Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

A New Bayesian Model for Survival Data with a Surviving Fraction

Ming-Hui Chen [a] , Joseph G. Ibrahim [b] & Debajyoti Sinha [c]

[a] Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA, 01609

[b] Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA, 02115

[c] Department of Mathematics, University of New Hampshire, Durham, NH, 03824

Available online: 17 Feb 2012

To cite this article: Ming-Hui Chen, Joseph G. Ibrahim & Debajyoti Sinha (1999): A New Bayesian Model for Survival Data with a Surviving Fraction, Journal of the American Statistical Association, 94:447, 909-919

To link to this article: http://dx.doi.org/10.1080/01621459.1999.10474196

A New Bayesian Model For Survival Data With a Surviving Fraction

Ming-Hui CHEN, Joseph G. IBRAHIM, and Debajyoti SINHA

We consider Bayesian methods for right-censored survival data for populations with a surviving (cure) fraction. We propose a model that is quite different from the standard mixture model for cure rates. We provide a natural motivation and interpretation of the model and derive several novel properties of it. First, we show that the model has a proportional hazards structure, with the covariates depending naturally on the cure rate. Second, we derive several properties of the hazard function for the proposed model and establish mathematical relationships with the mixture model for cure rates. Prior elicitation is discussed in detail, and classes of noninformative and informative prior distributions are proposed. Several theoretical properties of the proposed priors and resulting posteriors are derived, and comparisons are made to the standard mixture model. A real dataset from a melanoma clinical trial is discussed in detail.

KEY WORDS: Cure rate model; Gibbs sampling; Historical data; Latent variables; Posterior distribution; Weibull distribution.

1. INTRODUCTION

Survival models incorporating a cure fraction, often referred to as *cure rate models,* are becoming increasingly popular in analyzing data from cancer clinical trials. The cure rate model has been used for modeling time-to-event data for various types of cancers for which a significant proportion of patients are "cured," including breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma, and head and neck cancer. Perhaps the most popular type of cure rate model is the mixture model introduced by Berkson and Gage (1952). In this model, we assume that a certain fraction π of the population is "cured" and the remaining $1 - \pi$ are not cured. For 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 $S^*(t)$ denotes the survivor function for the noncured group in the population. Common choices for $S^*(t)$ are the exponential and Weibull distributions. This model, which we call the *standard cure rate model,* has been extensively discussed in the statistical literature by several authors, including Ewell and Ibrahim (1997), Farewell (1982, 1986), Goldman (1984), Gray and Tsiatis (1989), Greenhouse and Wolfe (1984), Halpern and Brown (1987a, 1987b), Kuk and Chen (1992), Laska and Meisner (1992), Sposto, Sather, and Baker (1992), Stangl and Greenhouse (1998), Taylor (1995), and Yamaguchi (1992). Although the standard cure rate model appears to be attractive and is widely used, it has several drawbacks. First, in the presence of covariates, it cannot have a proportional hazards structure, which is a desirable property for survival models, especially from a frequentist perspective, because many asymptotic and computational results require a proportional hazards structure. Second, 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) because it implies that Bayesian inference with (1) essentially requires a proper prior. We elaborate more on this issue in Sections 4.1 and 4.2. Third, (1) does not appear to describe the underlying biological process generating the failure time, at least in the context of cancer relapse, where cure rate models are frequently used.

In this article we discuss a different type of cure rate model, which overcomes the drawbacks just mentioned. The model that we propose is quite attractive in several respects. First, the model is derived from a natural biological motivation. Second, it has a proportional hazards structure through the cure rate parameter, and thus has an appealing interpretation. Third, it is very computationally attractive. In particular, by introducing latent variables, we are able to efficiently sample from the posterior distribution of the parameters. Fourth, the model has a mathematical relationship with the standard cure rate model. Specifically, we show that any standard cure rate model can be written as the proposed model and vice versa. Fifth, the model yields proper posterior distributions under a wide class of noninformative improper priors for the regression coefficients, including an improper uniform prior. This is an especially solid feature of our model, as it readily allows for noninformative Bayesian inference and facilitates comparisons with frequentist methods.

In addition, our model leads to a straightforward informative prior elicitation scheme based on historical data. This elicitation procedure yields proper priors that are technically and computationally convenient, which are not available using the formulation in (1). Specifically, in this article

Ming-Hui Chen is Associate Professor, Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609 (E-mail: *mhchen@wpi.edu*). Joseph G. Ibrahim is Associate Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115 (E-mail: *ibrahim@jimmy.harvard.edu*). Debajyoti Sinha is Associate Professor, Department of Mathematics, University of New Hampshire, Durham, NH 03824 (E-mail: *sinha@purabi.unh.edu*).

The authors wish to thank the editor, associate editor, and two referees for several suggestions that have greatly improved the article. Chen's research was partially supported by National Science Foundation (NSF) grant DMS-9702172 and National Institutes of Health (NIH) grant CA 74015, Ibrahim's research was partially supported by NIH grant CA 70101 and CA 74015, and Sinha's research was partially supported by National Cancer Institute grant R29-CA69222.

we propose novel classes of noninformative priors that lead to proper posterior distributions. We also propose a novel class of informative priors that are based on the notion of the availability of historical data and are constructed within this framework. Availability of historical data appears to be more relevant for populations with surviving fractions, because the very notion that the population may have a surviving fraction is probably based on previous studies or similar problems. We give theorems that characterize the propriety of the proposed priors under some very general settings, and also derive theoretical properties of the resulting posteriors. We discuss elicitation of prior parameters and sensitivity issues in detail, and demonstrate the proposed priors with real data in Section 5.

The article is organized as follows. In Section 2 we motivate the proposed model, give its general form, and derive several of its properties. In Section 3 we derive the likelihood function with covariates, and in Section 4 we propose a useful class of informative prior distributions and derive some of its theoretical properties. We also derive several properties of the resulting posterior distribution. In Section 5 we present a melanoma dataset from an actual clinical trial to illustrate the proposed methodologies. We conclude with a brief discussion in Section 6, and give proofs of the theorems of the proposed model in the Appendix.

2. THE MODEL

The proposed model can be derived as follows. Suppose that for an individual in the population, we let N denote the number of carcinogenic cells (often called clonogens) for that individual left active after the initial treatment, and assume that N has a Poisson distribution with mean θ. Also let Z_i denote the random time for the ith clonogenic cell to produce a detectable cancer mass. That is, Z_i can be viewed as an incubation time for the ith clonogenic cell. The variables $Z_i, i = 1, 2, \ldots$ are assumed to be iid with a common distribution function $F(t) = 1 - S(t)$ and are independent of N. The time to relapse of cancer can be defined by the random variable $T = \min\{Z_i, 0 \le i \le N\}$, where $P(Z_0 = \infty) = 1$ and N is independent of the sequence $Z_1, Z_2, \ldots$. The survival function for T, and hence the survival function for the population, is given by

$$\begin{aligned} S_p(t) &= P(\text{no cancer by time } t) \\ &= P(N = 0) + P(Z_1 > t, \ldots, Z_N > t, N \ge 1) \\ &= \exp(-\theta) + \sum_{k=1}^{\infty} S(t)^k \frac{\theta^k}{k!} \exp(-\theta) \\ &= \exp(-\theta + \theta S(t)) = \exp(-\theta F(t)). \end{aligned} \tag{2}$$

Because $S_p(\infty) = \exp(-\theta) > 0$, (2) is not a proper survival function. We see that (2) shows explicitly the contribution to the failure time of two distinct characteristics of tumor growth: the initial number of carcinogenic 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 failure time data that has a surviving fraction. Thus failure time data that do not "fit" the biological definition given earlier can still certainly be modeled by (2) as long as the data have a surviving fraction and can be thought of as being generated by an unknown number N of latent competing risks (Z_i's). Thus the model can be useful for modeling various types of failure time data, including time to relapse, time to death, time to first infection, and so forth. Yakovelev et al. (1993) discussed a similar modeling technique for modeling tumor latency.

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

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

As $\theta \to \infty$, the cure fraction tends to 0, whereas as $\theta \to 0$, the cure fraction tends to 1. The "density" corresponding to (2) is given by

$$f_p(t) = \theta f(t) \exp(-\theta F(t)), \tag{4}$$

where $f(t) = (d/dt)F(t)$. We emphasize here that $f_p(t)$ is *not* a proper probability density, because $S_p(t)$ is not a proper survival function; that is, $S_p(\infty) \neq 0$. But $f(t)$ appearing on the right side of (4) is a proper probability density function. The hazard function is given by

$$h_p(t) = \theta f(t). \tag{5}$$

Again, $h_p(t)$ is not a hazard function corresponding to a probability distribution, because $S_p(t)$ is not a proper survival function. We note, however, that $h_p(t) \to 0$ at a fast rate as $t \to \infty$ and $\int_0^\infty h_p(t)\, dt < \infty$.

We see that the cure rate model (2) yields an attractive form for the hazard in (5). Specifically, we see that $h_p(t)$ is multiplicative in θ and $f(t)$ and thus has the proportional hazards structure, with the covariates modeled through θ. This form of the hazard is more appealing than the one from the standard cure rate model in (1), which does not have the proportional hazards structure if π is modeled as a function of covariates. The proportional hazards property in (5) is also computationally attractive, as Markov Chain Monte Carlo (MCMC) methods are relatively easy to implement. The survival function for the "noncured" population is given by

$$S^*(t) = P(T > t | N \ge 1) = \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)}. \tag{6}$$

We note that $S^*(0) = 1$ and $S^*(\infty) = 0$, so that $S^*(t)$ is a proper survival function. The survival density for the noncured population (a proper density function) is given by

$$f^*(t) = -\frac{d}{dt} S^*(t) = \left(\frac{\exp(-\theta F(t))}{1 - \exp(-\theta)} \right) \theta f(t), \tag{7}$$

and the hazard function for the noncured population is given by

$$\begin{aligned} h^*(t) &= \frac{f^*(t)}{S^*(t)} = \left(\frac{\exp(-\theta F(t))}{\exp(-\theta F(t)) - \exp(-\theta)} \right) h_p(t) \\ &= \left(\frac{1}{P(T < \infty | T > t)} \right) h_p(t). \end{aligned} \tag{8}$$

Thus (8) is magnified by the factor $1/[P(T < \infty | T > t)] > 1$ compared to the hazard function $h_p(t)$ of the entire population. Clearly, (8) does not have a proportional hazards structure, because $1/[P(T < \infty | T > t)]$ can never be free of t for any $f(t)$ with support on $(0, \infty)$. Furthermore, $h^*(t) \to [f(t)]/[S(t)]$ as $t \to \infty$, and thus $h^*(t)$ converges to the hazard function of the incubation time random variable Z as $t \to \infty$. Finally, it can be shown that the hazard function $h^*(t)$ is an increasing function of θ.

There is a mathematical relationship between the models in (1) and (2). We can write $S_p(t) = \exp(-\theta) + (1 - \exp(-\theta))S^*(t)$, where $S^*(t)$ is given by (6). Thus $S_p(t)$ is a standard cure rate model with cure rate equal to $\pi = \exp(-\theta)$ and survival function for the noncured population given by $S^*(t)$. This shows that every model defined by (2) can be written as a standard cure rate model. This result also implies that every standard cure rate model corresponds to some model of the form (2) for some θ and $F(\cdot)$. We mention that if the covariates enter through θ, then $S_p(t)$ can be taken to have a Cox proportional hazards structure, but then in this case $h^*(t)$ will not have a proportional hazards structure. Equations (2)–(8) and the foregoing discussion summarize the fundamental modeling differences between our model and the standard cure rate model. In our model, we model the entire population as a proportional hazards model, whereas in the standard cure rate model, only the noncured group can be modeled with a proportional hazards structure.

In model (1), we let the covariates depend on θ through the relationship $\theta = \exp(\mathbf{x}'\boldsymbol{\beta})$, where $\mathbf{x}$ is a $k \times 1$ vector of covariates and $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients. We demonstrate in the next section that entering the covariates in this fashion corresponds to a canonical link in a Poisson regression model. Using $\theta = \exp(\mathbf{x}'\boldsymbol{\beta})$, (3), and (8), we can interpret the role of the regression coefficients for the cured and noncured groups. For the cured group, the sign of the regression coefficients affects the cure fraction. Thus a negative regression coefficient, for example, leads to a larger cure fraction, when the corresponding covariate takes a positive value. For the noncured group, the regression coefficients affect the hazard function in (8). Specifically, a negative regression coefficient, for example, leads to a larger hazard, whereas a positive regression coefficient leads to a smaller hazard, when the corresponding covariate takes a positive value.

3. THE LIKELIHOOD FUNCTION

Suppose that we have n subjects, and let N_i denote the number of carcinogenic cells for the ith subject. Further, assume that the N_i's are iid Poisson random variables with mean $\theta, i = 1, \ldots, n$. We emphasize here that the N_i's are not observed and can be viewed as latent variables in our model formulation. Further, suppose that $Z_{i1}, \ldots, Z_{i,N_i}$ are the iid incubation times for the N_i carcinogenic cells for the ith subject, which are unobserved, and all have cdf $F(\cdot), i = 1, \ldots, n$. In this article we specify a parametric form for $F(\cdot)$, such as a Weibull or a gamma distribution. We denote the indexing parameter (possibly vector valued) by γ, and thus write $F(\cdot|\gamma)$ and $S(\cdot|\gamma)$. For example, if $F(\cdot|\gamma)$ corresponds to a Weibull distribution, then $\gamma = (\alpha, \lambda)$, where α is the shape parameter and λ is the scale parameter. Let t_i denote the failure time for subject i, where t_i may be right censored. Let c_i denote the censoring time, so that we observe $y_i = \min(t_i, c_i)$, where the censoring indicator $\delta_i = I(t_i \le c_i)$ equals 1 if t_i is a failure time and 0 if it is right censored. We can represent the observed data by the vector $(n, \mathbf{y}, \boldsymbol{\delta})$, where $\mathbf{y} = (y_1, \ldots, y_n)$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$. Also, let $\mathbf{N} = (N_1, \ldots, N_n)$. The "complete data" are given by $\mathbf{D} = (n, \mathbf{y}, \boldsymbol{\delta}, \mathbf{N})$, where $\mathbf{N}$ is an unobserved vector of latent variables. The complete-data likelihood function of the parameters (γ, θ) can then be written as

$$L(\gamma, \theta | \mathbf{D}) = \left(\prod_{i=1}^{n} S(y_i|\gamma)^{N_i - \delta_i} (N_i f(y_i|\gamma))^{\delta_i} \right) \times \exp\left\{ \sum_{i=1}^{n} (N_i \log(\theta) - \log(N_i!)) - n\theta \right\}. \quad (9)$$

Throughout the remainder of this article, we assume a Weibull density for $f(y_i|\gamma)$, so that $f(y_i|\gamma) = \alpha y_i^{\alpha-1} \exp\{\lambda - y_i^\alpha \exp(\lambda)\}$, where $\gamma = (\alpha, \lambda)$.

We incorporate covariates for the cure rate model (2) through the cure rate parameter θ. When covariates are included, we have a different cure rate parameter, θ_i, for each subject, $i = 1, \ldots, n$. Let $\mathbf{x}_i' = (x_{i1}, \ldots, x_{ik})$ denote the $k \times 1$ vector of covariates for the ith subject, and let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ denote the corresponding vector of regression coefficients. We relate θ to the covariates by $\theta_i \equiv \theta(\mathbf{x}_i'\boldsymbol{\beta}) = \exp(\mathbf{x}_i'\boldsymbol{\beta})$, so that the cure rate for subject i is $\exp(-\theta_i) = \exp(-\exp(\mathbf{x}_i'\boldsymbol{\beta})), i = 1, \ldots, n$. This relationship between θ_i and $\boldsymbol{\beta}$ is equivalent to a canonical link for θ_i in the setting of generalized linear models. This model requires that either the covariate affects the expected number of carcinogenic cells in the patient or is negligible compared to the effect of the number of carcinogenic cells. With this relation, we can write the complete-data likelihood of $(\boldsymbol{\beta}, \gamma)$ as

$$L(\boldsymbol{\beta}, \gamma | \mathbf{D}) = \left(\prod_{i=1}^{n} S(y_i|\gamma)^{N_i - \delta_i} (N_i f(y_i|\gamma))^{\delta_i} \right) \times \exp\left\{ \sum_{i=1}^{n} [N_i \mathbf{x}_i'\boldsymbol{\beta} - \log(N_i!) - \exp(\mathbf{x}_i'\boldsymbol{\beta})] \right\}, \quad (10)$$

where $\mathbf{D} = (n, \mathbf{y}, \mathbf{X}, \boldsymbol{\delta}, \mathbf{N})$, $\mathbf{X}$ is the $n \times k$ matrix of covariates, $f(y_i|\gamma)$ is Weibull density given previously, and $S(y_i|\gamma) = \exp(-y_i^\alpha \exp(\lambda))$. If we assume independent priors for $(\boldsymbol{\beta}, \gamma)$, then the posterior distributions of $(\boldsymbol{\beta}, \gamma)$ are also independent. We mention that the part of the complete-data likelihood in (10) involving $\boldsymbol{\beta}$ looks exactly like a Poisson generalized linear model with a canonical link, with the N_i's the observables. Thus once the N_i's are given, inference about the parameters $\boldsymbol{\beta}$ and γ is straightforward.

4. THE PRIOR DISTRIBUTIONS

In this section we propose novel classes of noninformative and informative priors for $(\boldsymbol{\beta}, \gamma)$, and discuss some

theoretical properties of the proposed priors and resulting posteriors.

4.1 Noninformative Priors

Suppose that we consider a joint noninformative prior for $\pi(\boldsymbol{\beta},\boldsymbol{\gamma})$ of the form $\pi(\boldsymbol{\beta},\boldsymbol{\gamma}) \propto \pi(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = (\alpha,\lambda)$ are the Weibull parameters in $f(y|\boldsymbol{\gamma})$. This noninformative prior implies that $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are independent a priori and $\pi(\boldsymbol{\beta}) \propto 1$ is a uniform improper prior. We assume throughout that $\pi(\boldsymbol{\gamma}) = \pi(\alpha|\nu_0,\tau_0)\pi(\lambda)$, where $\pi(\alpha|\nu_0,\tau_0) \propto \alpha^{\nu_0-1}\exp(-\tau_0\alpha)$, and ν_0 and τ_0 are two specified hyperparameters. With these specifications, the posterior distribution of $(\boldsymbol{\beta},\boldsymbol{\gamma})$ based on the observed data $\mathbf{D}_{\text{obs}} = (n,\mathbf{y},\mathbf{X},\boldsymbol{\delta})$ is given by

$$p(\boldsymbol{\beta},\boldsymbol{\gamma}|\mathbf{D}_{\text{obs}}) \propto \left(\sum_{\mathbf{N}} L(\boldsymbol{\beta},\boldsymbol{\gamma}|\mathbf{D})\right)\pi(\alpha|\nu_0,\tau_0)\pi(\lambda), \quad (11)$$

where the sum in (11) 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 (11) using the noninformative prior $\pi(\boldsymbol{\beta},\boldsymbol{\gamma}) \propto \pi(\boldsymbol{\gamma})$.

Theorem 1. Let $d = \sum_{i=1}^{n}\delta_i$ and $\mathbf{X}^*$ be an $n \times k$ matrix with rows $\delta_i\mathbf{x}_i'$. Then if (a) $\mathbf{X}^*$ is of full rank, (b) $\pi(\lambda)$ is proper, and (c) $\tau_0 > 0$ and $\nu_0 > -d$, the posterior given in (11) is proper.

The proof of Theorem 1 is quite technical and is given in the Appendix. Note that the conditions given in Theorem 1 are sufficient but *not* necessary for the propriety of the posterior distribution. However, the conditions stated in the theorem are quite general and are typically satisfied for most datasets. Also note that a proper prior for α is not required to obtain a proper posterior. This can be observed from condition (c), because $\pi(\alpha|\nu_0,\tau_0)$ is no longer proper when $\nu_0 < 0$. Based on condition (b), $\pi(\lambda)$ is required to be proper. Although several choices can be made, we use a normal density for $\pi(\lambda)$ in the remainder of the article.

Theorem 1 is a very solid feature of our proposed model because, under some very general conditions, it guarantees propriety of the posterior distribution of $\boldsymbol{\beta}$ using a improper uniform prior. This enables us to carry out noninformative Bayesian inference for the regression coefficients and facilitates comparisons with maximum likelihood. However, under the noninformative priors $\pi(\boldsymbol{\beta},\boldsymbol{\gamma}) \propto \pi(\boldsymbol{\gamma})$, the standard cure rate model in (1) always leads to an improper posterior distribution for $\boldsymbol{\beta}$. This result is stated in the following theorem.

Theorem 2. For the standard cure rate model given in (1), suppose that we relate the cure fraction π to the covariates via a standard binomial regression

$$\pi_i = G(\mathbf{x}_i'\boldsymbol{\beta}), \quad (12)$$

where $G(\cdot)$ is a continuous cdf. Assume that the survival function $S^*(\cdot)$ for the noncured group depends on the parameter $\boldsymbol{\gamma}^*$. Let $L_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*|\mathbf{D}_{\text{obs}})$ denote the resulting likelihood function based on the observed data. Then, if we take an improper uniform prior for $\boldsymbol{\beta}$ [i.e., $\pi(\boldsymbol{\beta}) \propto 1$], the posterior distribution

$$\pi_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*|\mathbf{D}_{\text{obs}}) \propto L_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*|\mathbf{D}_{\text{obs}})\pi(\boldsymbol{\gamma}^*) \quad (13)$$

is always improper regardless of the propriety of $\pi(\boldsymbol{\gamma}^*)$.

The proof of Theorem 2 is given in the Appendix.

4.2 Informative Priors

We now propose a class of informative priors for $(\boldsymbol{\beta},\boldsymbol{\gamma})$. Our prior construction is based on the notion of the existence of a previous similar study that measures the same response variable and covariates as the current study. For ease of exposition, we assume only one previous study, as the extension to multiple previous studies is straightforward. As mentioned earlier, we call the data from the similar previous study the *historical data.* Prior elicitation using historical data has been discussed by the authors for generalized linear models in earlier work (Chen, Ibrahim, and Yiannoutsos 1999; Ibrahim, Ryan, and Chen 1998). However, the informative priors proposed here are quite different from those in that work. Informative prior construction for the proposed cure rate model proceeds as follows. Let n_0 denote the sample size for the historical data, let $\mathbf{y}_0$ be an $n_0 \times 1$ of right-censored failure times for the historical data with censoring indicators $\boldsymbol{\delta}_0$, let $\mathbf{N}_0$ be the unobserved vector of latent counts of carcinogenic cells, and let $\mathbf{X}_0$ be a $n_0 \times k$ matrix of covariates corresponding to $\mathbf{y}_0$. Let $\mathbf{D}_0 = (n_0,\mathbf{y}_0,\mathbf{X}_0,\boldsymbol{\delta}_0,\mathbf{N}_0)$ denote the complete historical data. Further, let $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$ denote the *initial prior* distribution for $(\boldsymbol{\beta},\boldsymbol{\gamma})$ from the previous study; that is, $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$ is the prior distribution for $(\boldsymbol{\beta},\boldsymbol{\gamma})$ when the historical data were the "current data." Using this information, we wish to construct a prior distribution for $(\boldsymbol{\beta},\boldsymbol{\gamma})$ based on the current study. We propose a joint informative prior distribution of the form

$$\pi(\boldsymbol{\beta},\boldsymbol{\gamma}|\mathbf{D}_{0,\text{obs}},a_0) \propto \left[\sum_{\mathbf{N}_0} L(\boldsymbol{\beta},\boldsymbol{\gamma}|\mathbf{D}_0)\right]^{a_0}\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma}), \quad (14)$$

where $L(\boldsymbol{\beta},\boldsymbol{\gamma}|\mathbf{D}_0)$ is the complete-data likelihood given in (10) with $\mathbf{D}$ replaced by the historical data $\mathbf{D}_0$ and $\mathbf{D}_{0,\text{obs}} = (n_0,\mathbf{y}_0,\mathbf{X}_0,\boldsymbol{\delta}_0)$. Thus our informative prior is based on the observed historical data likelihood raised to the power a_0. General constructions of priors based on exponentiating functions of observables to a power have been discussed in different contexts by Diaconis and Ylvisaker (1979) and Morris (1982, 1983), but those are different from what we propose here. The parameter a_0 can be interpreted as a dispersion parameter for the historical data. It is reasonable to restrict the range of a_0 to be between 0 and 1, and thus we take $0 \le a_0 \le 1$. One of the main roles of a_0 is controlling the heaviness of the tails of the prior for $(\boldsymbol{\beta},\boldsymbol{\gamma})$. As a_0 becomes smaller, the tails of (14) become heavier, resulting in a less informative prior. Setting $a_0 = 1$, (14) corresponds to the usual Bayesian update of $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$ via Bayes's theorem. That is, with $a_0 = 1$, (14) corresponds to the posterior distribution of $(\boldsymbol{\beta},\boldsymbol{\gamma})$ from the previous study. When $a_0 = 0$, the prior does not depend on the histor-

ical data, and in this case $\pi(\boldsymbol{\beta},\boldsymbol{\gamma}|\mathbf{D}_{0,\text{obs}},a_0) \equiv \pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$. Therefore, the prior (14) can be viewed as a generalization of the usual Bayesian update of $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$. The parameter a_0 allows the investigator to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the previous and current study, or when the sample sizes of the two studies are quite different. In practice, it is reasonable to take a noninformative prior for $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$, such as $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma}) \propto \pi_0(\boldsymbol{\gamma})$, which implies $\pi_0(\boldsymbol{\beta}) \propto 1$. For $\boldsymbol{\gamma} = (\alpha,\lambda)$, we take a gamma prior for α with small shape and scale parameters, and an independent normal prior for λ with mean 0 and variance c_0.

The prior specification is completed by specifying a prior distribution for a_0. We take a beta prior for a_0, and thus we propose a joint prior distribution for $(\boldsymbol{\beta},\boldsymbol{\gamma},a_0)$ of the form

$$\pi(\boldsymbol{\beta},\boldsymbol{\gamma},a_0|\mathbf{D}_{0,\text{obs}}) \propto \left[\sum_{\mathbf{N}_0} L(\boldsymbol{\beta},\boldsymbol{\gamma}|\mathbf{D}_0)\right]^{a_0} \pi_0(\boldsymbol{\beta},\boldsymbol{\gamma}) a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}, \quad (15)$$

where (δ_0,λ_0) are specified prior parameters. The prior in (15) does not have a closed form but has several attractive theoretical properties. First, note that if $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$ is proper, then (15) is guaranteed to be proper. Further, (15) can be proper even if $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$ is improper. The following theorem characterizes the propriety of (15) when $\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma})$ is improper.

Theorem 3. Assume that

$$\pi_0(\boldsymbol{\beta},\boldsymbol{\gamma}) \propto \pi_0(\boldsymbol{\gamma}) \equiv \pi_0(\alpha|\nu_0,\tau_0)\pi_0(\lambda) \propto \alpha^{\nu_0-1}\exp(-\tau_0\alpha)\pi_0(\lambda),$$

where ν_0 and τ_0 are specified hyperparameters. Let $d_0 = \sum_{i=1}^{n_0}\delta_{0i}$ and $\mathbf{X}_0^*$ be an $n_0 \times k$ matrix with rows $\delta_{0i}\mathbf{x}_{0i}'$. If (a) $\mathbf{X}_0^*$ is of full rank, (b) $\nu_0 > 0$ and $\tau_0 > 0$, (c) $\pi_0(\lambda)$ is proper, and (d) $\delta_0 > k$ and $\lambda_0 > 0$, then the joint prior given in (15) is proper.

The proof of Theorem 3 is given in the Appendix. We mention that this type of prior construction based on the model (1) will lead to an improper prior as well as an improper posterior distribution. Thus the development of the historical data priors based on (1) will not work. This again shows a solid feature of our proposed model, as (2) enables us to construct proper historical data priors under some very general assumptions. Moreover, the priors are technically superior to those based on the model (1). This result can be summarized in the following theorem.

Theorem 4. For the standard cure rate model given in (1), suppose that we relate the cure fraction π to the covariates via a standard binomial regression given by (12). Assume that the survival function for the noncured group $S^*(\cdot)$ depends on the parameter $\boldsymbol{\gamma}^*$. Let $L_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*|\mathbf{D}_{0,\text{obs}})$ and $L_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*|\mathbf{D}_{\text{obs}})$ denote the likelihood functions based on the observed historical and current data. Suppose that we use an improper uniform initial prior for $\boldsymbol{\beta}$ [i.e., $\pi_0(\boldsymbol{\beta}) \propto 1$] to construct the joint prior as

$$\pi_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*,a_0|\mathbf{D}_{0,\text{obs}}) \propto [L_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*|\mathbf{D}_{0,\text{obs}})]^{a_0}\pi_0(\boldsymbol{\gamma}^*)a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}, \quad (16)$$

where δ_0 and λ_0 are specified hyperparameters. Then $\pi_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*,a_0|\mathbf{D}_{0,\text{obs}})$ is always improper regardless of the propriety of $\pi_0(\boldsymbol{\gamma}^*)$. In addition, if we use $\pi_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*,a_0|\mathbf{D}_{0,\text{obs}})$ as a prior, then the resulting posterior, given by

$$p_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*,a_0|\mathbf{D}_{\text{obs}},\mathbf{D}_{0,\text{obs}}) \propto L_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*|\mathbf{D}_{\text{obs}})\pi_1(\boldsymbol{\beta},\boldsymbol{\gamma}^*,a_0|\mathbf{D}_{0,\text{obs}}), \quad (17)$$

is also improper.

The proof of Theorem 4 is similar to that of Theorem 2, and thus we omit the details for brevity.

Finally, we mention that the prior in (15) can be used even if a previous study does not exist on which to base $\mathbf{D}_0$. In this case $\mathbf{y}_0$ can be obtained via a prior prediction, including specifications based on a theoretical prediction model, expert opinion, or case-specific information. For example, a theoretical model of the form $\mathbf{y}_0 = g(\mathbf{X}_0)$ may be available for obtaining the prior predictions, where g is a known function. Also, in these cases one may take $\mathbf{X}_0$ to be the covariate matrix of the current study; that is, $\mathbf{X}_0 = \mathbf{X}$ and $n_0 = n$. In any case, the existence of historical data from a similar previous study leads to the most natural specification of $\mathbf{D}_0$ and serves as the primary motivation for (15). Taking $\mathbf{D}_0$ to be the raw data from a previous study results

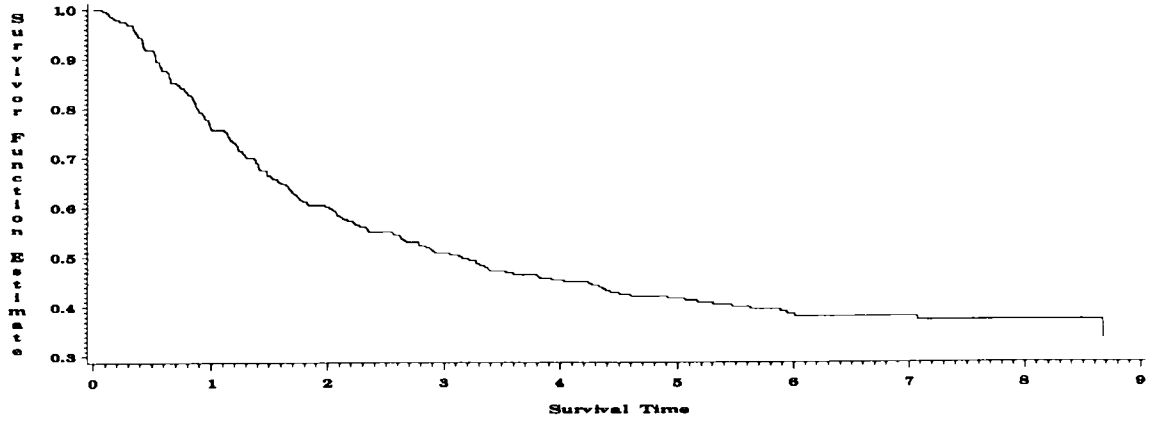

Figure 1. Kaplan–Meier Plot for E1684 Data.

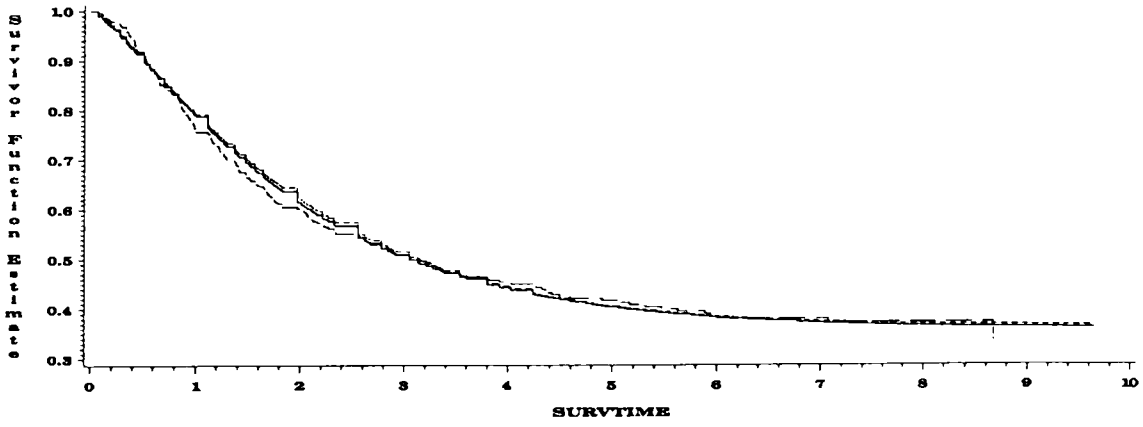

Figure 2. Superimposed Survival Curves for the E1684 Data.

in a more natural, interpretable, and automated specification for (15).

5. MELANOMA DATA

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 overall survival, which led to FDA approval of this regimen as a standard adjuvant therapy for high-risk melanoma patients. Here overall survival is defined as the time from randomization until death. One of our main goals in this example is to compare inferences between the standard cure rate model and the proposed model. First, we compare the maximum likelihood estimates (MLE's) of the cure rates between the two models (2) and (1). Second, using the cure rate model in (2), we carry out a Bayesian analysis with covariates using the proposed priors in (15), then compare the results to the estimates based on the standard cure rate model in (1). Three covariates and an intercept are included in the analyses. The covariates are age (x_1), gender (x_2; male, female), and performance status (PS) (x_3; fully active, other). After deleting missing observations, 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.

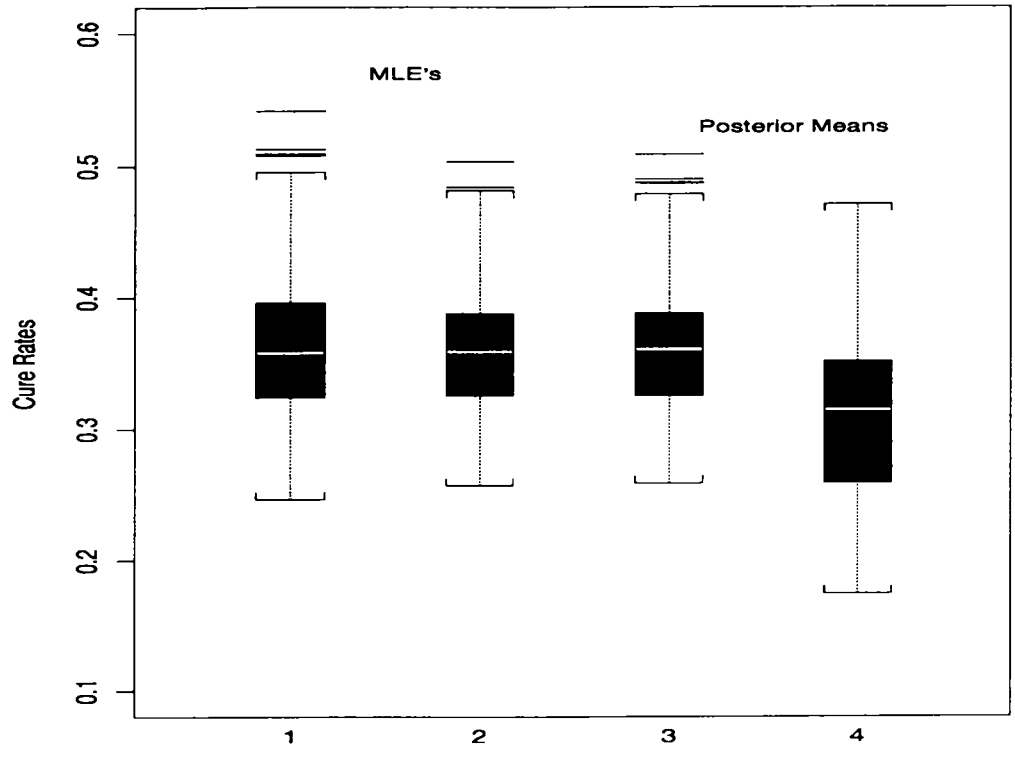

Figure 3. Boxplots of the Cure Rates for All Patients; 1, the standard cure rate model; 2, the proposed model; 3, $a_0 = 0$; and 4, $E(a_0|\mathbf{D}_0; \mathbf{D}_{0;obs}) = .29$.

Table 1. MLE's of the Model Parameters

Variable	MLE	SD	P value
Intercept	.09	.11	.38
Age	.09	.07	.21
Gender	−.12	.16	.44
PS	−.20	.26	.44
α	1.32	.09	.00
λ	−1.34	.12	.00

Figure 1 displays a Kaplan–Meier plot for overall survival, which shows a "plateau" in the survival curve, and thus a cure rate model appears to be suitable for these data. Figure 2 shows three superimposed plots of the survival curve based on the Kaplan–Meier method (dashed line), the standard cure rate model (1) (dotted line), and the proposed 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. Figure 3 shows a boxplot of the MLE's of the cure rates for all patients for the two models, where 1 denotes the standard cure rate model and 2 denotes the proposed model. We see that the two boxplots are very similar. The first, second, and third quartiles for the two boxplots are .32, .36, and .40 for the standard cure rate model and .33, .36, and .39 for the proposed model. We see from Figure 3 that the variation in the cure rate estimates from the standard model is greater than that of the proposed model. In fact, the standard deviations of the cure rate estimates are .06 for the standard cure rate model and .05 for the proposed model. Table 1 reports the MLE's, their standard deviations, and p values for the proposed model.

Several years earlier, a similar melanoma study with the same patient population was conducted by ECOG. This study, denoted E1673, serves as the historical data for our Bayesian analysis of E1684. A total of $n_0 = 650$ patients is used in the historical data. Tables 2 and 3 summarize the historical data E1673 and the current data E1684. Using the E1673 study as historical data, we consider an analysis with the proposed priors (15). For $\pi_0(\beta)$, we take an improper uniform prior, and for $\pi_0(\alpha|\nu_0, \tau_0)$, we take $\nu_0 = 1$ and $\tau_0 = .01$ to ensure a proper prior. We note that this choice for $\pi_0(\alpha|\nu_0, \tau_0)$ also guarantees log-concavity. The parameter λ is taken to have a normal distribution with mean 0 and variance 10,000.

Table 4 gives posterior estimates of β based on several values of (δ_0, λ_0) using the proposed model (2). In the table we obtain, for example, $E(a_0|\mathbf{D}_{\text{obs}}, \mathbf{D}_{0,\text{obs}}) = .03, .06$, and .14 by taking $(\delta_0, \lambda_0) = (50, 50), (100, 100)$, and $(200, 1)$. The case $a_0 = 0$ with probability 1 essentially yields the MLE's of β, α, and λ given in Table 1. 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 much 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, then 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. Thus giving more weight to the historical data, has the potential to affect our inference about β. 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, because it indicates that age is a potentially important prognostic factor for predicting survival in melanoma. Such a conclusion is not possible based on a frequentist or a Bayesian analysis of the current data alone.

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, demonstrating the importance of gender in predicting overall survival. Thus we see the potential impact of the historical data on the posterior analysis of β, and hence the potential impact on the posterior estimates of the cure rates. 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 is a strong feature of our model, as it demonstrates that incorporating 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 (.08, .24), whereas when historical data are not incorporated (i.e., $a_0 = 0$), these values are .09, .07, and (−.05, .23). There is 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 in the current data E1684, which has had less than 10 years of follow-up on 284 patients and has about 39% censoring.

Incorporation of historical data can also affect the posterior estimates of the cure rates. Figure 3 shows a boxplot of the posterior means of the cure rates based on no incorporation of historical data (coded 3) and a boxplot corresponding to $E(a_0|\mathbf{D}_0, \mathbf{D}_{0,\text{obs}}) = .29$ (coded 4). The posterior estimates in the cure rates are quite different in the model with $E(a_0|\mathbf{D}_0, \mathbf{D}_{0,\text{obs}}) = .29$ than in the model with no incorporation of historical data. The mean and standard deviations are .36 and .05 for boxplot 3 ($a_0 = 0$) and .31 and .06 for boxplot 4 $[E(a_0|\mathbf{D}_{\text{obs}}, \mathbf{D}_{0,\text{obs}}) = .29]$. Thus we

Table 2. Summary of E1684 Data

Survival time (y) (years)		Status (frequency)		Age (x_1) (years)		Gender (x_2) (frequency)		PS (x_3) (frequency)	
Median	1.38	Censored	110	Mean	47.03	Male	171	Fully active	253
IQR	1.90	Death	174	SD	13.00	Female	113	Other	31

Table 3. Summary of E1673 Data

Survival time (y_0) (years)		Status (frequency)		Age (x_{01}) (years)		Gender (x_{02}) (frequency)		PS (x_{03}) (frequency)	
Median	2.33	Censored	257	Mean	48.02	Male	375	Fully active	561
IQR	4.24	Death	393	SD	13.99	Female	275	Other	89

see that the mean cure rate drops from .36 to .31 when the historical data are incorporated. Again, a partial explanation of this result is due to the fact that the historical data are much more mature than the current data, with nearly 20 years of follow-up and a smaller fraction of censored cases. Thus the results in Figure 3 are not surprising, and in fact are appealing, as they give us a better estimate of the cure rate than an estimate based on the current data alone. Such a conclusion is not possible from a frequentist or Bayesian analysis using the current data alone. It is also interesting to mention that boxplot 3 is almost identical to boxplot 2. This is clear from the fact that when $a_0 = 0$ and vague proper priors are chosen for $\pi_0(\cdot)$, the posterior distribution of β (and hence θ) is essentially the same as the likelihood function in (9), and thus the estimates are quite similar. This again is a desirable feature of our model, as it implies that we can obtain MLE's via Gibbs sampling without any analytic maximizations. That is, if we take $a_0 = 0$ and choose vague proper priors for $\pi_0(\cdot)$, then the posterior means of the parameters are very close to the MLE's.

We also conducted a detailed sensitivity analysis for the regression coefficients by varying the hyperparameters for a_0 [i.e., (δ_0, λ_0)] and varying the hyperparameters for $\gamma = (\alpha, \lambda)$. Table 4 shows that the posterior estimates of the parameters are fairly robust as the hyperparameters (δ_0, λ_0) are varied. When we vary the hyperparameters for γ, the posterior estimates of β are also robust for a wide range of hyperparameter values. For example, when fixing the hyperparameters for a_0 so that $E(a_0|\mathbf{D}_0, \mathbf{D}_{0,\text{obs}}) = .29$ 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 we did a burn-in of 1,000 samples with autocorrelations disappearing after lag 5 for nearly all parameters, and used 50,000 Gibbs iterates after the burn-in for all of the posterior computations. Further, we computed all HPD intervals using an efficient Monte Carlo method of Chen and Shao (1999). 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 are quite novel, and 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 Gilks and Wild (1992) for Gibbs sampling.

Table 4. Melanoma Data: Posterior Estimates of the Model Parameters With $\alpha \sim$ gamma(1, .01) and $\lambda \sim$ N(0, 10,000)

$E(a_0\|\mathbf{D}_{obs}, \mathbf{D}_{0,obs})$	Variable	Posterior mean	Posterior SD	95% HPD interval
0	Intercept	.09	.11	(−.12, .30)
(With probability 1)	Age	.09	.07	(−.05, .23)
	Gender	−.12	.16	(−.44, .19)
	PS	−.23	.26	(−.73, .28)
	α	1.31	.09	(1.15, 1.48)
	λ	−1.36	.12	(−1.60, −1.11)
.03	Intercept	.17	.11	(−.04, .38)
	Age	.10	.07	(−.04, .24)
	Gender	−.14	.15	(−.44, .15)
	PS	−.19	.25	(−.68, .28)
	α	1.17	.07	(1.03, 1.32)
	λ	−1.45	.13	(−1.70, −1.20)
.06	Intercept	.21	.11	(.01, .43)
	Age	.11	.07	(−.03, .24)
	Gender	−.16	.15	(−.45, .13)
	PS	−.16	.24	(−.63, .29)
	α	1.12	.07	(.99, 1.25)
	λ	−1.53	.13	(−1.78, −1.28)
.14	Intercept	.25	.10	(.05, .45)
	Age	.12	.06	(−.00, .24)
	Gender	−.20	.14	(−.47, .07)
	PS	−.09	.22	(−.53, .31)
	α	1.06	.06	(.95, 1.17)
	λ	−1.62	.12	(−1.85, −1.39)
.21	Intercept	.26	.10	(.08, .45)
	Age	.13	.06	(.01, .24)
	Gender	−.22	.13	(−.48, .03)
	PS	−.05	.20	(−.44, .34)
	α	1.04	.05	(.94, 1.15)
	λ	−1.67	.11	(−1.89, −1.45)
.29	Intercept	.26	.09	(.08, .43)
	Age	.13	.06	(.02, .24)
	Gender	−.24	.12	(−.48, .00)
	PS	−.01	.19	(−.38, .35)
	α	1.03	.05	(.93, 1.13)
	λ	−1.70	.11	(−1.91, −1.50)
1	Intercept	.22	.06	(.11, .35)
(With probability 1)	Age	.16	.04	(.08, .24)
	Gender	−.32	.09	(−.50, −.15)
	PS	.14	.13	(−.11, .39)
	α	1.00	.04	(.93, 1.07)
	λ	−1.82	.08	(−1.97, −1.67)

6. DISCUSSION

We have proposed a novel cure rate model that has several advantages over the standard cure rate model. The new model has several appealing interpretations, has a proportional hazards structure, allows for incorporation of historical data, leads to prior and posterior distributions with

Table 5. Posterior Estimates of the Model Parameters With $E(a_0|\mathbf{D}_{obs}, \mathbf{D}_{0,obs}) = .29$, $\alpha \sim$ Gamma(1, 1) and $\lambda \sim$ N(0, 10)

Variable	Posterior mean	Posterior SD	95% HPD interval
Intercept	.26	.09	(.07, .42)
Age	.13	.06	(.02, .24)
Gender	−.24	.12	(−.48, .00)
PS	−.01	.19	(−.38, .35)
α	1.02	.05	(.93, 1.12)
λ	−1.69	.11	(−1.90, −1.48)

desirable properties, and is computationally attractive. Moreover, our melanoma example demonstrates that the proposed model yields MLE's and Kaplan–Meier estimates similar to the standard cure rate model. We mention that our proposed methods are quite different from those of Yakovlev et al. (1993), who did not discuss covariates, estimation, prior elicitation, or any form of Bayesian inference, and did not examine any data augmentation or MCMC computational methods for the model. In addition, they did not investigate the additional properties regarding the hazard function, proportional hazards, or connections with the standard cure rate model.

Several extensions to the proposed model (2) are possible and are currently under investigation. The first possible extension is to consider a slightly more general model to incorporate more patient heterogeneity. This can be done by introducing an additional parameter, w, in the distribution of N. Thus we can take N to have a Poisson distribution with mean θw, and then specify an appropriate distribution for w. This formulation for the distribution of N allows for more interpatient heterogeneity and can serve as a useful robust model. Another extension currently under investigation is a multivariate version of (2). Such an extension appears feasible by the form of (2) and could prove to be quite powerful, as a multivariate extension of the standard cure rate model appears to be analytically and computationally disastrous.

APPENDIX: PROOFS OF THEOREMS

A.1 PROOF OF THEOREM 1

By summing out the unobserved latent vector $\mathbf{N}$, the complete-data likelihood given in (10) reduces to

$$\sum_{\mathbf{N}} L(\boldsymbol{\beta}, \boldsymbol{\gamma}|\mathbf{D}) = \prod_{i=1}^{n} (\theta_i f(y_i|\boldsymbol{\gamma}))^{\delta_i} \exp\{-\theta_i(1 - S(y_i|\boldsymbol{\gamma}))\}. \quad \text{(A.1)}$$

To prove Theorem 1, we first show that there exists a constant $M > 1$ such that

$$(\theta_i f(y_i|\boldsymbol{\gamma}))^{\delta_i} \exp\{-\theta_i(1 - S(y_i|\boldsymbol{\gamma}))\} \le \alpha^{\delta_i} M. \quad \text{(A.2)}$$

When $\delta_i = 0$, (A.2) is obviously true, because $\exp\{-\theta_i(1 - S(y_i|\boldsymbol{\gamma}))\} \le 1$. For $\delta_i = 1$, the left side of (A.2) can be rewritten as

$$y_i^{-1} \frac{\alpha y_i^{\alpha} e^{\lambda} \exp(-e^{\lambda} y_i^{\alpha})}{1 - \exp(-e^{\lambda} y_i^{\alpha})} \times [(1 - S(y_i|\boldsymbol{\gamma}))\theta_i \exp(-\theta_i(1 - S(y_i|\boldsymbol{\gamma}))]. \quad \text{(A.3)}$$

Let

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

Then it can be shown that there exists a common constant $g_0 > 0$ such that

$$g_1(z) \le g_0,$$

and

$$g_2(z) \le g_0 \quad \forall z > 0. \quad \text{(A.4)}$$

Using (A.4), (A.3) is less than or equal to $y_i^{-1}\alpha g_0^2$. Thus, taking $M^* = g_0^2 \max_{i:\delta_i=1}\{y_i^{-1}\}$ and $M = \max\{1, M^*\}$, we obtain (A.2).

Because $\mathbf{X}^*$ is of full rank, there must exist k linearly independent row vectors $\mathbf{x}'_{i_1}, \mathbf{x}'_{i_2}, \ldots, \mathbf{x}'_{i_k}$ such that $\delta_{i_1} = \delta_{i_2} = \cdots = \delta_{i_k} = 1$. Using (A.1) and (A.2),

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{R^k} \sum_{\mathbf{N}} L(\boldsymbol{\beta}, \boldsymbol{\gamma}|\mathbf{D})\pi(\alpha|\nu_0, \tau_0)\pi(\lambda)\, d\boldsymbol{\beta}\, d\alpha\, d\lambda \\
&\le \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{R^k} (\alpha M)^{d-k} \prod_{j=1}^{k} f(y_{i_j}|\boldsymbol{\gamma}) \\
&\quad \times \exp\{\mathbf{x}'_{i_j}\boldsymbol{\beta} - (1 - S(y_{i_j}|\boldsymbol{\gamma}))\exp(\mathbf{x}'_{i_j}\boldsymbol{\beta})\} \\
&\quad \times \pi(\alpha|\nu_0, \tau_0)\pi(\lambda)\, d\boldsymbol{\beta}\, d\alpha\, d\lambda, \qquad \text{(A.5)}
\end{aligned}$$

where R^k denotes k-dimensional Euclidean space. Now we make the transformation $u_j = \mathbf{x}'_{i_j}\boldsymbol{\beta}$ for $j = 1, 2, \ldots, k$. This is a one-to-one linear transformation from $\boldsymbol{\beta}$ to $\mathbf{u} = (u_1, \ldots, u_k)'$. Thus (A.5) is proportional to

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{R^k} \alpha^{d-k} \prod_{j=1}^{k} f(y_{i_j}|\boldsymbol{\gamma}) \\
&\quad \times \exp\{u_j - (1 - S(y_{i_j}|\boldsymbol{\gamma}))\exp(u_j)\} \\
&\quad \times \pi(\alpha|\nu_0, \tau_0)\pi(\lambda)\, du\, d\alpha\, d\lambda \\
&= \int_{-\infty}^{\infty}\int_{0}^{\infty} \alpha^{d-k} \prod_{j=1}^{k} \\
&\quad \times \left[f(y_{i_j}|\boldsymbol{\gamma}) \int_{-\infty}^{\infty} \exp\{u_j - (1 - S(y_{i_j}|\boldsymbol{\gamma}))\exp(u_j)\}\, du_j\right] \\
&\quad \times \pi(\alpha|\nu_0, \tau_0)\pi(\lambda)\, d\alpha\, d\lambda. \qquad \text{(A.6)}
\end{aligned}$$

Integrating out $\mathbf{u}$, (A.6) reduces to

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \alpha^{d-k} \left[\prod_{j=1}^{k} \frac{f(y_{i_j}|\boldsymbol{\gamma})}{(1 - S(y_{i_j}|\boldsymbol{\gamma}))}\right] \times \pi(\alpha|\nu_0, \tau_0)\pi(\lambda)\, d\alpha\, d\lambda. \quad \text{(A.7)}$$

In (A.7), using (A.4), we have

$$\frac{f(y_{i_j}|\boldsymbol{\gamma})}{(1 - S(y_{i_j}|\boldsymbol{\gamma}))} = \frac{\alpha(y_{i_j})^{\alpha-1}e^{\lambda}\exp(-e^{\lambda}y_{i_j}^{\alpha})}{1 - \exp(-e^{\lambda}y_{i_j}^{\alpha})} \le K_0\alpha,$$

where $K_0 = g_0 \max_{1\le j\le k}\{y_{i_j}^{-1}\}$. Thus (A.7) is less than or equal to

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{0}^{\infty} (\alpha)^{d-k} \prod_{j=1}^{k} [K_0\alpha]\pi(\alpha|\nu_0, \tau_0)\pi(\lambda)\, d\alpha\, d\lambda \\
&\qquad = K_0^k \int_{-\infty}^{\infty}\int_{0}^{\infty} \alpha^{d}\pi(\alpha|\nu_0, \tau_0)\pi(\lambda)\, d\alpha\, d\lambda < \infty
\end{aligned}$$

by conditions (b) and (c). This completes the proof.

A.2 PROOF OF THEOREM 2

To prove Theorem 2, it suffices to show that

$$\int_{R^k} L_1(\boldsymbol{\beta}, \boldsymbol{\gamma}^*|\mathbf{D}_{\text{obs}})\, d\boldsymbol{\beta} = \infty. \quad \text{(A.8)}$$

Using (1) and (12), $L_1(\beta, \gamma^*|\mathbf{D}_{\text{obs}})$ can be written as

$$L_1(\beta,\gamma^*|\mathbf{D}_{\text{obs}}) = \prod_{i=1}^{n} [(1 - G(\mathbf{x}_i'\beta))f^*(y_i|\gamma^*)]^{\delta_i} \times [G(\mathbf{x}_i'\beta) + (1 - G(\mathbf{x}_i'\beta))S^*(y_i|\gamma^*)]^{1-\delta_i}, \quad \text{(A.9)}$$

where $f^*(y_i|\gamma^*)$ is the density function corresponding to $S^*(y_i|\gamma^*)$. From (A.9), it can be shown that

$$L_1(\beta,\gamma^*|\mathbf{D}_{\text{obs}}) \geq \prod_{i=1}^{n} ((1 - G(\mathbf{x}_i'\beta))[f^*(y_i|\gamma^*)]^{\delta_i}[S^*(y_i|\gamma^*)]^{1-\delta_i}. \quad \text{(A.10)}$$

Thus it is obvious that the integration of (A.10) over β equals ∞, because $L_1(\beta, \gamma^*|\mathbf{D}_{\text{obs}})$ is equivalent to a binomial regression likelihood that has all the responses "failures." This proves the theorem.

A.3 PROOF OF THEOREM 3

Similar to (A.2), we have

$$(\theta_{0i} f(y_{0i}|\gamma))^{\delta_{0i}} \exp\{-\theta_{0i}(1 - S(y_{0i}|\gamma))\} \leq \alpha^{\delta_{0i}} M_0, \quad \text{(A.11)}$$

where $M_0 > 1$ is a constant. Because $\mathbf{X}_0^*$ is of full rank, there must exist k linearly independent row vectors $\mathbf{x}_{0i_1}', \mathbf{x}_{0i_2}', \ldots, \mathbf{x}_{0i_k}'$ such that $\delta_{0i_1} = \delta_{0i_2} = \cdots = \delta_{0i_k} = 1$. Following the proof of Theorem 1, we have

$$\begin{aligned}
&\int_0^1 \int_{-\infty}^{\infty} \int_0^{\infty} \int_{R^k} \left[\sum_{\mathbf{N}_0} L(\beta,\gamma|\mathbf{D}_0)\right]^{a_0} \\
&\quad \times \pi_0(\alpha|\nu_0,\tau_0)\pi_0(\lambda)a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, d\beta\, d\alpha\, d\lambda\, da_0 \\
&\leq \int_0^1 \int_{-\infty}^{\infty} \int_0^{\infty} \int_{R^k} (\alpha M_0)^{a_0(d_0-k)} \\
&\quad \times \prod_{j=1}^{k} [f(y_{0i_j}|\gamma)]^{a_0} \exp\{a_0\mathbf{x}_{0i_j}'\beta - a_0(1 - S(y_{0i_j}|\gamma)) \\
&\quad \times \exp(\mathbf{x}_{0i_j}'\beta)\}\pi_0(\alpha|\nu_0,\tau_0)\pi_0(\lambda) \\
&\quad \times a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, d\beta\, d\alpha\, d\lambda\, da_0. \qquad \text{(A.12)}
\end{aligned}$$

Because $M_0 \geq 1$ and $0 < a_0 < 1$, $M_0^{a_0} \leq M_0$. Taking the transformation $u_{0j} = \mathbf{x}_{0i_j}'\beta$ for $j = 1, 2, \ldots, k$ and ignoring the constant, (A.12) is less than or equal to

$$\begin{aligned}
&\int_0^1 \int_{-\infty}^{\infty} \int_0^{\infty} \int_{R^k} \alpha^{a_0(d_0-k)} \prod_{j=1}^{k} [f(y_{0i_j}|\gamma)]^{a_0} \\
&\quad \times \exp\{a_0 u_{0j} - a_0(1 - S(y_{0i_j}|\gamma))\exp(u_{0j})\} \\
&\quad \times \pi_0(\alpha|\nu_0,\tau_0)\pi_0(\lambda)a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, du_0\, d\alpha\, d\lambda\, da_0 \\
&= \int_0^1 \int_{-\infty}^{\infty} \int_0^{\infty} \alpha^{a_0(d_0-k)} \prod_{j=1}^{k} \\
&\quad \times \left\{[f(y_{i_j}|\gamma)]^{a_0} \int_{-\infty}^{\infty} \exp\{a_0 u_{0j} - a_0(1 - S(y_{0i_j}|\gamma)) \right. \\
&\qquad \left. \times \exp(u_{0j})\}\, du_{0j}\right\} \\
&\quad \times \pi_0(\alpha|\nu_0,\tau_0)\pi_0(\lambda)a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, d\alpha\, d\lambda\, da_0, \qquad \text{(A.13)}
\end{aligned}$$

where $\mathbf{u}_0 = (u_{01}, \ldots, u_{0k})'$. Integrating out $\mathbf{u}_0$, (A.13) reduces to

$$\begin{aligned}
&\int_0^1 \int_{-\infty}^{\infty} \int_0^{\infty} \alpha^{a_0(d_0-k)} \prod_{j=1}^{k} \left[\frac{f(y_{0i_j}|\gamma)}{(1 - S(y_{0i_j}|\gamma))}\right]^{a_0} \frac{\Gamma(a_0)}{a_0^{a_0}} \\
&\quad \times \pi_0(\alpha|\nu_0,\tau_0)\pi_0(\lambda)a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, d\alpha\, d\lambda\, da_0, \qquad \text{(A.14)}
\end{aligned}$$

where $\Gamma(\cdot)$ denotes the gamma function. Using (A.4), it can be shown that

$$\left[\frac{f(y_{0i_j}|\gamma)}{(1 - S(y_{0i_j}|\gamma))}\right]^{a_0} \leq K_1\alpha^{a_0},$$

where K_1 is a positive constant. Because $0 < a_0 < 1$,

$$\frac{\Gamma(a_0)}{a_0^{a_0}} = \frac{a_0^{-1}\Gamma(a_0+1)}{a_0^{a_0}} \leq K_2 a_0^{-1},$$

where K_2 is a positive constant. Using the foregoing two inequalities, (A.14) is less than or equal to

$$\begin{aligned}
&\int_0^1 \int_{-\infty}^{\infty} \int_0^{\infty} \alpha^{a_0(d_0-k)} \prod_{j=1}^{k} [K_1\alpha^{a_0}K_2 a_0^{-1}] \\
&\quad \times \pi_0(\alpha|\nu_0,\tau_0)\pi_0(\lambda)a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, d\alpha\, d\lambda\, da_0 \\
&= (K_1K_2)^k \int_0^1 \int_{-\infty}^{\infty} \int_0^{\infty} \alpha^{a_0 d_0} a_0^{-k} \pi_0(\alpha|\nu_0,\tau_0) \\
&\quad \times \pi_0(\lambda)a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, d\alpha\, d\lambda\, da_0 \\
&\leq (K_1K_2)^k \int_0^1 \int_{-\infty}^{\infty} \int_0^{\infty} a_0^{-k}(1 + \alpha^{d_0})\pi_0(\alpha|\nu_0,\tau_0) \\
&\quad \times \pi_0(\lambda)a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, d\alpha\, d\lambda\, da_0 < \infty
\end{aligned}$$

by conditions (b)–(d). This completes the proof.

[Received March 1998. Revised January 1999.]

REFERENCES

Berkson, J., and Gage, R. P. (1952), "Survival Curve for Cancer Patients Following Treatment," *Journal of the American Statistical Association*, 47, 501–515.

Chen, M-H., Ibrahim, J. G., and Yiannoutsos, C. (1999), "Prior Elicitation, Variable Selection, and Bayesian Computation for Logistic Regression Models," *Journal of the Royal Statistical Society*, Ser. B, 61, 223–242.

Chen, M-H., and Shao, Q-M. (1999), "Monte Carlo Estimation of Bayesian Credible and HPD Intervals," *Journal of Computational and Graphical Statistics*, 8, 69–92.

Diaconis, P., and Ylvisaker, D. (1979), "Conjugate Priors for Exponential Families," *The Annals of Statistics*, 7, 269–281.

Ewell, M., and Ibrahim, J. G. (1997), "The Large Sample Distribution of the Weighted Log Rank Statistic Under General Local Alternatives," *Lifetime Data Analysis*, 3, 5–12.

Farewell, V. T. (1982), "The Use of Mixture Models for the Analysis of Survival Data With Long-Term Survivors," *Biometrics*, 38, 1041–1046.

——— (1986), "Mixture Models in Survival Analysis: Are They Worth the Risk?" *Canadian Journal of Statistics*, 14, 257–262.

Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.

Goldman, A. I. (1984), "Survivorship Analysis When Cure is a Possibility: A Monte Carlo Study," *Statistics in Medicine*, 3, 153–163.

Gray, R. J., and Tsiatis, A. A. (1989), "A Linear Rank Test for Use When the Main Interest is in Differences in Cure Rates," *Biometrics*, 45, 899–904.

Greenhouse, J. B., and Wolfe, R. A. (1984), "A Competing Risks Derivation of a Mixture Model for the Analysis of Survival," *Communications in Statistics, Part A—Theory and Methods*, 13, 3133–3154.

Halpern, J., and Brown, B. W., Jr. (1987a), "Cure Rate Models: Power of the Log Rank and Generalized Wilcoxon Tests," *Statistics in Medicine*, 6, 483–489.

——— (1987b), "Designing Clinical Trials With Arbitrary Specification of Survival Functions and for the Log Rank or Generalized Wilcoxon Test," *Controlled Clinical Trials*, 8, 177–189.

Ibrahim, J. G., Ryan, L. M., and Chen, M-H. (1998), "Use of Historical Controls to Adjust for Covariates in Trend Tests for Binary Data," *Journal of the American Statistical Association*, 93, 1282–1293.

Kirkwood, J. M., Strawderman, M. H., Ernstoff, M. S., Smith, T. J., Borden, E. C., and Blum, R. H. (1996), "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.

Kuk, A. Y. C., and Chen, C-H. (1992), "A Mixture Model Combining Logistic Regression With Proportional Hazards Regression," *Biometrika*, 79, 531–541.

Laska, E. M., and Meisner, M. J. (1992), "Nonparametric Estimation and Testing in a Cure Rate Model," *Biometrics*, 48, 1223–1234.

Morris, C. N. (1982), "Natural Exponential Families With Quadratic Variance Functions," *The Annals of Statistics*, 10, 65–80.

Morris, C. N. (1983), "Natural Exponential Families With Quadratic Variance Functions: Statistical Theory," *The Annals of Statistics*, 11, 515–529.

Sposto, R., Sather, H. N., and Baker, S. A. (1992), "A Comparison of Tests of the Difference in the Proportion of Patients Who are Cured," *Biometrics*, 48, 87–99.

Stangl, D. K., and Greenhouse, J. B. (1998), "Assessing Placebo Response Using Bayesian Hierarchical Survival Models," *Lifetime Data Analysis*, 4, 5–28.

Taylor, J. M. G. (1995), "Semi-Parametric Estimation in Failure Time Mixture Models," *Biometrics*, 51, 899–907.

Yakovlev, A. Y., Asselain, B., Bardou, V. J., Fourquet, A., Hoang, T., Rochefediere, A., and Tsodikov, A. D. (1993), "A Simple Stochastic Model of Tumor Recurrence and Its Applications to Data on Premenopausal Breast Cancer," in *Biometrie et Analyse de Donnees Spatio-Temporelles*, 12, eds. B. Asselain, M. Boniface, C. Duby, C. Lopez, J. P. Masson, and J. Tranchefort, France, Rennes: Société Francaise de Biométrie, ENSA, 66–82.

Yamaguchi, K. (1992), "Accelerated Failure-Time Regression Models With a Regression Model of Surviving Fraction: An Application to the Analysis of 'Permanent Employment' in Japan," *Journal of the American Statistical Association*, 87, 284–292.