Journal of Multivariate Analysis **80**, 101–126 (2002)
doi:10.1006/jmva.2000.1975, available online at http://www.idealibrary.com on IDEAL®

Bayesian Inference for Multivariate Survival Data with a Cure Fraction

Ming-Hui Chen

Worcester Polytechnic Institute

Joseph G. Ibrahim

Harvard School of Public Health and Dana-Farber Cancer Institute

and

Debajyoti Sinha

Medical University of South Carolina

Received January 20, 2000; published online June 19, 2001

We develop Bayesian methods for right censored multivariate failure time data for populations with a cure fraction. We propose a new model, called the *multivariate cure rate model*, and provide a natural motivation and interpretation of it. To create the correlation structure between the failure times, we introduce a frailty term, which is assumed to have a positive stable distribution. The resulting correlation structure induced by the frailty term is quite appealing and leads to a nice characterization of the association between the failure times. Several novel properties of the model are derived. First, conditional on the frailty term, it is shown that the model has a proportional hazards structure with the covariates depending naturally on the cure rate. Second, we establish mathematical relationships between the marginal survivor functions of the multivariate cure rate model and the more standard mixture model for modelling cure rates. With the introduction of latent variables, we show that 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. Specifically, we propose a modified version of the collapsed Gibbs technique (J. S. Liu, 1994, *J. Amer. Statist. Assoc.* **89**, 958–966) to sample from the posterior distribution. This development will lead to an efficient Gibbs sampling procedure, which would otherwise be extremely difficult. We characterize the propriety of the joint posterior distribution of the parameters using a class of noninformative improper priors. A real dataset from a melanoma clinical trial is presented to illustrate the methodology.

AMS 2000 subject classifications: 92B15; 62P10.

Key words and phrases: cure rate; frailty model; Gibbs sampling; latent variables; posterior distribution; stable law.

0047-259X/01 $35.00

1. INTRODUCTION

Survival data with a cure fraction are becoming increasingly common in clinical trials and epidemiological studies. For example, the univariate *cure rate model* has been used for modelling failure time 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 (univariate) cure rate model is the mixture model introduced by Berkson and Gage (1952). In this model, we assume a certain fraction π of the population are cured, and the remaining $1-\pi$ are not cured. The survivor function for the entire population, denoted by $S_1(t)$, for this model, is given by

$$S_1(t) = \pi + (1-\pi)\, S^*(t), \tag{1.1}$$

where $S^*(t)$ denotes the survivor function for the non-cured group in the population. Clearly, $S_1(\infty) = \pi$, and thus $S_1(t)$ is not a proper survival function if $\pi > 0$. We mention that any cure rate model has an improper survival function by definition. However, $S^*(t)$ is a proper survival function and common choices for $S^*(t)$ are the exponential and Weibull distributions. We shall refer to the model in (1.1) as the *standard univariate cure rate model*. The standard univariate cure rate model has been extensively discussed in the statistical literature by several authors, including Farewell (1982, 1986), Goldman (1984), Halpern and Brown (1987a, 1987b), Gray and Tsiatis (1989), Sposto *et al.* (1992), Laska and Meisner (1992), Kuk and Chen (1992), Yamaguchi (1992), and Taylor (1995). Although the standard cure rate model appears to be attractive and is widely used, it has several drawbacks. First, $S_1(t)$ does not have a proportional hazards structure if the covariates are entered through π via a binomial regression. However, as Kuk and Chen (1992), Sy and Taylor (2000) and Peng and Dear (2000) point out, a proportional hazards structure is often specified for $S^*(t)$ in (1.1). Second, it is computationally difficult to work with, as it is well known that mixture models often have multiple modes and have likelihoods that are computationally unstable. For example, finding maximum likelihood estimates via Newton–Raphson or some other iterative method can often fail, as discussed by Cantor and Shuster (1992) and Yakovlev (1994). Such problems are especially prevalent when covariates are included in the model. Even Markov chain Monte Carlo (MCMC) sampling from (1.1) can be quite tricky as noted by Müller and Rosner (1997) since the Markov chain can easily get stuck in certain parts of the parameter space. This is an especially disappointing feature since this situation is often encountered in practice. Moreover, one of the most

crucial drawbacks of (1.1) is that it lacks a simple and natural multivariate extension.

In survival analysis, it is often of interest to jointly model several types of failure time random variables, such as time to cancer relapse at two different organs, times to cancer relapse and death, times to first and second infection, and so forth. Another important source of this type of data is survival data following a major surgery, say a heart bypass, when a patient may die due to post-surgery complications (first failure time variable) and he/she may be at risk of organ rejection (second failure time variable). These types of failure time variables are typically of great importance in survival analysis, and thus developing multivariate models which yield suitable properties and which induce an appropriate correlation structure is of great interest. 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.

There does not appear to be a natural multivariate extension of the standard cure rate model in (1.1). 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 (1.1), we propose a new model in this paper, called the *multivariate cure rate model*, which proves to be quite useful for modelling 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. The model we propose has some relation to the univariate cure rate model discussed by Yakovlev *et al.* (1993) and Asselain *et al.* (1996). To induce the correlation structure between the failure times, we introduce a frailty term (Clayton 1978, Hougaard 1986, and Oakes 1989), which is assumed to have a positive stable distribution. A positive frailty assumes that we have Cox's (Cox 1972) proportional hazards structure conditionally (i.e., given the unobserved frailty). Thus the marginal and conditional hazards of each component have a proportional hazards structure, and thus remain in the same class of univariate cure rate models.

The multivariate cure rate model we propose 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. Second, the model is computationally feasible. In particular, by introducing latent variables, we develop MCMC algorithms that enable us to sample from the joint posterior distribution of the parameters. Specifically, we propose a modified version of the collapsed Gibbs technique of Liu (1994). Our computational development facilitates an efficient Gibbs

sampling scheme for the posterior distribution. Without the development of the modified collapsed Gibbs methodology, the Gibbs sampling for the proposed model would be extremely difficult. Third, the model has several desirable properties. Specifically, we show that the marginal survivor functions have a cure rate structure and have a mathematical relationship with the standard cure rate model. In addition, we discuss Bayesian analyses of this model with covariates, and propose a class of noninformative improper priors that guarantee the propriety of the joint posterior distribution. We note that we assume a noninformative censoring mechanism throughout. The multivariate models described here are much more complicated in the context of informative censoring and are not examined here.

The rest of this article is organized as follows. In Section 2, we derive the multivariate cure rate model, and obtain several of its properties. In Section 3, we derive the likelihood function with covariates, and in Section 4 we characterize the propriety of the resulting posterior distribution with a particular class of noninformative improper priors. In Section 5, we develop an MCMC algorithm by introducing latent variables and proposing a collapsed Gibbs methodology for efficient sampling from the joint posterior distribution. In Section 6, we present a melanoma data set from an actual clinical trial to illustrate the proposed methodology. We conclude the paper with a brief discussion.

2. THE MODEL

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. The proposed bivariate cure rate model can be derived as follows. Let $T=(T_1, T_2)$ be a bivariate failure time, such as $T_1=$ time to cancer relapse and $T_2=$ time to death, or $T_1=$ time to first infection, and $T_2=$ time to second infection, and so forth. In our methodological development here, we assume that (T_1, T_2) are not ordered and have support on the upper orthant of the plane. For an arbitrary patient in the population, let $N=(N_1, N_2)$ denote latent (unobserved) variables for (T_1, T_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 $w \sim \text{Stable}(\alpha)$, 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 for the proposed multivariate model.

Let $Z_i = (Z_{1i}, Z_{2i})$ denote the random time for the ith latent risk factor to cause an event. We call Z_i the latent time for $T_i = (T_{1i}, T_{2i})$. The random vectors Z_i, $i = 1, 2, \ldots$ are assumed to be independent and identically distributed. The cumulative distribution function of Z_{ki} is denoted by $F_k(t) = 1 - S_k(t)$, $k = 1, 2$, and F_k is independent of N. The observed survival time can be defined by the random variable $T_k = \min\{Z_{ki},\ 0 \leqslant i \leqslant N_k\}$, where $P(Z_{k0} = \infty) = 1$ and N_k is independent of the sequence $Z_{k1}, Z_{k2}, \ldots$, for $k = 1, 2$. The survival function for $T = (T_1, T_2)$ given w, and hence the survival function for the population given w, is given by

$$
\begin{aligned}
S_{\text{pop}}(t_1, t_2 \mid w) &= \prod_{k=1}^{2} \left(P(N_k = 0) + P(Z_{k1} > t_k, \ldots, Z_{kN} > t_k, N_k \geqslant 1)\right) \\
&= \prod_{k=1}^{2} \left(\exp(-w\theta_k) + \left(\sum_{r=1}^{\infty} S_k(t_k)^r \frac{(w\theta_k)^r}{r!} \exp(-w\theta_k)\right)\right) \\
&= \prod_{k=1}^{2} \left(\exp(-w\theta_k + \theta_k w S_k(t_k))\right) \\
&= \exp(-w[\theta_1 F_1(t_1) + \theta_2 F_2(t_2)]), \qquad (2.1)
\end{aligned}
$$

where $P(N_k = 0) = P(T_k = \infty) = \exp(-\theta_k)$, $k = 1, 2$. We emphasize here that the primary roles of N and 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 in Section 5, and thus are defined primarily for this purpose. The model in (2.1) is valid for *any* time-to-event data with a cure rate structure as implied by (2.1) and the subsequent development. Thus the model can be useful for modelling various types of failure time data, including time to relapse, time to death, 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 T_1 and T_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 $\theta_2 w$. A univariate cure rate model related to (2.1), but quite different, is examined in Chen *et al.* (1999).

Following Ibragimov and Chernin (1959), the Stable(α) density for w ($0 < \alpha < 1$) can be expressed in the form

$$
f_s(w \mid \alpha) = a w^{-(a+1)} \int_0^1 s(u) \exp\left\{-\frac{s(u)}{w^a}\right\} du, \qquad w > 0, \qquad (2.2)
$$

where

$$a = \frac{\alpha}{1-\alpha} \qquad \text{and} \qquad s(u) = \left(\frac{\sin(\alpha\pi u)}{\sin(\pi u)}\right)^a \left(\frac{\sin[(1-\alpha)\,\pi u]}{\sin(\pi u)}\right),$$

and the Laplace transform of w is given by $E(\exp(-sw)) = \exp(-s^\alpha)$. A useful reference on stable distributions is Samorodnitsky and Taqqu (1994). Using the Laplace transform of w, a straightforward derivation yields the unconditional survival function

$$S_{\text{pop}}(t_1, t_2) = \exp\{-[\theta_1 F_1(t_1) + \theta_2 F_2(t_2)]^\alpha\}. \tag{2.3}$$

It can be shown that (2.3) has a proportional hazards structure if the covariates enter the model through (θ_1, θ_2). This is a desirable feature of the proposed model that leads to attractive theoretical and computational properties as discussed in Sections 4 and 5.

The joint cure fraction implied by (2.3) is $S_{\text{pop}}(\infty, \infty) = \exp(-[\theta_1 + \theta_2]^\alpha)$. From (2.3), the marginal survival functions are

$$S_k(t) = \exp(-\theta_k^\alpha (F_k(t))^\alpha), \qquad k = 1, 2. \tag{2.4}$$

Equation (2.4) indicates that the marginal survival functions have a cure rate structure with probability of cure $\exp(-\theta_k^\alpha)$ for T_k, $k = 1, 2$. It is important to note in (2.4) that each marginal survival function has a proportional hazards structure as long as the covariates, x, only enter through θ_k. The marginal hazard function is given by $\alpha\theta_k^\alpha f_k(t)(F_k(t))^{\alpha-1}$, with attenuated covariate effect $(\theta_k(x))^\alpha$, and $f_k(t)$ is the survival density corresponding to $F_k(t)$. This property is similar to the earlier observations made by Oakes (1989) for the ordinary bivariate stable frailty survival model.

In addition, we can express the marginal survival functions in (2.4) in terms of standard cure rate models. We can write

$$\begin{aligned} S_k(t) &= \exp(-\theta_k^\alpha (F_k(t))^\alpha) \\ &= \exp(-\theta_k^\alpha) + (1 - \exp(-\theta_k^\alpha)) \left(\frac{\exp(-\theta_k^\alpha (F_k(t))^\alpha) - \exp(-\theta_k^\alpha)}{1 - \exp(-\theta_k^\alpha)}\right) \\ &= \exp(-\theta_k^\alpha) + (1 - \exp(-\theta_k^\alpha))\, S_k^*(t), \end{aligned} \tag{2.5}$$

where

$$S_k^*(t) = \frac{\exp(-\theta_k^\alpha (F_k(t))^\alpha) - \exp(-\theta_k^\alpha)}{1 - \exp(-\theta_k^\alpha)},\ k = 1, 2.$$

It is easily shown that $S_k^*(t)$ defines a proper survivor function. Thus (2.5) is a standard cure rate model with cure rate given by $\pi_k = \exp(-\theta_k^\alpha)$ and survivor function for the non-cured population given by $S_k^*(t)$, for $k = 1, 2$.

The parameter α $(0 < \alpha < 1)$ is a scalar parameter that is a measure of association between (T_1, T_2). Small values of α indicate high association between (T_1, T_2). As $\alpha \to 1$, this implies less association between (T_1, T_2) which can be seen from (2.3). Following Clayton (1978) and Oakes (1989), we can compute a local measure of association, denoted, $\theta^*(t_1, t_2)$, as a function of α. This measure of association is defined as

$$\theta^*(t_1, t_2) = \frac{S_{\text{pop}}(t_1, t_2) \dfrac{\partial^2}{\partial t_1 \partial t_2} S_{\text{pop}}(t_1, t_2)}{\left(\dfrac{\partial}{\partial t_1} S_{\text{pop}}(t_1, t_2)\right)\left(\dfrac{\partial}{\partial t_2} S_{\text{pop}}(t_1, t_2)\right)}. \tag{2.6}$$

The measure in (2.6), introduced by Clayton (1978), has the interpretation of a ratio of conditional hazard rate of the conditional distribution of T_1, given $T_2 = t_2$, to that of T_1 given $T_2 > t_2$. For more discussion of (2.6), see Clayton (1978) and Oakes (1989). For the multivariate cure rate model in (2.3), $\theta^*(t_1, t_2)$ is well defined and is given by

$$\theta^*(t_1, t_2) = \alpha^{-1}(1-\alpha)(\theta_1 F_1(t_1) + \theta_2 F_2(t_2))^{-\alpha} + 1. \tag{2.7}$$

We see that $\theta^*(t_1, t_2)$ in (2.7) decreases in (t_1, t_2). That is, the association between (T_1, T_2) is greater when (T_1, T_2) are small and the association decreases over time. Such a property is desirable, for example, when T_1 denotes time to relapse and T_2 denotes time to death. 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 (2.3) since no moments for cure rate models exist due to the improper survival function.

3. THE LIKELIHOOD FUNCTION

Suppose we have n subjects, and let N_{ki} denote the number of latent risks for the ith subject, $i = 1, ..., n$, $k = 1, 2$. Further, we assume that the N_{ki}'s are independent Poisson random variables with mean $w_i\theta_k$, $i = 1, ..., n$,

$k=1, 2$. We also assume the $w_i \sim \text{Stable}(\alpha)$, and the w_i's are *i.i.d.* We emphasize here that the N_{ki}'s are not observed, and can be viewed as latent variables in our model formulation. Further, suppose $Z_{ki1}, ..., Z_{ki, N_{ki}}$ are the independent latent times for the N_{ki} latent risks for the ith subject, which are unobserved, and all have cumulative distribution function $F_k(.)$, $i=1, ..., n$, $k=1, 2$. In this paper, we will specify a parametric form for $F_k(.)$, such as a Weibull or gamma distribution. We denote the indexing parameter (possibly vector valued) by γ_k, and thus write $F_k(.\,|\,\gamma_k)$ and $S_k(.\,|\,\gamma_k)$. For example, if $F_k(.\,|\,\gamma_k)$ corresponds to a Weibull distribution, then $\gamma_k=(\xi_k, \lambda_k)$, where ξ_k is the shape parameter and λ_k is the scale parameter. Let t_{ki} denote the failure time for subject i for the kth component, where t_{ki} may be right censored. Let c_{ki} denote the censoring time so that we observe $y_{ki}=\min(t_{ki}, c_{ki})$, where the censoring indicator $\delta_{ki}=I(t_{ki}\leqslant c_{ki})$ equals 1 if t_{ki} is a failure time and 0 if it is right censored. Let $y_k=(y_{k1}, ..., y_{kn})$, $\delta_k=(\delta_{k1}, ..., \delta_{kn})$, $N_k=(N_{k1}, ...N_{kn})$, $k=1, 2$, and $w=(w_1, ..., w_n)$. The "complete data" is given by $D=(n, y_1, y_2, \delta_1, \delta_2, N_1, N_2, w)$, where N_1, N_2, and w are unobserved random vectors, and the observed data is given by $D_{\text{obs}}=(n, y_1, y_2, \delta_1, \delta_2)$. Further, let $\theta=(\theta_1, \theta_2)$ and $\gamma=(\gamma_1, \gamma_2)$. The likelihood function of (θ, γ) based on the complete data D is given by

$$L(\theta, \gamma\,|\,D)=\left(\prod_{k=1}^{2}\prod_{i=1}^{n} S_k(y_{ki}\,|\,\gamma_k)^{N_{ki}-\delta_{ki}}(N_{ki}f_k(y_{ki}\,|\,\gamma_k))^{\delta_{ki}}\right)$$
$$\times\exp\left\{\sum_{i=1}^{n}(N_{ki}\log(w_i\theta_k)-\log(N_{ki}!)-w_i\theta_k)\right\}, \quad (3.1)$$

where $f_k(y_{ki}\,|\,\gamma_k)$ is the density corresponding to $F_k(y_{ki}\,|\,\gamma_k)$. Throughout the remainder of this paper, we shall assume a Weibull density for $f_k(y_{ki}\,|\,\gamma_k)$, so that

$$f_k(y\,|\,\gamma_k)=\xi_k y^{\xi_k-1}\exp\{\lambda_k-y^{\xi_k}\exp(\lambda_k)\}. \quad (3.2)$$

To construct the likelihood function of the observed data, we integrate (3.1) with respect to (N, w) assuming a Stable(α) density for each w_i, denoted by $f_s(w_i\,|\,\alpha)$. We refer the reader to Section 5 for an explicit expression for the probability density $f_s(w_i\,|\,\alpha)$. We are led to the following theorem.

THEOREM 3.1. *The likelihood function based on the observed data, denoted* $L(\theta, \gamma, \alpha\,|\,D_{\text{obs}})$, *is given by*

$$
\begin{aligned}
L(\theta, \gamma, \alpha \mid D_{\text{obs}}) &\equiv \int_{R^{+n}} L(\theta, \gamma \mid D) \times \left[\prod_{i=1}^{n} f_s(w_i \mid \alpha) \right] dw \\
&= \theta_1^{d_1} \theta_2^{d_2} \alpha^{d_1 + d_2} \left[\prod_{k=1}^{2} \prod_{i=1}^{n} f_k(y_{ki} \mid \gamma_k)^{\delta_{ki}} \right] \\
&\quad \times \prod_{i=1}^{n} \{ [\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2)]^{(\alpha - 1)(\delta_{1i} + \delta_{2i})} \} \\
&\quad \times \prod_{i=1}^{n} [\alpha^{-1}(1-\alpha)(\theta_1 F_1(y_{1i} \mid \gamma_1) \\
&\quad + \theta_2 F_2(y_{2i} \mid \gamma_2))^{-\alpha} + 1]^{\delta_{1i}\delta_{2i}} \\
&\quad \times \prod_{i=1}^{n} \exp\{ -(\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2))^{\alpha} \}, \qquad (3.3)
\end{aligned}
$$

where $f_s(w_i \mid \alpha)$ *denotes the probability density function of* w_i, $d_k = \sum_{i=1}^{n} \delta_{ki}$ *for* $k = 1, 2$, $R^{+n} = R^+ \times R^+ \times \cdots \times R^+$, *and* $R^+ = (0, \infty)$.

The proof is technical and is given in the Appendix.

We incorporate covariates for the cure rate model (2.3) through the cure rate parameter θ. When covariates are included, we have a different cure rate parameter, θ_{ki}, for each subject, $i = 1, ..., n$. Let $x_i' = (x_{i1}, ..., x_{ip})$ denote the $p \times 1$ vector of covariates for the ith subject, and let $\beta_k = (\beta_{k1}, ..., \beta_{kp})'$ denote the corresponding vector of regression coefficients for the failure time random variable T_k, $k = 1, 2$. We relate θ to the covariates by $\theta_{ki} \equiv \theta(x_i'\beta_k) = \exp(x_i'\beta_k)$, so that the cure rate for subject i is $\exp(-\theta_{ki}) = \exp(-\exp(x_i'\beta_k))$, $i = 1, ..., n$, $k = 1, 2$. This relationship between θ_{ki} and β_k is equivalent to a canonical link for θ_{ki} in the setting of generalized linear models. Letting $\beta = (\beta_1, \beta_2)$, we can write the observed data likelihood of (β, γ, α) as

$$
\begin{aligned}
L(\beta, \gamma, \alpha \mid D_{\text{obs}}) &= \left(\alpha^{d_1 + d_2} \prod_{k=1}^{2} \prod_{i \in \mathscr{D}_k} \exp(x_i'\beta_k) \right) \left[\prod_{k=1}^{2} \prod_{i=1}^{n} f_k(y_{ki} \mid \gamma_k)^{\delta_{ki}} \right] \\
&\quad \times \prod_{i=1}^{n} \{ [\exp(x_i'\beta_1)\, F_1(y_{1i} \mid \gamma_1) \\
&\quad + \exp(x_i'\beta_2)\, F_2(y_{2i} \mid \gamma_2)]^{(\alpha - 1)(\delta_{1i} + \delta_{2i})} \} \\
&\quad \times \prod_{i=1}^{n} [\alpha^{-1}(1-\alpha)(\exp(x_i'\beta_1)\, F_1(y_{1i} \mid \gamma_1) \\
&\quad + \exp(x_i'\beta_2)\, F_2(y_{2i} \mid \gamma_2))^{-\alpha} + 1]^{\delta_{1i}\delta_{2i}} \\
&\quad \times \prod_{i=1}^{n} \exp\{ -(\exp(x_i'\beta_1)\, F_1(y_{1i} \mid \gamma_1) \\
&\quad + \exp(x_i'\beta_2)\, F_2(y_{2i} \mid \gamma_2))^{\alpha} \}, \qquad (3.4)
\end{aligned}
$$

where $\mathscr{D}_k$ consists of those patients who failed according to T_k, $k=1, 2$, $D_{\text{obs}}=(n, y_1, y_2, X, \delta_1, \delta_2)$, X is the $n\times p$ matrix of covariates, $f_k(y_{ki}\,|\,\gamma_k)$ is given by (3.2) and $S_k(y_{ki}\,|\,\gamma_k)=\exp(-y_{ki}^{\xi_k}\exp(\lambda_k))$.

4. THE PRIOR AND POSTERIOR DISTRIBUTIONS

In this section, we propose a class of noninformative improper priors that guarantee the propriety of the joint posterior distribution. Clearly, if proper priors are specified for all parameters, then proper posterior distributions result, but the chore of informative prior elicitation is then required. For the proposed class of models, this can be a monumental task, and we do not discuss such priors here. Noninformative priors serve as a convenient device for doing Bayesian inference for this class of models. We consider a joint improper prior for $(\beta, \gamma)=(\beta_1, \beta_2, \gamma_1, \gamma_2)$ of the form

$$\begin{aligned}\pi(\beta, \gamma, \alpha) &\equiv \pi(\beta_1, \beta_2, \gamma_1, \gamma_2, \alpha) \propto \pi(\gamma_1)\,\pi(\gamma_2)\\ &= I(0<\alpha<1)\prod_{k=1}^{2}\pi(\xi_k, \lambda_k),\end{aligned} \tag{4.1}$$

where $I(0<\alpha<1)=1$ if $0<\alpha<1$, and 0 otherwise. Thus, (4.1) implies that β, γ, and α are independent a priori, (β_1, β_2) are independent a priori with an improper uniform prior, α has a proper uniform prior over the interval $(0, 1)$, and (γ_1, γ_2) are independent and identically distributed as $\pi(\gamma_k)$ a priori. We will assume throughout that

$$\pi(\xi_k, \lambda_k)=\pi(\xi_k\,|\,v_0, \tau_0)\,\pi(\lambda_k),$$

where

$$\pi(\xi_k\,|\,v_0, \tau_0)\propto \xi_k^{v_0-1}\exp\{-\tau_0\xi_k\}, \qquad \text{and} \qquad \pi(\lambda_k)\propto \exp\{-c_0\lambda_k^2\}\ ,$$

and v_0, τ_0, and c_0 are specified hyperparameters. With these specifications, the posterior distribution of (β, γ, α) based on the observed data $D_{\text{obs}}=(n, y_1, y_2, X, \delta_1, \delta_2)$ is given by

$$p(\beta, \gamma, \alpha\,|\,D_{\text{obs}})\propto L(\beta, \gamma, \alpha\,|\,D_{\text{obs}})\prod_{k=1}^{2}\pi(\xi_k\,|\,v_0, \tau_0)\,\pi(\lambda_k), \tag{4.2}$$

where $L(\beta, \gamma, \alpha\,|\,D_{\text{obs}})$ is given by (3.4). We are led to the following theorem concerning the propriety of the posterior distribution in (4.2) using the noninformative improper prior (4.1).

THEOREM 4.1. *Let* X_k^* *be an* $n \times p$ *matrix with rows* $\delta_{ki} x'_{ki}$ *for* $k = 1, 2,$ *and* $\psi = (\beta_1, \beta_2, \xi_1, \xi_2, \lambda_1, \lambda_2, \alpha)$. *Then if* (C1) X_k^* *is of full rank for* $k = 1, 2,$ (C2) $\pi(\lambda_k)$ *is proper, and* (C3) $\tau_0 > 0$ *and* $v_0 > -\min\{d_1, d_2\}$, *the posterior given in* (4.2) *is proper; that is,*

$$\int L(\beta, \gamma, \alpha \mid D_{\text{obs}}) \left[\prod_{k=1}^{2} \pi(\xi_k \mid v_0, \tau_0)\, \pi(\lambda_k) \right] d\psi < \infty. \tag{4.3}$$

The proof of Theorem 4.1 is quite technical and is given in the appendix. Note that the conditions given in Theorem 4.1 are sufficient but *not* necessary for the propriety of the posterior distribution. However, the conditions stated in the theorem are quite general and typically satisfied for most data sets. We also notice that a proper prior for ξ_k is not required in order to obtain a proper posterior. This can be observed from condition (C3) because $\pi(\xi_k \mid v_0, \tau_0)$ is no longer proper when $v_0 < 0$. We note that Theorem 4.1 only requires that $\pi(\lambda_k)$ be any proper prior. Although several choices can be made, we will take independent normal densities for $\pi(\lambda_k)$, $k = 1, 2,$ in the remainder of this paper.

5. COMPUTATIONAL IMPLEMENTATION

In this section, we propose a modified version of the collapsed Gibbs technique of Liu (1994) to sample from the posterior distribution. This technique results in an efficient Gibbs sampling scheme which reduces the correlations between the parameters and the latent variables. As a by-product of our overall methodology, we develop a Gibbs sampling scheme for positive stable law distributions. We note here that MCMC methods for multivariate survival data have also been examined by Qiou *et al.* (1999).

From (2.2), it can be shown that $f_s(w \mid \alpha)$ is obtained by marginalizing, with respect to u, the joint density

$$f(w, u \mid \alpha) = a w^{-(a+1)} s(u) \exp\left\{ -\frac{s(u)}{w^a} \right\}, \qquad w > 0, \qquad 0 < u < 1. \tag{5.1}$$

This relationship plays an important role in the implementation of the Gibbs sampler.

To facilitate the Gibbs sampler, we introduce several auxiliary (latent) variables. We note here that Gibbs sampling using auxiliary variables has

been used by many in the Bayesian literature, including Besag and Green (1993), and Higdon (1998). The auxiliary variables are $N=(N_1, N_2)$, where $N_k=(N_{k1}, ..., N_{kn})$ for $k=1, 2$, $w=(w_1, w_2, ..., w_n)$, and $u=(u_1, u_2, ..., u_n)$. The joint posterior distribution of $(\beta, \gamma, \alpha, N, w, u \mid D_{\text{obs}})$ is given by

$$
\begin{aligned}
p(\beta, \gamma, \alpha, N, w, u \mid D_{\text{obs}}) \propto & \left(\prod_{k=1}^{2} \prod_{i=1}^{n} S_k(y_{ki} \mid \gamma_k)^{N_{ki}-\delta_{ki}} (N_{ki} f_k(y_{ki} \mid \gamma_k))^{\delta_{ki}} \right) \\
& \times \exp\left\{ \sum_{i=1}^{n} (N_{ki} \log(w_i \theta_{ki}) - \log(N_{ki}!) - w_i \theta_{ki}) \right\} \\
& \times \prod_{i=1}^{n} \left[w_i^{-(a+1)} s(u_i) \exp\left\{ -\frac{s(u_i)}{w_i^a} \right\} \right] \\
& \times \prod_{k=1}^{2} (\pi(\xi_k \mid v_0, \tau_0)\, \pi(\lambda_k)), \qquad (5.2)
\end{aligned}
$$

where $\theta_{ki}=\exp(x_i'\beta_k)$, $v_0 > -\min\{d_1, d_2\}$, $\tau_0 > 0$, and $c_0 > 0$. To run the Gibbs sampler, we need to sample from the following conditional distributions: $[\gamma \mid \beta, \alpha, N, w, u, D_{\text{obs}}]$ and $[\beta, \alpha, N, w, u \mid \gamma, D_{\text{obs}}]$.

The conditional posterior density for $[\gamma \mid \beta, \gamma, \alpha, N, u, D_{\text{obs}}]$ is given by

$$
\begin{aligned}
p(\gamma \mid \beta, \gamma, \alpha, N, u, D_{\text{obs}}) \propto & \prod_{k=1}^{2} \xi_k^{d_k + v_0 - 1} \exp\left\{ d_k \lambda_k + \sum_{i=1}^{n} [\delta_{ki} \xi_k \log(y_{ki})] \right. \\
& \left. - N_{ki} e^{\lambda_k} y_{ki}^{\xi_k} - \tau_0 \xi_k - c_0 \lambda_k^2 \right\}. \qquad (5.3)
\end{aligned}
$$

Using a similar proof given by Berger and Sun (1993), we can show that $p(\gamma \mid \beta, \gamma, \alpha, N, u, D_{\text{obs}})$ is log-concave in ξ_k or λ_k for $k=1, 2$. Thus, the adaptive rejection algorithm of Gilks and Wild (1992) can be used here to sample γ.

Sampling from $[\beta, \alpha, N, w, u \mid \gamma, D_{\text{obs}}]$ is the most challenging and expensive part of this algorithm. Sampling from the five complete conditional distributions may result in high correlations between (β, α, N, w, u) due to the high dimension of the latent vectors. To remedy this potential problem, we apply the collapsed Gibbs procedure of Liu (1994). It is easy to observe that

$$
[\beta, \alpha, N, w, u \mid \gamma, D_{\text{obs}}] = [\beta, \alpha, w, u \mid \gamma, D_{\text{obs}}][N \mid \beta, \alpha, w, u, \gamma, D_{\text{obs}}]. \qquad (5.4)
$$

In (5.4), we draw (β, α, w, u) by collapsing N, which is crucial for achieving convergence of our MCMC algorithm. For $[\beta, \alpha, w, u \mid \gamma, D_{\text{obs}}]$, we draw $[\beta \mid \alpha, w, u, \gamma, D_{\text{obs}}]$ and $[\alpha, w, u \mid \beta, \gamma, D_{\text{obs}}]$.

Using the proof of Theorem 3.1, the density of $[\beta \mid \alpha, w, u, \gamma, D_{\text{obs}}]$ is given by

$$p(\beta \mid \alpha, w, u, \gamma, D_{\text{obs}}) \propto \exp\left\{\sum_{k=1}^{2}\sum_{i=1}^{n}[\delta_{ki}x_i'\beta_k - w_i F_k(y_{ki} \mid \gamma_k)\exp(x_i'\beta_k)]\right\}. \quad (5.5)$$

It is easy to see that $p(\beta \mid \alpha, w, u, \gamma, D_{\text{obs}})$ is log-concave in each component of β and thus we can use the adaptive rejection algorithm of Gilks and Wild (1992) to draw β.

To draw $[\alpha, w, u \mid \beta, \gamma, D_{\text{obs}}]$, we use the collapsed Gibbs procedure one more time. That is, we draw α from $[\alpha \mid \beta, \gamma, D_{\text{obs}}]$ by collapsing w and u, and then draw (w, u) from $[w, u \mid \alpha, \beta, \gamma, D_{\text{obs}}]$.

The conditional posterior density for $[\alpha \mid \beta, \gamma, D_{\text{obs}}]$ can be written as

$$p(\alpha \mid \beta, \gamma, D_{\text{obs}}) \propto L(\beta, \gamma, \alpha \mid D_{\text{obs}}), \quad (5.6)$$

where $L(\beta, \gamma, \alpha \mid D_{\text{obs}})$ is given by (3.4). Generating α from (5.6) is not trivial since $p(\alpha \mid \beta, \gamma, D_{\text{obs}})$ is not log-concave. Therefore, we consider the following Metropolis–Hastings algorithm with a "de-constraint" transformation to draw α. Since $0 < \alpha < 1$, we let

$$\alpha = \frac{e^{\eta}}{1+e^{\eta}}, \qquad -\infty < \eta < \infty. \quad (5.7)$$

Then

$$p(\eta \mid \beta, \gamma, D_{\text{obs}}) = p(\alpha \mid \beta, \gamma, D_{\text{obs}})\frac{e^{\eta}}{(1+e^{\eta})^2}.$$

Instead of directly sampling α, we generate η by choosing a normal proposal $N(\hat{\eta}, \hat{\sigma}_{\hat{\eta}}^2)$, where $\hat{\eta}$ is the maximizer of the logarithm of $p(\eta \mid \beta, \gamma, D_{\text{obs}})$ and $\hat{\sigma}_{\hat{\eta}}^2$ is the minus of the inverse of the second derivative of $\log p(\eta \mid \beta, \gamma, D_{\text{obs}})$ evaluated at $\eta = \hat{\eta}$; that is,

$$\hat{\sigma}_{\hat{\eta}}^{-2} = -\frac{d^2 \log p(\eta \mid \beta, \gamma, D_{\text{obs}})}{d\eta^2}\bigg|_{\eta=\hat{\eta}}.$$

The algorithm to generate η operates as follows: (a) let η be the current value; (b) generate a proposal value η^* from $N(\hat{\eta}, \hat{\sigma}^2_{\hat{\eta}})$; and (c) a move from η to η^* is made with probability

$$\min\left\{\frac{p(\eta^* \mid \beta, \gamma, D_{\text{obs}})\,\phi\left(\dfrac{\eta-\hat{\eta}}{\hat{\sigma}_{\hat{\eta}}}\right)}{p(\eta \mid \beta, \gamma, D_{\text{obs}})\,\phi\left(\dfrac{\eta^*-\hat{\eta}}{\hat{\sigma}_{\hat{\eta}}}\right)}, 1\right\},$$

where ϕ is the standard normal probability density function. After we obtain η, we compute α by using (5.7).

Following the proof of Theorem 4.1, the joint conditional density for (w, u) is given by

$$p(w, u \mid \alpha, \beta, \gamma, D_{\text{obs}}) \propto \prod_{i=1}^{n} w_i^{\delta_i} \exp\left\{-w_i \sum_{k=1}^{2} \exp(x_i'\beta_k)(1 - S_k(y_{ki} \mid \gamma_k))\right\} \times w_i^{-(a+1)} s(u_i) \exp\left\{-\frac{s(u_i)}{w_i^a}\right\}, \tag{5.8}$$

where $\delta_i = \delta_{1i} + \delta_{2i}$ and $a = \alpha/(1-\alpha)$. Now, we use the ratio of uniforms (ROU) method and a rejection algorithm (for example, see Devroye, 1986, pp. 40–65, 194–205) to draw (w_i, u_i) for $i = 1, 2, ..., n$. More specifically, the ROU algorithm for drawing w_i requires the following steps:

(i) Compute $a^* = \sup(p^*(w_i \mid u_i, \alpha, \beta, \gamma, D_{\text{obs}}))^{1/2}$ and $b^* = \sup w_i(p^*(w_i \mid u_i, \alpha, \beta, \gamma, D_{\text{obs}}))^{1/2}$, where

$$p^*(w_i \mid u_i, \alpha, \beta, \gamma, D_{\text{obs}}) = w_i^{\delta_i} \exp\left\{-w_i \sum_{k=1}^{2} \exp(x_i'\beta_k)(1 - S_k(y_{ki} \mid \gamma_k))\right\} \times w_i^{-(a+1)} \exp\left\{-\frac{s(u_i)}{w_i^a}\right\}. \tag{5.9}$$

(ii) Draw ζ from $U(0, a^*)$ and ω from $U(0, b^*)$.

(iii) Return $w_i = \zeta/\omega$ if $\zeta^2 \leqslant p^*(\zeta/\omega \mid u_i, \alpha, \beta, \gamma, D_{\text{obs}})$; otherwise, go to (ii).

The rejection algorithm for sampling u_i operates as follows:

(i) Independently generate u_i and v from $U(0, 1)$.

(ii) Return u_i if $v \leqslant \frac{s(u_i)}{w_i^a} \exp\{-\frac{s(u_i)}{w_i^a}\}$; otherwise, go to (i).

Finally, we draw N from $[N \mid \beta, \alpha, w, u, \gamma, D_{\text{obs}}]$. Since

$$N_{ki} \mid \beta, \alpha, w, u, \gamma, D_{\text{obs}} \sim \text{Poisson}(w_i S_k(y_{ki} \mid \gamma_k) \exp(x_i'\beta_k)) + \delta_{ki}$$

for $k = 1, 2$ and $i = 1, ..., n$, sampling N from its conditional posterior distribution is trivial.

The introduction of latent variables indeed converts an intractable and nearly impossible computational problem (which involves direct sampling from the posterior with the likelihood based on the observed data given in (3.4)), into an attractive one, in which the parameters are sampled from a posterior based on the complete-data likelihood. Throughout the entire MCMC implementational scheme, we use the collapsed Gibbs sampling technique of Liu (1994). Thus, instead of sampling α directly from its conditional distribution $\pi(\alpha \mid w, u, \beta, \gamma, \alpha, D_{\text{obs}})$ as in Buckle (1995), we sample α from its marginal posterior distribution $\pi(\alpha \mid \beta, \gamma, \alpha, D_{\text{obs}})$. Similarly, we draw from $\pi(\beta \mid w, u, \gamma, \alpha, D_{\text{obs}})$ instead of $\pi(\beta \mid N, w, u, \gamma, \alpha, D_{\text{obs}})$. By doing these two steps, we reduce the intra-correlations between α and (w, u), and β and N, respectively. Therefore the convergence of the induced Markov chain is improved.

6. ILLUSTRATIVE EXAMPLE

To illustrate the methodology, we consider data from a phase III melanoma clinical trial conducted by the Eastern Cooperative Oncology Group (ECOG). The study, denoted E1684, was a two-arm clinical trial involving patients randomized to one of two treatment arms: high-dose interferon (IFN) or observation. The results of this study (see Kirkwood *et al.*, 1996) suggested that IFN has a significant impact on time to relapse and time to death, which led to FDA approval of this regimen as a standard adjuvant therapy for high risk melanoma patients. Our purpose in this example is to illustrate the proposed multivariate cure rate model in (2.3) and demonstrate several of its properties. We emphasize here that our proposed model is valid only when sufficient follow-up is available on all of the time-to-event endpoints and the calendar date of entry is assumed to be noninformative on the outcome variables.

We consider the two failure-time random variables, $T_1 =$ time to relapse from randomization, and $T_2 =$ relapse to death. We note that all of the patients who died in this study had also relapsed. Three covariates and an intercept are included in the model. The covariates are age (x_1), sex (x_2) (male, female), and performance status (x_3) (fully active, other). Performance status is abbreviated by PS in the tables below. Tables I and II give statistical summaries for (T_1, T_2) and the covariates, respectively. A total of $n = 274$ observations are used in the analysis. In all of the

TABLE I

Summary of Survival and Relapse Times

Time to Relapse (T_1) (years)		Status (frequency)		Relapse to Death (T_2) (years)		Status (frequency)	
median	0.537	censored	88	median	0.660	censored	110
IQR	1.247	relapse	186	IQR	1.014	death	164

computations, we standardized all the covariates to have mean 0 and standard deviation 1 in order to improve the convergence of the MCMC algorithm. Specifically, standardizing the covariates greatly reduces the correlation between the intercept term and the other regression coefficients. We use the noninformative improper prior in (4.1), with $\pi(\beta) \propto 1$, $\lambda_k \sim N(0, 10, 000)$, $\xi_k \sim \text{gamma}(1, 0.01)$, and independent for each $k = 1, 2$. Also, we take a uniform prior for α on the interval $(0, 1)$. In this example, 50,000 MCMC iterations were used in all of the computations after a burn-in of 1,000 iterations. Convergence was checked using the methods discussed in Cowles and Carlin (1996). Specifically, trace plots, autocorrelations, and Gelman–Rubin statistics (Gelman and Rubin, 1992) were computed, and convergence was observed to occur before 500 iterations.

Table III gives posterior estimates of $\beta = (\beta_1, \beta_2)$, $\xi = (\xi_1, \xi_2)$, $\lambda = (\lambda_1, \lambda_2)$, and α, where $\beta_k = (\beta_{k0}, \beta_{k1}, \beta_{k2}, \beta_{k3})'$, $k = 1, 2$. We see from Table 3 that all of the highest posterior density (HPD) intervals for the regression coefficients of the covariates contain 0. Also, from Table 3, we see that the posterior mean of α is 0.709, with a 95% HPD interval of (0.585, 0.840). As discussed in Section 2, this indicates a moderate association between time to relapse and relapse to death for these data, as was expected. A plot of the marginal posterior distribution of α is given in Fig. 1. We see in this figure that the posterior distribution of α appears quite symmetric with a mode at 0.699. Fig. 2 shows a box plot of the posterior means of the cure rates for each failure time variable. We note that when covariates are included in the model, each subject has an individual cure

TABLE II

Summary of Covariates

Age (x_1) (years)		Gender (x_2) (frequency)		PS (x_3) (frequency)	
mean	46.663	Male	165	Fully Active	243
SD	12.818	Female	109	Other	31

TABLE III

Posterior Estimates

parameter	mean	SD	95% HPD interval
β_{10}	0.234	0.116	(0.004, 0.459)
β_{11}	0.072	0.101	(−0.122, 0.274)
β_{12}	0.008	0.104	(−0.202, 0.206)
β_{13}	0.020	0.105	(−0.191, 0.222)
β_{20}	0.922	0.170	(0.585, 1.256)
β_{21}	0.147	0.116	(−0.077, 0.383)
β_{22}	−0.195	0.122	(−0.434, 0.041)
β_{23}	−0.199	0.121	(−0.435, 0.039)
α	0.709	0.066	(0.585, 0.840)
ξ_1	1.258	0.101	(1.064, 1.457)
ξ_2	1.496	0.124	(1.253, 1.737)
λ_1	−0.852	0.178	(−1.209, −0.516)
λ_2	−1.421	0.240	(−1.895, −0.966)

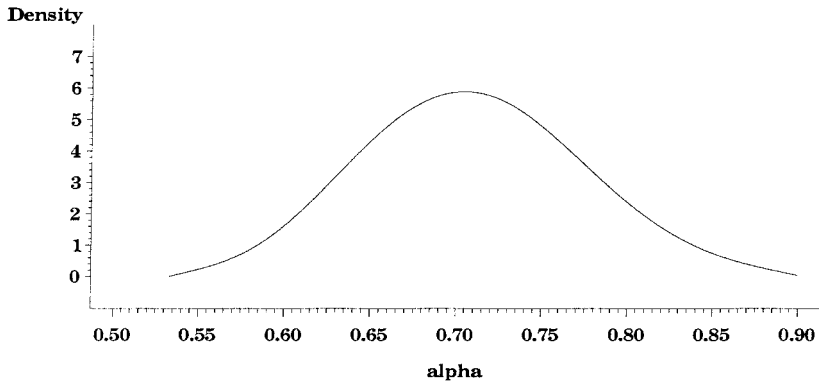

FIG. 1. The marginal posterior density of α.

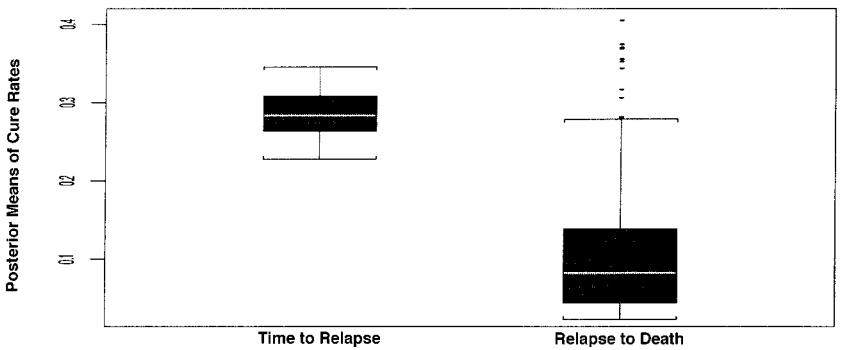

FIG. 2. Box plots of the posterior means of the cure rates for all patients.

TABLE IV

Summary of Box Plots

Failure time	mean	SD	median	IQR	min	max
time to relapse	0.285	0.026	0.283	0.045	0.228	0.346
relapse to death	0.103	0.078	0.082	0.094	0.023	0.405

rate. From Fig. 2, we see that the median cure rate for time to relapse (0.285) is much higher than the median cure rate for relapse to death (0.103). In general, there is much more variability in the estimated cure rates for the relapse to death variable. Table IV gives numerical summaries for both box plots of Fig. 2. In Table 4, IQR denotes interquartile range.

Figure 3 shows two superimposed plots, where plot (a) represents time to relapse and plot (b) represents relapse to death. The covariates are not used in constructing plots (a) and (b). In plot (a), the two superimposed plots correspond to the Kaplan–Meier estimate of survival and the maximum likelihood estimate of the marginal survival function based on the multivariate cure rate model. We see that the two curves in plot (a) are nearly identical and appear to plateau after approximately 6 years of follow-up. In plot (b), the relapse to death variable appears to plateau after approximately 4 years of follow-up. Figure 4 shows a three-dimensional plot of the posterior mean survival surface based on average age for males with fully active performance status. We see in this plot how the survival curve plateaus for each failure time variable. The joint survival function approaches a joint cure fraction, and the marginal survival functions each approach a cure fraction. From this figure, it is clear that the estimated cure rate for the time to relapse variable is larger than the estimated cure rate for the relapse to death variable.

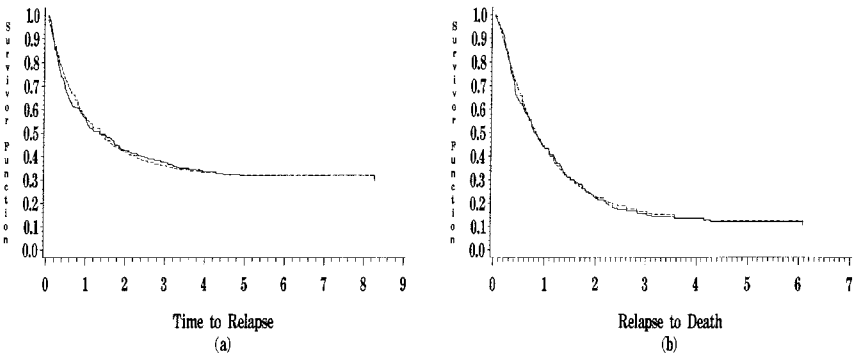

FIG. 3. Superimposed survival curves.

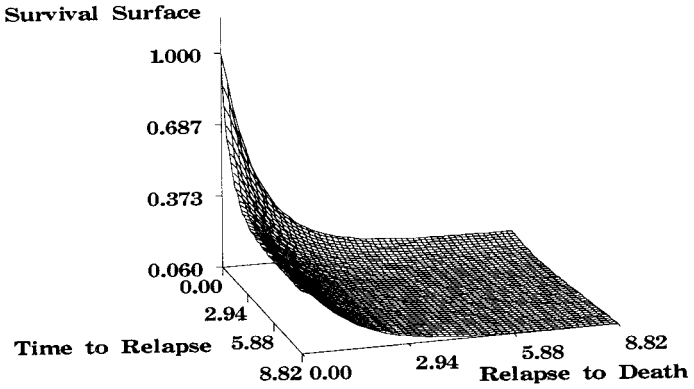

FIG. 4. The bivariate posterior survival surface.

Finally, we compared the individual fits of the univariate models for T_1 and T_2 to the marginal models induced by the bivariate cure rate model. The 95% HPD intervals for the univariate model were all narrower than those based on the multivariate model. For example, for T_1, the 95% HPD interval for β_1 was $(-0.083, 0.199)$, which is narrower than the 95% HPD interval for β_1 given in Table II. In general, there is no general trend to these HPD intervals. The width of the HPD intervals for the multivariate model depends on the frailty distribution and the data, and therefore these intervals can be narrower or wider than intervals based on the corresponding univariate model. Therefore, in general, it becomes difficult to assess the efficiency in the multivariate model, since the width of the HPD intervals and posterior standard deviations heavily depend on the frailty distribution and the dataset at hand.

7. DISCUSSION

We have proposed a new multivariate cure rate model and have examined several of its properties. This model is useful for jointly modelling any type of failure time data with a surviving fraction. We emphasize here that cure rate models should only be used when sufficient follow-up is available on all of the time-to-event endpoints. This is a critical practical issue that arises with the use of any cure rate model. We never recommend using a cure rate model when there is insufficient follow-up on the patients and/or when there is heavy censoring. The type of follow-up necessary certainly depends on the application and/or disease type, and thus we cannot give specific recommendations here, but only provide caution. In Section 6, we demonstrated that the model is computationally feasible with several

covariates included. We observed from Fig. 3 how well the proposed model matches the Kaplan–Meier non-parametric estimate of the marginal survival function. For ease and clarity of exposition, we have focused our development on the bivariate cure rate model. Extensions to the general multivariate case are quite straightforward, as all of the methodology and the theorems given remain valid in the general multivariate case. We mention here that model-checking techniques need to to be developed to investigate the fit of the model in Section 6. This is a very important issue for these types of highly parametric models. Future work with this model includes developing methods for hypothesis testing, model selection, and model adequacy.

APPENDIX: THE PROOFS OF THE THEOREMS

Proof of the Theorem 3.1

Throughout the derivation, we make use of the fact that the Laplace transform of w, where $w \sim \text{Stable}, (\alpha)$ is,

$$E(\exp(-sw)) = \exp(-s^{\alpha}). \tag{A.1}$$

To prove Theorem 3.1, we need to derive an expression of the likelihood after summing $L(\beta,\gamma \mid D)$ over the possible values of (N_{1i}, N_{2i}). From (3.1), it is easy to see that given w_i, N_{1i} and N_{2i} are conditionally independent. Thus, after summing over the possible values of (N_{1i}, N_{2i}), we obtain

$$L(\beta, \gamma \mid w, D_{\text{obs}}) = \prod_{k=1}^{2} \left[\prod_{i=1}^{n} (w_i \theta_k f_k(y_{ki} \mid \gamma_k))^{\delta_{ki}} \times \exp\{-w_i \theta_k (1 - S_k(y_{ki} \mid \gamma_k))\} \right]. \tag{A.2}$$

Simplification of (A.2) yields

$$L(\beta, \gamma \mid w, D_{\text{obs}}) = \theta_1^{d_1} \theta_2^{d_2} \left[\prod_{k=1}^{2} \prod_{i=1}^{n} f_k(y_{ki} \mid \gamma_k)^{\delta_{ki}} \right] \times \prod_{i=1}^{n} [w_i^{\delta_i} \exp\{-w_i(\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2))\}], \tag{A.3}$$

where $\delta_i = \delta_{1i} + \delta_{2i}$ for $i = 1, 2, ..., n$. Using (A.3), the likelihood based on the observed data can be expressed as

$$L(\beta, \gamma, \alpha \mid D_{\text{obs}}) = \theta_1^{d_1} \theta_2^{d_2} \left[\prod_{k=1}^{2} \prod_{i=1}^{n} f_k(y_{ki} \mid \gamma_k)^{\delta_{ki}} \right] \times \prod_{i=1}^{n} \int_0^\infty w_i^{\delta_i} \exp\{ -w_i(\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2)) \} f_s(w_i \mid \alpha)\, dw_i. \quad \text{(A.4)}$$

Now, we consider the following three cases: (i) $\delta_i = 0$, (ii) $\delta_i = 1$, and (iii) $\delta_i = 2$.

For case (i), using (A.1), we integrate out w_i for the ith observation, leading to

$$L_{i0}(\theta, \gamma, \alpha \mid D_{\text{obs}}) = \int_0^\infty \exp(-w_i(\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2))) f_s(w_i \mid \alpha)\, dw_i = \exp\{ -[\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2)]^\alpha \}. \quad \text{(A.5)}$$

For case (ii), the contribution to the last term of the right side of (A.4) for the ith observation is given by

$$L_{i1}(\theta, \gamma, \alpha \mid D_{\text{obs}}) = \int_0^\infty w_i \exp(-w_i(\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2)) f_s(w_i \mid \alpha)\, dw_i . \quad \text{(A.6)}$$

Now, we note that (A.1) implies that $E[w_i \exp(-s w_i)] = \alpha s^{\alpha - 1} \exp(-s^\alpha)$, so that (A.6) equals

$$L_{i1}(\theta, \gamma, \alpha \mid D_{\text{obs}}) = \alpha(\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2))^{\alpha - 1} \times \exp(-[\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2)]^\alpha). \quad \text{(A.7)}$$

For case (iii), for the ith observation, we write

$$L_{i2}(\theta, \gamma, \alpha \mid D_{\text{obs}}) = \int_0^\infty w_i^2 \exp(-w_i(\theta_1 F_1(y_{1i} \mid \gamma_1) + \theta_2 F_2(y_{2i} \mid \gamma_2))) f_s(w_i \mid \alpha)\, dw_i. \quad \text{(A.8)}$$

From (A.1), we have

$$E(w_i^2 \exp(-s w_i)) = \alpha(1 - \alpha)\, s^{\alpha - 2} \exp(-s^\alpha) + \alpha^2 s^{2(\alpha - 1)} \exp(-s^\alpha) = \alpha^2 s^{2(\alpha - 1)} (\alpha^{-1}(1 - \alpha)\, s^{-\alpha} + 1) \exp(-s^\alpha). \quad \text{(A.9)}$$

Now letting $s=\theta_1 F_1(y_{1i}\,|\,\gamma_1)+\theta_2 F_2(y_{2i}\,|\,\gamma_2)$ in (A.9), we get the closed-form expression of $L_{i2}(\theta, \gamma, \alpha\,|\,D_{\text{obs}})$. Multiplying the terms, we get

$$\begin{aligned} L(\beta, \gamma, \alpha\,|\,D_{\text{obs}}) &= \theta_1^{d_1}\theta_2^{d_2}\left[\prod_{k=1}^{2}\prod_{i=1}^{n} f_k(y_{ki}\,|\,\gamma_k)^{\delta_{ki}}\right] \\ &= \prod_{i=1}^{n} L_{i0}^{1_{\{\delta_i=0\}}}(\theta, \gamma, \alpha\,|\,D_{\text{obs}}) \\ &\quad \times L_{i1}^{1_{\{\delta_i=1\}}}(\theta, \gamma, \alpha\,|\,D_{\text{obs}})\, L_{i2}^{1_{\{\delta_i=2\}}}(\theta, \gamma, \alpha\,|\,D_{\text{obs}}), \end{aligned}$$

where $1_{\{\delta_i=j\}}$ is the indicator function for $j=0, 1, 2$, and this coincides with Eq. (3.3). This completes the proof.

Proof of Theorem 4.1

To prove Theorem 4.1, it suffices to show that

$$\iint L(\beta, \gamma\,|\,w, D_{\text{obs}})\left[\prod_{i=1}^{n} f_s(w_i\,|\,\alpha)\right]\left[\prod_{k=1}^{2}\pi(\xi_k\,|\,v_0, \tau_0)\,\pi(\lambda_k)\right] d\psi\, dw < \infty, \tag{A.10}$$

where $L(\beta, \gamma\,|\,w, D_{\text{obs}})$ is given in (A.2) with θ_k being replaced by θ_{ki}.

We first prove

$$(w_i\theta_{ki} f_k(y_{ki}\,|\,\gamma_k))^{\delta_{ki}} \exp\{-w_i\theta_{ki}(1-S_k(y_{ki}\,|\,\gamma_k))\} \leqslant M\xi_k^{\delta_{ki}}, \tag{A.11}$$

where $M \geqslant 1$ is a constant. When $\delta_{ki}=0$, (A.11) is obviously true since $\exp\{-w_i\theta_{ki}(1-S_k(y_{ki}\,|\,\gamma_k))\} \leqslant 1$. For $\delta_{ki}=1$, the left side of (A.11) can be rewritten as

$$\begin{aligned} & y_{ki}^{-1}\,\frac{\xi_k\, y_{ki}^{\xi_k} e^{\lambda_k}\exp(-e^{\lambda_k} y_{ki}^{\xi_k})}{1-\exp(-e^{\lambda_k} y_{ki}^{\xi_k})} \\ & \quad \times [(1-S_k(y_{ki}\,|\,\gamma_k))\, w_i\theta_{ki}\exp(-w_i\theta_{ki}(1-S_k(y_{ki}\,|\,\gamma_k)))]. \end{aligned} \tag{A.12}$$

Let

$$g_1(z)=\frac{ze^{-z}}{1-e^{-z}}, \qquad g_2(z)=ze^{-z}, \qquad \text{for} \quad z>0.$$

Thus, there exists a common constant g_0 such that $1 \leqslant g_0 < \infty$,

$$g_1(z) \leqslant g_0, \qquad \text{and} \qquad g_2(z) \leqslant g_0 \qquad \text{for all} \quad z>0. \tag{A.13}$$

Using (A.13), (A.12) is less than or equal to $y_{ki}^{-1}\xi_k g_0^2$. Thus, taking $M^* = g_0^2 \max_{(k,i):\,\delta_{ki}=1}\{y_{ki}^{-1}\}$ and $M = \max\{1, M^*\}$, we obtain (A.11).

Next we prove (A.10). Since X_k^* is of full rank, there must exist p linearly independent row vectors $x'_{ki_{k1}}$, $x'_{ki_{k2}}$,..., $x'_{ki_{kp}}$ such that $\delta_{ki_{k1}} = \delta_{ki_{k2}} = \cdots = \delta_{ki_{kp}} = 1$. Using (A.11), the left side of (A.10) is less than or equal to

$$\int_0^1 \int_{R^{+n}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \left\{ \prod_{k=1}^{2} \left[M^{n-p} \xi_k^{d_k - p} \int_{R^p} \prod_{j=1}^{p} f_k(y_{ki_{kj}} \mid \gamma_k) \right.\right.$$

$$\left.\left. w_{i_{kj}} \exp\{x'_{ki_{kj}} \beta_k - (1 - S_k(y_{ki_{kj}} \mid \gamma_k))\, w_{i_{kj}} \exp(x'_{ki_{kj}} \beta_k)\} \right] d\beta_k \right\}$$

$$\left[\prod_{i=1}^{n} f_s(w_i \mid \alpha) \right] \left[\prod_{k=1}^{2} \pi(\xi_k \mid v_0, \tau_0)\, \pi(\lambda_k) \right] d\xi_1\, d\xi_2\, d\lambda_1\, d\lambda_2\, dw\, d\alpha. \quad \text{(A.14)}$$

Now we make the transformation $u_{kj} = x'_{ki_{kj}} \beta_k + \log(w_{i_{kj}})$ for $j = 1, 2, ..., p$. This is a one-to-one linear transformation from β_k to $u_k = (u_{k1}, ..., u_{kp})'$. Thus, (A.14) is proportional to

$$\int_0^1 \int_{R^{+n}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \left\{ \prod_{k=1}^{2} \left[\xi_k^{d_k - p} \int_{R^p} \prod_{j=1}^{p} f_k(y_{ki_{kj}} \mid \gamma_k) \right.\right.$$

$$\left.\left. \times \exp\{u_{kj} - (1 - S_k(y_{ki_{kj}} \mid \gamma_k)) \exp(u_{kj})\} \right] du_k \right\}$$

$$\left[\prod_{i=1}^{n} f_s(w_i \mid \alpha) \right] \left[\prod_{k=1}^{2} \pi(\xi_k \mid v_0, \tau_0)\, \pi(\lambda_k) \right] d\xi_1\, d\xi_2\, d\lambda_1\, d\lambda_2\, dw\, d\alpha. \quad \text{(A.15)}$$

Integrating out u_k, (A.15) reduces to

$$\int_0^1 \int_{R^{+n}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \left\{ \prod_{k=1}^{2} \xi_k^{d_k - p} \prod_{j=1}^{p} \frac{f_k(y_{ki_{kj}} \mid \gamma_k)}{1 - S_k(y_{ki_{kj}} \mid \gamma_k)} \right\}$$

$$\left[\prod_{i=1}^{n} f_s(w_i \mid \alpha) \right] \left[\prod_{k=1}^{2} \pi(\xi_k \mid v_0, \tau_0)\, \pi(\lambda_k) \right] d\xi_1\, d\xi_2\, d\lambda_1\, d\lambda_2\, dw\, d\alpha. \quad \text{(A.16)}$$

In (A.16), using (A.13), we have

$$\frac{f_k(y_{ki_{kj}} \mid \gamma_k)}{1 - S_k(y_{ki_{kj}} \mid \gamma_k)} \leqslant M\xi_k.$$

Thus, (A.16) is less than or equal to

$$M^p \prod_{k=1}^{2} \left[\int_0^\infty \xi_k^{d_k} \pi(\xi_k \mid \nu_0, \tau_0)\, d\xi_k \int_{-\infty}^{\infty} \pi(\lambda_k)\, d\lambda_k \right] \times \int_0^1 \left[\prod_{i=1}^{n} \int_0^\infty f_s(w_i \mid \alpha)\, dw_i \right] d\alpha < \infty \tag{A.17}$$

by conditions (C2) and (C3) and the fact that $f_s(w_i \mid \alpha)$ is a proper density. This completes the proof.

ACKNOWLEDGMENTS

We thank the Editor, the Associate Editor, and the three referees for their helpful comments and suggestions, which have led to an improvement in this article. The order of authorship in this paper is alphabetical. Dr. Chen's research was partially supported by NSF grant DMS-9702172 and NIH Grant CA 74015-01. Dr. Ibrahim's research was partially supported by NIH Grants CA 70101-01 and CA 74015-01. Dr. Sinha's research was partially supported by NCI Grant R29-CA69222.

REFERENCES

B. Asselain, A. Fourquet, T. Hoang, A. D. Tsodikov, and A. Y. Yakovlev, A parametric regression model of tumor recurrence: An application to the analysis of clinical data on breast cancer, *Statist. Probab. Lett.* **29** (1996), 271–278.

J. O. Berger and D. Sun, Bayesian analysis for the poly-Weibull distribution, *J. Amer. Statist. Assoc.* **88** (1993), 1412–1418.

J. Berkson and R. P. Gage, Survival curve for cancer patients following treatment, *J. Amer. Statist. Assoc.* **47** (1952), 501–515.

J. Besag and P. J. Green, Spatial statistics and Bayesian computation (with discussion), *J. Roy. Statist. Soc., Ser. B* **16** (1993), 395–407.

D. J. Buckle, Bayesian inference for stable distributions, *J. Amer. Statist. Assoc.* **90** (1995), 605–613.

A. B. Cantor, and J. J. Shuster, Parametric versus non-parametric methods for estimating cure rates based on censored survival data, *Statist. Medicine* **11** (1992), 931–937.

M.-H. Chen, J. G. Ibrahim, and D. Sinha, A new Bayesian model for survival data with a surviving fraction, *J. Amer. Statist. Assoc.* **94** (1999), 909–919.

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

M. K. Cowles and B. P. Carlin, Markov chain Monte Carlo convergence diagnostics: A comparative review, *J. Amer. Statist. Assoc.* **91** (1996), 883–904.

D. R. Cox, Regression models and life tables, (with discussion), *J. Roy. Statist. Soc. Ser. B* **34** (1972), 187–220.

L. Devroye, "Non-Uniform Random Variate Generation," Springer-Verlag, New York, 1986.

V. T. Farewell, The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics* **38** (1982), 1041–1046.
V. T. Farewell, Mixture models in survival analysis: Are they worth the risk? *Canad. J. Statist.* **14** (1986), 257–262.
A. Gelman and D. B. Rubin, Inference from iterative simulation using multiple sequences, *Statist. Sci.* **7** (1992), 457–511.
W. R. Gilks and P. Wild, Adaptive rejection sampling for Gibbs sampling, *Appl. Statist.* **41** (1992), 337–348.
A. I. Goldman, Survivorship analysis when cure is a possibility: A Monte Carlo study, *Statist. Med.* **3** (1984), 153–163.
R. J. Gray and A. A. Tsiatis, A linear rank test for use when the main interest is in differences in cure rates, *Biometrics* **45** (1989), 899–904.
J. Halpern and B. W. Jr. Brown, Cure rate models: Power of the log rank and generalized Wilcoxon tests, *Statist. Med.* **6** (1987a), 483–489.
J. Halpern and B. W. Jr. Brown, Designing clinical trials with arbitrary specification of survival functions and for the log rank or generalized Wilcoxon test, *Controlled Clin. Trials* **8** (1987b), 177–189.
D. M. Higdon, Auxiliary variable methods for Markov chain Monte Carlo with applications, *J. Amer. Statist. Assoc.* **93** (1998), 585–595.
P. Hougaard, A class of multivariate failure time distributions, *Biometrika* **73** (1986), 671–678.
I. A. Ibragimov and K. E. Chernin, On the unimodality of stable laws, *Theory Probab. Appl.* **4** (1959), 417–419.
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, *J. Clin. Oncol.* **14** (1996), 7–17.
A. Y. C. Kuk and C.-H. Chen, A mixture model combining logistic regression with proportional hazards regression, *Biometrika* **79** (1992), 531–541.
E. M. Laska and M. J. Meisner, Nonparametric estimation and testing in a cure rate model, *Biometrics* **48** (1992), 1223–1234.
J. S. Liu, The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem, *J. Amer. Statist. Assoc.* **89** (1994), 958–966.
P. Müller and G. L. Rosner, A Bayesian population model with hierarchical mixture priors applied to blood count data, *J. Amer. Stat. Assoc.* **92** (1997), 1279–1292.
D. Oakes, Bivariate survival models induced by frailties, *J. Amer. Statist. Assoc.* **84** (1989), 487–493.
Y. Peng and K. B. G. Dear, A nonparametric mixture model for cure estimation, *Biometrics* **56** (2000), 237–243.
Z. Qiou, N. Ravishanker, and D. K. Dey, Multivariate survival analysis with positive stable frailties, *Biometrics* **55** (1999), 637–644.
G. Samorodnitsky and M. S. Taqqu, "Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance," Chapman and Hall, New York, 1994.
R. Sposto, H. N. Sather, and S. A. Baker, A comparison of tests of the difference in the proportion of patients who are cured, *Biometrics* **48** (1992), 87–99.
J. P. Sy and J. M. G. Taylor, Estimation in a Cox proportional hazards cure model, *Biometrics* **56** (2000), 227–236.
J. M. G. Taylor, Semi-parametric estimation in failure time mixture models, *Biometrics* **51** (1995), 899–907.
A. Y. Yakovlev, Letter to the editor, *Statist. Med.* **13** (1994), 983–986.
A. Y. Yakovlev, B. Asselain, V. J. Bardou, A. Fourquet, T. Hoang, A. Rochefediere, and A. D. Tsodikov (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," (B. Asselain, M. Boniface, C. Duby, C. Lopez, J. P. Masson, and J. Tranchefort, Eds.), Vol. 12, pp. 66–82.

K. Yamaguchi, Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of 'permanent employment' in Japan, *J. Amer. Statist. Assoc.* **87** (1992), 284–292.