Biometrics 61, 552–558
June 2005

DOI: 10.1111/j.1541-0420.2005.040336.x

Bayesian Cure Rate Frailty Models with Application to a Root Canal Therapy Study

Guosheng Yin

Department of Biostatistics and Applied Mathematics, M. D. Anderson Cancer Center,
The University of Texas, Houston, Texas 77030, U.S.A.
email: gyin@odin.mdacc.tmc.edu

Summary. Due to natural or artificial clustering, multivariate survival data often arise in biomedical studies, for example, a dental study involving multiple teeth from each subject. A certain proportion of subjects in the population who are not expected to experience the event of interest are considered to be "cured" or insusceptible. To model correlated or clustered failure time data incorporating a surviving fraction, we propose two forms of cure rate frailty models. One model naturally introduces frailty based on biological considerations while the other is motivated from the Cox proportional hazards frailty model. We formulate the likelihood functions based on piecewise constant hazards and derive the full conditional distributions for Gibbs sampling in the Bayesian paradigm. As opposed to the Cox frailty model, the proposed methods demonstrate great potential in modeling multivariate survival data with a cure fraction. We illustrate the cure rate frailty models with a root canal therapy data set.

Key words: Bayesian inference; Cure fraction; Frailty model; Gibbs sampling; Multivariate failure time data; Proportional hazards.

1. Introduction

As possibly the last endeavor to save a natural tooth before extraction, root canal therapy (RCT) corrects the disorder and dysfunction of the dental pulp when the caries or restorations are deep. RCT usually involves removing the tooth crown and the affected pulpal tissue, followed by cleaning the surrounding infected area to provide a healthy and bondable surface for a permanent filler. After filling, a crown is fabricated to complete the procedure. Many root-canal-filled (RCF) teeth last for a lifetime while some may be lost shortly after completion of endodontic therapy. Non-RCF teeth can be lost due to nonrestorable caries, advanced alveolar bone loss, or catastrophic fracture. Besides the above reasons, RCF teeth can be lost due to endodontic mishaps (e.g., perforation) or postendodontic restorations (e.g., vertical root fracture from intracanal posts).

To quantify the degree to which endodontic involvement affects tooth survival, a retrospective RCT study was conducted in the School of Dentistry at the University of North Carolina at Chapel Hill (Caplan et al., 2005). Using databases of the Kaiser Permanente Northwest Division Dental Care Program, and following the criteria of the study design, 202 eligible patients were identified. Each patient contributed data from one RCF tooth and one similar non-RCF tooth. If the contralateral tooth was present, it was selected as the matching non-RCF tooth. If that tooth was missing or was already endodontically treated, a tooth of the same type (i.e., anterior, premolar, or molar) adjacent to the contralateral tooth was selected. Follow-up for both the RCF and non-RCF teeth started on the index date and continued through the date of extraction or the end of the study, whichever came first. Time to extraction was the outcome of interest which could be censored by the study termination or by the RCT. From a dental scientific perspective, many RCF teeth are considered sound and will last for a lifetime after successful endodontical treatment, and thus can be viewed as "cured."

If a significant number of patients are "cured" and thus risk-free of the disease of interest, the population is then a mixture of susceptible and insusceptible subjects. In these cases, the Cox (1972) proportional hazards model may not be appropriate, because it inherently assumes that all the subjects have the same susceptibility to the disease and will eventually experience the event over a sufficiently long period of follow-up. Cure rate models are intended to model failure time data with a surviving fraction, which becomes increasingly important and popular in clinical trials and medical research, especially in various types of oncology studies, such as breast cancer, leukemia, and melanoma. Let $S^*(t \mid \mathbf{Z}_i^*)$ be a proper survival function (i.e., $\lim_{t\to\infty} S^*(t \mid \mathbf{Z}_i^*) = 0$), where $\mathbf{Z}_i$ and $\mathbf{Z}_i^*$ may share common components and the first component of $\mathbf{Z}_i$ is 1. The mixture cure model proposed by Berkson and Gage (1952) assumes that a certain probability $\theta(\mathbf{Z}_i)$ of being cured is mixed with the remaining $1 - \theta(\mathbf{Z}_i)$ of not being cured,

$$S_{\text{pop}}\big(t \mid \mathbf{Z}_i, \mathbf{Z}_i^*\big) = \theta(\mathbf{Z}_i) + \{1 - \theta(\mathbf{Z}_i)\} S^*\big(t \mid \mathbf{Z}_i^*\big), \qquad (1)$$

where $S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{Z}_i^*)$ is the population survival function and $S^*(t \mid \mathbf{Z}_i^*)$ is that for the uncured subjects. The cure fraction in model (1) is $\lim_{t\to\infty} S_{\text{pop}}(t \mid \mathbf{Z}_i, \mathbf{Z}_i^*) = \theta(\mathbf{Z}_i)$. A logistic regression structure is usually assumed so that

$\theta(\mathbf{Z}_i) = \exp(\boldsymbol{\beta}'\mathbf{Z}_i)/\{1 + \exp(\boldsymbol{\beta}'\mathbf{Z}_i)\}$, where $\boldsymbol{\beta}$ is the parameter vector of interest including an intercept. The mixture cure model (1) has been extensively studied in the literature, including Gray and Tsiatis (1989), Kuk and Chen (1992), Taylor (1995), Maller and Zhou (1996), Sy and Taylor (2000), Peng and Dear (2000), and Betensky and Schoenfeld (2001), among others. 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. In the Bayesian computation, if $\boldsymbol{\beta}$ takes an improper uniform prior, i.e., $\pi(\boldsymbol{\beta}) \propto 1$, the posterior distribution based on (1) is improper (Chen, Ibrahim, and Sinha, 1999). In particular, when modeling the heterogeneity in the population using frailty models, it would be more appealing and convenient to employ the proportional hazards modeling scheme.

An alternative definition of the univariate cure rate model, which has been investigated by Yakovlev and Tsodikov (1996), Tsodikov (1998), Chen et al. (1999), and Ibrahim, Chen, and Sinha (2001a) among others, is given by

$$S_{\text{pop}}(t \mid \mathbf{Z}_i) = \exp\{-\theta(\mathbf{Z}_i)F(t)\}, \tag{2}$$

where $F(t)$ is a proper cumulative distribution function. The corresponding cure rate in model (2) is $\lim_{t\to\infty} S_{\text{pop}}(t \mid \mathbf{Z}_i) = \exp\{-\theta(\mathbf{Z}_i)\}$. For subject i, $\theta(\mathbf{Z}_i)$ is linked to the covariates via the exponential relation $\theta(\mathbf{Z}_i) = \exp(\boldsymbol{\beta}'\mathbf{Z}_i)$. For a more detailed discussion of (2), see Chapter 5 of the book by Ibrahim, Chen, and Sinha (2001b).

The aforementioned methods are based on a critical assumption that the survival times are independent. However, we often encounter multivariate failure time data in biomedical research where the correlation may be induced by natural or artificial clustering effects. In family studies of genetic diseases, litter-matched mice experiments, or ophthalmologic research, observations in the same cluster or group may be correlated. The underlying correlation needs to be carefully adjusted to ensure valid estimation and inference. In the RCT example, each subject contributed a pair of observations from RCF and non-RCF teeth. Clearly, data collected on two teeth from the same individual cannot be assumed to be independent.

Extensive research has been carried out for multivariate failure time data. The frailty model and the marginal model are the most popular approaches. Focusing on subject-specific effects, the frailty model explicitly formulates the nature of the underlying dependence structure. The marginal model takes a population-average approach to model the marginal mean while treating the correlation as a nuisance. In this article, we concentrate on the frailty model which accommodates the intraclass correlation through an unobservable random effect, or a frailty.

For the lth subject in the ith cluster, with a covariate vector $\mathbf{Z}_{il}$ $(i = 1, \ldots, n;\ l = 1, \ldots, L_i)$, the usual Cox shared frailty model is given by

$$\lambda(t \mid \mathbf{Z}_{il}, W_i) = \lambda(t) W_i \exp(\boldsymbol{\beta}'\mathbf{Z}_{il}), \tag{3}$$

where $\lambda(t)$ is the unknown and unspecified baseline hazard function and W_i is the unobservable frailty induced by the ith cluster. Conditional on W_i, the failure times in cluster i are assumed to be independent. The most studied parametric assumption is that the W_i's are independent and identically distributed (i.i.d.) from a gamma density with mean 1 (Clayton, 1978). Another popular distribution for W_i is the positive stable distribution (Hougaard, 2000), which preserves the proportional hazards structure unconditionally (after integrating W_i out). Semiparametric Bayesian methods for multivariate failure time data have been proposed in various contexts (Clayton, 1991; Sinha, 1993; Sargent, 1998; Qiou, Ravishanker, and Dey, 1999; among others).

However, limited attention has been paid to the research in multivariate cure rate models. In the frequentist framework, Chatterjee and Shih (2001) proposed a marginal approach using bivariate Copula models. Price and Manatunga (2001) imposed frailty to account for correlation and conducted the maximum likelihood estimation under a parametric model assumption. Both methods were based on the mixture cure model (1). In the Bayesian paradigm, Chen, Ibrahim, and Sinha (2002) generalized the work of Chen et al. (1999) to bivariate failure time data by introducing a positive stable frailty, where an illustrative example was given for simultaneously modeling two distinct events, i.e., time to cancer relapse and time to death. A recent review paper by Tsodikov, Ibrahim, and Yakovlev (2003) gives a comprehensive treatment and discussion of the development of the cure rate model (2).

Motivated by the RCT study, in which "cure" is apparently a possibility and correlation naturally arises from the paired teeth of the same subject, we propose two new cure rate frailty models for multivariate failure time data with a surviving fraction. The proposed methods are closely related to the univariate cure rate model (2), because it might not be intuitively convenient to incorporate frailty to model (1), and the interpretation could be potentially cumbersome. We formulate the model in a Bayesian framework, so that the prior information can be easily incorporated through historical data.

The rest of this article is organized as follows. In Section 2, we motivate one form of a cure rate frailty model from a clonogenic tumor cell example, and propose the other by mimicking the Cox frailty model. In Section 3, we derive the likelihood functions for the proposed cure frailty models within the Bayesian paradigm, and obtain the full conditional distributions based on suitable prior distributions. In Section 4, we propose a model selection technique using the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002). We illustrate the proposed methods with the RCT example in Section 5, and provide concluding remarks in Section 6.

2. Cure Rate Frailty Models

We introduce a cure rate frailty model that is motivated by the following clonogenic tumor cell example. For the ith individual in the population, let N_i be the number of tumor cells that have the potential of metastasizing, i.e., N_i is the number of metastasis-competent tumor cells. Assume that N_i has a Poisson distribution with mean $\theta(\mathbf{Z}_i)$. Given $N_i = K$, let $(X_{i1}, \ldots, X_{iK})$ be the promotion times for all the K tumor cells in the ith subject. That is X_{ik} $(k = 1, \ldots, K)$ is the time for the kth metastasis-competent tumor cell in subject i to produce a detectable tumor mass. Because the N_i cells belong to the same subject, we assume a random effect W_i for subject i to account for the within-subject correlation among

the X_{ik}'s. Conditional on N_i and W_i, we assume that the X_{ik}'s are i.i.d. from $F(t)$. Here, we emphasize that both N_i and $(X_{i1}, \ldots, X_{iN_i})$ are unobservable random variables. The time to cancer relapse for the ith subject, which is observed, is defined as $T_i = \min(X_{i1}, \ldots, X_{iN_i})$. Therefore, the population survival function for the ith subject is given by

$$\begin{aligned}
&S_{\text{pop}}(t \mid \mathbf{Z}_i, W_i) \\
&\quad = \Pr(N_i = 0) + \sum_{K=1}^{\infty} \Pr(T_i > t \mid N_i = K, W_i) \Pr(N_i = K) \\
&\quad = \Pr(N_i = 0) + \sum_{K=1}^{\infty} \Pr(X_{i1} > t, \ldots, X_{iK} > t \mid N_i \\
&\qquad\qquad = K, W_i) \Pr(N_i = K) \\
&\quad = \exp\{-\theta(\mathbf{Z}_i)\} + \sum_{K=1}^{\infty} \exp\{-W_i \Lambda(t) K\} \\
&\qquad\qquad \times \frac{\theta(\mathbf{Z}_i)^K \exp\{-\theta(\mathbf{Z}_i)\}}{K!} \\
&\quad = \exp[-\theta(\mathbf{Z}_i) + \theta(\mathbf{Z}_i) \exp\{-W_i \Lambda(t)\}], \qquad (4)
\end{aligned}$$

where $\Lambda(t)$ is the common cumulative hazard function of the X_{ik}'s. Note that $\lambda(t) = d\Lambda(t)/dt$, then the population hazard function of (4) is

$$\lambda_{\text{pop}}(t \mid \mathbf{Z}_i, W_i) = \theta(\mathbf{Z}_i) W_i \lambda(t) \exp\{-W_i \Lambda(t)\}. \qquad (5)$$

Model (5), referred to as the promotion time cure rate frailty model, is a generalization of the work of Chen et al. (1999), which did not consider the heterogeneity of metastasizing tumor cells from different subjects. The formulation of model (5) is substantially different from that of Chen et al. (2002) where they incorporated the frailty through the Poisson means to model parallel distinct types of failures. In contrast, we focus on the time to the same type of event while correlation arises from the clustering effects.

We assume W_i in (5) to be a random variable from a gamma distribution with mean 1 and variance η^{-1}, i.e., $W_i \sim Ga(\eta, \eta)$. In the limit $\eta \to \infty$, model (5) reduces to the univariate cure rate model (2). Aside from the biological motivation, model (5) is suitable for correlated failure time data with a cure fraction in a wide variety of contexts. Thus, clustered survival data with a cure fraction, which may not "fit" the definition of a metastasizing tumor cell given above, can still be modeled by (5).

For ease of exposition, we formulate the cure rate frailty models in the following setup. Suppose that there are n clusters, and within cluster i, there are L_i subjects. For $i = 1, \ldots, n$, and $l = 1, \ldots, L_i$, let T_{il} be the failure time for the lth member in the ith cluster, C_{il} be the censoring variable, and $Y_{il} = \min(T_{il}, C_{il})$ be the observed time. Define the censoring indicator $\nu_{il} = I(T_{il} \le C_{il})$, where $I(\cdot)$ is the indicator function. Let $\mathbf{Z}_{il}$ be the $(p+1) \times 1$ vector of bounded covariates, where the first component of $\mathbf{Z}_{il}$ is 1 corresponding to the intercept. Failure times are assumed to be independent of censoring times conditional on $\mathbf{Z}_{il}$. Within cluster i, $\{(T_{il}, C_{il}, \mathbf{Z}_{il}), l = 1, \ldots, L_i\}$ may be dependent but exchangeable.

We first propose a promotion time cure rate frailty model, of which the population hazard is

$$\lambda^{(1)}_{\text{pop}}(t \mid \mathbf{Z}_{il}, W_i) = \lambda(t) W_i \exp\{-\Lambda(t) W_i\} \exp(\boldsymbol{\beta}' \mathbf{Z}_{il}). \qquad (6)$$

As an alternative, we then introduce a different form of cure rate frailty model that is analogous to the Cox frailty model (3). Model (2) can be rewritten as $\lambda_{\text{pop}}(t \mid \mathbf{Z}_i) = f(t) \exp(\boldsymbol{\beta}' \mathbf{Z}_i)$, where $f(t)$ is an unknown baseline density function. Hence, we propose the following cure gamma frailty model,

$$\lambda^{(2)}_{\text{pop}}(t \mid \mathbf{Z}_{il}, W_i) = f(t) W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il}). \qquad (7)$$

Both cure rate frailty models (6) and (7) are constructed to take the within-cluster correlation into consideration, which apparently reduces to the univariate case if $W_i \equiv 1$.

3. Likelihoods and Full Conditionals

We assume a piecewise exponential distribution for the baseline hazard function $\lambda(t)$. The piecewise exponential model is useful and simple for modeling survival data, which serves as a benchmark for comparisons with other semiparametric and fully parametric models. The likelihood function is constructed as follows. Let J be the finite number of partitions of the time axis, i.e., $0 < s_1 < \cdots < s_J$, with $s_J > y_{il}$ for $i = 1, \ldots, n$; $l = 1, \ldots, L_i$. Thus, we have J intervals, $(0, s_1]$, $(s_1, s_2], \ldots, (s_{J-1}, s_J]$, where each interval contains at least one failure and a reasonable way to allocate the data is to balance the number of events among intervals. The piecewise exponential model assumes that $\lambda(y) = \lambda_j$ for $y \in (s_{j-1}, s_j]$, $j = 1, \ldots, J$. Define $\delta_{ilj} = 1$ if the lth subject in the ith cluster fails or is censored in the jth interval, and 0 otherwise. When $J = 1$, namely with no partition, the baseline hazard reduces to that of an exponential distribution with $\lambda(t) \equiv \lambda_1$. By increasing J, we would obtain finer partitions of the time scale such that a more flexible structure of the underlying baseline hazard can be captured. Let D denote the observed data, $\mathbf{W} = (W_1, \ldots, W_n)'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$. The random effects W_i $(i = 1, \ldots, n)$ are usually assumed to follow a gamma distribution, $W_i \sim Ga(\eta, \eta)$, with mean 1 and variance η^{-1}. Thus, the conditional likelihood function concerning model (6) is given by $\mathcal{L}^{(1)}(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid \mathbf{W}, D) = \prod_{i=1}^{n} \mathcal{L}^{(1)}_i(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid W_i, D)$, where

$$\begin{aligned}
&\mathcal{L}^{(1)}_i(\boldsymbol{\beta}, \boldsymbol{\lambda} \mid W_i, D) \\
&\quad = \prod_{l=1}^{L_i} \left[\lambda(y_{il}) W_i \exp\{-\Lambda(y_{il}) W_i\} \exp(\boldsymbol{\beta}' \mathbf{Z}_{il})\right]^{\nu_{il}} \\
&\qquad \times \exp\{-[1 - \exp\{-\Lambda(y_{il}) W_i\}] \exp(\boldsymbol{\beta}' \mathbf{Z}_{il})\} \\
&\quad = \prod_{l=1}^{L_i} \prod_{j=1}^{J} \left\{ \lambda_j W_i \exp\left[- \left\{ \lambda_j (y_{il} - s_{j-1}) \right.\right.\right. \\
&\qquad\qquad \left.\left.\left. + \sum_{q=1}^{j-1} \lambda_q (s_q - s_{q-1}) \right\} W_i \right] \exp(\boldsymbol{\beta}' \mathbf{Z}_{il}) \right\}^{\nu_{il} \delta_{ilj}} \\
&\qquad \times \exp\left\{ -\delta_{ilj} \left(1 - \exp\left[- \left\{ \lambda_j (y_{il} - s_{j-1}) \right.\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.\left. + \sum_{q=1}^{j-1} \lambda_q (s_q - s_{q-1}) \right\} W_i \right] \right) \exp(\boldsymbol{\beta}' \mathbf{Z}_{il}) \right\}.
\end{aligned}$$

Similarly, for model (7), $\mathcal{L}^{(2)}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D) = \prod_{i=1}^{n} \mathcal{L}_i^{(2)}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, W_i, D)$, where

$$\begin{aligned}
&\mathcal{L}_i^{(2)}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, W_i, D) \\
&= \prod_{l=1}^{L_i} \{f(y_{il}) W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il})\}^{\nu_{il}} \\
&\quad \times \exp\{-F(y_{il}) W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il})\} \\
&= \prod_{l=1}^{L_i} \prod_{j=1}^{J} \left[\lambda_j \exp\left\{ -\lambda_j (y_{il} - s_{j-1}) \right.\right. \\
&\qquad \left.\left. - \sum_{q=1}^{j-1} \lambda_q (s_q - s_{q-1}) \right\} W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il}) \right]^{\nu_{il} \delta_{ilj}} \\
&\quad \times \exp\left\{ -\delta_{ilj} \left[1 - \exp\left\{ -\lambda_j (y_{il} - s_{j-1}) \right.\right.\right. \\
&\qquad \left.\left.\left. - \sum_{q=1}^{j-1} \lambda_q (s_q - s_{q-1}) \right\} \right] W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il}) \right\}.
\end{aligned}$$

We take noninformative priors for all the parameters such that the likelihood functions dominate the posterior distributions. Without loss of generality, we assume that $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are independent, and their components are independent, a priori. Specifically, we take $\beta_k \sim N(\mu, \sigma^2)$ for $k = 0, 1, \ldots, p$, and $\lambda_j \sim Ga(\alpha, \gamma)$ for $j = 1, \ldots, J$. Furthermore, we take $W_i \sim Ga(\eta, \eta)$ and assume that $\eta \sim Ga(a, b)$, where the hyperparameters a and b are chosen to yield a large prior variance for W_i.

Let $[U \,|\, V]$ denote the posterior distribution of U given V. For $m = 1, 2$; $k = 0, 1, \ldots, p$; $j = 1, \ldots, J$; and $i = 1, \ldots, n$, the full conditional distributions of the parameters are given as follows:

$$\begin{aligned}
\left[\beta_k \,|\, \boldsymbol{\beta}_{(-k)}, \boldsymbol{\lambda}, \mathbf{W}, D\right] &\propto \mathcal{L}^{(m)}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D) \pi(\beta_k), \\
\left[\lambda_j \,|\, \boldsymbol{\beta}, \boldsymbol{\lambda}_{(-j)}, \mathbf{W}, D\right] &\propto \mathcal{L}^{(m)}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D) \pi(\lambda_j), \\
\left[W_i \,|\, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{W}_{(-i)}, \eta, D\right] &\propto \mathcal{L}_i^{(m)}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, W_i, D) W_i^{\eta - 1} \exp(-\eta W_i), \\
[\eta \,|\, \mathbf{W}, D] &\propto \frac{\eta^{n\eta + a - 1} \left(\prod_{i=1}^{n} W_i \right)^{\eta - 1} \exp\left\{ -\eta \left(\sum_{i=1}^{n} W_i + b \right) \right\}}{\{\Gamma(\eta)\}^n},
\end{aligned}$$

where $\boldsymbol{\beta}_{(-k)}$ is the rest of $\boldsymbol{\beta}$ after deleting the kth component, $\boldsymbol{\lambda}_{(-j)}$ and $\mathbf{W}_{(-i)}$ are defined similarly, and $\pi(\beta_k)$ and $\pi(\lambda_j)$ are the prior densities. In particular for model (7), i.e., $m = 2$, due to the conjugate property, the complete conditional distribution of W_i has the closed form of

$$Ga\left(\eta + \sum_{l=1}^{L_i} \nu_{il}, \eta + \sum_{l=1}^{L_i} \sum_{j=1}^{J} \delta_{ilj} \left[1 - \exp\left\{ -\lambda_j (y_{il} - s_{j-1}) - \sum_{q=1}^{j-1} \lambda_q (s_q - s_{q-1}) \right\} \right] \exp(\boldsymbol{\beta}' \mathbf{Z}_{il}) \right).$$

4. Model Adequacy Evaluation

An important part of selecting regression models is evaluating the adequacy of the model fit. It is critical to compare several competing models for a given data set and select the one that best fits the data.

The DIC, recently proposed by Spiegelhalter et al. (2002), is a Bayesian model selection criterion, which is given by

$$\mathrm{DIC} = \overline{\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})} + p_{\mathrm{Dev}}\,.$$

The term p_{Dev} is a penalty term for model complexity, which is reflected by the effective number of parameters in the model. The deviance is obtained from the conditional likelihood, i.e., $\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = -2 \log \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D)$, $\overline{\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})}$ is the posterior mean of $\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})$, $p_{\mathrm{Dev}} = \overline{\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})} - \mathrm{Dev}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\lambda}})$ and thus

$$\mathrm{DIC} = 2\overline{\mathrm{Dev}(\boldsymbol{\beta}, \boldsymbol{\lambda})} - \mathrm{Dev}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\lambda}})\,,$$

where $\bar{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\lambda}}$ are the posterior means of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, respectively. With noninformative priors, DIC is approximately equivalent to the Akaike information criterion (AIC) proposed by Akaike (1973). Specifically, for the proposed cure frailty models, $m = 1$ and 2,

$$\begin{aligned}
\mathrm{DIC}_m = &-\frac{4}{G} \sum_{g=1}^{G} \log \mathcal{L}^{(m)}\left(\boldsymbol{\beta}_{[g]}, \boldsymbol{\lambda}_{[g]} \,\middle|\, \mathbf{W}_{[g]}, D\right) \\
&+ 2 \log \mathcal{L}^{(m)}(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\lambda}} \,|\, \bar{\mathbf{W}}, D),
\end{aligned}$$

where $\boldsymbol{\beta}_{[g]}$, $\boldsymbol{\lambda}_{[g]}$, and $\mathbf{W}_{[g]}$ are the corresponding posterior samples of the gth Gibbs iteration, $\bar{\mathbf{W}}$ is the posterior mean, and G is the number of Gibbs iterations after burn-in. The smaller the DIC value, the better the model fits.

5. Example

As an illustration, we applied the proposed methods to the RCT data. In this analysis, we had three covariates: the RCF tooth indicator, tooth type, and pocket variables. We combined the anterior and premolar teeth together (nonmolar), because there were relatively fewer anterior teeth. Among 404 teeth in the data set, there were 176 molars and 228 nonmolars (64 anteriors and 164 premolars). Pocket depths had been recorded at six sites for each tooth. If at least one of the six periodontal pockets was $\geq$5 mm, a binary variable took a value of 1 (31%), and was otherwise 0 (69%). Figure 1 shows the Kaplan–Meier survival curves for four groups stratified by the root canal treatment and tooth type. The appropriateness of the application of cure rate models needs to be examined cautiously. The survival curves level off and show plateaus at the tail parts, which suggests a possibility of cure. The length of the follow-up was 2916 days, and the last event occurred at the 2678th day for the RCF group and the 2662th day for the non-RCF group. There were 56 teeth censored between the last event and the end of the study for each group. A general guideline for the proper usage of cure rate models is to have a sufficient period of follow-up, and a strong biological justification for the cure or immune possibility, as in the RCT study.

We incorporated a gamma frailty W_i to account for the correlated observations from the paired teeth of the same patient. We took $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$ to

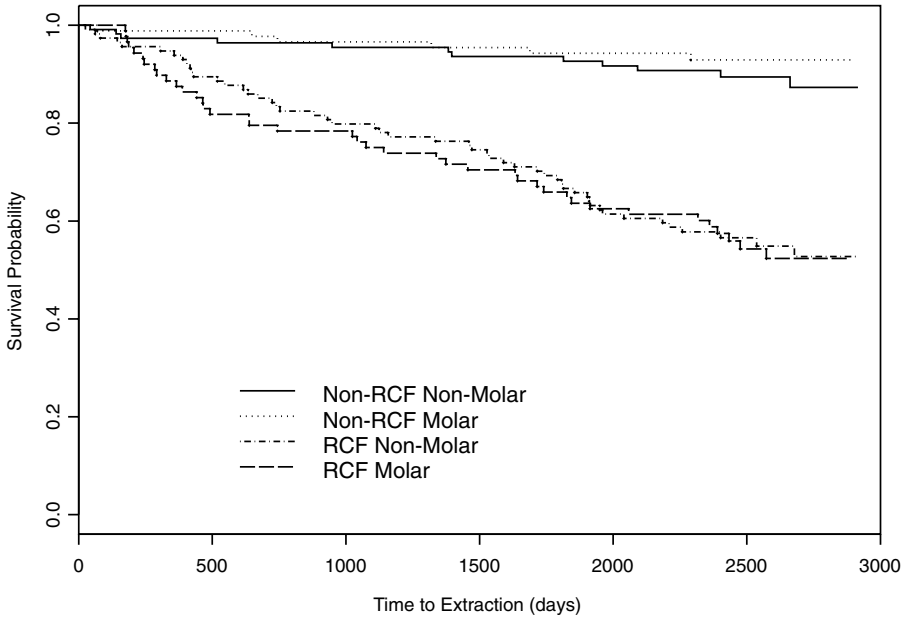

Figure 1. Kaplan–Meier curves stratified by the root canal treatment and tooth type.

be independent, a priori, and gave them noninformative prior distributions, for example, $\beta_k \sim N(0, 100)$ for $k = 0, 1, 2, 3$, and $\lambda_j \sim Ga(\alpha, \gamma)$ with $\alpha = 2$ and $\gamma = 0.1$, and independent for $j = 1, \ldots, J$. For the frailty, we took $W_i \sim Ga(\eta, \eta)$, where $\eta \sim Ga(a, b)$ and we specified $a = 2$ and $b = 100$. We chose priors in such a way that the likelihood functions clearly dominated the posterior distributions. We took the partitions of the time scale, $J = 1$ up to $J = 5$. A larger J gives more flexibility to model the baseline hazard, whereas at the same time it brings in more unknown parameters (the λ_j's) that need to be estimated.

For comparison, we also applied the Cox shared gamma frailty model (3) to the data, for which the corresponding conditional likelihood based on the piecewise constant hazards is $\mathcal{L}^{\text{Cox}}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, \mathbf{W}, D) = \prod_{i=1}^{n} \mathcal{L}_i^{\text{Cox}}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, W_i, D)$, where

$$
\begin{aligned}
&\mathcal{L}_i^{\text{Cox}}(\boldsymbol{\beta}, \boldsymbol{\lambda} \,|\, W_i, D) \\
&\quad = \prod_{l=1}^{L_i} \{\lambda(y_{il}) W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il})\}^{\nu_{il}} \\
&\qquad \times \exp\left[- \int_0^{y_{il}} \{\lambda(t) W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il})\} \, dt \right] \\
&\quad = \prod_{l=1}^{L_i} \prod_{j=1}^{J} \{\lambda_j W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il})\}^{\delta_{ilj} \nu_{il}} \\
&\qquad \times \exp\left[-\delta_{ilj} \left\{ \lambda_j (y_{il} - s_{j-1}) \right.\right. \\
&\qquad\qquad \left.\left. + \sum_{q=1}^{j-1} \lambda_q (s_q - s_{q-1}) \right\} W_i \exp(\boldsymbol{\beta}' \mathbf{Z}_{il}) \right].
\end{aligned}
$$

We ran 30,000 Gibbs samples for each Markov Chain Monte Carlo (MCMC) chain and recorded a sample every five iterations, after 3000 burn-ins. The chains appeared to mix well and the convergence could usually be achieved after 500 iterations. We used the diagnostic methods recommended by Cowles and Carlin (1996) to monitor the chains. Table 1 shows the DICs with respect to three competing models and five different J's. The DIC statistics clearly indicate that the promotion time frailty model with $J = 3$ is the best fitting one, with the smallest DIC = 1818.64. Table 2 summarizes the analysis of the RCT data under the Cox gamma frailty model and the proposed cure rate frailty models, using $J = 1$ and 3, respectively. We present the posterior mean, standard deviation, and 95% highest posterior density (HPD) interval for each parameter. The three different models consistently show that the root canal treatment significantly reduced tooth survival, whereas the tooth type and pocket variables were not important factors. We carried out sensitivity analyses on the prior distributions by varying the hyperparameters (σ and γ). The results in Table 3 demonstrate that the posterior estimation is very robust with respect to a wide range of priors.

Table 1
Model selection criterion based on DIC with respect to J, for the RCT data

	J				
Frailty model	1	2	3	4	5
Cox gamma	2016.99	1988.75	1982.56	1986.36	1988.53
Promotion time	2011.10	1950.43	1818.64	1858.01	1915.93
Cure gamma	2013.17	1997.15	1998.38	2003.41	2007.94

Table 2
The posterior mean, standard deviation, and 95% HPD interval for the RCT data

J	Frailty model	Covariate	Mean	Std. Dev.	95% HPD interval
1	Cox gamma	Root	2.1992	0.2633	(1.7106, 2.7392)
		Molar	−0.0265	0.3498	(−0.7342, 0.6218)
		Pocket	0.0680	0.3228	(−0.5574, 0.7046)
		Frailty (η)	0.2852	0.0353	(0.2192, 0.3545)
	Promotion time	Intercept	−0.1949	0.6382	(−1.3048, 1.1184)
		Root	2.2092	0.2772	(1.7108, 2.8003)
		Molar	0.0338	0.3354	(−0.6279, 0.6911)
		Pocket	0.1380	0.3164	(−0.4595, 0.7800)
		Frailty (η)	0.2733	0.0356	(0.2068, 0.3427)
	Cure gamma	Intercept	−0.0798	0.7629	(−1.4823, 1.5051)
		Root	2.3089	0.2737	(1.7791, 2.8420)
		Molar	0.0493	0.3700	(−0.6472, 0.8095)
		Pocket	0.0939	0.3239	(−0.5045, 0.7376)
		Frailty (η)	0.2646	0.0265	(0.2137, 0.3177)
3	Cox gamma	Root	2.3198	0.2772	(1.8133, 2.8878)
		Molar	−0.2749	0.3746	(−0.9586, 0.5092)
		Pocket	−0.0606	0.3361	(−0.7247, 0.5843)
		Frailty (η)	0.2551	0.0280	(0.2034, 0.3109)
	Promotion time	Intercept	−1.5714	0.3276	(−2.1989, −0.9273)
		Root	2.0613	0.2741	(1.5256, 2.5935)
		Molar	−0.0908	0.2765	(−0.6293, 0.4602)
		Pocket	0.2591	0.2731	(−0.2683, 0.8071)
		Frailty (η)	0.2425	0.0243	(0.1952, 0.2911)
	Cure gamma	Intercept	−1.8507	0.3834	(−2.5860, −1.0704)
		Root	2.5156	0.2958	(1.9098, 3.0688)
		Molar	0.0839	0.3575	(−0.5938, 0.8013)
		Pocket	0.1443	0.3316	(−0.5073, 0.7856)
		Frailty (η)	0.2711	0.0282	(0.2154, 0.3242)

Table 3
Sensitivity analysis with different priors, based on the promotion time frailty model with $J = 3$

σ	γ	Covariate	Mean	Std. Dev.	95% HPD interval
5	0.1	Intercept	−1.6155	0.3288	(−2.2359, −0.9502)
		Root	2.0712	0.2789	(1.4973, 2.5884)
		Molar	−0.0938	0.2829	(−0.6770, 0.4309)
		Pocket	0.2699	0.2746	(−0.2594, 0.8162)
		Frailty (η)	0.2414	0.0292	(0.1890, 0.3015)
100	0.1	Intercept	−1.6279	0.3384	(−2.3064, −0.9908)
		Root	2.0708	0.2747	(1.5857, 2.6549)
		Molar	−0.1156	0.2793	(−0.6547, 0.4321)
		Pocket	0.3237	0.2793	(−0.1996, 0.8959)
		Frailty (η)	0.2398	0.0240	(0.1947, 0.2873)
10	0.5	Intercept	−1.5815	0.3262	(−2.2497, −0.9679)
		Root	2.0855	0.2797	(1.5463, 2.6371)
		Molar	−0.1014	0.2876	(−0.6502, 0.4832)
		Pocket	0.2799	0.2760	(−0.2626, 0.8052)
		Frailty (η)	0.2265	0.0254	(0.1809, 0.2786)
10	0.01	Intercept	−1.6320	0.3318	(−2.2765, −0.9642)
		Root	2.0445	0.2754	(1.4965, 2.5683)
		Molar	−0.1054	0.2658	(−0.6242, 0.3978)
		Pocket	0.3077	0.2717	(−0.2359, 0.8179)
		Frailty (η)	0.2589	0.0287	(0.2064, 0.3174)

6. Remarks

We have proposed cure rate frailty models for multivariate failure time data incorporating a survival fraction in the Bayesian paradigm. The proposed methods have attractive features in model formulation and Bayesian computation. Model (6) has a strong biological motivation while model (7) is statistically and computationally desirable. It is particularly appealing that the full conditional distribution of W_i under model (7) has a closed form due to conjugacy. Both cure frailty models reduce to the same univariate case (2) when all the observations are independent. The existence of insusceptible or immune individuals in the population is the key condition for the applicability of cure rate models. In such situations, censored data are a mixture of cured subjects and uncured subjects who are censored due to incomplete follow-up. For i.i.d. survival data, Maller and Zhou (1995) have investigated the Kaplan–Meier estimator of the cumulative risk function for testing for sufficient follow-up and a heterogeneous population. The development of a corresponding formal test with correlated failure time data is important and requires further investigation.

Acknowledgements

We would like to thank the editor, an associate editor, and two referees for their constructive comments and suggestions, which led to substantial improvement of the article. We thank Donglin Zeng and Joseph Ibrahim for helpful discussions, and Daniel Caplan in the School of Dentistry at the University of North Carolina at Chapel Hill, for providing the RCT data.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *International*

Symposium on Information Theory, B. N. Petrov and F. Csaki (eds), 267–281. Budapest: Akademia Kiado.

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

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

Caplan, J. D., Cai, J., Yin, G., and White, B. A. (2005). Root canal filled versus non-root canal filled teeth: A retrospective comparison of survival times. *Journal of Public Health Dentistry*, in press.

Chatterjee, N. and Shih, J. (2001). A bivariate cure-mixture approach for modeling familial association in diseases. *Biometrics* **57,** 779–786.

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

Chen, M. H., Ibrahim, J. G., and Sinha, D. (2002). Bayesian inference for multivariate survival data with a cure fraction. *Journal of Multivariate Analysis* **80,** 101–126.

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

Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47,** 476–485.

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

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

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.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

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

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

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

Maller, R. A. and Zhou, S. (1995). Testing for the presence of immune or cured individuals in censored survival data. *Biometrics* **51,** 1197–1205.

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

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

Price, D. L. and Manatunga, A. K. (2001). Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine* **20,** 1515–1527.

Qiou, Z., Ravishanker, N., and Dey, D. K. (1999). Multivariate survival analysis with positive stable frailties. *Biometrics* **55,** 637–644.

Sargent, D. J. (1998). A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics* **54,** 1486–1497.

Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association* **88,** 979–998.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64,** 583–616.

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

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

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

Tsodikov, A., Ibrahim, J. G., and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association* **98,** 1063–1078.

Yakovlev, A. Y. and Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. Hackensack, New Jersey: World Scientific.

Received March 2004. *Revised June* 2004.
Accepted August 2004.