The Canadian Journal of Statistics
Vol. 33, No. 4, 2005, Pages 559–570
La revue canadienne de statistique

Cure rate models: a unified approach

Guosheng YIN and Joseph G. IBRAHIM

Key words and phrases: Bayesian inference; Box–Cox transformation; cure fraction; Gibbs sampling; mixture cure model; promotion time cure model.

MSC 2000: Primary 62N01; secondary 62F15.

Abstract: The authors propose a novel class of cure rate models for right-censored failure time data. The class is formulated through a transformation on the unknown population survival function. It includes the mixture cure model and the promotion time cure model as two special cases. The authors propose a general form of the covariate structure which automatically satisfies an inherent parameter constraint and includes the corresponding binomial and exponential covariate structures in the two main formulations of cure models. The proposed class provides a natural link between the mixture and the promotion time cure models, and it offers a wide variety of new modelling structures as well. Within the Bayesian paradigm, a Markov chain Monte Carlo computational scheme is implemented for sampling from the full conditional distributions of the parameters. Model selection is based on the conditional predictive ordinate criterion. The use of the new class of models is illustrated with a set of real data involving a melanoma clinical trial.

Une approche unifiée des modèles de survie à taux de guérison

Résumé : Les auteurs proposent une nouvelle classe de modèles pour des durées de vie censurées à droite dans lesquels une possibilité de guérison est prise en compte. Cette classe est définie par le biais d'une transformation de la fonction de survie théorique inconnue. Les modèles dans lesquels la guérison est incorporée par voie de mélange et par temps d'apparition de tumeurs détectables en sont des cas particuliers. Les auteurs proposent une forme générale de structure de covariables qui satisfait automatiquement à une contrainte paramétrique inhérente au problème et qui inclut les structures binomiales et exponentielles correspondant aux deux principales formulations des modèles à taux de guérison. En plus de lier de façon naturelle les modèles à taux de guérison par mélange et à temps d'apparition de tumeurs détectables, la classe proposée suggère un grand nombre de nouvelles structures de modèles. Dans le cadre du paradigme bayésien, un algorithme de calcul de Monte-Carlo à chaîne de Markov est implanté pour l'échantillonnage à partir des lois conditionnelles complètes des paramètres. La sélection de modèle s'appuie sur un critère faisant intervenir des prévisions conditionnelles. L'emploi de la nouvelle classe de modèles est illustré au moyen de données issues d'essais cliniques sur le mélanome.

1. INTRODUCTION

Cure rate models, which are used for modelling time-to-event data incorporating a cure fraction, have become increasingly important in clinical trials. Let $S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i)$ be the population survival function, which is improper (i.e., $\lim_{t\to\infty} S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) > 0$), and let $S(t \mid \mathbf{X}_i)$ be a proper survival function (i.e., $\lim_{t\to\infty} S(t \mid \mathbf{X}_i) = 0$), where $\mathbf{Z}_i$ and $\mathbf{X}_i$ are two covariate vectors for subject i ($i = 1, \ldots, n$). Note that $\mathbf{Z}_i$ includes 1 and may share common components with $\mathbf{X}_i$. The mixture cure model (Berkson & Gage 1952) is the mixture of a certain fraction $1 - \theta(\mathbf{Z}_i)$ of the population being cured and the remaining proportion $\theta(\mathbf{Z}_i)$ which are not cured, such that

$$S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) = 1 - \theta(\mathbf{Z}_i) + \theta(\mathbf{Z}_i) S(t \mid \mathbf{X}_i), \tag{1}$$

where $S(t \mid \mathbf{X}_i)$ is the survival function for the uncured population. A logistic regression formulation is usually assumed for $\theta(\mathbf{Z}_i)$ so that

$$\theta(\mathbf{Z}_i) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)}{1 + \exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)}.$$

The mixture cure model (1) has been extensively studied in the literature, including Gray & Tsiatis (1989), Kuk & Chen (1992), Taylor (1995), Sy & Taylor (2000), Peng & Dear (2000) and Betensky & Schoenfeld (2001), among others. Extensive discussion of frequentist methods for the mixture cure model is given in the book by Maller & Zhou (1996). Although (1) is intuitively attractive and widely used, it does not have a proportional hazards structure in the presence of covariates, which is an undesirable feature when doing covariate analysis as noted in Ibrahim, Chen & Sinha (2001a, ch. 5).

An alternative definition of a cure rate model has been proposed and investigated by Yakovlev et al. (1993), Yakovlev & Tsodikov (1996), Tsodikov (1998), Chen, Ibrahim & Sinha (1999) and Ibrahim, Chen & Sinha (2001b), among others, which we refer to as the promotion time cure model. This model, which is strongly motivated by biological considerations, is given a Bayesian formulation by Chen, Ibrahim & Sinha (1999). For the ith individual with covariate $\mathbf{Z}_i$ in the population, let N_i be the number of tumour cells which have the potential of metastasizing, that is, N_i is the number of metastasis-competent tumour cells. Assume that N_i has a Poisson distribution with mean $\theta(\mathbf{Z}_i)$. Denote the promotion time for the kth tumour cell by $\tilde{t}_k$, $k = 1, \ldots, N_i$, which is the time for the kth metastasis-competent tumour cell to produce a detectable tumour mass. Conditional on N_i, assume that the $\tilde{t}_k$ are independent and identically distributed with cumulative distribution function $F(t)$, and $S(t) = 1 - F(t)$. Both N_i and $\tilde{t}_k$ are unobservable latent variables. The time to relapse of cancer, which is observed, is defined as $T_i = \min(\tilde{t}_1, \ldots, \tilde{t}_{N_i})$. Therefore, the survival function for the population is given by

$$\begin{aligned} S_{\text{pop}}(t \mid \mathbf{Z}_i) &= \mathrm{P}(N_i = 0) + \mathrm{P}(\tilde{t}_1 > t, \ldots, \tilde{t}_{N_i} > t \mid N_i \geq 1)\,\mathrm{P}(N_i \geq 1) \\ &= \exp\{-\theta(\mathbf{Z}_i)\} + \sum_{k=1}^{\infty} S(t)^k \frac{\theta(\mathbf{Z}_i)^k \exp\{-\theta(\mathbf{Z}_i)\}}{k!} \\ &= \exp\{-\theta(\mathbf{Z}_i)F(t)\}. \end{aligned} \tag{2}$$

The corresponding population hazard function of (2) is $\lambda_{\text{pop}}(t \mid \mathbf{Z}_i) = \theta(\mathbf{Z}_i)f(t)$, where $f(t) = dF(t)/dt$ is the density function corresponding to $F(t)$ and $\theta(\mathbf{Z}_i) = \exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)$. The cure rate for subject i under model (2) is $\lim_{t\to\infty} S_{\text{pop}}(t \mid \mathbf{Z}_i) = \exp\{-\theta(\mathbf{Z}_i)\}$. To make (2) consistent with (1), a natural generalization of (2) is to allow $F(t)$ to depend on a set of covariates $\mathbf{X}_i$, leading to

$$S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) = \exp\{-\theta(\mathbf{Z}_i)F(t \mid \mathbf{X}_i)\}. \tag{3}$$

The mixture and promotion time cure models are the most widely used cure rate models, which may be seen as competitors. Each model offers its own advantages as well as its disadvantages (Ibrahim, Chen & Sinha 2001a). Specifically, the posterior distribution for $\boldsymbol{\beta}$ is always proper for the promotion time cure model, but is always improper for the mixture cure rate model, when a uniform improper prior is given for $\boldsymbol{\beta}$. In this paper, we establish a general class of cure rate models that contain these two cure rate models as special cases. The unified family that we construct is indexed by a link parameter, where a particular value of the parameter yields the mixture cure model, while another value yields the promotion time cure model. This class of models is built through a transformation (Box & Cox 1964) on the population survival function.

The transformation we consider here is conceptually different from the usual Box–Cox transformation applied to the response variable or the covariates, as in linear regression. For example, when the normality assumption of the errors is not satisfied in linear models, the transformed response variable is defined as $Y^{(a)} = (Y^a - 1)/a$ if $a \neq 0$, and $\log(Y)$ if $a = 0$, where a is the transformation parameter and $a \in R^1$ (the real line). Transformation models have been studied for survival data as well. For example, Aranda-Ordaz (1983) proposed to impose a transformation on a conditional probability related to the hazard for grouped failure time data. Breslow & Storer (1985) and Barlow (1985) applied power transformations to the covariate structure to model the relative risk. More recently, Yin & Ibrahim (2005) have proposed a class of transformation models for proper and improper survival functions.

Here, we propose a novel family of cure rate models by imposing the Box–Cox transformation on the population survival function. This class of transformation models is very general, and it links the standard mixture cure model and the promotion time cure model in an attractive and elegant fashion. By adding an extra transformation parameter, the two main formulations of cure rate models are unified together and the resulting modelling structure allows a much wider class of cure rate structures.

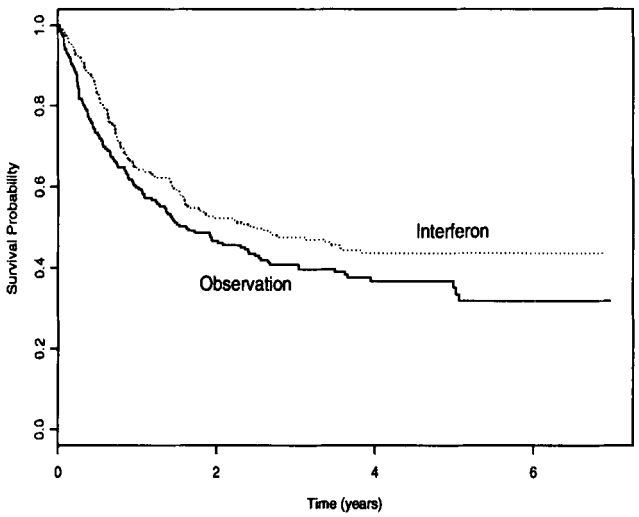

FIGURE 1: Kaplan–Meier curves of high-dose interferon and observation groups in E1690.

Failure time data with a cure fraction is quite common in oncology studies. The analysis of a phase III melanoma clinical trial (E1690) conducted by the Eastern Cooperative Oncology Group (Kirkwood et al. 2000) provides an interesting illustration of this class of cure rate models. The objective of this study was to compare treatment with high-dose interferon to observation (control). Relapse-free survival (in years) was the outcome of interest, which was defined as the time from randomization to progression of tumour or death. There were a total of $n = 427$ patients on these combined treatment arms. Figure 1 shows the estimated survival curves for the interferon and the observation groups. An obvious plateau can be observed after about 5 years of follow-up, which offers empirical evidence for a cure possibility in E1690. If indeed a certain percentage of patients were cured and risk-free, then the conventional proportional hazards assumption (Cox 1972) would be violated in this data set.

The rest of this article is organized as follows. In Section 2, we introduce notation and a new class of cure rate models based on the transformed population survival function. In Section 3, we propose a covariate structure which is general and versatile. The covariate structure naturally satisfies the underlying model constraint and simplifies the proposed method, yielding an unconstrained parameter problem. In Section 4, we formulate the piecewise exponential likelihood function within the Bayesian paradigm, study prior specifications and derive the full conditional distributions needed for Gibbs sampling. In Section 5 we introduce a model selection criterion based on the conditional predictive ordinate (Geisser 1993). We present numerical studies in Section 6, including an in-depth analysis of the E1690 data as well as simulation studies, and give a brief discussion in Section 7.

2. A NEW CLASS OF CURE MODELS

Let T_i $(i = 1, \ldots, n)$ be the failure time for the ith subject, and let $\mathbf{Z}_i$ $(p \times 1)$ and $\mathbf{X}_i$ $(q \times 1)$ be the covariate vectors which may have elements (covariates) in common. The first component

of $\mathbf{Z}_i$ is 1. Let C_i be the censoring variable and $Y_i = \min(T_i, C_i)$ is the observed time. The failure time indicator $\nu_i = 1$ if T_i is observed, and $\nu_i = 0$ otherwise. Assume that T_i and C_i are conditionally independent given $\mathbf{Z}_i$ and $\mathbf{X}_i$.

Analogous to the Box–Cox transformation, we impose a transformation on the population survival function, and propose a class of cure models of the form

$$\frac{S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i)^a - 1}{a} = -\theta(a, \mathbf{Z}_i)F(t \mid \mathbf{X}_i), \quad a \in [0, 1]. \tag{4}$$

It is easy to see that as $a \to 0$, (4) becomes $\log\{S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i)\} = -\theta(0, \mathbf{Z}_i)F(t \mid \mathbf{X}_i)$, and thus reduces to model (3). When $a = 1$, (4) becomes model (1),

$$S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) = 1 - \theta(1, \mathbf{Z}_i)F(t \mid \mathbf{X}_i),$$

where $F(t \mid \mathbf{X}_i) = 1 - S(t \mid \mathbf{X}_i)$.

Our primary interest for a lies in $[0, 1]$, since it yields an intermediate modelling structure between the promotion time cure model ($a = 0$) and the standard mixture cure model ($a = 1$), although a could mathematically take any value on the real line. We can rewrite (4) as,

$$S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) = \left\{1 - a\theta(a, \mathbf{Z}_i)F(t \mid \mathbf{X}_i)\right\}^{1/a}, \tag{5}$$

where the corresponding cure rate for subject i is $\lim_{t\to\infty} S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) = \{1 - a\theta(a, \mathbf{Z}_i)\}^{1/a}$. Thus, we can model a broad class of improper survival functions with a variety of cure fractions for different values of a. The cure rate is jointly determined by $\theta(a, \mathbf{Z}_i)$ and a, but it does not depend on $\mathbf{X}_i$. In (5), we need the constraint $0 \le a\theta(a, \mathbf{Z}_i)F(t \mid \mathbf{X}_i) \le 1$ to be satisfied for all i and t, which can be simplified to $0 \le a\theta(a, \mathbf{Z}_i) \le 1$, since $0 \le F(t \mid \mathbf{X}_i) \le 1$. Constrained parameter problems typically involve unknown normalizing constants in the posterior distribution and thus complicate Bayesian computation and analysis (Gelfand, Smith & Lee 1992; Chen & Shao 1998). If the normalizing constant in the posterior density contains analytically intractable integrals, it is very difficult to implement Gibbs sampling or Metropolis–Hastings algorithms (Chen, Shao & Ibrahim 2000). The nonnegativity of the survival function constraint in our model is very different from and substantially more complicated than the usual order constraints in other Bayesian constrained parameter problems. If the survival function is negative, the likelihood function and the posterior density are not well defined.

3. A GENERAL COVARIATE STRUCTURE

To accommodate different covariate structures for long-term survivors in models (1) and (2), we propose the general form

$$\theta(a, \mathbf{Z}_i) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)}{1 + a\exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)}. \tag{6}$$

When $a = 0$, (6) has the exponential form as in the promotion time model (2), i.e., $\theta(0, \mathbf{Z}_i) = \exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)$, and when $a = 1$, it has the binomial structure of the mixture cure model (1), i.e., $\theta(1, \mathbf{Z}_i) = \exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)/\{1 + \exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)\}$. The interpretation of $\theta(a, \mathbf{Z}_i)$ varies with respect to a. For $a = 0$, $\theta(0, \mathbf{Z}_i)$ has an exponential form, which is analogous to the Cox (1972) proportional hazards model, and $\theta(0, \mathbf{Z}_i)$ can take any positive value. For $a = 1$, $\theta(1, \mathbf{Z}_i)$ is a probability and lies in [0, 1]. For a general a, the cure rate for subject i is $\{1 + a\exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)\}^{-1/a}$. Furthermore, with the class of covariate structures (6), the constraint arising from (5) is automatically satisfied, since

$$0 \le a\theta(a, \mathbf{Z}_i) = \frac{a\exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)}{1 + a\exp(\boldsymbol{\beta}^\top \mathbf{Z}_i)} \le 1, \quad a \in [0, 1],\ i = 1, \ldots, n.$$

Thus, (6) reduces a complex constrained parameter problem to an unconstrained one and facilitates the implementation of Markov chain Monte Carlo (MCMC) in a relatively straightforward fashion.

Based on the proportional hazards model, we incorporate the covariates $\mathbf{X}_i$ through

$$F(t \mid \mathbf{X}_i) = 1 - S(t)^{\exp(\gamma^\top \mathbf{X}_i)}, \tag{7}$$

where $S(t)$ is the baseline survival function, and γ can be viewed as the associated parameter vector for short-term survivors. From (4), the population density function is given by

$$f_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) = -\frac{dS_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i)}{dt} = \theta(a, \mathbf{Z}_i) f(t \mid \mathbf{X}_i) \{1 - a\theta(a, \mathbf{Z}_i) F(t \mid \mathbf{X}_i)\}^{1/a-1},$$

and the population hazard function by

$$\lambda_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) = \frac{\theta(a, \mathbf{Z}_i) f(t \mid \mathbf{X}_i)}{1 - a\theta(a, \mathbf{Z}_i) F(t \mid \mathbf{X}_i)}.$$

4. LIKELIHOOD, PRIOR AND POSTERIOR

We assume a piecewise exponential distribution for the baseline survival function $S(t)$ in (7). The likelihood function is constructed as follows. Let y_i be the observed time for the ith subject, and let J denote the number of partitions of the time axis, i.e., $0 < s_1 < \cdots < s_J$, $s_J > y_i$ for $i = 1, \ldots, n$, and $\lambda_0(y) = \lambda_j$ for $y \in (s_{j-1}, s_j]$, $j = 1, \ldots, J$. When $J = 1$, we obtain a parametric exponential model. There is a trade-off between model flexibility and the number of partitions. By increasing J, the piecewise exponential model can essentially capture any shape of the underlying hazard; this approach is flexible and is commonly used. A reasonable way to partition the time scale is to balance the number of failure times in the time intervals and also guarantee that at least one failure is in each interval. Define $\delta_{ij} = 1$ if subject i fails or is censored in interval j, and 0 otherwise. Let D denote the observed data and let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)^\top$. Then the likelihood function based on the piecewise exponential assumption is given by

$$\begin{aligned} L(a, \beta, \gamma, \boldsymbol{\lambda} \mid D) &= \prod_{i=1}^{n} \prod_{j=1}^{J} \left\{ \frac{\exp(\beta^\top \mathbf{Z}_i)}{1 + a\exp(\beta^\top \mathbf{Z}_i)} f_j(y_i \mid \mathbf{X}_i) \right\}^{\nu_i \delta_{ij}} \\ &\quad \times \left\{ 1 - \frac{a\exp(\beta^\top \mathbf{Z}_i)}{1 + a\exp(\beta^\top \mathbf{Z}_i)} F_j(y_i \mid \mathbf{X}_i) \right\}^{\delta_{ij}(1/a - \nu_i)}, \end{aligned}$$

where

$$f_j(y_i \mid \mathbf{X}_i) = \lambda_j \exp(\gamma^\top \mathbf{X}_i) \exp\left[-\left\{ \sum_{k=1}^{j-1} \lambda_k (s_k - s_{k-1}) + \lambda_j (y_i - s_{j-1}) \right\} \exp(\gamma^\top \mathbf{X}_i) \right], \tag{8}$$

and

$$F_j(y_i \mid \mathbf{X}_i) = 1 - \exp\left[-\left\{ \sum_{k=1}^{j-1} \lambda_k (s_k - s_{k-1}) + \lambda_j (y_i - s_{j-1}) \right\} \exp(\gamma^\top \mathbf{X}_i) \right]. \tag{9}$$

Although not required for the development, we can take a, β, γ and $\boldsymbol{\lambda}$ to be independent a priori, i.e., $\pi(a, \beta, \gamma, \boldsymbol{\lambda}) = \pi(a)\pi(\beta)\pi(\gamma)\pi(\boldsymbol{\lambda})$. We take a discrete uniform prior for $a \in [0, 1]$, for example, $a = a_k$, $k = 1, \ldots, K$ with probability $1/K$, where $a_1 = 0$ and $a_K = 1$. It is appealing to assign discrete probability masses on a set of a values, which can explicitly include the two most popular cure models arising from $a = 0$ and $a = 1$. If a continuous prior on a is used, the model becomes numerically unstable when a is close to 0. This instability can be

avoided with a discrete prior on a. We also assume that the components of $\boldsymbol{\lambda}$ are independent a priori, and each λ_j has a gamma prior distribution, denoted Gamma$(\alpha_\lambda, \xi_\lambda)$, where α_λ and ξ_λ are the hyperparameters. The components of $\boldsymbol{\lambda}$ can be easily correlated a priori by imposing a first-order autoregressive structure or Markovian relation on the λ_j (Arjas & Gasbarra 1994; Ibrahim, Chen & Sinha 2001a). The joint posterior distribution of a, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ is thus given by

$$\pi(a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda} \mid D) \propto L(a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda} \mid D)\pi(a)\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\lambda}). \tag{10}$$

Through a transformation on $S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i)$, we build a general, flexible, and encompassing class of cure rate models that unifies the various cure modelling approaches (i.e., the mixture and promotion time cure models and intermediate structures), and also provides a better fit to the data than a particular member of the class, e.g., models with $a = 0$ and $a = 1$. This class of models provides estimates of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ that are averaged over the cure rate models arising from the different values of a in [0, 1], and thus this extended class of models provides "model-averaged" estimates of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$. Therefore, from a data analytic perspective, the general class in (4) along with (6) provides an extended cure rate modelling structure that is fit to the data once, and by specifying a discrete prior on a, we are able to conduct posterior inference for $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ for several values of a simultaneously. In such a framework, the interpretations of the parameters are based on averaging over the possible values of a, and the posterior estimates of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ are based on the various cure rate models generated by (4) and (6) for a in [0, 1]. As shown in Section 6, there is typically little information in most real applications that would strongly favor a particular value of a.

Assuming a priori independence between $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ as well as a priori independence in the components of $\boldsymbol{\lambda}$, the full conditionals of the parameters are given as follows:

$$\begin{aligned}
\pi(\boldsymbol{\beta} \mid a, \boldsymbol{\gamma}, \boldsymbol{\lambda}, D) &\propto L(a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda} \mid D)\pi(\boldsymbol{\beta}), \\
\pi(\boldsymbol{\gamma} \mid a, \boldsymbol{\beta}, \boldsymbol{\lambda}, D) &\propto L(a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda} \mid D)\pi(\boldsymbol{\gamma}), \\
\pi(\lambda_j \mid a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}_{(-j)}, D) &\propto L(a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda} \mid D)\pi(\lambda_j),
\end{aligned}$$

$$\pi(a = a_k \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, D) = \frac{L(a_k, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda} \mid D)}{\sum_{k=1}^{K} L(a_k, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda} \mid D)}, \quad k = 1, \ldots, K$$

where $\boldsymbol{\lambda}_{(-j)}$ is the $(J-1) \times 1$ vector after the jth component removed, and $\pi(\boldsymbol{\beta}) \sim \mathsf{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, $\pi(\boldsymbol{\gamma}) \sim \mathsf{N}(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma)$, and $\lambda_j \sim$ Gamma$(\alpha_\lambda, \xi_\lambda)$, $j = 1, \ldots, J$. Due to the non-log-concavity of the full conditionals of the parameters, a metropolis step is required within the Gibbs sampler (Gilks, Best & Tan 1995).

5. CHOICE OF J

Misspecification of the modelling structure may bring severe bias into the estimation and lead to incorrect statistical inference. Comparing a set of competing models is critical for a given data set. For the proposed cure models, we use the conditional predictive ordinate statistic for model selection. It is a Bayesian cross-validation approach and measures the adequacy of a given model. The conditional predictive ordinate has been well studied in a variety of contexts (Geisser 1993; Gelfand, Dey & Chang 1992; Dey, Chen & Chang 1997). We use the conditional predictive ordinate (CPO) here *only* to determine the appropriate number of intervals, J. To obtain the CPO, let $\mathbf{y}^{(-i)}$ denote the $(n-1) \times 1$ response vector with y_i deleted, and let $\boldsymbol{\nu}^{(-i)}$, $\mathbf{Z}^{(-i)}$ and $\mathbf{X}^{(-i)}$ be defined similarly. The observed data that result after we delete the ith case can be written as $D^{(-i)} = \{(n-1), \mathbf{y}^{(-i)}, \mathbf{Z}^{(-i)}, \mathbf{X}^{(-i)}, \boldsymbol{\nu}^{(-i)}\}$. Let $g(y_i \mid \mathbf{Z}_i, \mathbf{X}_i, a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ be the density function of y_i, and let $\pi(a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda} \mid D^{(-i)})$ be the posterior density of $(a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})$ given $D^{(-i)}$, $i = 1, \ldots, n$. Then, CPO_i is the marginal posterior predictive density of y_i given $D^{(-i)}$, which can be written as

$$\text{CPO}_i = g(y_i \mid \mathbf{Z}_i, \mathbf{X}_i, D^{(-i)})$$

$$= \int\int\int\int g(y_i \mid \mathbf{Z}_i, \mathbf{X}_i, a, \boldsymbol{\beta}, \gamma, \boldsymbol{\lambda}) \pi(a, \boldsymbol{\beta}, \gamma, \boldsymbol{\lambda} \mid D^{(-i)}) da\, d\boldsymbol{\beta}\, d\gamma\, d\boldsymbol{\lambda}$$

$$= \left\{ \int\int\int\int \frac{\pi(a, \boldsymbol{\beta}, \gamma, \boldsymbol{\lambda} \mid D)}{g(y_i \mid \mathbf{Z}_i, \mathbf{X}_i, a, \boldsymbol{\beta}, \gamma, \boldsymbol{\lambda})} da\, d\boldsymbol{\beta}\, d\gamma\, d\boldsymbol{\lambda} \right\}^{-1}.$$

Let M be the number of Gibbs samples after burn-in, and $a_{[m]}$ $\boldsymbol{\beta}_{[m]}$, $\gamma_{[m]}$ and $\boldsymbol{\lambda}_{[m]}$ are the posterior samples of the mth Gibbs iteration. A Monte Carlo approximation of CPO_i (Chen, Shao & Ibrahim 2000) is given by

$$\widehat{\mathrm{CPO}}_i = \left\{ \frac{1}{M} \sum_{m=1}^{M} \frac{1}{g(y_i \mid \mathbf{Z}_i, \mathbf{X}_i, a_{[m]}, \boldsymbol{\beta}_{[m]}, \gamma_{[m]}, \boldsymbol{\lambda}_{[m]})} \right\}^{-1}.$$

For a failure time, $g(y_i \mid \mathbf{Z}_i, \mathbf{X}_i, a_{[m]}, \boldsymbol{\beta}_{[m]}, \gamma_{[m]}, \boldsymbol{\lambda}_{[m]})$ is given by

$$\prod_{j=1}^{J} \left\{ \frac{\exp(\boldsymbol{\beta}_{[m]}^{\top} \mathbf{Z}_i)}{1 + a_{[m]} \exp(\boldsymbol{\beta}_{[m]}^{\top} \mathbf{Z}_i)} f_{j[m]}(y_i \mid \mathbf{X}_i) \right\}^{\delta_{ij}}$$

$$\times \left\{ 1 - \frac{a_{[m]} \exp(\boldsymbol{\beta}_{[m]}^{\top} \mathbf{Z}_i)}{1 + a_{[m]} \exp(\boldsymbol{\beta}_{[m]}^{\top} \mathbf{Z}_i)} F_{j[m]}(y_i \mid \mathbf{X}_i) \right\}^{\delta_{ij}(1/a_{[m]} - 1)},$$

and for a censoring time, it takes the form of

$$\prod_{j=1}^{J} \left\{ 1 - \frac{a_{[m]} \exp(\boldsymbol{\beta}_{[m]}^{\top} \mathbf{Z}_i)}{1 + a_{[m]} \exp(\boldsymbol{\beta}_{[m]}^{\top} \mathbf{Z}_i)} F_{j[m]}(y_i \mid \mathbf{X}_i) \right\}^{\delta_{ij}/a_{[m]}},$$

where $f_{j[m]}(y_i \mid \mathbf{X}_i)$ and $F_{j[m]}(y_i \mid \mathbf{X}_i)$ are given in (8) and (9) with γ and $\boldsymbol{\lambda}$ replaced by $\gamma_{[m]}$ and $\boldsymbol{\lambda}_{[m]}$, respectively. A common summary statistic based on the CPO_i is $B = \sum_{i=1}^{n} \log(\mathrm{CPO}_i)$. A larger value of B indicates a better fitting model.

6. NUMERICAL STUDIES

6.1. Example.

As an illustration, we applied the class of cure rate models to the E1690 data. The covariates included in this analysis were treatment (high-dose interferon = 1, observation = 0), age (a continuous variable ranging from 19.13 to 78.05 with a mean of 47.93 years, which was standardized) and sex (female = 1, male = 0).

All the Bayesian computations were based on posterior samples recorded every 10th iteration from 100,000 Gibbs samples after a burn-in of 2,000 samples. We chose a uniform discrete prior for a and noninformative priors for the β_ℓ, γ_k and λ_j, independent for $\ell = 0, 1, 2, 3$, $k = 1, 2, 3$ and $j = 1, \ldots, J$. We assigned the prior probabilities for $a = (0, .25, .5, .75, 1)$ as $(.2, .2, .2, .2, .2)$, $\beta_\ell \sim \mathrm{N}(0,\ 10{,}000)$, $\gamma_k \sim \mathrm{N}(0,\ 10{,}000)$, and $\lambda_j \sim \mathrm{Gamma}(2,\ .01)$ which has mean 200 and variance 20,000. Markov chain Monte Carlo convergence was monitored according to the methods recommended by Cowles & Carlin (1996). The Markov chains converged very quickly and the parameters mixed very well.

We considered the class (4) using $J = (1, 2, 3, 4)$. Table 1 summarizes the B statistics, posterior means, standard deviations, and 95% highest probability density (HPD) intervals for the regression parameters in the model. We use the B statistic to help us determine an appropriate partition of the time axis J. Based on the B statistic, the model with $J = 1$ is deemed as the best fitting model. In fact, the posterior estimates are quite close for different values of J and age is clearly an important predictor for long-term survival.

TABLE 1: The B statistics, posterior means, standard deviations, and 95% HPD intervals for the continuous parameters in the E1690 data.

			β			γ		
J	B	Covariate	Mean	SD	HPD Interval	Mean	SD	HPD Interval
1	−522.08	Intercept	.5255	.2594	(.0505, 1.0295)			
		Treatment	−.2380	.2129	(−.6528, .1784)	.0696	.1957	(−.4485, .3129)
		Age	.2356	.1137	(.0241, .4693)	−.1176	.1193	(−.3584, .1126)
		Sex	−.2246	.2138	(−.6368, .1945)	.0231	.2080	(−.3890, .4241)
		λ_1	.7441	.1370	(.4641, .9923)			
2	−523.05	Intercept	.6075	.2717	(.0769, 1.1300)			
		Treatment	−.2431	.2295	(−.6825, .2174)	−.1051	.1949	(−.4874, .2820)
		Age	.2504	.1237	(.0104, .4970)	−.1111	.1142	(−.3480, .1004)
		Sex	−.2384	.2308	(−.7049, .2069)	.0071	.2039	(−.3954, .4026)
		λ_1	.8265	.1506	(.5305, 1.1124)			
		λ_2	.7493	.1372	(.4888, 1.0272)			
3	−522.60	Intercept	.6927	.2856	(.1342, 1.2569)			
		Treatment	−.2194	.2494	(−.6962, .2917)	−.1577	.2036	(−.5593, .2402)
		Age	.2816	.1442	(.0137, .5765)	−.1149	.1182	(−.3518, .1127)
		Sex	−.2558	.2563	(−.7513, .2488)	.0002	.2095	(−.4166, .4046)
		λ_1	.8157	.1506	(.5216, 1.1115)			
		λ_2	.9595	.1708	(.6314, 1.2981)			
		λ_3	.6720	.1469	(.4078, .9754)			
4	−524.18	Intercept	.6994	.2824	(.1264, 1.2388)			
		Treatment	−.2091	.2547	(−.7313, .2706)	−.1838	.2007	(−.5555, .2305)
		Age	.2833	.1492	(.0100, .5837)	−.1124	.1137	(−.3397, .1036)
		Sex	−.2520	.2592	(−.7773, .2380)	.0284	.2074	(−.4578, .3706)
		λ_1	.8200	.1551	(.5220, 1.1250)			
		λ_2	1.0025	.1852	(.6372, 1.3665)			
		λ_3	.8741	.1602	(.5707, 1.1914)			
		λ_4	.7006	.1733	(.3821, 1.0555)			

To evaluate the robustness of our cure rate model with respect to the prior hyperparameters, we conducted a set of sensitivity analyses using $J = 1$. We varied the value of one hyperparameter while keeping others fixed, e.g., the prior standard deviation of β and γ, $\sigma_{\beta/\gamma} = 1{,}000$, and the scale parameter $\xi_\lambda = .001$. Table 2 shows that the proposed model is very robust for a wide range of noninformative priors.

6.2. Simulation.

To examine the performance of the proposed class of cure rate models, we conducted a simulation study. We assumed an exponential distribution for the baseline survival function, i.e., $S(t) = \exp(-\lambda t)$ where we took $\lambda = 1$. We independently generated two covariates: Z_1 had a Bernoulli distribution taking values of 0 or 1 with probability .5; and Z_2 had the standard normal distribution. The failure time data were simulated from model (4) with $a = .5$, and

$$S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i) = \left[1 - \frac{.5\exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)}{1 + .5\exp(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2)}\left[1 - \exp\{-\lambda t \exp(\gamma_1 X_1 + \gamma_2 X_2)\}\right]\right]^2,$$

where $X_1 \equiv Z_1$, $X_2 \equiv Z_2$, and the true parameter values were $\beta_0 = .4$, $\beta_1 = .5$, $\beta_2 = .1$, $\gamma_1 = 1$ and $\gamma_2 = .2$. To generate failure time data with a cure fraction, we calculated the cure rate for each subject, $\theta(.5, \mathbf{Z}_i) = \{1 + .5\exp(\beta^\top \mathbf{Z}_i)\}^{-2}$, and created an indicator denoting whether the subject was cured or not. For uncured subject i, we generated a uniform random variable from Uniform$(\theta(.5, \mathbf{Z}_i), 1)$ and set it equal to $S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{X}_i)$. We took the sample sizes to be $n = 300$, $n = 500$, $n = 1000$ and $n = 2000$. The censoring times were generated from a uniform distribution yielding a censoring rate of 30%.

TABLE 2: Sensitivity analysis with different hyperparameters on the priors for the E1690 data, with $J = 1$.

			β			γ		
$\sigma_{\beta/\gamma}$	ξ_λ	Covariate	Mean	SD	HPD Interval	Mean	SD	HPD Interval
1000	.01	Intercept	.5302	.2584	(.0405, 1.0196)			
		Treatment	−.2453	.2119	(−.6488, .1909)	−.0685	.1984	(−.4542, .3253)
		Age	.2348	.1144	(.0251, .4664)	−.1172	.1197	(−.3477, .1195)
		Sex	−.2243	.2139	(−.6319, .2132)	.0208	.2084	(−.3917, .4312)
		λ_1	.7451	.1373	(.4782, 1.0066)			
100	.001	Intercept	.5186	.2609	(.0363, 1.0089)			
		Treatment	−.2385	.2098	(−.6594, .1627)	−.0688	.1961	(−.4656, .3088)
		Age	.2333	.1127	(.0199, .4623)	−.1189	.1195	(−.3561, .1141)
		Sex	−.2279	.2120	(−.6443, .1901)	.0249	.2098	(−.3765, .4488)
		λ_1	.7418	.1377	(.4790, 1.0102)			

We specified noninformative priors for all the parameters. For each simulated data set, we obtained 3,000 posterior samples from the full conditional distributions after a burn-in of 200, using Gibbs sampling. For each configuration, we conducted 500 replicates. The results are summarized in Table 3. We can see that the posterior means are close to the true values and the posterior standard deviations decrease as the sample size increases. This clearly demonstrates the well-behaved model convergence and estimation properties of this model within the Bayesian framework. The data are usually not very informative on estimating a. We are mainly interested in inference on β and γ. Incorporating a uniform discrete prior on a yields an encompassing model across different values of $a \in [0, 1]$, as opposed to fixing a at a point mass.

7. DISCUSSION

We have carried out an analysis assuming that a is random. Allowing a to be random thus facilitates a full Bayesian solution to the transformation cure model. We conducted a similar Bayesian analysis of the E1690 data based on model (4) while fixing a at a certain value in $[0, 1]$. There seems to be substantial posterior dependence between a and the regression parameters, and the interval estimates for the models treating a as random are generally wider than those under the models with a fixed. Our experience, however, shows that allowing a to be random might not offer additional advantages over the fixed a case. This issue of a random versus a fixed is a general issue that arises for power parameters in parametric models, where the power parameter is the power of an unobservable quantity (the population survival function in our case). When such power transformations are taken, the general principle is that an analysis with the power parameter fixed is as advantageous as an analysis with the power parameter random. This issue was mentioned and elaborated upon in the context of the power priors in Ibrahim & Chen (2000).

We have proposed a class of cure rate models by imposing the Box–Cox transformation on the population survival function. Even though a nonlinear parameter constraint naturally arises from

the new model, our proposed general covariate form in (6) removes the constraint completely. Hence, Gibbs sampling can be easily implemented for the resulting unconstrained model. This class of transformation models makes the cure rate modelling scheme much more flexible and more general than other methods. It nicely links the two main formulations of cure rate models, i.e., the mixture cure model and the promotion time cure model. This family of cure rate models has great potential in clinical trials and for modelling survival data with a cure fraction.

TABLE 3: Simulation results with 500 replications, and the true regression parameter values are $\beta_0 = .4, \beta_1 = .5, \beta_2 = .1, \gamma_1 = 1, \gamma_2 = .2, \lambda_1 = 1$.

		Posterior Estimates						Posterior Probabilities of a				
n	Estimate	β_0	β_1	β_2	γ_1	γ_2	λ_1	$\pi_{0\|D}$	$\pi_{.25\|D}$	$\pi_{.5\|D}$	$\pi_{.75\|D}$	$\pi_{1\|D}$
300	Mean	.54	.58	.11	1.00	.20	1.11	.06	.11	.18	.27	.37
	SD	.24	.26	.13	.16	.08	.16					
500	Mean	.50	.58	.12	.99	.20	1.09	.06	.13	.21	.28	.32
	SD	.20	.20	.10	.13	.06	.13					
1000	Mean	.48	.55	.12	1.01	.20	1.06	.04	.15	.28	.31	.22
	SD	.16	.14	.07	.09	.04	.10					
2000	Mean	.46	.53	.10	1.01	.20	1.04	.00	.11	.49	.31	.09
	SD	.10	.10	.04	.06	.03	.07					

ACKNOWLEDGEMENTS

We would like to thank the Editor, an Associate Editor and two referees for their critical and insightful suggestions, which led to a great improvement of the earlier version. This research was partially supported by the National Cancer Institute Grant 74015 and National Institute of General Medical Sciences Grant GM 070335.

REFERENCES

F. J. Aranda-Ordaz (1983). An extension of the proportional-hazards model for grouped data. *Biometrics*, 39, 109–117.

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

W. E. Barlow (1985). General relative risk models in stratified epidemiologic studies. *Applied Statistics*, 34, 246–257.

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

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

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

N. E. Breslow & B. E. Storer (1985). General relative risk functions for case-control studies. *American Journal of Epidemiology*, 122, 149–162.

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

M. H. Chen & Q. Shao (1998). Monte Carlo methods for Bayesian analysis of constrained parameter problems. *Biometrika*, 85, 73–87.

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

M. K. Cowles & B. P. Carlin (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91, 883–904.

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

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

S. Geisser (1993). *Predictive Inference: an Introduction*. Chapman & Hall, London.

A. E. Gelfand, D. K. Dey & H. Chang (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, April 15–20, 1991* (J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds.), Oxford University Press, pp. 147–167.

A. E. Gelfand, A. F. M. Smith & T. Lee (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, 87, 523–532.

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

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

J. G. Ibrahim & M. H. Chen (2000). Power prior distributions for regression models. *Statistical Science*, 15, 46–60.

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

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

J. M. Kirkwood, J. G. Ibrahim, V. K. Sondak, J. Richards, L. E. Flaherty, M. S. Ernstoff, T. J. Smith, U. Rao, M. Steele & R. H. Blum (2000). High- and low-Dose interferon Alfa-2b in high-risk melanoma: first analysis of intergroup trial E1690/S9111/C9190. *Journal of Clinical Oncology*, 18, 2444–2458.

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

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

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

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

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

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

A. Yu. Yakovlev, B. Asselain, V.-J. Bardou, A. Fourquet, T. Hoang, A. Rochefedière & A. D. Tsodikov (1993). A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. In *Biométrie et Analyse de Données Spatio-Temporelles: 12* (B. Asselain, M. Boniface, C. Duby, C. Lopez, J.-P. Masson & J. Tranchefort, eds.), Société française de biométrie, ENSA Rennes, France, pp. 66–82.

A. Yu. Yakovlev & A. D. Tsodikov (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.

G. Yin & J. Ibrahim (2005). A general class of Bayesian survival models with zero and non-zero cure fractions. *Biometrics*, 61, 403–412.

Received 12 May 2004
Accepted 26 December 2004

Guosheng YIN: gsyin@mdanderson.org
Department of Biostatistics & Applied Mathematics
M. D. Anderson Cancer Center, The University of Texas
Houston, TX 77030, USA

Joseph G. IBRAHIM: ibrahim@bios.unc.edu
Department of Biostatistics, The University of North Carolina
Chapel Hill, NC 27599, USA