

Bayesian cure rate model accommodating multiplicative and additive covariates

GUOSHENG YIN* AND LUIS E. NIETO-BARAJAS

We propose a class of Bayesian cure rate models by incorporating a baseline density function as well as multiplicative and additive covariate structures. Our model naturally accommodates zero and non-zero cure rates, which provides an objective way to examine the existence of a survival fraction in the failure time data. An inherent parameter constraint needs to be incorporated into the model formulation due to the additive covariates. Within the Bayesian paradigm, we take a Markov gamma process prior to model the baseline hazard rate, and mixture prior distributions for the parameters in the additive component of the model. We implement a Markov chain Monte Carlo computational scheme to sample from the full conditional distributions of the posterior. We conduct simulation studies to assess the estimation and inference properties of the proposed model, and illustrate it with data from a bone marrow transplant study.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62N01; secondary 62N03.

KEYWORDS AND PHRASES: Additive hazards model, Cure rate model, Markov gamma process, Mixture prior, Proportional hazards model, Semiparametric method, Survival fraction.

1. INTRODUCTION

In many clinical trials, particularly in oncology research, a certain percentage of patients may be cured and thus become risk-free of a specific event of interest or disease relapse [36]. On the other hand, a study may involve a proportion of subjects who are insusceptible to the event of interest, while the rest are susceptible. In either case, the risk set at a given time point during the follow-up needs to exclude those cured or insusceptible subjects [19]. Many interesting examples of such data are from bone marrow transplant (BMT) research. BMT is a procedure to replace bone marrow that has been destroyed by treatment with high doses of anticancer drugs or radiation. Transplantation can be classified into three types: autologous (the patient's own marrow saved before treatment), allogeneic (marrow donated by someone else), or syngeneic (marrow donated by an identical twin). During the transplant, a common complication, known as the graft-versus-host disease (GVHD), is a reaction of donated stem cells against patients' tissue. In general, patients are less

likely to develop GVHD if the stem cells of the donor and patient are closely matched. After treatment, patients receive the harvested stem cells, which travel to the bone marrow and begin to produce new blood cells. To minimize potential side effects, doctors most often use transplanted stem cells that match the patient's own stem cells as closely as possible. The matching is usually based on the proteins on the surface of the cells, namely the human leukocyte-associated antigens (HLA). Szydlo et al. [38] studied leukemia patients with BMT from HLA-identical siblings, HLA-matched and mismatched unrelated donors. The primary end point was the time to cancer relapse or death while in remission. Although death would eventually occur if the follow-up is long enough, the definition for cure depends on the study duration and the nature of the disease. In fact, due to the GVHD, most of the BMT studies share the common feature: the GVHD typically causes the occurrence of death quickly as shown by the estimated Kaplan-Meier survival curves for the three donor groups in Figure 1. Patients may be considered "cured" if they could survive the risk of the GVHD after the transplant. After approximately five years follow-up, we observe a stable plateau for each arm, which indicates a possible cure fraction in this patient population.

Most of the survival models in the literature are based on a fundamental assumption that all the "alive" subjects are at risk, under which a cure fraction is typically not allowed. However, cure rate models are explicitly constructed to incorporate a survival fraction. In particular, for subject i with covariate vector \mathbf{Z}_i , let $S_{\text{pop}}(t|\mathbf{Z}_i)$ denote an improper population survival function, i.e., $\lim_{t \rightarrow \infty} S_{\text{pop}}(t|\mathbf{Z}_i) > 0$. Let $S(t)$ denote a proper survival function corresponding to those uncured or susceptible subjects, i.e., $\lim_{t \rightarrow \infty} S(t) = 0$. By mixing a certain fraction $1 - \theta(\mathbf{Z}_i)$ of the population as being cured and the rest $\theta(\mathbf{Z}_i)$ not cured, Berkson and Gage [3] proposed the two-component mixture cure rate model,

$$(1.1) \quad S_{\text{pop}}(t|\mathbf{Z}_i) = 1 - \theta(\mathbf{Z}_i) + \theta(\mathbf{Z}_i)S(t).$$

The covariates \mathbf{Z}_i characterize the long-term survival effects, which can be linked via $\theta(\cdot)$ in a logistic form, $\theta(\mathbf{Z}_i) = \exp(\beta'\mathbf{Z}_i) / \{1 + \exp(\beta'\mathbf{Z}_i)\}$. Extensive research has been carried out with the mixture cure rate model [18, 25, 31, 33, 37, 2].

Although the mixture cure rate model is quite intuitive, it lacks the proportional hazards structure [11] and other desirable properties as noted in Chapter 5 of [19]. In an

*Corresponding author.

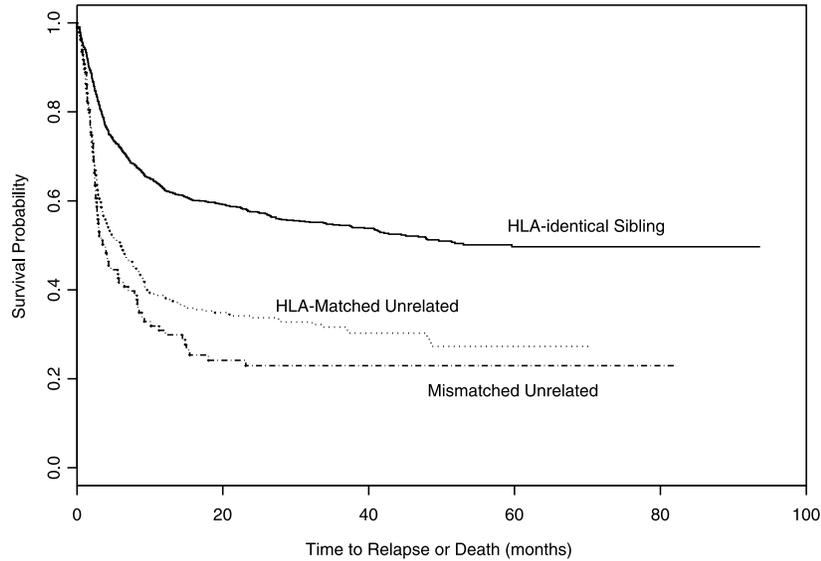


Figure 1. Estimated Kaplan-Meier survival curves stratified by the type of the donor in the BMT data.

alternative formulation, the cure rate model proposed by Yakovlev et al. [41] maintains the multiplicative structure of the proportional hazards model [39, 6, 20]. The population survival function is given by

$$(1.2) \quad S_{\text{pop}}(t|\mathbf{Z}_i) = \exp\{-\theta(\mathbf{Z}_i)F(t)\},$$

where $F(t)$ is a proper cumulative distribution function (c.d.f.), and the covariate structure takes the form of $\theta(\mathbf{Z}_i) = \exp(\beta'\mathbf{Z}_i)$. Ignoring the impropriety of $S_{\text{pop}}(t|\mathbf{Z}_i)$ in (1.2), we can compute the population hazard function as

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

where $f(t)$ is the density function corresponding to $F(t)$. Chen et al. [6] showed that under certain mild conditions, a uniform improper prior distribution for the regression parameters in (1.2) still leads to a proper posterior distribution, while an improper prior typically yields an improper posterior under model (1.1). Tsodikov et al. [40] provided a comprehensive review of the recent developments of model (1.2). More recently, cure rate models have been studied in a variety of contexts, such as the interval censored data [26], jointly modeling longitudinal and cure survival data [5] and cure rate modeling through latent activation schemes [9].

In this article, we consider an encompassing model

$$(1.3) \quad \lambda_{\text{pop}}(t|\mathbf{Z}_i, \mathbf{X}_i) = f(t) \exp(\beta'\mathbf{Z}_i) + \gamma'\mathbf{X}_i,$$

in which either β or γ contains an intercept, but not both, for identifiability. As the hazard function must be nonnegative for all t , and $f(t) \geq 0$, the parameter constraint, $\gamma'\mathbf{X}_i \geq 0$, needs to be satisfied throughout. This constraint is in accordance with other additive hazards models [27].

Model (1.3) has a non-zero cure rate only when $\gamma = 0$, and all the other values of γ correspond to a proper survival function, because $\Lambda_{\text{pop}}(\infty|\mathbf{Z}_i, \mathbf{X}_i) = \exp(\beta'\mathbf{Z}_i) + \int_0^\infty \gamma'\mathbf{X}_i dt$ is not bounded if $\gamma'\mathbf{X}_i > 0$.

It is critical to assess whether there exists a survival fraction in the data, since it is a fundamental assumption for the application of cure rate models [29, 30, 34, 22]. In practice, the validity of cure rate modeling is often examined empirically by the Kaplan-Meier survival curve. The ability to accommodate both cure and non-cure formulations in model (1.3) is particularly interesting, which offers an objective way to distinguish the model fitting between cure and non-cure models. Many flexible hazard-based regression models have been studied by incorporating multiplicative and additive covariate structures [4, 27, 28, 35]. However, in the context of cure rate models, very limited research has been conducted along this direction.

The rest of this article is organized as follows. In Section 2, we introduce notation and derive the likelihood function. Moreover, we specify mixture prior distributions for the parameters in the additive component of the model, and a Markov gamma process to model the baseline hazard rate. In Section 3, we obtain the full conditional distributions of the posterior to implement the Gibbs sampling. In Section 4, we conduct simulation studies and present a full Bayesian analysis of the BMT data. Finally, we give a brief discussion in Section 5.

2. LIKELIHOOD AND PRIOR

For $i = 1, \dots, n$, let T_i be the failure time, C_i be the censoring time, and $\nu_i = I(T_i \leq C_i)$ be the censoring indicator for the i th subject. The covariates are denoted by \mathbf{Z}_i and

\mathbf{X}_i , which may share common components. The observed data $D = \{Y_i = \min(T_i, C_i), \nu_i, \mathbf{Z}_i, \mathbf{X}_i\}$ are independent and identically distributed for $i = 1, \dots, n$. We assume that T_i is conditionally independent of C_i given covariates \mathbf{Z}_i and \mathbf{X}_i . For notational brevity, let $\text{Ga}(\alpha, \xi)$ denote a gamma distribution with mean α/ξ , $\text{Poi}(c)$ a Poisson distribution with mean c and $\text{Ber}(a)$ a Bernoulli probability function of taking a value of 1 with probability a .

Under model (1.3), the likelihood function is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, F|D) = \prod_{i=1}^n \{f(y_i) \exp(\boldsymbol{\beta}'\mathbf{Z}_i) + \boldsymbol{\gamma}'\mathbf{X}_i\}^{\nu_i} \\ \times \exp\{-F(y_i) \exp(\boldsymbol{\beta}'\mathbf{Z}_i) - \boldsymbol{\gamma}'\mathbf{X}_i y_i\},$$

where $f(t)$ and $F(t)$ are the baseline density and distribution functions, respectively. More explicitly, $F(t) = 1 - \exp\{-\int_0^t \lambda(s) ds\}$, and $\lambda(t)$ is the corresponding baseline hazard function. We discretize the time axis into J non-overlapping intervals such that $0 = s_0 < s_1 < \dots < s_J$, and $s_J > y_i$ for all $i = 1, \dots, n$, we then define the baseline hazard rate $\lambda(t) = \sum_{j=1}^J \lambda_j I(s_{j-1} < t \leq s_j)$. We define the partition in such a way that the number of failures in each interval is balanced.

To define the prior on $\lambda(t)$, we introduce latent Poisson variables $\{u_j\}$ to generate a Poisson-gamma alternating process to model $f(t)$ and $F(t)$. Let $\boldsymbol{\lambda}$ and \mathbf{u} denote the vectors of λ_j and u_j , and $\boldsymbol{\lambda}_{(-j)}$ and $\mathbf{u}_{(-j)}$ denote the rest of the vectors after deleting the j th component, respectively. We define $\delta_{ij} = 1$ if subject i fails or is censored in interval j , and 0 otherwise. The likelihood function based on the Markov gamma process is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}|D) = \prod_{i=1}^n \prod_{j=1}^J \{f_j(y_i) \exp(\boldsymbol{\beta}'\mathbf{Z}_i) + \boldsymbol{\gamma}'\mathbf{X}_i\}^{\nu_i \delta_{ij}} \\ \times \exp[-\delta_{ij} \{F_j(y_i) \exp(\boldsymbol{\beta}'\mathbf{Z}_i) + \boldsymbol{\gamma}'\mathbf{X}_i y_i\}],$$

where

$$f_j(y_i) = \lambda_j \exp\left\{-\sum_{k=1}^{j-1} \lambda_k (s_k - s_{k-1}) - \lambda_j (y_i - s_{j-1})\right\},$$

and

$$F_j(y_i) = 1 - \exp\left\{-\sum_{k=1}^{j-1} \lambda_k (s_k - s_{k-1}) - \lambda_j (y_i - s_{j-1})\right\}.$$

Instead of assigning independent gamma prior distributions to the λ_j 's, we employ a correlated Markov gamma process to model $\lambda(t)$:

$$\lambda_1 \sim \text{Ga}(\alpha_1, \xi_1) \\ u_j | \lambda_j \sim \text{Poi}(c_j \lambda_j) \\ \lambda_{j+1} | u_j \sim \text{Ga}(\alpha_{j+1} + u_j, \xi_{j+1} + c_j),$$

where the c_j 's are the smoothing parameters to induce correlations between the λ_j 's [32]. In the special case when all the c_j 's are zero, $u_j = 0$ with probability one, and the λ_j 's become independent. By introducing the latent Poisson process $\{u_j\}$, the components of $\{\lambda_j\}$ are naturally correlated a priori, and each λ_j has a gamma prior distribution given u_{j-1} . The advantages of introducing a latent Poisson process are that it formulates dependence between the λ_j 's, which is more realistic in practice and facilitates modeling the trend of the hazard; the Poisson latent variables simplify posterior simulations due to conjugacy; and when the α_j 's and ξ_j 's are all equal, the prior process is stationary. Other prior specifications on $\{\lambda_j\}$ can impose correlations as well, such as using a first-order autoregressive structure or a Markovian relation [12, 1, 21].

Although not required for the development, we take $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $(\boldsymbol{\lambda}, \mathbf{u})$ to be independent a priori, i.e., $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{u}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\lambda}, \mathbf{u})$. The joint prior distribution of $(\boldsymbol{\lambda}, \mathbf{u})$ can be written as

$$\pi(\boldsymbol{\lambda}, \mathbf{u}) = \text{Ga}(\lambda_1 | \alpha_1, \xi_1) \left\{ \prod_{j=1}^{J-1} \text{Poi}(u_j | c_j \lambda_j) \right. \\ \left. \times \text{Ga}(\lambda_{j+1} | \alpha_{j+1} + u_j, \xi_{j+1} + c_j) \right\}.$$

Without loss of generality, we assume that all the covariates are positive, which can be easily shifted and rescaled if this is not the case in practice. The prior distribution for each component of $\boldsymbol{\beta}$ can simply be a zero-mean normal distribution with a large variance. However, we need to be more cautious with the prior specification on $\boldsymbol{\gamma}$ due to the constraint $\boldsymbol{\gamma}'\mathbf{X}_i \geq 0$ ($i = 1, \dots, n$). Parameter constraints often make Bayesian computation more complicated [14, 15, 8]. In order to calibrate the posterior probability of $\boldsymbol{\gamma} = 0$, we consider a mixture of two zero-mean normal distributions (one with a small variance and the other with a large variance) for each component of $\boldsymbol{\gamma}$, as in the stochastic search variable selection (SSVS) developed by George and McCulloch [16].

More specifically, we take that

$$\pi(\boldsymbol{\gamma}) \propto \prod_{k=1}^q \{a_0 N(\gamma_k | 0, \sigma^2) + (1 - a_0) N(\gamma_k | 0, h^2 \sigma^2)\} \\ \times \prod_{i=1}^n I(\boldsymbol{\gamma}'\mathbf{X}_i \geq 0),$$

where q is the length of vector $\boldsymbol{\gamma}$, $0 < a_0 < 1$, $h > 1$, and σ is relatively small. This implies that if γ_k takes a sample from the posterior distribution corresponding to $N(\gamma_k | 0, \sigma^2)$ in the mixture, we can ‘‘safely’’ set $\gamma_k = 0$.

This construction implies that, a priori, $\gamma_k | \boldsymbol{\gamma}_{(-k)}$ has a mixture of two truncated normal densities with a small vari-

ance of σ^2 and a large variance of $h^2\sigma^2$,

$$\begin{aligned} \pi(\gamma_k|\gamma_{(-k)}) &= a_0 N^+(\gamma_k|\gamma_{(-k)}, 0, \sigma^2) \\ &\quad + (1 - a_0) N^+(\gamma_k|\gamma_{(-k)}, 0, h^2\sigma^2). \end{aligned}$$

The two truncated normal distributions, one with a small variance ($h = 1$) and the other with a large variance ($h > 1$), can be written as

$$\begin{aligned} N^+(\gamma_k|\gamma_{(-k)}, 0, h^2\sigma^2) &= \frac{1}{c(\gamma_{(-k)}, 0, h^2\sigma^2)} N(\gamma_k|0, h^2\sigma^2) \\ &\quad \times I(\gamma_k \geq -\eta_k), \end{aligned}$$

where

$$\eta_k = \min_{i=1, \dots, n} \left\{ \frac{\sum_{j=1, j \neq k}^q \gamma_j X_{ji}}{X_{ki}} \right\},$$

and the normalizing constant

$$c(\gamma_{(-k)}, 0, h^2\sigma^2) = 1 - \Phi\left(-\frac{\eta_k}{h\sigma}\right)$$

with $\Phi(\cdot)$ the c.d.f. of the standard normal distribution.

To facilitate the implementation of the Markov chain Monte Carlo (MCMC) method, we introduce a latent Bernoulli random variable w_k . We rewrite the conditional prior of γ_k given w_k and $\gamma_{(-k)}$ as

$$\pi(\gamma_k|w_k, \gamma_{(-k)}) = \begin{cases} N^+(\gamma_k|\gamma_{(-k)}, 0, \sigma^2), & \text{if } w_k = 1 \\ N^+(\gamma_k|\gamma_{(-k)}, 0, h^2\sigma^2), & \text{if } w_k = 0 \end{cases}$$

with $w_k \sim \text{Ber}(a_0)$ for $k = 1, \dots, q$. By mixing the two truncated normal densities, we are able to assess the existence of a cure through the posterior samples of $\mathbf{w} = (w_1, \dots, w_q)'$. If the proportion of $\mathbf{w} = 0$ is high in the posterior samples, it tends to support the fit of a cure rate model.

3. POSTERIOR COMPUTATION

We derive the full conditional distributions for the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{u}, \mathbf{w})$ to implement the Gibbs sampling. The full conditional distribution of $\boldsymbol{\beta}$ does not depend on the latent Poisson or Bernoulli variable, thus,

$$\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{u}, \mathbf{w}, D) \propto L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}|D)\pi(\boldsymbol{\beta}),$$

where $\pi(\boldsymbol{\beta})$ is a normal prior distribution. The full conditional distribution for $\boldsymbol{\lambda}$ simplifies to the following expression,

$$\begin{aligned} \pi(\lambda_j|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}_{(-j)}, \mathbf{u}, \mathbf{w}, D) &\propto L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}|D) \\ &\quad \times \text{Ga}(\lambda_j|\alpha_j + u_{j-1} + u_j, \xi_j + c_{j-1} + c_j), \end{aligned}$$

for $j = 1, \dots, J$, and $c_0 = c_J = 0$ and $u_0 = u_J = 0$ with probability one. Since \mathbf{u} does not depend on the observed data,

$$\pi(u_j|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{u}_{(-j)}, \mathbf{w}, D) \propto \frac{\{c_j(\xi_{j+1} + c_j)\lambda_j\lambda_{j+1}\}^{u_j}}{\Gamma(1 + u_j)\Gamma(\alpha_{j+1} + u_j)},$$

for $j = 1, \dots, J - 1$, where u_j takes an integer value from $\{0, 1, 2, \dots\}$, and $\Gamma(\cdot)$ is the usual gamma function. Finally, the full conditional distributions of $\boldsymbol{\gamma}$ and \mathbf{w} are respectively given by

$$\begin{aligned} \pi(\boldsymbol{\gamma}_k|\boldsymbol{\beta}, \boldsymbol{\gamma}_{(-k)}, \boldsymbol{\lambda}, \mathbf{u}, w_k, D) \\ \propto \begin{cases} L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}|D)N^+(\gamma_k|\gamma_{(-k)}, 0, \sigma^2), & \text{if } w_k = 1 \\ L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}|D)N^+(\gamma_k|\gamma_{(-k)}, 0, h^2\sigma^2), & \text{if } w_k = 0 \end{cases} \end{aligned}$$

and $\pi(w_k|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{u}, D) \propto \text{Ber}(a_1)$, where

$$a_1 = \frac{a_0}{a_0 + (1 - a_0) \frac{N^+(\gamma_k|\gamma_{(-k)}, 0, h^2\sigma^2)}{N^+(\gamma_k|\gamma_{(-k)}, 0, \sigma^2)}}.$$

Even though the mixture prior has a similar flavor as that in the SSVS [16], our focus is quite different. In fact, the fundamental model structure in our case changes during the Gibbs iterations. The Markov chain alternately visits between a non-cure model with a proper survival function (when $\gamma \neq 0$) and a cure model with an improper survival function (when $\gamma = 0$). The posterior probability of supporting a cure rate model can be estimated by averaging the indicator variables w_k across all the MCMC iterations, i.e., to obtain the proportion of $\gamma_k = 0$. Due to the non-log-concavity of the full conditional distributions, a Metropolis step is required within the Gibbs sampler [17].

4. NUMERICAL STUDIES

4.1 Simulation

We carried out two simulation studies to examine the performance of the proposed method. We generated the failure times from the multiplicative-additive cure model (1.3). In particular we considered,

$$(4.1) \quad \lambda_{\text{pop}}(t|Z, X) = f(t) \exp(\beta_0 + \beta_1 Z) + \gamma X.$$

The baseline density function was chosen as an exponential distribution with mean one, i.e., the baseline hazard function is one. Both covariates Z and X were simulated from a Bernoulli random variable, taking a value of 0 or 1 with probability 0.5. The censoring times were independently generated from a uniform distribution. The true parameter values were $\beta_0 = -0.2$ and $\beta_1 = 1$. For the first simulation study, we took $\gamma = 0$, i.e., the underlying true model did not have an additive covariate component, which led to a cure rate model. On the other hand, for the second simulation study, we took $\gamma = 1$, which led to a cure rate model when $X = 0$ and to a non-cured model when $X = 1$. Such a covariate-dependent cure rate model structure typically cannot be accommodated by the available methods in the literature. We took the sample size $n = 300$ and replicated 500 simulations.

For the model with $\gamma = 0$, the cure rate is given by $\exp\{-\exp(\beta_0 + \beta_1 Z)\}$. The true cure rate for the group with $Z = 0$ is 44% and that for $Z = 1$ is 11%. In Figure 2, we show

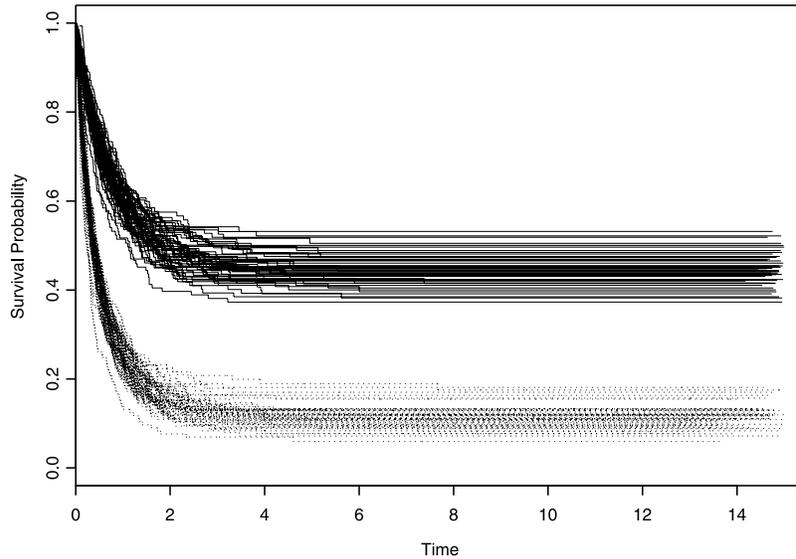


Figure 2. First 50 simulated data sets, when $\gamma = 0$ in model (4.1). Estimated Kaplan-Meier survival curves stratified by the indicator covariate (solid curves with $Z = 0$ and dotted curves with $Z = 1$).

the estimated Kaplan-Meier survival curves of each arm for the first 50 simulated data sets. We observe a common pattern with sharp drops at the beginning of the follow-up, and stable plateaus at the tails of the curves, which resembles the survival curves of the BMT data in Figure 1. For the model with $\gamma = 1$, our model specifies four groups by the combination of the two values of covariates Z and X . The Kaplan-Meier survival curves for these four groups of the first 50 simulated data sets are shown in Figure 3, from which we clearly see cure rate patterns (the first column) as well as non-cure ones (the second column) arising from the same model.

For each data realization, we fit the proposed model with different values of (σ, h) to examine the percentage of the posterior samples of γ being zero. We took vague prior distributions for all of the model parameters so that the likelihood would dominate the posterior estimation. Under each configuration, we recorded 5,000 MCMC iterations after the first 100 burn-in samples in the Gibbs sampling. The posterior means and standard deviations of the parameters were computed based on the 5,000 posterior samples.

In Table 1, we report the posterior mean and standard deviation averaged over 500 simulations for each regression parameter. We can see that the posterior means of the parameters are quite close to the true values, indicating that the MCMC chains converged properly. In simulation 1 with the true $\gamma = 0$, even under the misspecified model, the posterior estimates for β_0 and β_1 were close to the true values. Across different pairs of (σ, h) , there were high posterior probabilities that the additive part of model (4.1) did not exist, because a high percentage of the posterior samples of γ were zero. In simulation 2 with the true $\gamma = 1$, the percentage of the posterior samples of γ being zero was very low.

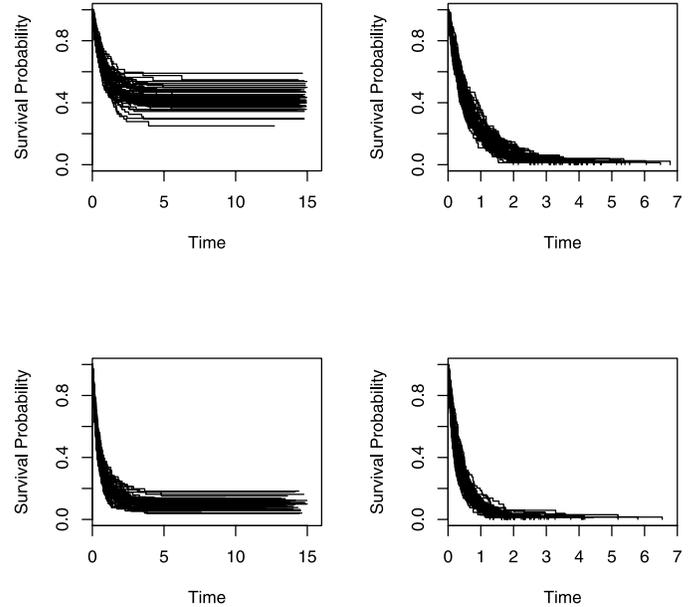


Figure 3. First 50 simulated data sets, when $\gamma = 1$ in model (4.1). Estimated Kaplan-Meier survival curves: top-left panel with $(Z, X) = (0, 0)$, top-right panel with $(Z, X) = (0, 1)$, bottom-left panel with $(Z, X) = (1, 0)$ and bottom-right panel with $(Z, X) = (1, 1)$.

In addition, when the variance of the sharp normal distribution was very small and that of the flat normal distribution was very large, $\widehat{\Pr}(\gamma = 0|D)$ was very close to zero. This demonstrates the usefulness of the proposed model for detecting the survival fraction in the data, which provides an

Table 1. Average of posterior means, empirical standard deviation of posterior means (ESD) and average of posterior standard deviations (ASD) of regression parameters over 500 simulated data sets

(σ, h)	$\beta_0 = -0.2$			$\beta_1 = 1$			γ				
	Mean	ESD	ASD	Mean	ESD	ASD	Mean	ESD	ASD	$\Pr(\gamma = 0 D)$	
	Model with $\gamma = 0$										
(.5, 10)	-0.210	0.130	0.124	0.994	0.151	0.149	0.022	0.015	0.019	0.909	
(.33, 10)	-0.209	0.123	0.125	0.994	0.151	0.150	0.022	0.014	0.019	0.908	
(.1, 10)	-0.212	0.120	0.123	0.996	0.151	0.149	0.020	0.011	0.018	0.905	
(.1, 33)	-0.212	0.118	0.122	0.996	0.151	0.150	0.020	0.010	0.018	0.969	
(.1, 100)	-0.207	0.135	0.126	0.996	0.151	0.149	0.020	0.010	0.017	0.989	
(.01, 100)	-0.209	0.119	0.123	1.012	0.152	0.147	0.007	0.005	0.007	0.972	
(.01, 1000)	-0.208	0.120	0.124	1.013	0.152	0.147	0.007	0.001	0.006	0.997	
	Model with $\gamma = 1$										
(.5, 10)	-0.183	0.158	0.168	1.007	0.179	0.184	0.983	0.161	0.162	0.570	
(.33, 10)	-0.186	0.159	0.170	1.009	0.179	0.184	1.004	0.176	0.170	0.188	
(.1, 10)	-0.188	0.161	0.170	1.011	0.180	0.185	1.008	0.153	0.159	0.00003	
(.1, 33)	-0.196	0.159	0.170	1.015	0.181	0.186	1.033	0.162	0.163	0.00009	
(.1, 100)	-0.196	0.161	0.170	1.015	0.180	0.185	1.036	0.163	0.163	0.0002	
(.01, 100)	-0.191	0.157	0.169	1.011	0.179	0.185	1.008	0.153	0.159	0	
(.01, 1000)	-0.197	0.159	0.171	1.015	0.181	0.185	1.035	0.163	0.163	0	

easy and objective way to examine the existence of a cure in the Bayesian paradigm.

4.2 BMT example

As an illustration, we applied the proposed method to analyze the BMT data. Transplants were performed between 1985 and 1991, and reported to the International Bone Marrow Transplant Registry [23]. In this analysis, we had 1,715 patients with acute lymphoblastic leukemia (ALL), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML). The time to cancer relapse or cancer death while in remission was the primary end point, and was recorded in months. The covariates of interest included three donor types: HLA-identical sibling, HLA-matched unrelated and mismatched unrelated; three disease types: ALL, AML and CML; three stages of disease: early, intermediate and advanced; and Karnofsky score (taking a value of 1 if the Karnofsky score ≥ 90 , 0 otherwise). We created two indicator variables for each of the three categorical variables by taking the HLA-identical sibling, ALL, and early stage of disease as the baseline, respectively. In our analysis, we were particularly interested in whether a cure rate model was appropriate for the BMT data, for which we studied the model

$$(4.2) \quad \lambda_{\text{pop}}(t|\mathbf{Z}) = f(t) \exp\left(\sum_{k=1}^7 \beta_k Z_k\right) + \gamma,$$

where (Z_1, \dots, Z_7) corresponded to the covariates in Table 2.

Table 2. Posterior means, standard deviations and 95% HPD intervals for the regression parameters with the BMT data with $c_j = 100$ and $(\sigma = .1, h = 100)$

J	Covariate	Mean	SD	95% HPD
5	HLA-match	.546	.081	(.385, .700)
	Mis-match	.640	.125	(.385, .862)
	AML	.110	.100	(-.078, .312)
	CML	.058	.088	(-.112, .230)
	Inter Stage	.469	.087	(.301, .639)
	Adv Stage	1.040	.102	(.845, 1.245)
	Karnofsky	-.520	.083	(-.676, -.352)
10	HLA-match	.569	.085	(.402, .735)
	Mis-match	.659	.126	(.412, .901)
	AML	.158	.104	(-.047, .360)
	CML	.129	.092	(-.052, .311)
	Inter Stage	.536	.095	(.350, .724)
	Adv Stage	1.148	.113	(.910, 1.354)
	Karnofsky	-.445	.087	(-.627, -.288)
20	HLA-match	.576	.085	(.409, .740)
	Mis-match	.668	.131	(.403, .919)
	AML	.157	.105	(-.054, .365)
	CML	.118	.093	(-.060, .300)
	Inter Stage	.550	.097	(.363, .734)
	Adv Stage	1.151	.109	(.947, 1.368)
	Karnofsky	-.466	.086	(-.630, -.297)

We partitioned the time axis from zero up to 94 months into a total of J time intervals, and we took $J = 5, 10$ and 20, for comparison. We balanced the number of failures

in each interval. Therefore, the intervals are shorter during the earlier follow-up and longer at the later times, since events occurred more frequently in the first 30 months. To determine the best fitting model with different J values, we computed the log-pseudo marginal likelihood (LPML) as the model selection criterion to choose the best J . In particular, the LPML is related to the conditional predictive ordinate (CPO) statistic [13, 14]. CPO is a Bayesian cross-validation statistic by computing the conditional predictive distribution after deleting a single observation, which measures the adequacy of a model.

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

$$\begin{aligned} \text{CPO}_i &= f(y_i|\mathbf{Z}_i, \mathbf{X}_i, D^{(-i)}) \\ &= \left\{ \int \int \int \frac{\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}|D)}{f(y_i|\mathbf{Z}_i, \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})} d\boldsymbol{\beta} d\boldsymbol{\gamma} d\boldsymbol{\lambda} \right\}^{-1}. \end{aligned}$$

A Monte Carlo approximation of CPO_i [8] is given by

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

where M is the number of Gibbs samples after burn-in, and $\boldsymbol{\beta}_{[m]}$, $\boldsymbol{\gamma}_{[m]}$ and $\boldsymbol{\lambda}_{[m]}$ are the samples of the m th Gibbs iteration. Note that $L_i(\boldsymbol{\beta}_{[m]}, \boldsymbol{\gamma}_{[m]}, \boldsymbol{\lambda}_{[m]}|y_i, \Delta_i, \mathbf{Z}_i, \mathbf{X}_i)$ is a density function if $\nu_i = 1$ and a survival function if $\nu_i = 0$. A summary statistic based on the CPO_i 's is $\text{LPML} = \sum_{i=1}^n \log(\text{CPO}_i)$. The larger the value of the LPML, the better fit of the model.

In the Bayesian analysis, we took vague priors for all the model parameters so that the likelihood dominated the posterior inference. In particular, we took $\alpha_j = \xi_j = 0.01$ for $j = 1, \dots, J$. We specified a normal prior with mean zero and variance 100 for each component of $\boldsymbol{\beta}$. As to the prior distribution of $\boldsymbol{\gamma}$, we took a mixture of two truncated normal densities with the corresponding variances σ^2 and $h^2\sigma^2$. We took $a_0 = 0.5$ so that the prior probability of $\boldsymbol{\gamma}$ from each normal distribution in the mixture was the same. We monitored the MCMC convergence using the methods recommended by Cowles and Carlin [10], and observed that the chain converged fast and mixed well. We recorded one posterior sample every five iterations to reduce the autocorrelation from 20,000 Gibbs samples after 2,000 burn-ins.

We first fixed $c_j = 100$ and $(\sigma = 0.1, h = 100)$, but varied $J = 5, 10$ and 20 . We obtained the corresponding LPML statistics of -3416.2 , -3411.5 and -3405.6 , and thus the model with $J = 20$ was deemed as the best fit. Table 2 summarizes the posterior mean, standard deviation and 95% highest posterior density (HPD) interval for each model parameter. Moreover, we carried out a sensitivity analysis by

Table 3. Posterior means, standard deviations and 95% HPD intervals for the regression parameters with the BMT data with $J = 20$ and $c_j = 100$

(σ, h)	Covariate	Mean	SD	95% HPD
(.33, 10)	HLA-match	.578	.085	(.416, .746)
	Mis-match	.668	.131	(.406, .920)
	AML	.162	.104	(-.044, .370)
	CML	.125	.091	(-.055, .302)
	Inter Stage	.552	.097	(.358, .735)
	Adv Stage	1.157	.108	(.939, 1.372)
(.01, 100)	Karnofsky	-.462	.086	(-.639, -.303)
	HLA-match	.576	.086	(.417, .757)
	Mis-match	.666	.130	(.410, .910)
	AML	.157	.104	(-.045, .360)
	CML	.118	.091	(-.064, .295)
	Inter Stage	.554	.095	(.368, .737)
(.001, 100)	Adv Stage	1.147	.110	(.930, 1.363)
	Karnofsky	-.467	.087	(-.632, -.292)
	HLA-match	.562	.084	(.407, .732)
	Mis-match	.648	.124	(.419, .901)
	AML	.152	.104	(-.053, .353)
	CML	.109	.092	(-.062, .396)
	Inter Stage	.522	.090	(.343, .698)
	Adv Stage	1.118	.103	(.922, 1.324)
	Karnofsky	-.464	.084	(-.626, -.294)

choosing different values of (σ, h) in the mixture prior of two normal densities, see Table 3. We can see that the posterior estimates are quite similar for different paired values of (σ, h) . Patients treated with the BMT from HLA-matched unrelated donors or mis-matched unrelated donors had significantly worse survival than those with HLA-identical sibling donors. The stage of disease significantly affected survival: more advanced disease stages clearly led to shorter survival. There was not much difference between survival across the three disease types. For a wide range of paired values of $(\sigma, h) = (.33, 10), (.01, 100)$, and $(.001, 100)$, the posterior probabilities of $\boldsymbol{\gamma} = 0$: $\widehat{\text{Pr}}(\boldsymbol{\gamma} = 0|D) = .910, .990$ and $.966$, respectively. Therefore, there are consistently high posterior probabilities, above 90%, indicating that $\boldsymbol{\gamma} = 0$. Our model strongly supports a cure rate model structure by eliminating the additive component from model (4.2).

Moreover, we fixed $J = 20$, and took $c_j = 0, 10, 50$ and 100 to smooth and correlate the baseline hazard rates, respectively. Figure 4 presents the posterior means of the baseline hazard rate $\lambda(t)$ for $c_j = 0, 10, 50$ and 100 , respectively. The hazard estimate with $c_j = 0$ imposes an independent gamma process prior, and a larger value of c_j induces a higher correlation in the gamma process. These four plots exhibit a very similar pattern. During the early follow-up, the baseline hazard appeared to be much higher, which resulted in more occurrences of events, that would correspond

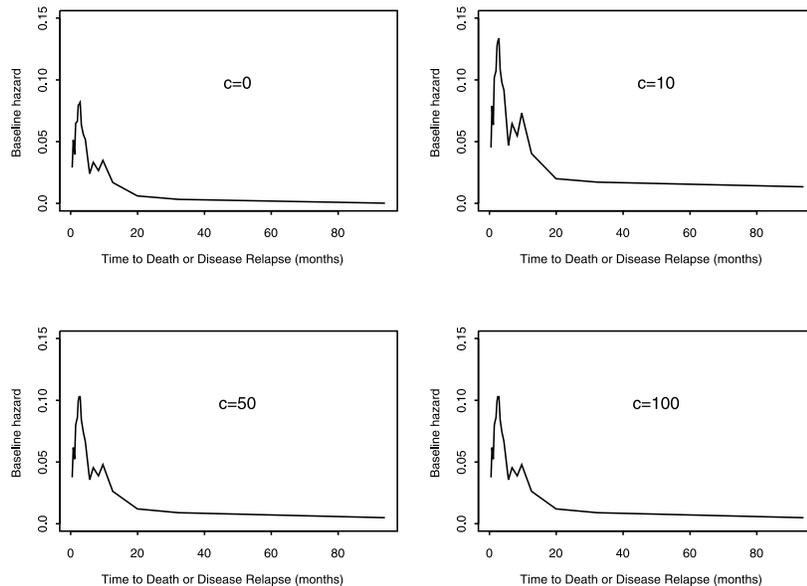


Figure 4. Posterior estimates of the baseline hazard function for the BMT data, with $(\sigma = 0.1, h = 100)$, $c_j = 0, 10, 50$ and 100 , respectively.

to the sharp drops of the survival curves in Figure 1. In the later follow-up period, the hazard decayed to a lower level and became more stable, matching with the plateaus at the tails of the Kaplan-Meier survival curves.

5. DISCUSSION

We have proposed a class of flexible Bayesian hazard regression models by incorporating the multiplicative and additive covariate structures. It naturally provides an objective assessment of the existence of a survival fraction in the data. Due to the nonnegativity of a hazard function, we impose a parameter constraint on the additive component of our model, that is, $\gamma'X_i \geq 0$. As pointed out by a referee, to avoid the constraint we could define the additive part to be of the form $|\gamma'X_i|$. However, we need to be cautious with the interpretation of the parameter γ because a positive or negative change in the covariate X_i would imply the same change in the hazard. Furthermore, we use a mixture prior distribution for the parameters associated with the additive component in the model. An alternative mixture prior involving a point mass at zero instead of a sharp density can also be considered. In our numerical experience, posterior inference is more stable with the mixture of two normal distributions.

A Markov gamma process prior is specified for the baseline hazard to facilitate the full Bayesian semiparametric estimation. Gibbs sampling can be easily implemented for the resulting model, from which it is straightforward to compute the posterior probability of a cure rate model. The proposed method links the proper and improper population survival functions through the multiplicative and additive

components of the covariate structure. This family of cure rate models is widely applicable due to its flexibility, particularly for modeling two treatment groups where a cure fraction is only present in one group but not in the other. Such a situation has been illustrated in the simulation study.

ACKNOWLEDGEMENTS

We would like to thank the referees, associate editor and editor for helpful comments that substantially improved the paper. We also thank Peter Müller and Ying Yuan at M. D. Anderson Cancer Center for very helpful discussions.

Received 3 March 2009

REFERENCES

- [1] ARJAS, E. and GASBARRA, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica* **4**, 505–524.
- [2] BETENSKY, R. A. and SCHOENFELD, D. A. (2001). Nonparametric estimation in a cure model with random cure times. *Biometrics* **57**, 282–286.
- [3] BERKSON, J. and GAGE, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* **47**, 501–515.
- [4] BRESLOW, N. E. and DAY, N. E. (1987). *Statistical Methods in Cancer Research, 2, The Design and Analysis of Case-Control Studies*. Lyon: IARC.
- [5] BROWN, E. R. and IBRAHIM, J. G. (2003). Bayesian approaches to joint cure-rate and longitudinal models with applications to cancer vaccine trials. *Biometrics* **59**, 686–693.
- [6] 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.

- [7] CHEN, M. H. and SHAO, Q. (1998). Monte Carlo methods for Bayesian analysis of constrained parameter problems. *Biometrika* **85**, 73–87.
- [8] CHEN, M. H., SHAO, Q. and IBRAHIM, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.
- [9] COONER, F., BANERJEE, S., CARLIN, B. P. and SINHA, D. (2007). Flexible cure rate modelling under latent activation schemes. *Journal of the American Statistical Association* **102**, 560–572.
- [10] COWLES, M. K. and CARLIN, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91**, 883–904.
- [11] COX, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.
- [12] GAMERMAN, D. (1991). Dynamic Bayesian models for survival data. *Applied Statistics* **40**, 63–79.
- [13] GEISSER, S. (1993). *Predictive Inference: an Introduction*. London: Chapman and Hall.
- [14] 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*, (Eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford: Oxford University Press, pp. 147–167.
- [15] GELFAND, A. E., SMITH, A. F. M. and LEE, T. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association* **87**, 523–532.
- [16] GEORGE, E. I. and MCCULLOCH, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.
- [17] GILKS, W. R., BEST, N. G. and TAN, K. K. C. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics* **44**, 455–472.
- [18] 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.
- [19] IBRAHIM, J. G., CHEN, M. H. and SINHA, D. (2001a). *Bayesian Survival Analysis*. New York: Springer.
- [20] IBRAHIM, J. G., CHEN, M. H. and SINHA, D. (2001b). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics* **57**, 383–388.
- [21] KIM, S., CHEN, M. H., DEY, D. K. and GAMERMAN, D. (2007). Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Analysis* **13**, 17–35.
- [22] KLEBANOV, L. and YAKOVLEV, A. (2007). A new approach to testing for sufficient follow-up in cure-rate analysis. *Journal of Statistical Planning and Inference* **137**, 3557–3569.
- [23] KLEIN, J. P. and ANDERSEN, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* **61**, 223–229.
- [24] KLEIN, A. Y. C. and CHEN, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531–541.
- [25] KUK, A. Y. C. and CHEN, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531–541.
- [26] LAM, K. F. and XUE, H. (2005). A semiparametric regression cure model with current status data. *Biometrika* **92**, 573–586.
- [27] LIN, D. Y. and YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.
- [28] LIN, D. Y. and YING, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics* **23**, 1712–1734.
- [29] MALLER, R. A. and ZHOU, S. (1994). Testing for sufficient follow-up and outliers in survival data. *Journal of the American Statistical Association* **89**, 1499–1506.
- [30] MALLER, R. A. and ZHOU, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics* **51**, 1197–1205.
- [31] MALLER, R. and ZHOU, X. (1996). *Survival Analysis with Long-Term Survivors*. New York: Wiley.
- [32] NIETO-BARAJAS, L. E. and WALKER, S. G. (2002). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics* **29**, 413–424.
- [33] PENG, Y. and DEAR, K. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56**, 237–243.
- [34] PENG, Y., DEAR, K. B. G. and CARRIERE, K. C. (2001). Testing for the presence of cured patients: a simulation study. *Statistics in Medicine* **20**, 1783–1796.
- [35] SCHEIKE, T. H. and ZHANG, M.-J. (2002). An additive-multiplicative Cox-Aalen regression model. *Scandinavian Journal of Statistics* **29**, 75–88.
- [36] SPOSTO, R. (2002). Cure model analysis in cancer: An application to data from the Children’s Cancer Group. *Statistics in Medicine* **21**, 293–312.
- [37] SY, J. P. and TAYLOR, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56**, 227–236.
- [38] SZYDLO, R., GOLDMAN, J. M., KLEIN, J. P., GALE, R. P., ASH, R. C., BACH, F. H., BRADLEY, B. A., CASPER, J. T., FLOMENBERG, N., GAJEWSKI, J. L., GLUCKMAN, E., HENSLEE-DOWNEY, P. J., HOWS, J. M., JACOBSEN, N., KOLB, H. J., LOWENBERG, B., MASAOKA, T., ROWLINGS, P. A., SONDEL, P. M., VAN BEKKUM, D. W., VAN ROOD, J. J., VOWELS, M. R., ZHANG, M. J. and HOROWITZ, M. M. (1997). Results of allogeneic bone marrow transplants for leukemia using donors other than HLA-identical siblings. *Journal of Clinical Oncology* **15**, 1767–1777.
- [39] TSODIKOV, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics* **54**, 1508–1516.
- [40] 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.
- [41] 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*, **12**, (eds. B. Asselain, M. Boniface, C. Duby, C. Lopez, J. P. Masson and J. Tranchefort). Société Française de Biométrie, ENSA Rennes, France, pp. 66–82.

Guosheng Yin

Department of Statistics and Actuarial Science

The University of Hong Kong

Pokfulam Road, Hong Kong

E-mail address: gyin@hku.hk

Luis E. Nieto-Barajas

Department of Statistics, ITAM

Rio Hondo 1, Progreso Tizapan, 01080 Mexico, D.F.

E-mail address: lnieto@itam.mx