Bayesian Inference For Survival Data With a Surviving Fraction

BY DEBAJYOTI SINHA * , MING-HUI CHEN †, AND JOSEPH G. IBRAHIM ‡

Medical University of South Carolina, University of Connecticut, and University of North Carolina

We develop Bayesian methods for right censored survival time data for populations with a surviving fraction, i.e., with a positive probability of cure. Our model, based on a two-stage process for the development of cancer, has a proportional hazards structure with the covariates depending naturally on the cure rate. We derive several properties of the model and physical interpretation of the model parameters. We also establish mathematical relationships of our model with other existing survival and cure rate models. The new model is computationally appealing, and novel computational Markov chain Monte Carlo (MCMC) methods are developed to sample from the posterior distribution of the parameters. We characterize the propriety of the joint posterior distribution of the parameters using a class of noninformative improper priors. Our model is flexible enough for implementing parametric as well as semiparametric Bayesian inference. We discuss novel nonparametric prior process and very practical prior elicitation for such purposes. A multivariate extension of this model is also possible. A real dataset from a melanoma clinical trial is presented to illustrate our methodology.

1. Introduction. Survival models incorporating a cure fraction, often referred to as *cure rate models*, have been used for modeling time-to-event data for various types of cancers, including breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma, and head and neck cancer, where for these diseases, a significant proportion of patients are "cured". Perhaps the most popular type of cure rate model is the mixture model discussed by Berkson and Gage [1]. In this model, the survivor function for the entire population, denoted by $S_1(t)$, is given by

$$S_1(t) = \pi + (1 - \pi)S^*(t), \tag{1}$$

where a fraction π of the population are considered "cured", and the remaining $1 - \pi$ are not cured. $S^*(t)$ denotes the survivor function for the non-cured group in the population. Popular choices for $S^*(t)$ are the exponential and Weibull distributions. We shall refer to the model in (1) as the *BG model.* The BG model has been extensively discussed in the statistical literature by several authors, including recently by Taylor [21], Ewell and Ibrahim [7], Stangl and Greenhouse [19], and Sy and Taylor [20]. We refer the reader interested in details about the frequentist methods for the BG model to these articles and to [15] and the references therein.

*Research was partially supported by NCI grant #R29-CA69222.

†Research was partially supported by NSF grant #DMS-9702172 and NIH grant #CA 74015-01.

‡Research was partially supported by NIH grants #CA 70101-01 and #CA 74015-01

Although the BG model is attractive and widely used, it has some drawbacks. Farewell [8, 9] and Cantor and Shuster [2] discuss the difficulties of the BG model within the frequentist paradigm. In the presence of covariates, it cannot have a proportional hazards structure if the covariates are modeled through π via a binomial regression model. Also, when including covariates through the parameter π via a standard binomial regression model, (1) yields improper posterior distributions for many types of noninformative improper priors, including the uniform prior for the regression coefficients. This is a crucial drawback of (1) since it implies that Bayesian inference with (1) essentially requires a proper prior. Also, there is no simple multivariate extension of the BG model. These drawbacks can be overcome with an alternative parameterization of a cure rate model which we discuss in the next section. Later in this paper we discuss this new model's properties, parametric Bayesian inference under the model, associated prior elicitation and multivariate extensions. We also find that this alternative model is also convenient for developing an appropriate semiparametric Bayesian inference procedure for cure-rate survival data. In the process of developing such a semiparametric Bayesian methodology, we explore deeply the very notion of prior processes and introduce a new class of prior processes. This class is quite different from classes such as gamma and Dirichlet, currently popular in semiparametric Bayes. We hope that these later sections demonstrate the obvious numerous advantages of our model over the BG model. This problem forces us to stretch our imagination of semiparametric Bayes and Bayesian survival analysis and survival model building. We consider this review article as an appropriate contribution to celebrate the 70th birthday of Prof.Hall., as he has always inspired us to stretch our imagination in statistics, explore new ideas and try unconventional problems in application and methodology and theory of statistics. We begin again with a congratulation to our dear mentor.

2. An Alternative Cure Rate Model. We present a formulation of the cure rate model discussed in [23, 22, 24, 3]. The alternative model can be derived as follows. Suppose that for an individual in the population, we let N denote the number of *metastasis-competent* tumor (MCT) cells for that individual left active after the initial treatment. A metastasis-competent tumor (MCT) cell has the potential of metastasizing, though it is not a full-blown cancer cell yet. Further, we assume that N has a Poisson distribution with mean θ. We let Z_i denote the random time for the MCT cell i to produce detectable metastatic disease. That is, Z_i can be viewed as a promotion time for the MCT cell i. Given N, the random variables Z_i, $i = 1, 2, \ldots$, are assumed to be independent and identically distributed with a common distribution function $F(t) = 1 - S(t)$ that does not depend on N. The time to relapse of cancer can be defined by the random variable $Y = \min\{Z_i, 0 \leq i \leq N\}$, where $P(Z_0 = \infty) = 1$. The survival function for Y, and hence the survival function for the population, with some straightforward algebra, can be shown as

$$S_{pop}(t) = P(\text{no metastatic cancer by time t}) = \exp(-\theta F(t)) . \tag{2}$$

Since $S_{pop}(\infty) = \exp(-\theta) > 0$, (2) is not a proper survival function. As (2) shows explicitly the contribution to the relapse time of two distinct characteristics of tumor growth: the initial number of MCT cells and the rate of their progression. Thus the

model incorporates parameters bearing clear biological meaning. We emphasize here that aside from the biological motivation, the model in (2) is suitable for any type of survival data which has a surviving fraction. Thus, survival data which do not "fit" the biological definition given above can still certainly be modeled by (2) as long as the data has a surviving fraction and can be thought of as being generated by a random unknown number (N) of latent competing risks with i.i.d. promotion times, the Z_is. Thus the model can be useful for modeling various types of survival data, including time to relapse, time to first infection, and so forth.

We also see from (2) that the cure fraction (i.e., cure rate) is given by

$$S_{pop}(\infty) \equiv P(N = 0) = \exp(-\theta). \tag{3}$$

As $\theta \to \infty$, the cure fraction tends to 0, whereas as $\theta \to 0$, the cure fraction tends to 1. The sub-density corresponding to (2) is given by $f_{pop}(t) = \theta f(t) \exp(-\theta F(t))$, where $f(t) = \frac{d}{dy} F(t)$. We emphasize here that $f_{pop}(t)$ is not a proper probability density since $S_{pop}(t)$ is not a proper survival function. However, $f(t)$ is a proper probability density function. The hazard function is given by

$$h_{pop}(t) = \theta f(t) \ . \tag{4}$$

Note that $\int_0^\infty h_{pop}(y)\, dy = \theta < \infty$.

The cure rate model (2) yields an attractive form for the hazard in (4). Specifically, we see that $h_{pop}(t)$ has the proportional hazards structure when the covariates x are modeled through $\theta = \theta(x)$ but $F(t)$ is modeled free of x. This form of the hazard is more appealing than the one from the BG model in (1), which does not have the proportional hazards structure if $\pi = \pi(x)$ is modeled as a function of covariates. The proportional hazards property in (4) is also computationally attractive, as MCMC methods are relatively easy to implement.

For the model in (2), the survival function of the "non-cured" population is given by

$$S^*(t) = P(Y > t | N \geq 1) = \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)} \ . \tag{5}$$

We note that $S^*(0) = 1$ and $S^*(\infty) = 0$ so that $S^*(t)$ is a proper survival function. For the non-cured population, the survival density (a proper density function) is given by $f^*(t) = [\exp(-\theta F(t))/\{1 - \exp(-\theta)\}]\theta f(t)$, and the hazard function is given by

$$\begin{aligned} h^*(t) = \frac{f^*(t)}{S^*(t)} &= \left(\frac{\exp(-\theta F(t))}{\exp(-\theta F(t)) - \exp(-\theta)} \right) h_p(t) \\ &= \left(\frac{1}{P(Y < \infty | Y > t)} \right) h_p(t). \end{aligned} \tag{6}$$

Thus (6) is magnified by the factor $\frac{1}{P(Y<\infty|Y>t)} > 1$ compared to the hazard function $h_{pop}(t)$ of the entire population. Now, it can be shown from (6) that $h^*(t|x)$ does not have a proportional hazards structure for any $f(t)$ with support on $(0, \infty)$. Furthermore, $h^*(t) \to \frac{f(t)}{S(t)}$ as $t \to \infty$, and thus $h^*(t)$ converges to the hazard function of the promotion time random variable Z as $y \to \infty$. Finally, $h^*(t)$ is an increasing function of θ.

There is a relationship between the model in (2) and the univariate survival model with multiplicative frailty. Suppose we have a model with conditional hazard $h_p(t|w) = wh(t)$ where $h(t) = f(t)/S(t)$ is the hazard function of the promotion time Z and w is distributed as Poisson with mean θ. Then, we get the model in (2) when we take the unconditional hazard $h_p(t)$. Similarly, this model has a relationship to the random-m-site model of Oakes [16]. The survival function of the time to first event in random-m-site model is the same as (2) when the promotion times of the sites are i.i.d. with common cdf $F(t)$.

There is a mathematical relationship between the model in (1) and (2), as $S_{pop}(t)$ is a standard cure rate model with cure rate equal to $\pi = \exp(-\theta)$ and survival function for the non-cured population given by $S^*(t)$ in (5). This result also implies that every BG model corresponds to some model of the form (2) for some θ and $F(.)$. In model (2), we model the entire population as a proportional hazards model, whereas in the BG model, only the non-cured group is typically modeled with a proportional hazards structure.

In model (2), we let the covariates depend on θ through the relationship $\theta = \exp(x'\beta)$, where x is a $p \times 1$ vector of covariates and β is a $p \times 1$ vector of regression coefficients. Entering the covariates in this fashion corresponds to a canonical link in a Poisson regression model. Using $\theta = \exp(x'\beta)$, (3), and (6), we can interpret the role of the regression coefficients for the cured and non-cured group.

Following [3], we can now construct the likelihood function as follows. Suppose we have n subjects, and let N_i denote the number of MCT cells for the subject i. Further, we assume that the N_i's are i.i.d. Poisson random variables with mean $\theta_i = \theta(\mathbf{x}_i)$ for $i = 1, \ldots, n$. We emphasize here that the N_i's are not observed, and can be viewed as latent variables in the model formulation. Further, suppose $Z_{i1}, \ldots, Z_{i,N_i}$ are the i.i.d. unobserved promotion times for the N_i MCT cells for the subject i, with common cumulative distribution function $F(.)$, $i = 1, \cdots, n$. For now, we specify a parametric $F(.) = F(.|\psi)$, such as a Weibull or gamma distribution, with the indexing parameter (possibly vector valued) ψ. For example, if $F(.|\psi)$ corresponds to a Weibull distribution, then $\psi = (\alpha, \lambda)$, where $f(t|\psi) \propto t^{\alpha-1}\exp(-\lambda t)$. Let y_i denote the observed survival time for subject i, which may be right censored, and let ν_i denote the censoring indicator, which equals 1 if y_i is a relapse time and 0 if it is right censored. The observed data is $D_{obs} = (n, \mathbf{y}, \nu)$, where $\mathbf{y} = (y_1, \ldots, y_n)'$, and $\nu = (\nu_1, \ldots, \nu_n)'$. Also, let $\mathbf{N} = (N_1, \ldots N_n)'$. The complete data is given by $D = (n, y, \nu, \mathbf{N})$, where $\mathbf{N}$ is an unobserved vector of latent variables. We now assume a Weibull density for $f(y_i|\psi)$, so that $f(y|\psi) = \alpha y^{\alpha-1} \exp\{\lambda - y^\alpha \exp(\lambda)\}$, where $\psi = (\alpha, \lambda)$.

Let $\mathbf{x}_i' = (x_{i1}, \ldots, x_{ip})$ denote the $p \times 1$ vector of covariates for the subject i, and let $\beta = (\beta_1, \ldots, \beta_p)'$ denote the corresponding vector of regression coefficients. We relate θ to the covariates by $\theta_i \equiv \theta(\mathbf{x}_i') = \exp(\mathbf{x}_i'\beta)$, so that the cure rate for subject i is $\exp(-\theta_i) = \exp(-\exp(\mathbf{x}_i'\beta))$. This relationship between θ_i and β is equivalent to a canonical link for θ_i in the setting of generalized linear models. With this relation, we can write the complete data likelihood of (β, ψ) as

$$L(\beta, \psi|D) = \left(\prod_{i=1}^{n} S(y_i \mid \psi)^{N_i - \nu_i} \left(N_i f(y_i \mid \psi)\right)^{\nu_i} \right) \times$$

$$\exp\left\{\sum_{i=1}^{n}\left[N_i\mathbf{x}_i'\beta - \log(N_i!) - \exp(\mathbf{x}_i'\beta)\right]\right\}, \tag{7}$$

where $D = (n, \mathbf{y}, X, \nu, \mathbf{N})$, X is the $n \times p$ matrix of covariates, $f(y_i|\psi)$ is Weibull density given above, and $S(y_i|\psi) = \exp(-y_i^{\alpha}\exp(\lambda))$. If we assume independent priors for (β, ψ), then the posterior distributions of (β, ψ) are also independent. Note that the part of the complete data likelihood in (7) involving β looks exactly like a Poisson generalized linear model with a canonical link, with the N_i's being the observables. The likelihood given observed data $D_{obs} = (n, \mathbf{y}, X, \nu)$ is obtained by summing (7) over possible values of $\mathbf{N}$. The posterior distribution given the prior $\pi(\beta, \psi)$ is given by

$$p(\beta, \psi|D_{obs}) \propto \left(\sum_{\mathbf{N}} L(\beta, \psi|D)\right)\ \pi(\beta, \psi). \tag{8}$$

In the next section, we investigate various methods of eliciting priors and the behavior of the posterior in (8) under such priors.

3. Prior Distributions. We discuss classes of noninformative priors as well as the power priors for (β, ψ), and examine some of their properties. Consider the joint noninformative prior $\pi(\beta, \psi) \propto \pi(\psi)$ where $\psi = (\alpha, \lambda)$ are the Weibull parameters in $f(y|\psi)$. This noninformative prior implies that β and ψ are independent a priori and $\pi(\beta) \propto 1$ is a uniform improper prior. We will assume throughout this subsection that $\pi(\psi) = \pi(\alpha|\delta_0, \tau_0)\pi(\lambda)$, where $\pi(\alpha|\delta_0, \tau_0) \propto \alpha^{\delta_0 - 1}\exp(-\tau_0\alpha)$, and δ_0 and τ_0 are two specified hyperparameters. With these specifications, the posterior distribution of (β, ψ) based on the observed data $D_{obs} = (n, \mathbf{y}, X, \nu)$ is given by

$$p(\beta, \psi|D_{obs}) \propto \left(\sum_{\mathbf{N}} L(\beta, \psi|D)\right)\ \pi(\alpha|\delta_0, \tau_0)\pi(\lambda)\ , \tag{9}$$

where the sum in (9) extends over all possible values of the vector $\mathbf{N}$. We are led to the following theorem concerning the propriety of the posterior distribution in (9) using the noninformative prior $\pi(\beta, \psi) \propto \pi(\psi)$.

THEOREM 1. *Let $d = \sum_{i=1}^{n}\nu_i$ and X^* be an $n \times p$ matrix with rows $\nu_i\mathbf{x}_i'$. Then if (i) X^* is of full rank, (ii) $\pi(\lambda)$ is proper, and (iii) $\tau_0 > 0$ and $\delta_0 > -d$, the posterior given in (9) is proper.*

The proof of Theorem 3.1 is given in [3]. Note that the conditions given in the theorem are sufficient but *not* necessary for the propriety of the posterior distribution, and they are quite general and typically satisfied for most data sets. A proper prior for α is not required in order to obtain a proper posterior. Based on condition (ii), $\pi(\lambda)$ is required to be proper. Although several choices can be made, we will use a normal density for $\pi(\lambda)$. Theorem 3.1 guarantees propriety of the posterior distribution of β using a improper uniform prior on β. This enables us to carry out Bayesian inference with improper priors for the regression coefficients

and facilitates comparisons with maximum likelihood. However, under the improper priors $\pi(\beta,\psi) \propto \pi(\psi)$, the BG model in (1) always leads to an improper posterior distribution for β. This detailed result is stated in a theorem from [3].

We get similar results for the historical-data based power priors for (β,ψ). Let n_0 denote sample size for the historical data, $\mathbf{y}_0$ be an $n_0 \times 1$ of right censored failure times for the historical data with censoring indicators ν_0, $\mathbf{N}_0$ is the unobserved vector of latent counts of metastasis-competent cells, and X_0 is an $n_0 \times p$ matrix of covariates corresponding to $\mathbf{y}_0$. Let $D_0 = (n_0, \mathbf{y}_0, X_0, \nu_0, N_0)$ denote the complete historical data. Further, let $\pi_0(\beta,\psi)$ denote the initial prior distribution for (β,ψ). The power prior is given as

$$\pi(\beta,\psi|D_{0,obs},a_0) \propto \left[\sum_{\mathbf{N}_0} L(\beta,\psi|D_0)\right]^{a_0} \pi_0(\beta,\psi), \tag{10}$$

where $L(\beta,\psi|D_0)$ is the complete data likelihood given in (7) with D being replaced by the historical data D_0, and $D_{0,obs} = (n_0, \mathbf{y}_0, X_0, \nu_0)$. We take a noninformative prior for $\pi_0(\beta,\psi)$, such as $\pi_0(\beta,\psi) \propto \pi_0(\psi)$, which implies $\pi_0(\beta) \propto 1$. For $\psi = (\alpha,\lambda)$, we take a gamma prior for α with small shape and scale parameters, and an independent informative normal prior for λ with mean 0 and the variance c_0. A beta prior is chosen for a_0 leading to the joint prior distribution

$$\pi(\beta,\psi,a_0 \mid D_{0,obs}) \propto \left[\sum_{\mathbf{N}_0} L(\beta,\psi|D_0)\right]^{a_0} \pi_0(\beta,\psi)\ a_0^{\gamma_0-1}(1-a_0)^{\lambda_0-1}, \tag{11}$$

where (γ_0,λ_0) are specified prior parameters. The prior in (11) does not correspond to any standard multivariate density, but, it has several attractive properties. First, we note that if $\pi_0(\beta,\psi)$ is proper, then (11) is guaranteed to be proper. Further, (11) can be proper even if $\pi_0(\beta,\psi)$ is improper. The following theorem characterizes the propriety of (11) when $\pi_0(\beta,\psi)$ is improper.

THEOREM 2. *Assume that*

$$\pi_0(\beta,\psi) \propto \pi_0(\psi) \equiv \pi_0(\alpha|\delta_0,\tau_0)\pi_0(\lambda) \propto \alpha^{\delta_0-1}\exp(-\tau_0\alpha)\pi_0(\lambda),$$

where δ_0 and τ_0 are specified hyperparameters. Let $d_0 = \sum_{i=1}^{n_0} \nu_{0i}$ and X_0^* be an $n_0 \times p$ matrix with rows $\nu_{0i}\mathbf{x}_{0i}'$. If (i) X_0^* is of full rank, (ii) $\delta_0 > 0$ and $\tau_0 > 0$, (iii) $\pi_0(\lambda)$ is proper, and (iv) $\gamma_0 > p$ and $\lambda_0 > 0$, then the joint prior given in (11) is proper.

We mention that the power prior for β based on the BG model in (1) will lead to an improper prior as well as an improper posterior distribution. Thus, the power priors based on (1) will not work. This result can be summarized in the following theorem.

THEOREM 3. *For the BG model given in (1), suppose we relate the cure fraction π to the covariates via a standard binomial regression given by $\pi(x) = \exp(\beta x)/(1+\exp(\beta x))$. Assume that the survival function for the non-cured group $S^*(.)$ depends on the parameter ψ^*. Let $L_1(\beta,\psi^*|D_{0,obs})$ and $L_1(\beta,\psi^*|D_{obs})$ denote the likelihood*

Table 1: Summary of E1684 Data

Survival Time (y) (years)		Status (frequency)		Age (x_1) (years)	
Median	1.38	censored	110	Mean	47.03
IQR	1.90	dead	174	Std Dev	13.00

Gender (x_2) (frequency)		PS (x_3) (frequency)	
Male	171	Fully Active	253
Female	113	Other	31

functions based on the observed historical and current data. Suppose we use an improper uniform initial prior for β (i.e., $\pi_0(\beta) \propto 1$) to construct the joint prior as

$$\pi_1(\beta, \psi^*, a_0 \mid D_{0,obs}) \propto [L_1(\beta, \psi^* | D_{0,obs})]^{a_0} \ \pi_0(\gamma^*) \ a_0^{\gamma_0 - 1}(1 - a_0)^{\lambda_0 - 1}, \tag{12}$$

where γ_0 and λ_0 are specified hyperparameters. Then, $\pi_1(\beta, \psi^*, a_0 \mid D_{0,obs})$ is always improper regardless of the propriety of $\pi_0(\psi^*)$. In addition, if we use $\pi_1(\beta, \psi^*, a_0 \mid D_{0,obs})$ as a prior, the resulting posterior, given by

$$p_1(\beta, \psi^*, a_0 \mid D_{obs}) \propto L_1(\beta, \psi^* | D_{obs}) \pi_1(\beta, \psi^*, a_0 \mid D_{0,obs}) \tag{13}$$

is also improper.

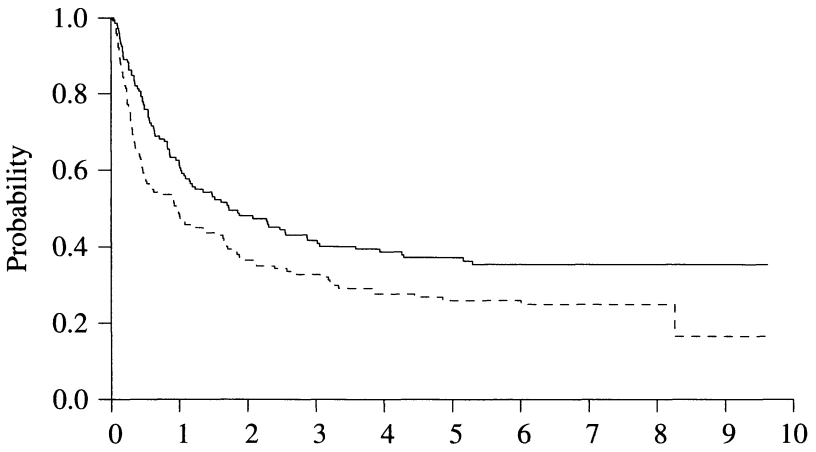
4. Example of Melanoma Data. We consider the E1684 data discussed in [3]. This trial, organized by Eastern Cooperative Oncology Group (ECOG), was a two-arm Phase-III clinical trial to compare high-dose interferon (IFN) group with the control group. The response variable is the relapse-free survival, which is defined as the time from randomization until death or relapse. See [13] for a more detailed description of this trial. We focus here only on the comparison of the maximum likelihood estimates and various other Bayesian estimates (under different priors) of the parameters for the cure rate model in (2). Three covariates and an intercept are included in the analyses. The covariates are age (x_1), gender (x_2) (male, female), and performance status (x_3) (fully active, other). Performance status is abbreviated by PS in the tables below. After deleting the observations with missing covariate values, a total of $n = 284$ observations are used in the analysis. In all of the analyses, we standardized the age covariate to stabilize the posterior computations. Table 1 gives the summary statistics of the the E1684 data. Figure 1 shows three superimposed plots of the survival curve based on the Kaplan-Meier method (dashed line), the BG model (1) (dotted line), and the alternative model (2) (solid line). We see that the three plots are nearly identical, giving essentially the same results. We now consider several analyses with the covariates included. The maximum likelihood estimates, their standard errors and p-values for the alternative model in (2) are computed and the results are reported in Table 2.

Several years earlier, a similar melanoma study with the same patient population was conducted by ECOG. This study, denoted by E1673, serves as the historical data for our Bayesian analysis of E1684. A total of $n_0 = 650$ patients are used in

Table 2: Maximum Likelihood Estimates (MLE) of the Model Parameters

Variable	MLE	Std Dev	P-value
intercept	0.09	0.11	0.38
age	0.09	0.07	0.21
gender	-0.12	0.16	0.44
ps	-0.20	0.26	0.44
α	1.32	0.09	0.00
λ	-1.34	0.12	0.00

Figure 1: Kaplan-Meier plot for E1684 data



Group	Time Interval 0-2	2-4	4-6	6-8	8-10
------ OBS	89/140	12/50	2/38	1/24	1/5
——— IFN	75/146	13/68	4/53	0/35	0/14

(# events/# at risk)

the historical data. Table 3 summarize the historical data E1673. Using the E1673 study as historical data, we consider an analysis with the power prior in (11). For initial prior for β, we take an improper uniform prior, and for $\pi_0(\alpha|\nu_0, \tau_0)$, we take $\nu_0 = 1$ and $\tau_0 = 0.01$ to ensure a proper prior. We note that this choice for $\pi_0(\alpha|\nu_0, \tau_0)$ also guarantees log-concavity of the conditional posterior of β. The parameter λ is taken to have a normal distribution with mean $\mu_0 = 0$ and variance $\sigma_0^2 = 10,000$. Table 4 gives posterior estimates of β based on several values of (γ_0, λ_0) using the model (2). In Table 4 we obtain, for example, $E(a_0|D_{obs}) = 0.03$, 0.06, and 0.14 by taking $(\gamma_0, \lambda_0) = (50, 50)$, $(100, 100)$, $(200, 1)$ respectively. The case $a_0 = 0$ with probability 1 gives the Bayesian analysis of E1673 data under

Table 3: Summary of E1673 Data

Survival Time (y_0) (years)		Status (frequency)		Age (x_{01}) (years)	
Median	2.33	censored	257	Mean	48.02
IQR	4.24	death	393	Std Dev	13.99

Gender (x_{02}) (frequency)		PS (x_{03}) (frequency)	
Male	375	Fully Active	561
Female	275	Other	89

the improper prior $\pi(\beta, \psi) \propto \pi_0(\psi) \propto \pi_0(\alpha|\nu_0, \tau_0) \times \pi_0(\lambda|\mu_0, \sigma_0)$. Given the above mentioned values of ν_0, τ_0, μ_0 and σ_0, we have chosen essentially very 'flat' or locally non-informative priors for the parameter ψ. So, the posterior estimates for $a_0 = 0$ given in the top part of Table 4 essentially yield the MLE's of β, α, and λ given in Table 2. This is a desirable feature of this model since it implies that we can obtain MLE's via Gibbs sampling, without doing any analytic maximizations. That is, if we take $a_0 = 0$ and choose vague proper priors for $\pi_0(.)$, the posterior means of the parameters are very close to the maximum likelihood estimates. But, for some of the parameters the HPD regions are not symmetric around the corresponding posterior means correctly reflecting the skewness of the posterior (or in turn the likelihood) surface. For those parameters, the posterior means are not going to be same as the posterior modes (or the MLEs). Even for these parameters the HPD intervals provide the narrow and honest interval estimates. The rest of Table 4 indicates a fairly robust pattern of behavior. The estimates of the posterior mean, standard deviation, or highest posterior density (HPD) intervals of β do not change a great deal if a low or moderate weight is given to the historical data. However, if a higher than moderate weight is given to the historical data, these posterior summaries can change a lot. For example, when the posterior mean of a_0 is less than .06, we see that all of the HPD intervals for β include 0, and when the posterior mean of a_0 is greater than or equal to .06, some HPD intervals for β do not include 0. The HPD interval for age does not include 0 when the posterior mean of a_0 is .21, and it includes 0 when less weight is given to the historical data. This finding is interesting, since it indicates that age is a potentially important prognostic factor for predicting survival in melanoma.

In addition, when the historical data and the current data are equally weighted (i.e., $a_0 = 1$ with probability 1), the HPD intervals for age and gender both do not include 0, thus demonstrating the importance of gender in predicting overall survival. Another feature of Table 4 is that the posterior standard deviations of the β_j's become smaller and the HPD intervals become narrower as the posterior mean of a_0 increases. This demonstrates that incorporation of historical data can yield more precise posterior estimates of β. For example, we see that when $a_0 = 1$, the posterior mean, standard deviation, and HPD interval for the age coefficient are .16, .04, and (0.08, 0.24), respectively, whereas when we do not incorporate any historical data (i.e., $a_0 = 0$), these values are .09, 0.07, and $(-0.05, 0.23)$. We see that there is

Table 4: Melanoma Data: Posterior Estimates of the Model Parameters with $\alpha \sim Gamma(1, 0.01)$ *and* $\lambda \sim N(0, 10,000)$

$E(a_0 \mid D_{obs}, D_{0,obs})$	Variable	Posterior Mean	Posterior Std Dev	95% HPD Interval
0	intercept	0.09	0.11	(-0.12, 0.30)
(with prob. 1)	age	0.09	0.07	(-0.05, 0.23)
	gender	-0.12	0.16	(-0.44, 0.19)
	ps	-0.23	0.26	(-0.73, 0.28)
	α	1.31	0.09	(1.15, 1.48)
	λ	-1.36	0.12	(-1.60,-1.11)
0.03	intercept	0.17	0.11	(-0.04, 0.38)
	age	0.10	0.07	(-0.04, 0.24)
	gender	-0.14	0.15	(-0.44, 0.15)
	ps	-0.19	0.25	(-0.68, 0.28)
	α	1.17	0.07	(1.03, 1.32)
	λ	-1.45	0.13	(-1.70,-1.20)
0.06	intercept	0.21	0.11	(0.01, 0.43)
	age	0.11	0.07	(-0.03, 0.24)
	gender	-0.16	0.15	(-0.45, 0.13)
	ps	-0.16	0.24	(-0.63, 0.29)
	α	1.12	0.07	(0.99, 1.25)
	λ	-1.53	0.13	(-1.78,-1.28)
0.14	intercept	0.25	0.10	(0.05, 0.45)
	age	0.12	0.06	(-0.00, 0.24)
	gender	-0.20	0.14	(-0.47, 0.07)
	ps	-0.09	0.22	(-0.53, 0.31)
	α	1.06	0.06	(0.95, 1.17)
	λ	-1.62	0.12	(-1.85,-1.39)
0.21	intercept	0.26	0.10	(0.08, 0.45)
	age	0.13	0.06	(0.01, 0.24)
	gender	-0.22	0.13	(-0.48, 0.03)
	ps	-0.05	0.20	(-0.44, 0.34)
	α	1.04	0.05	(0.94, 1.15)
	λ	-1.67	0.11	(-1.89,-1.45)
0.29	intercept	0.26	0.09	(0.08, 0.43)
	age	0.13	0.06	(0.02, 0.24)
	gender	-0.24	0.12	(-0.48, 0.00)
	ps	-0.01	0.19	(-0.38, 0.35)
	α	1.03	0.05	(0.93, 1.13)
	λ	-1.70	0.11	(-1.91,-1.50)
1	intercept	0.22	0.06	(0.11, 0.35)
(with prob. 1)	age	0.16	0.04	(0.08, 0.24)
	gender	-0.32	0.09	(-0.50,-0.15)
	ps	0.14	0.13	(-0.11, 0.39)
	α	1.00	0.04	(0.93, 1.07)
	λ	-1.82	0.08	(-1.97,-1.67)

Table 5: Posterior Estimates of the Model Parameters with $E(a_0|D_{obs}, D_{0,obs}) = 0.29$, $\alpha \sim Gamma(1,1)$ *and* $\lambda \sim N(0,10)$

Variable	Posterior Mean	Posterior Std Dev	95% HPD Interval
intercept	0.26	0.09	(0.07, 0.42)
age	0.13	0.06	(0.02, 0.24)
gender	-0.24	0.12	(-0.48, 0.00)
ps	-0.01	0.19	(-0.38, 0.35)
α	1.02	0.05	(0.93, 1.12)
λ	-1.69	0.11	(-1.90,-1.48)

a large difference in these estimates, especially in the standard deviations and the HPD intervals. A partial explanation of these results is that the E1673 study has had nearly 20 years of follow-up on 650 patients, and thus the potential impact of age and gender on overall survival is much more apparent in these data than the current data E1684, which has had less than 10 years of follow-up on 284 patients, and has about 39% censoring.

We conducted a detailed sensitivity analysis for the regression coefficients by varying the hyperparameters for a_0 (i.e. (γ_0, λ_0)) and varying the hyperparameters for $\psi = (\alpha, \lambda)$. Table 4 shows that the posterior estimates of the parameters are fairly robust as the hyperparameters (γ_0, λ_0) are varied. When we vary the hyperparameters for ψ, the posterior estimates of β are also robust for a wide range of hyperparameters values. For example, when fixing the hyperparameters for a_0 so that $E(a_0|D_{obs}) = .29$ and taking $\alpha \sim gamma(1,1)$ and $\lambda \sim N(0,10)$, we obtain the posterior estimates shown in Table 5. We see that these priors for (α, λ) are fairly informative relative to those of Table 4. Other moderate to informative choices of hyperparameters for (α, λ) also led to fairly robust posterior estimates of β.

Finally, we mention that we used the Gibbs sampler to sample from the posterior distribution, in the Gibbs sampler a burn-in of 1000 samples was used, with autocorrelations disappearing after lag 5 for nearly all parameters, and we used 50,000 Gibbs iterates after the burn-in for all of the posterior computations. Further, all HPD intervals were computed by using an efficient Monte Carlo method of Chen and Shao [5]. In summary, we see the powerful advantages of the cure rate model (2) and the desirable features of incorporating historical data into a Bayesian analysis. Our priors allow us to control the impact of the historical data on the overall analysis. In addition, our proposed model is computationally attractive, requiring only a straightforward adaptive rejection algorithm of [10] for Gibbs sampling.

5. Semiparametric Cure Rate Model. From point of view of a survival analyst, there is a lot of interest in developing a semi-parametric version of the parametric cure rate model in (2). The parametric assumptions about the distributions of the promotion times are often considered very restrictive and it is difficult to justify any particular parametric distribution to model the distribution Z's. A

crucial issue with semiparametric cure rate modeling, and semi-parametric survival models in general, is the behavior of the model in the right tail of the survival distribution. Due to censoring and failures over time, there are typically few subjects at risk in the tail of the survival curve after sufficient follow-up, and therefore estimation of the cure rate can be quite sensitive to the choice of the semi-parametric model [2]. Thus there is a need to carefully model the right tail of the survival curve, and allow the model to be more parametric in the right tail, while also allowing the model to be nonparametric in other parts of the curve. Ibrahim, Chen, and Sinha [12] construct such a model by defining a smoothing parameter κ, $0 < \kappa < 1$, which does not depend on the data, and this κ controls the degree of parametricity in the right tail of the survival curve.

Following [12], we construct a finite partition of the time axis, $0 < a_1 < \ldots < a_J$, with $a_J > y_i$ for all $i = 1, \ldots, n$. Thus, we have the J intervals $(0, a_1], (a_1, a_2], \ldots, (a_{J-1}, a_J]$. We thus assume that the hazard for $F(y)$ (the cdf of the promotion time) is equal to λ_j for the j^{th} interval, $j = 1, \ldots, J$, leading to

$$F(y) = 1 - \exp\left\{-\lambda_j(y - a_{j-1}) - \sum_{g=1}^{j-1} \lambda_g(a_g - a_{g-1})\right\} . \tag{14}$$

When $J = 1$, $F(y)$ reduces to the parametric exponential model. With this assumption, the complete data likelihood can be written as

$$\begin{aligned} L(\beta, \lambda | D_{comp}) = & \prod_{i=1}^{n} \prod_{j=1}^{J} \exp\left\{-(N_i - \nu_i)\delta_{ij}\left[\lambda_j(y_i - a_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(a_g - a_{g-1})\right]\right\} \\ & \times \prod_{i=1}^{n} \prod_{j=1}^{J} (N_i\lambda_j)^{\delta_{ij}\nu_i} \exp\left\{-\nu_i\delta_{ij}\left[\lambda_j(y_i - a_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(a_g - a_{g-1})\right]\right\} \\ & \times \exp\left\{\sum_{i=1}^{n} [N_i x_i'\beta - \log(N_i!) - \exp(x_i'\beta)]\right\} , \end{aligned} \tag{15}$$

where $\lambda = (\lambda_1, \ldots, \lambda_J)$ and $\delta_{ij} = 1$ if the i^{th} subject failed or was censored in the j^{th} interval, and 0 otherwise. The model in (15) is a semi-parametric version of the model in (2). There are several attractive features of the model in (15). First, we note the degree of the non-parametricity is controlled by J. The larger the J, the more non-parametric the model is. However, by picking a small to moderate J, we get more of a parametric shape for $F(y)$. In practice, we recommend doing analyses for several values of J to see the sensitivity of the posterior estimates of the regression coefficients. The semiparametric cure rate model (15) is quite flexible, as it allows us to model general shapes of the hazard function, as well as choose the degree of parametricity in $F(y)$ through suitable choices of J. Again, since $\mathbf{N}$ is not observed, the observed data likelihood, $L(\beta, \lambda | D)$ is obtained by summing out $\mathbf{N}$ from (15). For semiparametric Bayesian inference purposes, we now need to put a suitable prior process of $\mathbf{\Lambda} = (\lambda_1, \cdots, \lambda_J)$. In general, this is routinely done by using a prior process with parameters for the prior mean and prior confidence around the mean. A popular example of such a prior process in survival analysis

is the gamma process – see [18] for a review on semiparametric Bayesian survival analysis. But, for this particular problem, we need to use a more innovative prior process based on the parameter κ which controls the rate at which the tail of the survival curve under the prior process converges to a parametric function.

The important aspect of the semiparametric cure-rate model is the parameter κ controlling the degree of parametric nature of the hazard of $F(y)$ at the right tail. This is important to avoid the high dependence of the estimates of cure-rate and right tail of survival population survival curve on the few failures at the right tails. See [2] for details about the problem of high dependence of classical semiparametric estimates of cure-rate on observations with failures at the right tails. Specifically, the prior for λ_j used by us depends on κ, such that the model converges to a parametric model in the right tail of $F(t)$ as $t \to \infty$. Because the κ is assumed known, the degree of the parametric nature of the model in the right tail does not depend on the data. Also, κ will allow us some control over the degree of parametric nature in the beginning and middle part of the survival distribution at *a priori* as well as *a posteriori*. A more parametric shape of the model in the right tail facilitates more stable and precise estimates of the cure-rate parameters and right tail of $S_{pop}(t)$. This approach is fundamentally very different from previous approaches for semi-parametric Bayesian survival analysis, which primarily focus on specifying a prior process with a mean function and possibly a prior precision parameter, in which posterior properties of both of them depend on the data.

Let $F_0(t|\lambda_0)$ be the prior mean of the common cdf $F(t)$ of the promotion times. Note that $F_0(t|\lambda_0)$ is a known parametric function of t with some associated unknown parameter λ_0. Also, $H_0(t)$ denotes the corresponding cumulative baseline hazard function of $F_0(t)$. For example, $F_0(t|\lambda_0) = 1 - \exp(-\lambda_0 t)$ with $H_0(t|\lambda_0) = \lambda_0 t$.

We take the λ_j's to be independent a priori, each having a gamma prior distribution with mean

$$\mu_j = E(\lambda_j|\lambda_0) = \frac{H_0(a_j) - H_0(a_{j-1})}{a_j - a_{j-1}}\,, \tag{16}$$

and variance

$$\sigma_j^2 = \mathrm{Var}(\lambda_j|\lambda_0, \kappa) = \mu_j \kappa^j\,, \tag{17}$$

where $0 < \kappa < 1$ is the smoothing parameter. We see that as $\kappa \to 0$, $\sigma_j^2 \to 0$, so that small values of κ imply a more parametric model in the right tail. In addition, we observe that as $j \to \infty$, $\sigma_j^2 \to 0$, implying that the degree of parametricity is increased at a rate governed by κ as the number of intervals increases. This property also implies that as $j \to \infty$, the survival distribution in the right tail becomes more parametric regardless of any fixed value of κ. The properties of this model are attractive. For example, if $F_0(.|\lambda_0)$ is an exponential distribution, then $F_0(y|\lambda_0) = 1 - \exp(-\lambda_0 y)$, so that $\mu_j = \lambda_0$ and $\sigma_j^2 = \lambda_0 \kappa^j$. If $F_0(.|\lambda_0)$ is a Weibull distribution, then $F_0(y|\lambda_0) = 1 - \exp(-\gamma_0 y^{\alpha_0})$, $\lambda_0 = (\alpha_0, \gamma_0)$, so that $\mu_j = \gamma_0 \frac{(a_j^{\alpha_0} - a_{j-1}^{\alpha_0})}{a_j - a_{j-1}}$ and $\sigma_j^2 = \gamma_0 \frac{(a_j^{\alpha_0} - a_{j-1}^{\alpha_0})}{a_j - a_{j-1}} \kappa^j$.

The intervals $(a_{j-1}, a_j]$, $j = 1, \ldots, J$ need to be chosen so that with the combined data sets from the historical and current data, at least one failure observation

falls in each interval. This technique for choosing J is quite reasonable and results in a stable Gibbs sampler. For the melanoma data, we later conduct sensitivity analyses on the construction of the intervals, $(a_{j-1}, a_j]$, $j = 1, \ldots, J$. Three different constructions of $(a_{j-1}, a_j]$ were considered. We chose the subintervals $(a_{j-1}, a_j]$ with (i) equal numbers of failures or censored observations; (ii) approximately equal lengths subject to the restriction that at least one failure observation occurs in each interval; (iii) decreasing numbers of failures or censored observations. More specifically, in case (iii) we took a_j to be the $((1 - e^{(-j/J)})/(1 - e^{-1}))^{th}$ quantile of the y_j's. We found that the posterior estimates were quite robust with respect to these constructions. We now formally state several properties for this model. The proofs are omitted for the sake of reducing the length of the article.

Assume that $\frac{a_j + a_{j-1}}{2} \to t$ as $a_j - a_{j-1} \to 0$. Then for any j, according to this prior process, $E(\lambda_j|\lambda_0) \to h_0(t)$ as $a_j - a_{j-1} \to 0$, where $h_0(t) = \frac{d}{dt} H_0(t)$.

For example, when $F_0(y|\lambda_0) = 1 - \exp(-\lambda_0 y)$, then $E(\lambda_j|\lambda_0) = \lambda_0$ regardless of our choices of $a_1, \ldots, a_J$. When $F_0(y|\lambda_0) = 1 - \exp(-\gamma_0 y^{\alpha_0})$, then $E(\lambda_j|\lambda_0) \to \gamma_0 \alpha_0 t^{\alpha_0 - 1}$ as $a_j - a_{j-1} \to 0$. This assures that as j becomes large and $a_j - a_{j-1} \to 0$, then this prior process approximates any prior process with prior mean $h_0(t)$ defined on the promotion time hazard $h^*(t|\lambda)$ corresponding to (14).

Let $S_p*(y|\lambda) = \exp(-\theta F^*(y|\lambda))$, where $F^*(y|\lambda)$ is given by (14). Then, $S_p*(y|\lambda) \to S_p(y|\lambda_0)$ as $\kappa \to 0$, where $S_p(y|\lambda_0) = \exp(-\theta F_0(y|\lambda_0))$.

Let $f^*(y|\lambda) = \frac{d}{dy} F^*(y|\lambda)$, and $h_p^*(y|\lambda) = \theta f^*(y|\lambda)$ denote the corresponding hazard function. Then $h_p^*(y|\lambda) \to \theta f_0(y|\lambda_0)$ as $\kappa \to 0$, where $f_0(y|\lambda_0) = \frac{d}{dy} F_0(y|\lambda_0)$.

In practice, we recommend doing analyses for several values of κ, J, and $F_0(.|\lambda_0)$ to examine the sensitivity of the posterior estimates to various choices of these parameters.

5.1. *Prior Distributions.* We give joint prior specifications for the semiparametric model in (16) and (17). We specify a hierarchical model and first consider a joint (improper) noninformative prior distribution for $(\beta, \lambda, \lambda_0)$. We specify the joint prior of these parameters as

$$\pi(\beta, \lambda, \lambda_0) = \pi(\beta)\pi(\lambda|\lambda_0)\pi(\lambda_0) \propto \pi(\beta) \left[\prod_{j=1}^{J} \pi(\lambda_j|\lambda_0) \right] \pi(\lambda_0). \tag{18}$$

As noted earlier, we take each $\pi(\lambda_j|\lambda_0)$ to be independent gamma densities with mean μ_j and variance σ_j^2. If $F_0(.)$ is an exponential distribution, then λ_0 is a scalar, and we specify a gamma prior for it, i.e., $\pi(\lambda_0) \propto \lambda_0^{\zeta_0 - 1} \exp(-\tau_0 \lambda_0)$, where ζ_0 and τ_0 are specified hyperparameters. If $F_0(.)$ is a Weibull distribution, then $\lambda_0 = (\gamma_0, \alpha_0)$. In this case, we take a prior of the form

$$\pi(\lambda_0) = \pi(\alpha_0, \gamma_0) \propto \alpha_0^{\zeta_{\alpha_0} - 1} \exp(-\tau_{\alpha_0} \alpha_0) \gamma_0^{\zeta_{\gamma_0} - 1} \exp(-\tau_{\gamma_0} \gamma_0), \tag{19}$$

where ζ_{α_0}, τ_{α_0}, ζ_{γ_0} and τ_{γ_0} are specified hyperparameters. For β, we consider a uniform improper prior. The next theorem establishes the propriety of the joint posterior distribution of $(\beta, \lambda, \lambda_0)$, when using an exponential distribution or a Weibull distribution for $F_0(.)$.

THEOREM 4. *Suppose (i) when $\nu_i = 1$, $y_i > 0$, (ii) there exists $i_1, i_2, \ldots, i_J$ such that $\nu_{i_j} = 1$, and $a_{j-1} < y_{i_j} \leq a_j$, $j = 1, \ldots, J$, (iii) the design matrix X^* with i^{th} row equal to $\nu_i x_i'$ is of full rank, iv) if $F_0(.|\lambda_0)$ is an exponential distribution, $\zeta_0 > 0$ and $\tau_0 > \sum_{j=1}^{J} \frac{1}{\kappa^j} \log[(1/\kappa^j)/((y_{i_j} - a_{j-1})/2 + 1/\kappa^j)]$, and if $F_0(.|\lambda_0)$ is a Weibull distribution, $\zeta_{\gamma_0} > 0$, $\tau_{\gamma_0} \geq 0$, $\zeta_{\alpha_0} > 0$, and $\tau_{\alpha_0} > -\zeta_{\gamma_0} \log(a_J)$. Then the posterior distribution of $(\beta, \lambda, \lambda_0)$ is proper, i.e., $\int L(\beta, \lambda|D)\pi(\beta, \lambda, \lambda_0)\, d\beta\, d\lambda\, d\lambda_0 < \infty$, where $L(\beta, \lambda|D)$ is the likelihood function based on the observed data D.*

A proof of Theorem (4) is left as an exercise. Theorem (4) provides a very general class of improper noninformative priors for $(\beta, \lambda, \lambda_0)$. First we mention that in condition (iv) of Theorem 4, τ_0 can be negative, thus resulting in an improper prior for λ_0 when $F_0(.|\lambda_0)$ is exponential. Second, τ_{α_0} is also allowed to be negative, resulting in a joint improper prior for (γ_0, α_0) when $F_0(.|\lambda_0)$ is Weibull.

The power prior for this model takes the form

$$\pi(\beta, \lambda, \lambda_0, a_0|D_0) \propto L(\beta, \lambda|D_0)^{a_0} \pi_0(\beta, \lambda|\lambda_0)\pi_0(\lambda_0) a_0^{\xi_0 - 1}(1 - a_0)^{\psi_0 - 1}, \tag{20}$$

where $L(\beta, \lambda|D_0)$ is the likelihood function based on the observed historical data, and ξ_0 and ψ_0 are prespecified hyperparameters. The initial prior for $(\beta, \lambda, \lambda_0)$, is given by (18) with $\pi_0(\lambda_0)$ taking the form given by $\pi_0(\lambda_0) \propto \lambda_0^{\zeta_0 - 1} \exp(-\tau_0 \lambda_0)$ or by (19), depending on the form of $F_0(.|\lambda_0)$. Following the proofs of Theorem 4 and Theorem 3 of [3], it can be shown that the prior distribution $\pi(\beta, \lambda, \lambda_0, a_0|D_0)$ given by (20) is proper under some very general conditions.

5.2. *Example - Melanoma Data.* We revisit the E1684 and E1690 trials discussed in the previous section. Our main purpose in this example is to examine the tail behavior of our proposed model as κ, a_0, F_0, and J are varied. Of particular interest is the sensitivity of the posterior estimates of β, λ, and $S^*(t|\lambda) = 1 - F^*(t|\lambda)$, as these parameters are varied, where $F^*(t|\lambda)$ is defined in (14). The E1690 study is quite suitable for our purposes here since the median follow-up for E1690 (4.33 years) is considerably smaller than E1684 (6.9 years). Thus, cure rate estimation based on the E1690 study alone, i.e., $a_0 = 0$, may be more sensitive than that of an analysis which incorporates the historical data E1684. In our example, three covariates were used, and an intercept was included in the model. The three covariates are treatment (IFN, OBS), age, which is continuous, and gender (male, female). Let $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ be the regression coefficient vector corresponding to an intercept and the three covariates, respectively.

Table 6 gives posterior means, standard deviations, and 95% Highest Posterior Density (HPD) intervals of β for several values of κ using the exponential and Weibull models for F_0 with $J = 10$ intervals, and when $E(a_0|D, D_0) = 0.33$. As κ is varied for a given a_0 using an exponential or Weibull F_0, we see small to moderate changes in the posterior estimates of β. As a_0 is varied, more substantial changes occur in the posterior estimates of β across values of a_0. For example, using an exponential F_0 and $\kappa = 0.05$, the posterior means, standard deviations, and 95% HPD intervals for the treatment coefficient, i.e., β_2, are -0.209, 0.130, and $(-0.461, 0.050)$ when $a_0 = 0$ with probability 1; -0.242, 0.115, and $(-0.469, -0.018)$

when $E(a_0|D, D_0) = 0.33$; and -0.277, 0.079, and $(-0.462, -0.087)$ when $a_0 = 1$ with probability 1. In general, the posterior standard deviations for $E(a_0|D, D_0) > 0$ are smaller than those for $a_0 = 0$, therefore resulting in narrower 95% HPD intervals. A partial explanation of this is that, by incorporating the historical data, more precise estimates of the regression coefficients and right tail of the survival curve are obtained. Overall, for given a_0, we conclude that the estimates of β are reasonably robust as κ is varied, but change substantially as a_0 is varied.

Table 7 shows posterior summaries of the cure rates and survival function $S^*(t)$ for varying κ and F_0 when $E(a_0|D, D_0) = 0.33$. For a given F_0, we see moderate changes in the cure rates as κ is varied. When a_0 and κ remain fixed and F_0 is changed, we see that the estimates are quite robust. Also from Table 7, we can see that a monotonic increase in the mean of the cure rate estimates occurs as κ is increased. A similar phenomenon occurs with other values of a_0. In summary, Table 7 shows that small to moderate changes can occur in the cure rates as the degree of parametricity in the right tail of the survival curve, κ, is changed. We have also computed estimates of λ for several values of κ and a_0, assuming that F_0 is exponential. We observe that for a given a_0, the posterior estimates of λ can change moderately to considerably as κ varies. For example, with $a_0 = 0$, the posterior mean of λ_{10} is 0.617 for $\kappa = 0.05$, and 0.788 when $\kappa = 0.95$. A Similar phenomenon occurs when $a_0 = 1$. These changes in λ can be summarized better by examining the estimated survival function for the non-cured patients, denoted by $S^*(t|\lambda)$. Figure 2 shows the posterior estimates of $S^*(t|\lambda)$ for $E(a_0|D, D_0) = 0.33$ using several values of κ. Figure 2(a) corresponds to an exponential F_0 and Figure 2(b) corresponds to a Weibull F_0. We see from Figure 2 that small to moderate changes in the survival estimates occur as κ is varied. The biggest changes occur in the interval $1 \leq t \leq 5$. Table 7 summarizes $S^*(t|\lambda)$ at $t = 3.5$ for several values of κ and F_0. We see from Table 7 that for fixed F_0 and a_0, moderate changes in $S^*(3.5|\lambda)$ occur as κ is varied. When F_0 is Weibull, bigger differences are seen. Thus, $S^*(t|\lambda)$ can be moderately sensitive to the choice of κ. We also observe that as more weight is given to the historical data (i.e., a_0 is increased), $S^*(3.5|\lambda)$ increases. For example, for a Weibull F_0 with $a_0 = 0$, $S^*(3.5|\lambda) = 0.124$ for $\kappa = 0.05$ and 0.119 for $\kappa = 0.6$, while from able 7 with $E(a_0|D, D_0) = 0.33$, we have $S^*(3.5|\lambda) = 0.136$ for $\kappa = 0.05$ and 0.129 for $\kappa = 0.6$. This phenomenon is consistent with the notion that the cure rate decreases as more weight is given to the historical data. Finally, we note that sensitivity analyses were also carried out using several different values of J, and similar results were obtained.

6. Multivariate Cure Rate Models. It is often of interest to jointly model several types of failure time random variables in survival analysis, such as time to cancer relapse at two different organs, times to cancer relapse and time to other adverse events, times to first and second infections, and so forth. In addition, these random variables typically have joint and marginal survival curves that "plateau" beyond a certain period of follow-up, and therefore it is of great importance in these situations to develop a joint cure rate model for inference.

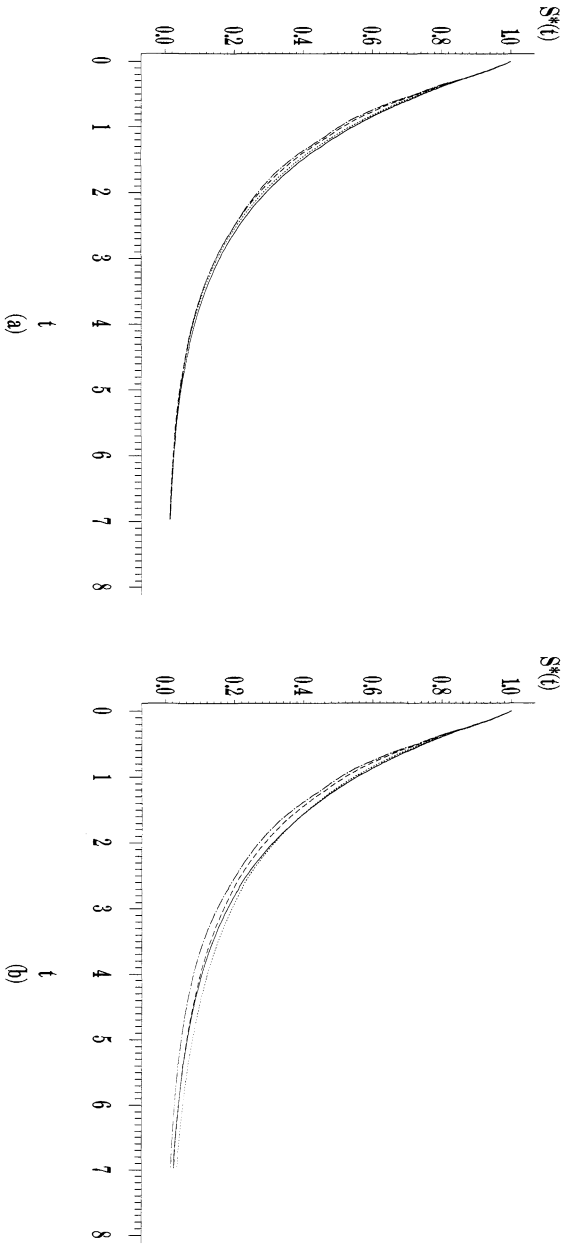
Table 6: Posterior Estimates of β

F_0	κ	Variable	Mean	Std Dev	95% HPD Interval
Exponential	0.05	intercept	0.183	0.096	(-0.009, 0.367)
		treatment	-0.242	0.115	(-0.469, -0.018)
		age	0.099	0.058	(-0.011, 0.214)
		gender	-0.118	0.120	(-0.361, 0.110)
	0.60	intercept	0.164	0.094	(-0.019, 0.348)
		treatment	-0.245	0.115	(-0.470, -0.020)
		age	0.098	0.058	(-0.016, 0.211)
		gender	-0.115	0.120	(-0.352, 0.119)
	0.95	intercept	0.157	0.098	(-0.038, 0.345)
		treatment	-0.242	0.116	(-0.472, -0.017)
		age	0.097	0.058	(-0.015, 0.211)
		gender	-0.113	0.121	(-0.348, 0.128)
Weibull	0.05	intercept	0.202	0.102	(0.002, 0.404)
		treatment	-0.242	0.115	(-0.471, -0.019)
		age	0.100	0.057	(-0.012, 0.213)
		gender	-0.118	0.121	(-0.358, 0.114)
	0.60	intercept	0.180	0.098	(-0.011, 0.371)
		treatment	-0.244	0.115	(-0.472, -0.021)
		age	0.098	0.058	(-0.018, 0.209)
		gender	-0.113	0.120	(-0.355, 0.117)
	0.95	intercept	0.160	0.097	(-0.034, 0.345)
		treatment	-0.244	0.116	(-0.473, -0.022)
		age	0.097	0.058	(-0.013, 0.213)
		gender	-0.115	0.120	(-0.351, 0.119)

Table 7: Posterior Summaries of Cure Rates and $S^*(t|\lambda)$

		Cure Rate		
F_0	κ	Mean	Std Dev	$S^*(3.5\|\lambda)$
Exponential	0.05	0.361	0.065	0.115
	0.60	0.368	0.065	0.106
	0.95	0.370	0.064	0.107
Weibull	0.05	0.354	0.065	0.136
	0.60	0.362	0.064	0.129
	0.95	0.369	0.065	0.108

Figure 2: Plots of Survival Function $S^*(t|\lambda)$ for non-cured patients with $J = 10$ and $E(a_0|D, D_0) = 0.33$, where (a) $F_0(.|\lambda_0)$ is an exponential distribution, (b) $F_0(.|\lambda_0)$ is a Weibull distribution, and the solid, dotted, dashed, and dot-dashed curves correspond to $\kappa = 0.05, 0.30, 0.60, 0.95$, respectively.



There does not appear to be a natural multivariate extension of the BG model in (1). There does not appear to be a natural multivariate extension of the BG model in (1). Even if such an extension was available, it appears that a multivariate mixture model would be extremely cumbersome to work with from a theoretical and computational perspective. As an alternative to a direct multivariate extension of (1), [4] present a *multivariate cure rate model* which is a natural multivariate extension of the univariate model in (2). This model proves to be quite useful for modeling multivariate data in which the joint failure random variables have a surviving fraction and each marginal failure time random variable also has a surviving fraction. To induce the correlation structure between the failure times, a frailty term [6, 11, 16] is used.

For clarity and ease of exposition, we will focus our discussion on the bivariate cure rate model, as extensions to the general multivariate case are quite straightforward. Let $\mathbf{Y} = (Y_1, Y_2)$ be a bivariate failure time, such as Y_1 = time to cancer relapse in location 1 and Y_2 = time to cancer relapse in location 2, or Y_1 = time to first infection, and Y_2 = time to second infection, and so forth. We assume that (Y_1, Y_2) are not ordered. For an arbitrary patient in the population, let $\mathbf{N} = (N_1, N_2)$ denote latent (unobserved) variables for (Y_1, Y_2), respectively. We assume throughout that N_k has a Poisson distribution with mean $\theta_k w$, $k = 1, 2$, and (N_1, N_2) are independent. The quantity w is a frailty component in the model which induces a correlation between the latent variables (N_1, N_2). Here we take w to have a positive stable law distribution indexed by the parameter α, denoted by $w \sim \mathit{Stable}(\alpha)$, where $0 < \alpha < 1$. Although several choices can be made for the distribution of w, the positive stable law distribution is quite attractive, common, and flexible in the multivariate survival setting. In addition, it will yield several desirable properties.

Using the same latent promotion times arguments for both components, after some algebra, the survival function for $\mathbf{Y} = (Y_1, Y_2)$ given w can be shown as

$$S_{pop}(y_1, y_2|w) = \exp(-w\,[\theta_1 F_1(y_1) + \theta_2 F_2(y_2)])\,, \tag{21}$$

where $P(N_k = 0) = P(Y_k = \infty) = \exp(-\theta_k)$, $k = 1, 2$. We emphasize here that the primary roles of $\mathbf{N}$ and $\mathbf{Z}_i$ is that they only facilitate the construction of the model and need not have any physical or biological interpretation at all for the model to be valid. They are quite useful for the computational implementation of the model via the Gibbs sampler as discussed below and thus are defined primarily for this purpose. The model in (21) is valid for *any* time-to-event data with a cure rate structure as implied by (21) and the subsequent development. Thus the model can be useful for modeling various types of failure time data, including time to relapse, time to infection, time to complication, time to rejection, and so forth. In addition, the frailty variable w serves a dual purpose in the model - it induces the correlation between Y_1 and Y_2 and at the same time relaxes the Poisson assumption of N_1 and N_2 by adding the same extra Poisson variation through their respective means $\theta_1 w$ and $\theta_2 w$.

The $\mathit{Stable}(\alpha)$ density for w $(0 < \alpha < 1)$ can be expressed through the Laplace transform of w, given by $E(\exp(-sw)) = \exp(-s^\alpha)$. A useful reference on stable distributions is [17]. A straightforward derivation yields the unconditional survival

function

$$S_{pop}(y_1, y_2) = \exp\{-[\theta_1 F_1(y_1) + \theta_2 F_2(y_2)]^\alpha\}. \tag{22}$$

Naturally, the marginal distribution of each component in (22) has a proportional hazards structure if the covariates enter the model only through (θ_1, θ_2).

The joint cure fraction implied by (22) is $S_{pop}(\infty, \infty) = \exp(-[\theta_1 + \theta_2]^\alpha)$. From (22), the marginal survival functions are

$$S_k(y) = \exp(-\theta_k^\alpha (F_k(y))^\alpha), \quad k = 1, 2. \tag{23}$$

Equation (23) indicates that the marginal survival functions have a cure rate structure with probability of cure $\exp(-\theta_k^\alpha)$ for Y_k, $k = 1, 2$. It is important to note in (23) that each marginal survival function has a proportional hazards structure as long as the covariates, $\mathbf{x}$, only enter through θ_k. The marginal hazard function is given by $\alpha\theta_k^\alpha f_k(y)(F_k(y))^{\alpha-1}$, with attenuated covariate effect $(\theta_k(x))^\alpha$, and $f_k(y)$ is the survival density corresponding to $F_k(y)$.

In addition, we can express the marginal survival functions in (23) in terms of BG models, as we can write

$$S_k(y) = \exp(-\theta_k^\alpha) + (1 - \exp(-\theta_k^\alpha)) S_k^*(y)\ , \tag{24}$$

where $S_k^*(y) = \frac{\exp(-\theta_k^\alpha (F_k(y))^\alpha) - \exp(-\theta_k^\alpha)}{1 - \exp(-\theta_k^\alpha)}$, $k = 1, 2$. Note that $S_k^*(y)$ defines a proper survivor function.

The parameter α $(0 < \alpha < 1)$ is a scalar parameter that is a measure of association between (Y_1, Y_2). Small values of α indicate high association between (Y_1, Y_2). As $\alpha \to 1$, this implies less association between (Y_1, Y_2) which can be seen from (22). Following [6] and [16], we can compute a local measure of association, denoted, $\theta^*(y_1, y_2)$, as a function of α. This measure of association is defined as

$$\theta^*(y_1, y_2) = \frac{S_{pop}(y_1, y_2) \frac{\partial^2}{\partial y_1 \partial y_2} S_{pop}(y_1, y_2)}{(\frac{\partial}{\partial y_1} S_{pop}(y_1, y_2))(\frac{\partial}{\partial y_2} S_{pop}(y_1, y_2))}\ . \tag{25}$$

For the multivariate cure rate model in (22), $\theta^*(t_1, t_2)$ is well defined, and is given by

$$\theta^*(y_1, y_2) = \alpha^{-1}(1 - \alpha)\,(\theta_1 F_1(y_1) + \theta_2 F_2(y_2))^{-\alpha} + 1\ . \tag{26}$$

We see that $\theta^*(t_1, t_2)$ in (26) decreases in (y_1, y_2). That is, the association between (Y_1, Y_2) is greater when (Y_1, Y_2) are small and the association decreases over time. Such a property, which is due to the stable frailty distribution, is quite desirable, for example, when Y_1 and Y_2 denote times to relapse in two locations. As $\alpha \to 0$, $S_{pop}(y_1, y_2)$ in (22) approaches the minimum of $(S_1(y_1), S_2(y_2))$, and thus $S_{pop}(y_1, y_2)$ achieves the Fréchet bound of maximal dependence (see [16]). Finally, we mention that a global measure of dependence such as Kendall's τ or the Pearson correlation coefficient is not well defined for the multivariate cure rate model (22) since no moments for cure rate models exist due to the improper survival function.

The multivariate cure rate model presented here is attractive in several respects. First, the model has a proportional hazards structure for the population hazard, conditionally as well as marginally, when covariates are entered through the cure

rate parameter, and thus has an appealing interpretation. Also, the model is computationally feasible. In particular, by introducing latent variables, efficient MCMC algorithms can be developed that enable us to sample from the joint posterior distribution of the parameters. Specifically, a modified version of the collapsed Gibbs technique of [14] can be in this case used for efficient Gibbs sampling from the posterior distribution. See [4] for details.

Acknowledgments. We would also like to thank Dr.Oakes for his encouragement and help.

REFERENCES

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

[2] A.B. Cantor and J.J. Shuster. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statistics in Medicine*, 11:931–937, 1992.

[3] M.-H. Chen, J.G. Ibrahim, and D. Sinha. A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94:909–919, 1999.

[4] M.-H. Chen, J.G. Ibrahim, and D. Sinha. Bayesian inference for multivariate survival data with a surviving fraction. *Journal of Multivariate Statistics*, 2000. In press.

[5] M.-H. Chen and Q.-M. Shao. Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8:69–92, 1999.

[6] D.G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151, 1978.

[7] M. Ewell and J.G. Ibrahim. The large sample distribution of the weighted log rank statistic under general local alternatives. *Lifetime Data Analysis*, 3:5–12, 1997.

[8] V.T. Farewell. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38:1041–1046, 1982.

[9] V.T. Farewell. Mixture models in survival analysis: Are they worth the risk? *Canadian Journal of Statistics*, 14:257–262, 1986.

[10] W.R. Gilks and P. Wild. Adaptive rejection sampling for gibbs sampling. *Applied Statistics*, 41:337–348, 1992.

[11] P. Hougaard. A class of multivariate failure time distributions. *Biometrika*, 73:671–678, 1986.

[12] J.G. Ibrahim, M.-H. Chen, and D. Sinha. Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, 2000. In press.

[13] J.M. Kirkwood, M.H. Strawderman, M.S. Ernstoff, T.J. Smith, E.C. Borden, and R.H. Blum. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: The Eastern Cooperative Oncology Group Trial EST 1684. *Journal of Clinical Oncology*, 14:7–17, 1996.

[14] J.S. Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89:958–966, 1994.

[15] R. Maller and X. Zhou. *Survival Analysis with Long-Term Survivors*. Wiley, New York, 1996.

[16] D. Oakes. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84:487–493, 1989.

[17] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, 1994.

[18] D. Sinha and D.K. Dey. Semiparametric bayesian analysis of survival data. *Journal of the American Statistical Association*, 92:1195–1212, 1997.

[19] D.K. Stangl and J.B. Greenhouse. Assessing placebo response using bayesian hierarchical survival models. *Lifetime Data Analysis*, 4:5–28, 1998.

[20] J.P. Sy and J.M.G. Taylor. Estimation in a proportional hazards cure model. *Biometrics*, 56:227–336, 2000.

[21] J.M.G. Taylor. Semi-parametric estimation in failure time mixture models. *Biometrics*, 51:899–907, 1995.

[22] A.Y. Yakovlev. Letter to the editor. *Statistics in Medicine*, 13:983–986, 1994.

[23] A.Y. Yakovlev, B. Asselain, V.J. Bardou, A. Fourquet, T. Hoang, A. Rochefediere, and A.D. Tsodikov. 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, pages 66–82. World Scientific, New Jersey, 1993.

[24] A.Y. Yakovlev and A.D. Tsodikov. *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, New Jersey, 1996.

DEBAJYOTI SINHA
DEPARTMENT OF BIOMETRYAND EPIDEMIOLOGY
MEDICAL UNIVERSITY OFSOUTH CAROLINA
135 CANNON ST., STE 303
CHARLESTON, SC 29425 USA
sinhad@musc.edu

JOSEPH IBRAHIM
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA
MCGAVRAN GREENBERG HALL,
CB #7420
CHAPEL HILL, NC 27599 USA

MING-HUI CHEN, PH.D.
ASSOCIATE PROFESSOR
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
215 GLENBROOK ROAD, U-4120
STORRS, CT 06269 USA