose interferon for prolonging relapse-free and overall survival in melanoma patients. Cancer vaccines are specifically designed to stimulate the immune system to recognize and attack cancer cells. Such vaccines are usually less toxic than chemotherapies and are used as adjuvant therapy following surgery to build up the antibody levels against a possible relapse of the tumor. The primary endpoint in E1694 was relapse-free survival time (in months), which was right censored. Relapse-free survival time was defined as the time from randomization to progression of tumor or death. The vaccine, labeled GMK, is a combination of GM2 (a ganglioside that is serologically defined to be a melanoma antigen) and KLH (keyhole limpet hemocyanin), administered with the adjuvant QS-21. Figure 2 shows the Kaplan–Meier survival plots for the two treatment arms.

3. Likelihood, Priors, and Gibbs Sampling

We consider a piecewise exponential model, which is a flexible and widely used modeling scheme for survival data. We formulate the likelihood function as follows. Let y_i be the observed time for subject i, and we partition the time scale into J intervals, i.e., $0 < s_1 < , \ldots, < s_J$, $s_J > y_i$ for $i = 1, \ldots, n$, where $\lambda_0(y) = \lambda_j$ for $y \in (s_{j-1}, s_j]$, $j = 1, \ldots, J$. By increasing J, the piecewise constant hazard model can essentially model any shape of the underlying hazard. A larger J allows more flexibility but it also introduces more unknown parameters, namely the λ_j's. Thus, there is a trade-off in determining the optimal J. The best J usually lies between 5 and 10. Define $\delta_{ij} = 1$ if the ith subject fails or is censored in the jth interval, and 0 otherwise. Under the piecewise exponential assumption, the promotion time density function in the jth interval is

$$f_j(t) = \lambda_j \exp\left[-\left\{ \sum_{k=1}^{j-1} \lambda_k (s_k - s_{k-1}) + \lambda_j (t - s_{j-1}) \right\} \right].$$

Let D denote the observed data, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$. The likelihood function is given by

$$
\begin{aligned}
&L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, D) \\
&= \prod_{i=1}^{n} \left\{ f(y_i)^{\gamma} + \gamma \boldsymbol{\beta}' \mathbf{Z}_i \right\}^{\nu_i/\gamma} \\
&\qquad \times \exp\left[- \int_0^{y_i} \left\{ f(t)^{\gamma} + \gamma \boldsymbol{\beta}' \mathbf{Z}_i \right\}^{1/\gamma} dt \right] \\
&= \prod_{i=1}^{n} \prod_{j=1}^{J} \left(\lambda_j^{\gamma} \exp\left[-\gamma \left\{ \sum_{k=1}^{j-1} \lambda_k (s_k - s_{k-1}) \right.\right.\right. \\
&\qquad \left.\left.\left. + \lambda_j (y_i - s_{j-1}) \right\} \right] + \gamma \boldsymbol{\beta}' \mathbf{Z}_i \right)^{\nu_i \delta_{ij}/\gamma} \\
&\qquad \times \exp\{ -\delta_{ij} \Lambda_j (y_i \,|\, \mathbf{Z}_i) \}.
\end{aligned}
$$

When $1/\gamma$ is an integer, we can obtain an explicit closed form for the integral associated with the population cumulative hazard as follows,

$$
\begin{aligned}
&\Lambda_j (y_i \,|\, \mathbf{Z}_i) \\
&= \sum_{k=1}^{j} I(y_i > s_{k-1}) \int_{s_{k-1}}^{\min(s_k, y_i)} \\
&\qquad \times \left(\lambda_k^{\gamma} \exp\left[-\gamma \left\{ \sum_{q=1}^{k-1} \lambda_q (s_q - s_{q-1}) \right.\right.\right. \\
&\qquad \left.\left.\left. + \lambda_k (t - s_{k-1}) \right\} \right] + \gamma \boldsymbol{\beta}' \mathbf{Z}_i \right)^{1/\gamma} dt \\
&= \sum_{k=1}^{j} I(y_i > s_{k-1}) \int_{s_{k-1}}^{\min(s_k, y_i)} \sum_{g=0}^{1/\gamma} \binom{1/\gamma}{g} \lambda_k^{\gamma g} \\
&\qquad \times \exp\left[-\gamma g \left\{ \sum_{q=1}^{k-1} \lambda_q (s_q - s_{q-1}) + \lambda_k (t - s_{k-1}) \right\} \right] \\
&\qquad \times (\gamma \boldsymbol{\beta}' \mathbf{Z}_i)^{1/\gamma - g} \, dt \\
&= \sum_{k=1}^{j} \sum_{g=0}^{1/\gamma} \binom{1/\gamma}{g} \lambda_k^{\gamma g} \exp\left[-\gamma g \left\{ \sum_{q=1}^{k-1} \lambda_q (s_q - s_{q-1}) \right.\right. \\
&\qquad \left.\left. - \lambda_k s_{k-1} \right\} \right] (\gamma \boldsymbol{\beta}' \mathbf{Z}_i)^{1/\gamma - g} \\
&\qquad \times I(y_i > s_{k-1}) \int_{s_{k-1}}^{\min(s_k, y_i)} \exp(-\gamma g \lambda_k t) \, dt.
\end{aligned}
$$

If $1/\gamma$ is not an integer, $\Lambda_j(y_i \,|\, \mathbf{Z}_i)$ does not have a closed form, and in this case, we use Gaussian quadrature to approximate the integral (Press et al., 1992). The Gaussian quadrature approximation provides flexibility in choosing the weights and abscissae where the functions are to be evaluated. In particular, we can take the 10 tabulated abscissae and weighting coefficients based on the Gauss–Legendre formula, which in fact yields satisfactory accuracy in our real data example. Let $(\zeta_1, \ldots, \zeta_{10})$ be the chosen abscissae, and let $\{w(\zeta_1), \ldots, w(\zeta_{10})\}$ be the corresponding weights. Then, when $1/\gamma$ is not an integer, the Gauss–Legendre quadrature approximation is given by

$$
\begin{aligned}
&\Lambda_j (y_i \,|\, \mathbf{Z}_i) \\
&\approx \sum_{k=1}^{j} \frac{I(y_i > s_{k-1}) \{ \min(s_k, y_i) - s_{k-1} \}}{2} \sum_{l=1}^{10} w(\zeta_l) \\
&\times \left\{ \lambda_k^{\gamma} \exp\left(-\gamma \left[\sum_{q=1}^{k-1} \lambda_q (s_q - s_{q-1}) + \lambda_k \left\{ \frac{\min(s_k, y_i) - s_{k-1}}{2} \zeta_l \right.\right.\right.\right. \\
&\qquad \left.\left.\left.\left. + \frac{\min(s_k, y_i) + s_{k-1}}{2} - s_{k-1} \right\} \right] \right) + \gamma \boldsymbol{\beta}' \mathbf{Z}_i \right\}^{1/\gamma}.
\end{aligned}
$$

Although not required for our development, we can take $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ to be independent a priori, where $\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\lambda})$. We also assume that the components of $\boldsymbol{\lambda}$ are independent a priori, and that each λ_j has a $Gamma(\alpha, \xi)$ distribution. One can easily construct priors to make the components of $\boldsymbol{\lambda}$ dependent a priori by considering first-order autoregressive structures or Markovian relations on the λ_j's, as in Arjas and Gasbarra (1994) and Ibrahim et al. (2001a). The constraint $\beta_0 + \beta_1 Z_{1i} + \cdots + \beta_p Z_{pi} \geq 0$, $(i = 1, \ldots, n)$, needs to be incorporated in the prior specification and each step of the Gibbs iterations. We can reduce the multiple dimensional constraint into a univariate one through a conditional-marginal specification of the joint prior (Yin and Ibrahim, 2005). In such a prior formulation, the constraint is completely absorbed in the conditional part of the prior specification, while the marginal part is free of constraints. For ease of exposition, we let $\boldsymbol{\beta}_{(-0)} = (\beta_1, \ldots, \beta_p)'$, and let $\boldsymbol{\lambda}_{(-j)}$ be the vector $\boldsymbol{\lambda}$ with the jth component removed. In particular, we propose a joint prior for $\boldsymbol{\beta}$ of the form

$$
\begin{aligned}
\pi(\boldsymbol{\beta}) = \pi(\beta_0 \,|\, \boldsymbol{\beta}_{(-0)}) I(\beta_0 \geq -\{\beta_1 Z_{1i} + \cdots + \beta_p Z_{pi}\}, \\
i = 1, \ldots, n) \pi(\boldsymbol{\beta}_{(-0)}),
\end{aligned} \tag{11}
$$

where the intercept β_0 is constrained. A natural prior for $\beta_0 \,|\, \boldsymbol{\beta}_{(-0)}$ is a truncated normal distribution. Let $\Phi(\cdot)$ denote the c.d.f. of the standard normal distribution. Thus, we have

$$
\begin{aligned}
\pi(\beta_0 \,|\, \boldsymbol{\beta}_{(-0)}) &= c(\boldsymbol{\beta}_{(-0)})^{-1} \exp\left(-\frac{\beta_0^2}{2\sigma_0^2} \right) \\
&\quad \times I\left(\beta_0 \geq - \min_{i=1,\ldots,n} \{\beta_1 Z_{1i} + \cdots + \beta_p Z_{pi}\} \right),
\end{aligned} \tag{12}
$$

where the normalizing constant is given by

$$
c(\boldsymbol{\beta}_{(-0)}) = \sqrt{2\pi}\sigma_0 \left[1 - \Phi\left(- \min_{i=1,\ldots,n} \left\{ \frac{\beta_1 Z_{1i} + \cdots + \beta_p Z_{pi}}{\sigma_0} \right\} \right) \right].
$$

Therefore, the full conditionals of the parameters are given by

$$
\begin{aligned}
&\pi(\beta_0 \,|\, \boldsymbol{\beta}_{(-0)}, \boldsymbol{\lambda}, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, D) \pi(\beta_0 \,|\, \boldsymbol{\beta}_{(-0)}), \\
&\pi(\boldsymbol{\beta}_{(-0)} \,|\, \beta_0, \boldsymbol{\lambda}, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, D) \pi(\boldsymbol{\beta}_{(-0)}) c(\boldsymbol{\beta}_{(-0)})^{-1}, \\
&\pi(\lambda_j \,|\, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-j)}, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, D) \pi(\lambda_j),
\end{aligned}
$$

where $\boldsymbol{\beta}_{(-0)} \sim N_p(\mathbf{0}, \Sigma)$, and $\lambda_j \sim Gamma(\alpha, \xi)$ for $j = 1, \ldots, J$. Based on the full conditional distributions, we use the ARMS (Adaptive Rejection Metropolis Sampling) algorithm of Gilks, Best, and Tan (1995) to obtain the posterior samples in each Gibbs iteration. The C code for implementing the proposed method is available upon request from the first author.

4. Model Comparison

Model selection plays an important role in survival analysis. After fitting the proposed models for a set of prespecified γ's, we can compute the Conditional Predictive Ordinate (CPO) statistic (Gelfand, Dey, and Chang, 1992; Geisser, 1993; Dey, Chen, and Chang, 1997). CPO is a Bayesian crossvalidation statistic and measures the adequacy of a given model. The CPO statistic validates the conditional predictive distribution from a single observation deletion against the observed responses.

Let $D^{(-i)}$ denote the data with the ith observation deleted. We denote the density function of y_i by $f(y_i \,|\, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{Z}_i)$, and the posterior density of $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ given $D^{(-i)}$ by $\pi(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, D^{(-i)})$, $i = 1, \ldots, n$. CPO_i is the marginal posterior predictive density of y_i given $D^{(-i)}$, which can be written as

$$\mathrm{CPO}_i = f\big(y_i \,|\, \mathbf{Z}_i, D^{(-i)}\big) = \left\{ \int\!\!\int \frac{\pi(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, D)}{f(y_i \,|\, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{Z}_i)} \, d\boldsymbol{\beta} \, d\boldsymbol{\lambda} \right\}^{-1}.$$

For the proposed transformation model, a Monte Carlo approximation of CPO_i (Chen, Shao, and Ibrahim, 2000) is given by

$$\widehat{\mathrm{CPO}}_i = \left\{ \frac{1}{M} \sum_{m=1}^{M} \frac{1}{L_i(\boldsymbol{\beta}_{[m]}, \boldsymbol{\lambda}_{[m]} \,|\, y_i, \mathbf{Z}_i, \nu_i)} \right\}^{-1},$$

where M is the number of Gibbs samples after burn-in, and $\boldsymbol{\beta}_{[m]}$ and $\boldsymbol{\lambda}_{[m]} = (\lambda_{1,[m]}, \ldots, \lambda_{J,[m]})'$ are the samples corresponding to the mth Gibbs iteration. Note that $L_i(\boldsymbol{\beta}_{[m]}, \boldsymbol{\lambda}_{[m]} \,|\, y_i, \mathbf{Z}_i, \nu_i)$ is a density function if $\nu_i = 1$ and a survival function if $\nu_i = 0$. Specifically,

$$\begin{aligned} &L_i(\boldsymbol{\beta}_{[m]}, \boldsymbol{\lambda}_{[m]} \,|\, y_i, \mathbf{Z}_i, \nu_i) \\ &= \prod_{j=1}^{J} \left(\lambda_{j,[m]}^{\gamma} \exp\left[-\gamma \left\{ \sum_{k=1}^{j-1} \lambda_{k,[m]}(s_k - s_{k-1}) + \lambda_{j,[m]}(y_i - s_{j-1}) \right\} \right] + \gamma \boldsymbol{\beta}_{[m]}' \mathbf{Z}_i \right)^{\nu_i \delta_{ij}/\gamma} \\ &\quad \times \exp\left\{ -\delta_{ij} \sum_{k=1}^{j} I(y_i > s_{k-1}) \right. \\ &\quad \left. \times \int_{s_{k-1}}^{\min(s_k, y_i)} \left(\lambda_{k,[m]}^{\gamma} \exp\left[-\gamma \left\{ \sum_{q=1}^{k-1} \lambda_{q,[m]}(s_q - s_{q-1}) + \lambda_{k,[m]}(t - s_{k-1}) \right\} \right] + \gamma \boldsymbol{\beta}_{[m]}' \mathbf{Z}_i \right)^{1/\gamma} dt \right\}. \end{aligned}$$

A summary statistic based on the CPO_i's is $B = \sum_{i=1}^{n} \log(\mathrm{CPO}_i)$. The larger the value of B, the better the fit of the model.

The Deviance Information Criterion (DIC) recently proposed by Spiegelhalter et al. (2002) is a Bayesian measure of fit and complexity for model selection, defined as

$$\mathrm{DIC} = \overline{\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})} + p_{\mathrm{Dev}}.$$

The deviance $\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = -2 \log L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, D)$, $p_{\mathrm{Dev}} = \overline{\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})} - \mathrm{Dev}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\lambda}})$, and thus $\mathrm{DIC} = 2\overline{\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})} - \mathrm{Dev}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\lambda}})$, where $\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\lambda}}$ and $\overline{\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})}$ are the corresponding posterior means. The penalty term p_{Dev} reflects the effective number of parameters in the model. In the proposed model,

$$\mathrm{DIC} = -\frac{4}{M} \sum_{m=1}^{M} \log L(\boldsymbol{\beta}_{[m]}, \boldsymbol{\lambda}_{[m]} \,|\, D) + 2 \log L(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\lambda}} \,|\, D).$$

The smaller the DIC, the better the fit of the model. As shown in the E1694 example, the CPO and DIC statistics are quite consistent with each other in discriminating between different models.

5. Application

To illustrate the methodology, we applied the proposed class of survival models to the E1694 trial, in which the primary objective was to evaluate whether the GMK vaccine was superior to interferon-α2b (IFN) with respect to the endpoint of relapse-free survival. In this analysis, there were 876 subjects in the combined treatment arms, of which 517 were right censored. The covariates of interest were treatment (GMK = 1 or IFN = 2), age (a continuous variable which ranged from 19.24 to 84.85 years with a mean of 51.15 years), and sex (female = 1, male = 2).

We examined $J = 1$ (a parametric exponential model) and $J = 5$ with $\gamma = (0, .25, .5, .75, 1)$. All the parameters were assumed independent a priori and assigned noninformative priors. For example, we took $\beta_k \sim N(0, 10, 000)$, for $k = 0, 1, 2, 3$, where β_0 had a truncated normal prior in (12), $\lambda_j \sim Gamma(2, .01)$, and independent for $j = 1, \ldots, J$. We monitored the convergence of the Gibbs chain using the method proposed by Geweke (1992). The posterior samples, especially for the β_k's, showed some autocorrelations among the iterations, and thus we used every 10th sample in the chain for obtaining posterior estimates. After a burn-in of 5000 samples and thinning by 10, we had 5000 MCMC posterior samples on which the analyses were based.

We fit a piecewise exponential proportional hazards (PEPH) model to the E1694 data for comparison. The likelihood function for the PEPH model (see Ibrahim et al., 2001a) is given by

$$\begin{aligned} &L(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, D) \\ &= \prod_{i=1}^{n} \{\lambda_0(y_i) \exp(\boldsymbol{\beta}' \mathbf{Z}_i)\}^{\nu_i} \exp\left[-\int_0^{y_i} \{\lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{Z}_i)\} \, dt \right] \\ &= \prod_{i=1}^{n} \prod_{j=1}^{J} \{\lambda_j \exp(\boldsymbol{\beta}' \mathbf{Z}_i)\}^{\delta_{ij} \nu_i} \\ &\quad \times \exp\left[-\delta_{ij} \left\{ \sum_{k=1}^{j-1} \lambda_k (s_k - s_{k-1}) + \lambda_j (y_i - s_{j-1}) \right\} \exp(\boldsymbol{\beta}' \mathbf{Z}_i) \right], \end{aligned}$$

Table 1
Posterior means, standard deviations, and 95% *HPD intervals for the E*1694 *data with* $J = 1$

Model	B/DIC	Covariate	Mean	SD	95% HPD interval
$\gamma = 0$	−1667.83/3335.68	Intercept	−.4397	.3357	(−1.0616, .2424)
		Treatment	−.2765	.1065	(−.4879, −.0717)
		Age	.1132	.0537	(.0080, .2172)
		Sex	.1514	.1118	(−.0657, .3654)
$\gamma = .25$	−1673.26/3346.02	Intercept	.1080	.0883	(−.0584, .2793)
		Treatment	−.0571	.0373	(−.1285, .0114)
		Age	.0082	.0104	(−.0104, .0292)
		Sex	.0177	.0291	(−.0346, .0822)
$\gamma = .5$	−1673.51/3346.56	Intercept	.0365	.0317	(−.0220, .0963)
		Treatment	−.0190	.0139	(−.0480, .0040)
		Age	.0026	.0036	(−.0035, .0103)
		Sex	.0060	.0103	(−.0124, .0279)
$\gamma = .75$	−1674.01/3347.52	Intercept	.0139	.0126	(−.0090, .0377)
		Treatment	−.0070	.0055	(−.0176, .0026)
		Age	.0009	.0012	(−.0013, .0036)
		Sex	.0019	.0040	(−.0057, .0105)
$\gamma = 1$	−1674.24/3347.98	Intercept	.0047	.0047	(−.0033, .0143)
		Treatment	−.0022	.0020	(−.0066, .0011)
		Age	.0003	.0004	(−.0005, .0012)
		Sex	.0006	.0015	(−.0023, .0036)
PEPH	−1694.84/3388.12	Treatment	−.3330	.1075	(−.5335, −.1170)
		Age	.0905	.0512	(.0098, .1917)
		Sex	.1320	.1100	(.0876, .3436)

Table 2
Posterior means, standard deviations, and 95% *HPD intervals for the E*1694 *data with* $J = 5$

Model	B/DIC	Covariate	Mean	SD	95% HPD interval
$\gamma = 0$	−1667.84/3335.66	Intercept	−.5039	.3420	(−1.1947, .1620)
		Treatment	−.2768	.1087	(−.4967, −.0774)
		Age	.1116	.0534	(.0035, .2139)
		Sex	.1432	.1126	(−.0819, .3588)
$\gamma = .25$	−1665.90/3331.07	Intercept	.1367	.1045	(−.0667, .3357)
		Treatment	−.0836	.0420	(−.1701, −.0073)
		Age	.0218	.0175	(−.0097, .0550)
		Sex	.0330	.0366	(−.0335, .1083)
$\gamma = .5$	−1665.36/3329.85	Intercept	.0444	.0400	(−.0326, .1213)
		Treatment	−.0307	.0158	(−.0620, −.0006)
		Age	.0104	.0073	(−.0024, .0246)
		Sex	.0139	.0144	(−.0136, .0421)
$\gamma = .75$	−1665.03/3329.10	Intercept	.0148	.0154	(−.0168, .0435)
		Treatment	−.0111	.0060	(−.0228, .0005)
		Age	.0045	.0028	(−.0005, .0099)
		Sex	.0052	.0054	(−.0047, .0166)
$\gamma = 1$	−1664.92/3329.01	Intercept	.0048	.0057	(−.0058, .0159)
		Treatment	−.0038	.0023	(−.0080, .0005)
		Age	.0017	.0009	(−.0001, .0034)
		Sex	.0017	.0019	(−.0017, .0059)
PEPH	−1670.97/3340.22	Treatment	−.4452	.1061	(−.6571, −.2401)
		Age	.0044	.0514	(−.0961, .1020)
		Sex	−.0419	.1066	(−.2437, .1676)

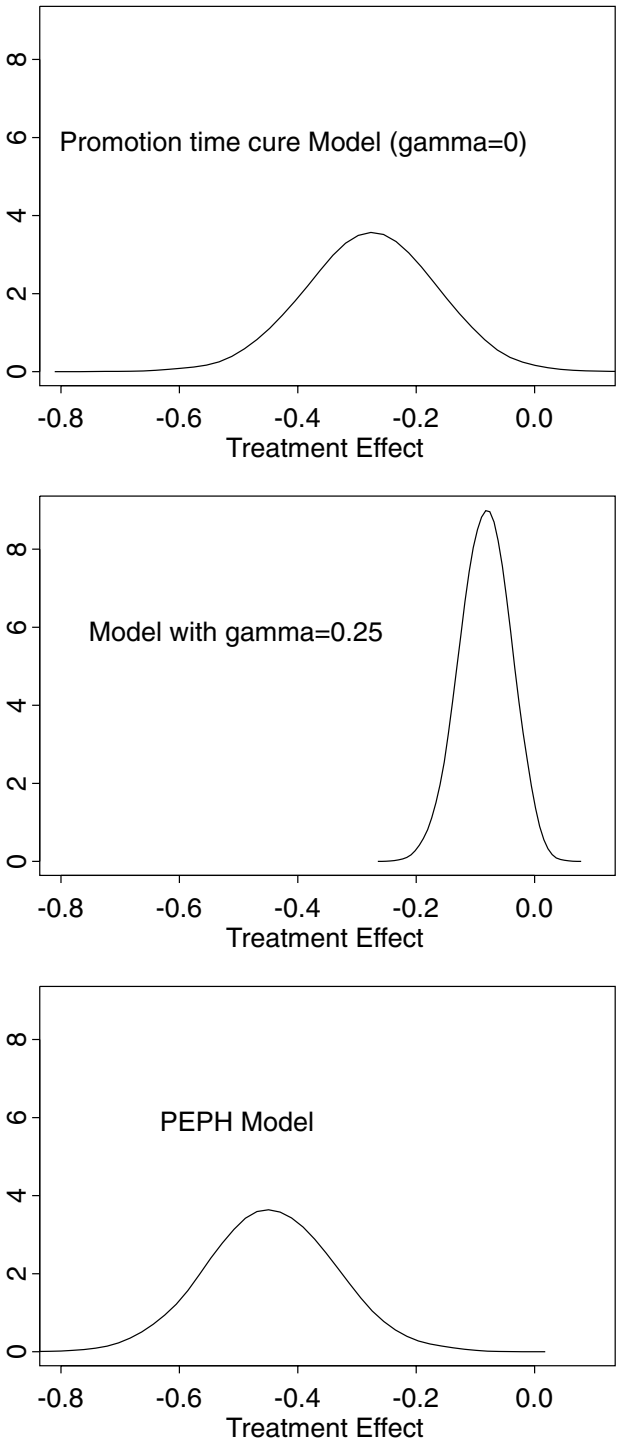

Figure 3. Posterior densities for the treatment effect under different models using $J = 5$.

Table 3
Sensitivity analysis with different regression parameters having truncated normal priors using $J = 5$ and $\gamma = 1$

Truncated covariate	Regression coefficient	Mean	SD	95% HPD interval
Treatment	Intercept	.0044	.0057	(−.0065, .0148)
	Treatment	−.0036	.0022	(−.0080, .0004)
	Age	.0017	.0009	(−.0002, .0034)
	Sex	.0017	.0020	(−.0021, .0056)
Age	Intercept	.0052	.0059	(−.0060, .0167)
	Treatment	−.0039	.0023	(−.0083, .0004)
	Age	.0016	.0009	(−.0002, .0033)
	Sex	.0016	.0020	(−.0019, .0058)
Sex	Intercept	.0044	.0059	(−.0065, .0159)
	Treatment	−.0037	.0023	(−.0083, .0005)
	Age	.0017	.0009	(−.0001, .0035)
	Sex	.0018	.0019	(−.0017, .0058)

where the unknown baseline hazard $\lambda_0(t)$ is modeled as a piecewise constant hazard function. Tables 1 and 2 summarize the B statistics, DICs, posterior means, posterior standard deviations, and 95% highest posterior density (HPD) intervals for each regression parameter in the model. There is not much difference between the competing models associated with different γ's in terms of the B and DIC statistics. The log-rank test for the treatment effect yields a p value of .0063 for $\rho = 0$ and .0038 for $\rho = 1$ in the G^ρ family of tests (Harrington and Fleming, 1982), which is consistent with the PEPH model. In fact, all the models with different values of γ demonstrate a similar trend in which the treatment effect favors IFN over GMK for relapse-free survival. A kernel-based estimate of the posterior density for the treatment effect is shown in Figure 3. We see that most of the mass is concentrated on the negative side of the horizontal axis (β_1) for each case.

6. Discussion

We have proposed a class of Bayesian survival models by applying the Box–Cox transformation to the population hazard function and the baseline density function. This class of regression models allows a zero as well as a nonzero cure fraction. It unifies a class of proper and improper population survival functions into a single family through the transformation parameter, γ. In the prior specification, for simplicity, we constrained the intercept (β_0) and allowed the other parameters to be free of constraints. We carried out a sensitivity analysis in which we constrained other parameters, but not β_0. In this case, we let $\boldsymbol{\beta}_{(-k)}$ denote $\boldsymbol{\beta}$ without the kth component β_k, then we take β_k to have a truncated normal prior,

$$\pi(\beta_k \mid \boldsymbol{\beta}_{(-k)}) = c(\boldsymbol{\beta}_{(-k)})^{-1} \exp\left(-\frac{\beta_k^2}{2\sigma_k^2}\right) I\left(\beta_k \geq -\min_{i=1,\ldots,n}\left\{\frac{\beta_0 + \cdots + \beta_{k-1}Z_{(k-1)i} + \beta_{k+1}Z_{(k+1)i} + \cdots + \beta_p Z_{pi}}{Z_{ki}}\right\}\right).$$

As shown in Table 3, the posterior estimates were very robust with respect to the choice of the constrained parameter β_k in the prior distribution.

We took the power parameter γ as fixed, since treating γ as random causes numerical difficulties and it generally results in a computationally intractable model. The reason for this intractability is that the parameter $\boldsymbol{\beta}$ is very sensitive to the values of γ. For different $\gamma \in [0, 1]$, the posterior samples of $\boldsymbol{\beta}$ can be quite different, and thus often cause the constraint (10) to be easily violated in the Gibbs sampling algorithm. For instance, the previous posterior sample quite often lies outside the parameter range of the current iteration under the constraints. Moreover, in most applications, there is typically not enough information in the data to precisely estimate γ and thus it is very sensitive to the choice of the prior for these proposed models. In this sense, γ is weakly identified in the model. Thus, in practice, it might not be feasible to allow γ to be random and assign a prior distribution to it because of the inherent parameter constraints and the complexity of the model.

Acknowledgements

We would like to thank an associate editor and an anonymous referee for their critical and insightful suggestions. This research was partially supported by National Cancer Institute grant 74015 and National Institute of General Medical Sciences grant GM 070335.

References

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica* **4,** 505–524.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* **47,** 501–515.

Betensky, R. A. and Schoenfeld, D. A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics* **57,** 282–286.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B* **26,** 211–252.

Chen, M. H., Ibrahim, J. G., and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association* **94,** 909–919.

Chen, M. H., Shao, Q., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Dey, D. K., Chen, M., and Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* **53,** 1239–1252.

Geisser, S. (1993). *Predictive Inference: An Introduction*. London: Chapman and Hall.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling based methods (with discussion). In *Bayesian Statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 147–167. Oxford: Oxford University Press.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*, J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith (eds), 169–193. Oxford: Oxford University Press.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics* **44,** 455–472.

Gray, R. J. and Tsiatis, A. A. (1989). A linear rank test for use when the main interest is in differences in cure rates. *Biometrics* **45,** 899–904.

Harrington, D. P. and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69,** 553–566.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001a). *Bayesian Survival Analysis*. New York: Springer.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001b). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics* **57,** 383–388.

Kirkwood, J. M., Ibrahim, J., Sosman, J. A., Sondak, V. K., Agarwala, S. S., Ernstoff, M. S., and Rao, U. (2001). High-dose interferon alfa-2b significantly prolonged relapse-free and overall survival compared with the GM2-KLH/QS-21 vaccine in patients with resected stage IIB-III melanoma: Results of intergroup trial E1694/S9512/C509801. *Journal of Clinical Oncology* **19,** 2370–2380.

Kuk, A. Y. C. and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79,** 531–541.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81,** 61–71.

Maller, R. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. New York: Wiley.

Peng, Y. and Dear, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56,** 237–243.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge University Press.

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* **64,** 583–616.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56,** 227–236.

Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics* **51,** 899–907.

Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics* **54,** 1508–1516.

Tsodikov, A. (2002). Semi-parametric models of long- and short-term survival: An application to the analysis of breast cancer survival in Utah by age and stage. *Statistics in Medicine* **21,** 895–920.

Tsodikov, A., Ibrahim, J. G., and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association* **98,** 1063–1078.

Yakovlev, A. Y., Asselain, B., Bardou, V. J., Fourquet, A., Hoang, T., Rochefediere, A., and Tsodikov, A. D. (1993). A simple stochastic model of tumor recurrence and its applications to data on premenopausal breast cancer. In *Biometrie et Analyse de Dormees Spatio-Temporelles*, Volume 12, B. Asselain, M. Boniface, C. Duby, C. Lopez, J. P. Masson, and J. Tranchefort (eds), 66–82. France: Société Francaise de Biométrie, ENSA Renned.

Yin, G. and Ibrahim, J. (2005). A class of Bayesian shared gamma frailty models with multivariate failure time data. *Biometrics* **61,** 209–217.

Received March 2004. *Revised August* 2004.
Accepted October 2004.