Journal of Statistical Research *https://doi.org/10.47302/jsr.2021550108*
2021, Vol. 55, No. 1, pp. 101-125 *ISSN 0256 - 422 X*

A BAYESIAN SEMIPARAMETRIC ACCELERATED FAILURE TIME CURE RATE MODEL FOR CENSORED DATA

SUJIT K. GHOSH*
Department of Statistics, North Carolina State University, NC, USA
Email: sujit.ghosh@ncsu.edu

ELIZABETH KRACHEY
Independent Researcher

SUMMARY

In the modern era of advanced medicine, often a fraction of patients might be cured from a disease and hence the survival probability may plateau at a non-zero value and a cure rate model is needed to capture such survival fractions. A semiparametric accelerated failure time (AFT) cure model is developed for time-to-event data with a positive surviving fraction. The error distribution of the AFT model for susceptible subjects is expressed as a nonparametric mixture of normal densities which can approximate an arbitrary distribution satisfying mild regularity conditions. A Bayesian inferential framework leads to efficient estimation of the posterior distribution of parameters. Posterior consistency of the proposed estimator is established under some regularity conditions providing large sample justification of the proposed model. Markov chain Monte Carlo methods are used to generate samples from the posterior distribution of the regression coefficients to aid statistical inference. Simulation studies are conducted to evaluate the performance of the proposed model in finite samples and an analysis of breast cancer data is also presented to illustrate the method.

Keywords and phrases: Long-term survival; Markov chain Monte Carlo method; Mixture density; Posterior consistency.

AMS Classification: 62C10; 62N02

1 Introduction

A cure model is useful for modeling failure/survival times when a proportion of subjects may never fail and effectively are cured from a disease. As an example, consider a clinical trial for adjuvant therapy for breast cancer originally analyzed by Farewell (1986). Time to relapse or death is used as a failure endpoint and patients are randomized to one of the three treatments. As shown in Figure 1, both the Kaplan–Meier survival curves and the estimated curves based on our proposed model which will be discussed later level off significantly above zero for each treatment group after an extended follow-up over 9 years. Because of this 'plateau' feature, long-term cure may be interpreted as occurring among those patients who remain alive and no longer experiencing

* Corresponding author

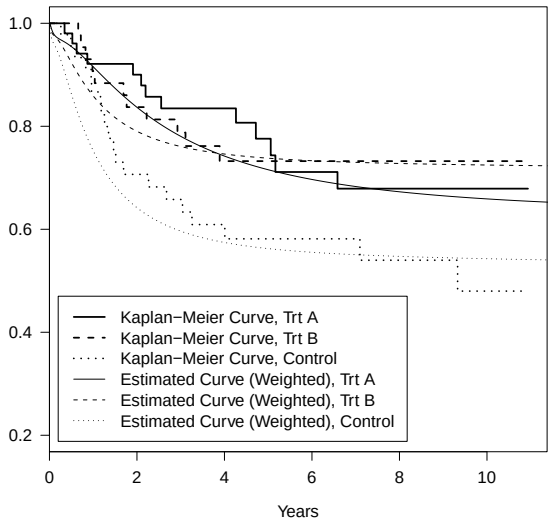

Figure 1: Breast cancer data set: Estimated survival curves using (i) nonparametric Kaplan-Meier estimator and (ii) posterior median of the survival estimates based on the proposed CRAFT model (weighted by lymph nodes and clinical stage) for each of the three treatments

excess mortality due to breast cancer relative to the general population. Cure models provide a means of adjusting the standard survival model to account for this cured fraction. Some popular applications of cure models include Burnett et al. (2000), Secasan et al. (2005), and Dickman and Adami (2006).

There are two commonly used modeling strategies for survival data with a cure fraction. One approach is to use a promotion time cure model of bounded cumulative hazard functions or its variant (Yakovlev and Tsodikov, 1996; Tsodikov, 1998; Chen et al., 1999; Zeng et al., 2006). Such cure models are motivated from a nice biological interpretation, but the short-term and long-term effects generally cannot be separated. Another popular approach is to consider a two-component mixture cure model (Berkson and Gage, 1952), where the short-term (latency) and long-term (incidence) effects are modeled separately with a natural interpretation. For example, the incidence rate is commonly modeled by logistic regression. However, many other link functions such as the probit or complementary log-log can also be used. In addition, quite a few methods have been explored to model the latency. Farewell (1982) originally suggested parametric modeling of the latency by a Weibull distribution. Taylor (1995) used a nonparametric approach in the absence of covariates for the latency component, while Kuk and Chen (1992), Sy and Taylor (2000), and Peng and Dear (2000) considered Cox's proportional hazards (PH) model (Cox, 1972). However, it is quite common that the proportional hazards assumption may be violated in practice. For example, as given in Figure 1, the Kaplan-Meier curves for treatments A and B crossed each other, which is an indication of violation of the PH assumption. We will give a detailed analysis of this breast cancer data in Section 6. When the PH assumption is in question, other semiparametric models, such as the linear transformation model and the AFT model can provide useful alternatives and possibly a better fit to the data. For mixture cure modeling, Lu and Ying (2004) studied the linear transformation cure models using martingale-based estimating equations while (Yamaguchi, 1992) considered the parametric

AFT model with the generalized gamma distribution, and (Li and Taylor, 2002) and Zhang and Peng (2007) explored the semiparameytric AFT cure model using EM-type estimation methods. However, it is to be noted that none of transformation models (which includes PH, AFT, proportional odds etc.) account for the crossing of survival functions. More general conditional hazard function models (e.g., HARE or Bernstein polynomials hazard functions proposed by Osman and Ghosh (2012)) are better suited to capture various features of survival functions. In particular, consider the simplest case of a binary covariate $z \in \{0, 1\}$ and survival time T and let $S(t|z) = \Pr(T > t|z)$ denote the conditional survival function. Under PH assumption, $S_1(t) = S_0(t)^\eta$ for some $\eta > 0$ where $S_0(t) = S(t|z = 0)$ is the baseline survival function. Clearly, $S_1(t) \gtrless S_0(t),\ \forall t$ if and only if $\eta \lessgtr 1$, and thus, the two survival functions can't cross for any $\eta \neq 1$. Similarly, under AFT which assumes $S_1(t) = S_0(t\eta)$ or under PO which assumes $(1 - S_1(t))/S_1(t) = \eta(1 - S_0(t))/S_0(t)$ for some $\eta > 0$, by similar arguments as shown above will not allow for the crossing of survival curves for any $\eta \neq 1$. So, the choice of semi-parametric models needs to be done carefully as illustrated by Sheng and Ghosh (2019). However, in this paper we limit the scope to only AFT models with a cure rate fraction case.

For survival data with cure, a challenging problem is the identifiability of the cure fraction in finite samples since a cure is never observed with censoring. Due to this latent feature, it is more natural to study cure rates within a Bayesian framework where prior information is assumed for both short-term and long-term parameters. It is most common that Bayesian methods have been mainly studied for the bounded cumulative hazards modeling of survival data with cure (Chen et al., 1999; Ibrahim et al., 2001; Zeng et al., 2006; Chi and Ibrahim, 2007; Cooner et al., 2007; Kim et al., 2007; Nieto-Barajas and Yin, 2008; Yin, 2008). However, relatively fewer Bayesian methods have been explored for mixture modeling of cure. Moreover, even less attention is paid to establish posterior consistency in a rigorous manner.

In this paper, we study a Bayesian mixture Cure Rate AFT model, or "CRAFT" model, where the error distribution in the AFT component is modeled as a flexible mixture of normal densities. For survival data without cure, several authors have investigated the standard AFT model using various priors on the error distribution (Kuo and Mallick, 1997; Walker and Mallick, 1999; Campolieti, 2001; Hansen and Johnson, 2004), which have provided very flexible fits of the data and preserved the semiparametric nature of the model. With suitable prior specifications, we establish the consistency of the posterior distribution for the CRAFT model and Markov chain Monte Carlo methodology is used to obtain estimates from the posterior distribution.

The rest of the paper will proceed as follows. Section 2 introduces the CRAFT model and specifies the prior distributions. Section 2.1 provides results on the consistency of the posterior distributions. Implementation methods for obtaining estimates based on the posterior distributions are presented in Section 2.2. Simulation studies are conducted to evaluate the performance of our method in Section 3 and an analysis of breast cancer data is given in Section 4. The technical details and some additional figures are presented in the Appendices.

2 The CRAFT Model with Unspecified Error Distribution

Let η_i indicate whether the ith subject is susceptible ($\eta_i = 1$) or not ($\eta_i = 0$) to the event of interest. Let T_i be the time to occurrence of the event which can be represented as $T_i = T_i^* \eta_i + \infty \cdot (1 - \eta_i)$ where T_i^* denotes the latent survival time when the ith subject is susceptible and we interpret $0 \cdot \infty = 0$. Also, let C_i denote the random censoring time for the i-th patient. Given a p-dimensional vector of covariates Z_i, we assume that T_i is independent of C_i. However, in practice we may not observe the T_i's due to censoring, but

instead observe $X_i = \min(T_i, C_i)$ and the censoring indicator $\Delta_i = I(T_i \leq C_i)$. Thus, we observe the triplet $\{X_i, \Delta_i, Z_i\}$ which are assumed to be independent across $i = 1, \ldots, n$. We denote the set of observed data as $O = \{X_i, \Delta_i, Z_i, i = 1, \ldots, n\}$. We use a logistic model for the incidence and a semiparametric accelerated failure time (AFT) model for the latency. The incidence rate can be modeled as

$$P(\eta = 1 \mid z) = p(\tilde{\gamma}, z) = F_0(\tilde{\gamma}^{\intercal}\tilde{z}), \tag{2.1}$$

where $\tilde{\gamma} = (\gamma_0, \gamma^{\intercal})^{\intercal}$ is a $(p+1)$-dimensional vector of unknown parameters, $\tilde{z} = (1, z^{\intercal})^{\intercal}$, and $F_0(\cdot)$ is a known link function, usually chosen to be a cumulative distribution function. We use the logistic link function where $F_0(u) = \{1 + \exp(-u)\}^{-1}$. Notice that many other link functions such as the probit or complementary log-log could have been used in place of the above logistic link or $p(\tilde{\gamma}, z)$ could also have been modeled using basis expansion non-parametrically. Next, the distribution of the latent survival time T^* can be expressed by the following AFT model:

$$\log T^* = \beta^{\intercal} z + \epsilon, \tag{2.2}$$

where β is a p-dimensional vector of unknown latency regression coefficient parameters, ϵ represents measurement error, and $\text{var}(\epsilon) = \sigma_\epsilon^2$. For simplicity, we assume that the same vector of covariates z is present in the incidence and latency components, but some of variables can be dropped by setting corresponding regression coefficients to zero if needed or by using suitable variable selection priors (e.g., spike-n-slab priors). Also, we assume that the error ϵ follows an infinite mixture of normals with an unknown mixing distribution $H(\cdot)$, where $H(\cdot)$ is an unknown cumulative distribution function satisfying

$$\int (\mu - \mu_\epsilon)^2 dH(\mu) = \left(\frac{k_0}{k_0 + 1}\right)\sigma_\epsilon^2, \tag{2.3}$$

$\mu_\epsilon = \int \mu dH(\mu)$, and $k_0 > 0$ is chosen arbitrarily. For identifiability we assume that k_0 is fixed and can be suitably chosen. More explicitly, the probability density function, $g(\epsilon)$ of ϵ is given by

$$g(\epsilon) = \int \frac{(k_0 + 1)^{1/2}}{\sigma_\epsilon} \phi\left(\frac{\epsilon - \mu}{\sigma_\epsilon/(k_0 + 1)^{1/2}}\right) dH(\mu), \tag{2.4}$$

where $\phi(\cdot)$ is the standard normal density. Notice that the density in Equation (2.4) satisfies the condition $\text{var}(\epsilon) = \sigma_\epsilon^2$ for any cumulative density function $H(\cdot)$ and $k_0 > 0$.

In terms of prior specification, we assume that the mixing distribution $H(\cdot) \sim \Pi_H$. Given H, $\mu \sim H$. Additionally, we assume the precision parameter $\sigma_\epsilon^2 \sim \Pi_{\sigma_\epsilon^2}$, the regression parameter $\beta \sim \Pi_\beta$, and the incidence parameter $\tilde{\gamma} \sim \Pi_{\tilde{\gamma}}$. Let Π stand for product measure $\Pi_\beta \times \Pi_{\tilde{\gamma}} \times \Pi_H \times \Pi_{\sigma_\epsilon^2}$. Define the set of parameters to be estimated as $\theta = \{\beta, \tilde{\gamma}, H(\cdot), \sigma_\epsilon^2\}$. Additional assumptions are discussed in the next section which are used to establish posterior consistency. Having defined both the incidence and latency, the conditional survival function of the CRAFT model is

$$\begin{aligned} P(T > t \mid Z = z) &= S\{t \mid \theta, z\} \\ &= 1 - p(\tilde{\gamma}, z) + p(\tilde{\gamma}, z) P(T^* > t \mid Z = z), \end{aligned} \tag{2.5}$$

where

$$\begin{aligned} P(T^* > t \mid Z = z) &= S_{T^*}\{t \mid \theta, z\} \\ &= \int \left\{1 - \Phi\left(\frac{\log t - \mu - \beta^{\intercal} z}{\sigma_\epsilon/(k_0 + 1)^{1/2}}\right)\right\} dH(\mu), \end{aligned} \tag{2.6}$$

and $\Phi(\cdot)$ is the standard normal cumulative distribution function. The corresponding density function for survival times T is given by

$$f\{t \mid \theta, z\} = p(\tilde{\gamma}, z) \int \phi\left(\frac{\log t - \mu - \beta^{\mathsf{T}} z}{\sigma_\epsilon/(k_0+1)^{1/2}}\right) \frac{(k_0+1)^{1/2}}{\sigma_\epsilon t} dH(\mu). \tag{2.7}$$

Based on right-censored data, the observed likelihood can then be written as

$$L(\theta; O) = \prod_{i=1}^{n} \left\{ f(x_i \mid \theta, z_i) \right\}^{\delta_i} \left\{ S(x_i \mid \theta, z_i) \right\}^{1-\delta_i}, \tag{2.8}$$

where $S(\cdot \mid \theta, z)$ and $f(\cdot \mid \theta, z)$ are defined above in Equations (2.5)-(2.7). Having specified the observed likelihood in Equation (2.8) and prior distributions above, the posterior distribution of the parameters θ given the observed data O is

$$\pi(\theta \mid O) \propto L(\theta; O)\Pi_\beta(\beta)\Pi_{\tilde{\gamma}}(\tilde{\gamma})\Pi_H(H)\Pi_{\sigma_\epsilon^2}(\sigma_\epsilon^2).$$

In practice, we can choose the prior distributions arbitrarily as long as it satisfies some mild regularity conditions as described in the next section.

2.1 Consistency of Posterior Distribution

Asymptotic consistency is a desirable large sample property of the posterior distribution. As Ghosal (2000) discusses, it guarantees that the posterior distribution will concentrate in arbitrarily small neighborhoods of the true value of the parameter, and hence with a sufficiently large amount of data, the truth may be discovered accurately. Diaconis and Freedman (1986) and Ghosh and Ramamoorthi (2003) provide nice discussions and examples of posterior consistency. Moreover, once a consistency is established for a broad class of prior distributions, it also provide robustness of the posterior distribution showing that the inference is not sensitive to the choice of the prior distributions if we had relatively large sample sizes. A formal definition of posterior consistency is as follows.

Definition 2.1 (Posterior Consistency)**.** Suppose $X_1, X_2, \cdots$ are independent and identically distributed according to an unknown density f_*. We take the parameter space as $\mathcal{F}$- a set of probability densities on the space of the observations and consider a prior distribution Π on $\mathcal{F}$. Then the posterior distribution $\Pi(\cdot|X_1, \ldots, X_n)$ of $f \in \mathcal{F}$ given a sample $X_1, \ldots, X_n$ is obtained as,

$$\Pi(A \mid X_1, \ldots, X_n) = \frac{\int_A \prod_{i=1}^n f(X_i) d\Pi(f)}{\int_{\mathcal{F}} \prod_{i=1}^n f(X_i) d\Pi(f)}.$$

We say that the posterior achieves weak posterior consistency at f_* if for any weak neighborhood U of f_*, $\Pi(U|X_1, \ldots, X_n) \to 1$ almost surely as $n \to \infty$.

Sufficient conditions for posterior consistency involving appropriate tests and the prior positivity of a neighborhood defined by the Kullback-Leibler divergence are presented in a theorem by Schwartz (1965). Barron et al. (1999), Ghosal et al. (1999), Ghosh and Ramamoorthi (2003), and Amewou-Atisso et al. (2003) explore useful extensions of Schwartz's theorem for density estimation in mixture models with and without covariates. In-depth justifications and proofs of consistency for the semiparametric accelerated failure time model with infinite Weibull density mixture using Dirichlet priors for the mixing distribution have been presented by Ghosh

and Ghosal (2006). We extend Ghosh and Ghosal's methodology to show consistency for our CRAFT model with infinite normal mixture and hence provide a large sample justification of our Bayesian analysis.

The following assumptions are sufficient to establish posterior consistency.

(A1) The domains of $Z, \beta, \tilde{\gamma}$, and σ_ϵ^2, and the support of H are compact.

(A2) The zero vector is a possible value of the covariate Z, which if not, the covariates may be shifted to satisfy this condition.

(A3) The true density f_* of T given $Z = z$ is a mixture of normal densities

$$f_*(t \mid z) = p(\tilde{\gamma}_*, z) \int_0^\infty \frac{(k_0+1)^{1/2}}{\sigma_{\epsilon_*} t} \phi\left(\frac{\log t - \beta_*^{\mathrm{T}} z - \mu}{\sigma_{\epsilon_*}/(k_0+1)^{1/2}}\right) dH_*(\mu)$$

where $\beta_*, \tilde{\gamma}_*, \sigma_{\epsilon_*}^2$, and H_* are the true values of the parameters $\beta, \tilde{\gamma}, \sigma_\epsilon^2$, and H, respectively.

Notice that the above assumption can be relaxed using the some of the results obtained by Wu and Ghosal (2008). However, since a mixture of normal densities can be used to approximate any bounded continuous density in total variation norm, the assumption about $f_*(t \mid z)$ is not too restrictive. We have fixed, independently distributed variables, with absolutely continuous distribution supporting the vector 0 in $\mathbb{R}^p$. Denote the density of Z at z by $q(z)$. Let $h_\theta(x, \delta, z)$ be the joint density of (X, Δ, Z), so

$$h_\theta(x, \delta, z) = \begin{cases} f\{x \mid \theta, z\} q(z) & \delta = 1 \\ S\{x \mid \theta, z\} q(z) & \delta = 0. \end{cases} \tag{2.9}$$

where $S\{x \mid \theta, z\}$ and $f\{x \mid \theta, z\}$ are defined in Equations (2.5)–(2.7). The class of distributions that are supported in a given compact domain is also compact with respect to the weak topology on the space of probability measures. So the parameter space of $(\beta, \tilde{\gamma}, \sigma_\epsilon^2, H)$ with respect to the product of Euclidean and weak topology is also compact. Hence, the following main theorem applies which verifies that the posterior distribution is consistent.

Theorem 1. *Suppose that the prior densities $\pi(\beta), \pi(\tilde{\gamma})$, and $\pi(\sigma_\epsilon^2)$ for $\beta, \tilde{\gamma}$, and σ_ϵ^2 have compact supports containing $\beta_*, \tilde{\gamma}_*$, and $\sigma_{\epsilon_*}^2$, the base measure of H has compact support that contains the support of H_*, and H_* satisfies the constraint in Equation (2.3). Then under the assumptions (A1)-(A3), the posterior distribution $\Pi((\beta, \tilde{\gamma}, \sigma_\epsilon^2, H) \in \cdot \mid (X_1, \Delta_1), \ldots, (X_n, \Delta_n))$ of $(\beta, \tilde{\gamma}, \sigma_\epsilon^2, H)$ given $(X_1, \Delta_1), \ldots (X_n, \Delta_n)$ is consistent with respect to the Euclidean distances on $\beta, \tilde{\gamma}$, and σ_ϵ^2 and the weak topology on H, that is, given any $\epsilon > 0$ and a weak neighborhood $\mathcal{N}$ of H_*,*

$$\begin{aligned} &\Pi\{(\beta, \tilde{\gamma}, \sigma_\epsilon^2, H) : \\ &|\beta - \beta_*| < \varepsilon, |\tilde{\gamma} - \tilde{\gamma}_*| < \varepsilon, |\sigma_\epsilon^2 - \sigma_{\epsilon_*}^2| < \varepsilon, H \in \mathcal{N} \mid (X_1, \Delta_1), \ldots, (X_n, \Delta_n)\} \to 1 \end{aligned} \tag{2.10}$$

almost surely in $P^\infty_{(\beta_*, \tilde{\gamma}_*, \sigma_{\epsilon_*}^2, H_*)}$- probability.

In showing consistency we have provided a large sample justification of our Bayesian method. The proof of Theorem 1 which establishes the Kullback–Leibler property for the CRAFT model is given in Appendix A.

2.2 Posterior Estimation using MCMC

In practice, the mixing distribution is often well approximated by a finite, possibly sample size-dependent, discrete distribution (Li and Barron, 2000; Komárek et al., 2005). Also, for this CRAFT model the posterior distribution of the parameters cannot be obtained in closed form. However, using a finite discrete approximation to the infinite normal mixture allows off-the-shelf openware programming tools like JAGS (Plummer (2003)) to perform Markov chain Monte Carlo sampling from the posterior distribution. So for practical implementation, we approximate the error density in Equation (2.4) by

$$g_{L_n}(\epsilon) = \sum_{l=1}^{L_n} w_l \frac{(k_0+1)^{1/2}}{\sigma_\epsilon} \phi\Big(\frac{\epsilon - \mu_l}{\sigma_\epsilon/(k_0+1)^{1/2}} \Big), \tag{2.11}$$

where L_n is the number of normal mixtures, $\mu_1 < \cdots < \mu_{L_n}$ is a suitably chosen completely known ordered sequence of knots in $\mathbb{R}$, and $w = (w_1, \ldots, w_{L_n})^\intercal$ are unknown non-negative mixture coefficients satisfying the restrictions (i) $\sum_{l=1}^{L_n} w_l = 1$ and (ii) $\sum_{l=1}^{L_n} \mu_l^2 w_l - (\sum_{l=1}^{L_n} \mu_l w_l)^2 = \{k_0/(k_0+1)\}\sigma_\epsilon^2$, where k_0 is a suitably chosen positive number. The first restriction guarantees that $g_{L_n}(\epsilon)$ is a density function while the second restriction ensures that the variance constraint from Equation (2.3) holds in the finite mixture. We define L with a subscript n to reflect the dependency between the number of normal densities and the sample size, which will be discussed in more detail later on.

Prior distributions may now be specified for a finite number of mixture coefficients w as well as β, $\tilde{\gamma}$, and σ_ϵ^2. If we choose to use a suitable Dirichlet distribution to model the weight vector w then Equation (2.11) provides an approximation to a mixture of Dirichlet process priors (Ishwaran and Zarepour, 2002). However, in this paper we develop a more flexible prior for w which is required to satisfy the additional variance constraint for model identifiability while also automatically penalizing for poor choices of L_n. Extending Komárek et al. (2005)'s penalty term to our Bayesian framework, a prior for w is constructed which avoids over-fitting by utilizing finite highter order differences between adjacent mixture coefficients. A fine grid of knots ensures accurate estimation of the error density while the penalty term restricts the flexibility of the curve. We represent the $w_l's$ as a multivariate logit model; as

$$w_l = \frac{e^{\alpha_l}}{\sum_{l=1}^{L} e^{\alpha_l}} \quad (l = 1, \ldots, L_n - 1),$$
$$w_{L_n} = 1 - \sum_{l=1}^{L_n - 1} w_l,$$

where $\alpha_{L_n} = 0$ and $\alpha_l \in \mathbb{R}$ for $l = 1 \ldots, L_n - 1$. Thus there is no restriction on the α_l's for $l = 1, \ldots, L_n - 1$ and this parameterization allows restriction (i) above to inherently hold.

We derive a prior for $\alpha = (\alpha_1, \ldots, \alpha_{L_n - 1})^\intercal$ which in turn induces a prior on w in a similar spirit as the penalty term used by Komárek et al. (2005). We assume that $\alpha \sim N_{L_n-1}\{0, \lambda^{-1}(D_m^\intercal D_m)^-\}$ where D_m is a $(L_n - m - 1) \times (L_n - 1)$ matrix to represent mth order differences (see Appendix B for more details) and $(D_m^\intercal D_m)^-$ is the Moore-Penrose generalized inverse. Notice that D_m is a matrix of rank $(L_n - m - 1)$ and hence, α has a singular normal distribution (Rao, 1973). But if we define $\tau = D_m \alpha$ then $\tau_l \overset{iid}{\sim} N(0, \lambda^{-1})$ for $l = 1, \ldots, L_n - m - 1$.

Finally, we consider priors for β and $\tilde{\gamma}$ with large variances, and use $\beta \sim N(0, D_\beta)$ and $\tilde{\gamma} \sim N(0, D_{\tilde{\gamma}})$. Each of the variances D_β and $D_{\tilde{\gamma}}$ are adjusted to provide reasonable parameter spaces for each of the incidence

and latency terms. Also, a noninformative inverse Gamma prior is set for the error variance, $\sigma_\epsilon^2 \sim IG(a_0, b_0)$ where $a_0 > 0$ and $b_0 > 0$ are suitably chosen to obtain a prior with large variance.

Ghosh and Ghosal (2006) show explicitly how to perform Markov chain Monte Carlo sampling from the posterior distribution. Latent variables $L^* = (L_1^*, \ldots, L_n^*)$ which indicate the group membership to the μ_l node along with probability vector $w = (w_1, \ldots, w_{L_n})^\intercal$ are introduced. The CRAFT model can then be re-written in hierarchical format as the following:

$$
\begin{aligned}
\log T_i \mid L_i^*, \eta_i &\sim \begin{cases} \text{Normal}\{\beta^\intercal z_i + \mu_{L_i^*}, \sigma_\epsilon^2/(k_0+1)\}, & \delta_i = 1, \eta_i = 1 \\ \text{Normal}\{\beta^\intercal z_i + \mu_{L_i^*}, \sigma_\epsilon^2/(k_0+1)\} I(\log x_i, \infty), & \delta_i = 0, \eta_i = 1 \end{cases} \\
\eta_i \mid \tilde{z}_i &\sim \text{Bernoulli}\{p(\tilde{\gamma}, z_i)\} \\
L_i^* \mid w &\sim \text{Multinomial}\{(1, \ldots, L_n), w\} \\
w &= \text{Multivariate Logit}(\alpha), \quad (\tau = D_m \alpha) \\
\tau &\sim \text{Normal}(0, \lambda^{-1} I_{L_n - m - 1}) \\
\beta &\sim \text{Normal}(0, D_\beta) \\
\tilde{\gamma} &\sim \text{Normal}(0, D_{\tilde{\gamma}}) \\
\sigma_\epsilon^2 &\sim \text{Inverse Gamma}(a_0, b_0)
\end{aligned}
$$

Note that $P(L_i^* = l) = w_l$ for $l = 1, \ldots, L_n$. This hierarchical format is easier to implement in JAGS (`http://mcmc-jags.sourceforge.net/`). We select knots ranging from $\mu_1 = -M$ to $\mu_{L_n} = M$ where $\mu_l = -M + 2M(l-1)/(L_n - 1), l = 1, \ldots, L_n$ for some suitably chosen large $M > 0$. This should be a wide enough range to account for densities that may have large tails, such as the extreme value or logistic distributions. Ishwaran and Zarepour (2002) suggest using $L_n = n^{1/2}$ for large n and $L_n = n$ for small n while Ghosh and Ghosal (2006) say that more work is necessary to determine an optimal number. We will select an order of L_n in a similar fashion, that should allow each normal density to overlap with a few of its neighborhoods and may increase with n. Alternatively, a reversible jump MCMC procedure which has recently become available in WinBUGS could have been used to select L_n (Lunn et al., 2008). Selecting $m = 2, 3,$ or 4 in the difference operator matrix seems to provide good smoothing of the density in simulations. Also, we set $\lambda = 1$ but in the future it may be determined by a cross-validation procedure suggested by Komárek et al. (2005) or estimated using a prior distribution.

With this hierarchical formulation of our model specifications, Markov chain Monte Carlo sampling may be easily performed to generate samples from the marginal posterior distributions of each of the parameters. The posterior mean and standard deviation are used for the measure of center and spread for the mixture coefficients, while the posterior median and standard deviation are used for all other estimated parameters. One can implement this by using the `jags.model` function available in the `rjags` package of R which calls JAGS to generate samples from the posterior distributions using Markov chain Monte Carlo methods. Convergence diagnostics were performed using the "CODA" package available in R. See Appendix C for a snippet of the JAGS code. Many other R packages (e.g., `runjags`) can be used (e.g., see `https://bayessm.wordpress.ncsu.edu/` for various examples).

3 Numerical Illustrations using Simulated Data

We conduct several numerical studies using simulated data to explore how well the estimators obtained from the posterior distributions are performing under a variety of settings. We consider two types of covariates, a continuous valued z_1 uniformly distributed between 0 and 1 and binary-valued z_2 variable obtained from a Bernoulli distribution with mean 0.5. Covariates are centered by their population mean, which is 0.5 for both variables. Such centering is known to reduce the posterior cross correlations between the regression coefficients and hence, leads to more efficient Markov chain Monte Carlo sampling mixing (Roberts and Sahu, 2001).

The incidence portion of the model is $\log\{p(\tilde{\gamma}, z)/[1 - p(\tilde{\gamma}, z)]\} = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2$. We set $\gamma_0 = 0.5$ or 1.0, $\gamma_1 = 1.0$, and $\gamma_2 = -1.0$. Cure fraction and censoring percentages correspond to 39% and 42% when $\gamma_0 = 0.5$ and 28% and 32% when $\gamma_0 = 1.0$. The latency portion of the model is $\log T^* = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \epsilon$ where $\beta_0 = -1$, $\beta_1 = -1.0$ and $\beta_2 = 1.0$. We consider symmetric, skewed, and bi-modal distributions for ϵ, each of which has mean 0 and variance 0.25. These include (i) Logistic$(0, 0.28)$ with density function $g(\epsilon; a, b) = \exp\{-(\epsilon-a)/b\}/(b[1+\exp\{-(\epsilon-a)/b\}]^2)$, (ii) EV$(0.23, 0.39)$ where EV represents the extreme value distribution with density function $g(\epsilon; a, b) = b^{-1}\exp\{(\epsilon - a)/b\}\exp[-\exp\{(\epsilon - a)/b\}]$, and (iii) mixture of two normal densities: $N(-0.45, 0.04)$ and $N(0.45, 0.055)$, with density function $g(\epsilon; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = (2\pi\sigma_1^2)^{-1/2}\exp\{(\epsilon - \mu_1)^2/(2\sigma_1^2)\} + (2\pi\sigma_2^2)^{-1/2}\exp\{(\epsilon - \mu_2)^2/(2\sigma_2^2)\}$.

The censoring times are generated from a uniform$(0, 10)$ distribution. The observed times are hence the minimum of the failure and censoring times. We use sample sizes of $n = 100$ or 200 subjects. In terms of prior specification, we set β_1, β_2, γ_1, $\gamma_2 \sim N(0, 3)$ and the incidence intercept $\gamma_0 \sim N(0, 1.5)$, which leaves prior densities to be relatively flat compared to that of the posterior density, making the choice of priors insensitive to posterior inference. However, recall that any prior distribution that has relatively much larger dispersion compared to that of posterior can be used as indicated by the theoretical result (see Section 2.1), any prior with full support of the parameter space is sufficient for asymptotic consistency. Also, we set $a_0 = 0.1$ and $b_0 = 10(k_0 + 1)$ so $\sigma_\epsilon^2 \sim IG(0.1, 10(k_0 + 1))$. The mixture of normal densities used to estimate the error term range from $\mu_l = -4, \ldots, 4$ with $L_n = 17$ nodes which roughly scales like $O(\sqrt{n})$. We set $m = 2$ in the difference operator matrix to control the smoothing of the normal mixtures. The prior on τ_l used in the error densities is $\tau_l \sim N(0, \lambda^{-1})$ for $l = 1, \ldots, 14$ where $\lambda = 1$. For each simulation, the posterior generation in WinBUGS is based on 4000 burn-ins and an additional 2000 runs from three parallel chains, for a total of 6000 Markov chain Monte Carlo samples per simulated data.

For all these simulation scenarios, 1000 Monte Carlo samples were generated to study the sampling variability of the posterior estimates of $\beta_1, \beta_2, \gamma_0, \gamma_1,$ and γ_2. In order to measure the empirical performance of these posterior estimates, we computed the bias of the posterior median, the Monte Carlo standard error of the posterior median (MCSE), the Monte Carlo average of the posterior standard deviation (ESE), and the nominal coverage probabilities of covering the true value based on the 95% posterior interval (CP). Original simulations resulted in considerable biases associated with the γ_0 estimates, even with a more precise prior. To increase precision, we consider censored subjects with event times greater than the maximum of the uncensored failure times as cured.

Tables 1-3 provide results of the Monte Carlo simulations for different density types ranging from symmetric to skewed to bi-modal. The posterior medians perform well in all simulations except for the γ_1 estimates, corresponding to the continuous covariate. For all error densities, these estimates tend to be biased when $n = 100$. In the higher cure fraction case, the bias diminishes for $n = 200$. In the lower cure fraction cases, the

Table 1: Results of the simulation study based on a CRAFT model with symmetric error distribution (using 1000 MC replications).

Symmetric		$n = 100$				$n = 200$			
	TRUE	BIAS	MCSE	ESE	CP	BIAS	MCSE	ESE	CP
β_1	-1.0	0.01	0.23	0.24	0.95	0.01	0.16	0.16	0.95
β_2	1.0	0.00	0.14	0.14	0.97	0.00	0.09	0.10	0.96
γ_0	0.5	0.02	0.22	0.23	0.96	0.00	0.16	0.16	0.94
γ_1	1.0	-0.14	0.68	0.73	0.96	-0.04	0.50	0.53	0.96
γ_2	-1.0	0.00	0.44	0.45	0.95	0.01	0.31	0.32	0.95
β_1	-1.0	0.01	0.20	0.22	0.96	0.01	0.15	0.15	0.95
β_2	1.0	-0.01	0.12	0.13	0.97	0.00	0.09	0.09	0.96
γ_0	1.0	0.01	0.25	0.25	0.95	0.02	0.17	0.17	0.95
γ_1	1.0	-0.19	0.70	0.77	0.96	-0.08	0.54	0.57	0.96
γ_2	-1.0	0.02	0.46	0.48	0.97	0.03	0.34	0.34	0.95

biases are statistically insignificant at a sample size of $n = 200$ and becomes almost zero for larger sample size simulations (not shown) corroborating the large sample consistency. Overall, the posterior medians perform well and any biases are due to small sample sizes. Coverage probabilities are close to their nominal values in all settings. The Monte Carlo standard errors of the posterior medians and Monte Carlo averages of the posterior standard deviations are fairly close and any differences reduce to a minimal amount when $n = 200$. Again, the posterior estimates seem to be performing well, with the only complication being some bias in the continuous covariate effects in small sample sizes. Also, the estimated mixture densities capture the true underlying error distributions very well for all simulation scenarios. As illustration, Figure 2 shows the estimated and true survival distributions when $n = 200$ and $\gamma_0 = 1$ for each of the three error distributions. It may be noted that above findings are limited by the case studies that we have performed under different design settings and in general, it is not possible to determine sample sizes to achieve a desired reduction in biases.

4 Analysis of Breast Cancer Data

Farewell (1986) analyzed the breast cancer data set described in the introduction to demonstrate the effectiveness of a Weibull-cure mixture model. Kuk and Chen (1992), Peng and Dear (2000), and Lu and Ying (2004) have each re-analyzed the same data set for different proportional hazards cure models. Lu and Ying (2004) also analyzed the breast cancer data set for a proportional odds cure model.

Table 2: Results of the simulation study based on a CRAFT model with skewed error distribution (using 1000 MC replications).

Skewed		$n = 100$				$n = 200$			
	TRUE	BIAS	MCSE	ESE	CP	BIAS	MCSE	ESE	CP
β_1	-1.0	0.01	0.22	0.24	0.96	0.01	0.16	0.16	0.95
β_2	1.0	0.00	0.13	0.14	0.96	0.00	0.09	0.09	0.96
γ_0	0.5	0.02	0.22	0.23	0.96	0.00	0.16	0.16	0.94
γ_1	1.0	-0.13	0.68	0.73	0.96	-0.04	0.50	0.53	0.96
γ_2	-1.0	0.00	0.44	0.45	0.95	0.01	0.31	0.32	0.95
β_1	-1.0	0.01	0.20	0.22	0.97	0.00	0.14	0.15	0.95
β_2	1.0	0.00	0.12	0.13	0.97	0.00	0.08	0.09	0.97
γ_0	1.0	0.02	0.25	0.25	0.95	0.02	0.17	0.17	0.96
γ_1	1.0	-0.18	0.70	0.77	0.97	-0.08	0.54	0.57	0.96
γ_2	-1.0	0.02	0.46	0.48	0.96	0.03	0.34	0.34	0.94

The data set consists of $n = 139$ patients. In addition to the two treatment indicator variables, two additional binary covariates are considered: a clinical stage indicator and an indicator for the number of lymph nodes. In the original data set as presented by Farewell (1986) two more covariates were included, pathological stage and histological stage, but unfortunately these data are no longer available. The censoring percentage is 68%, with 95 patients censored and 44 uncensored. Recall Figure 1 shows the Kaplan–Meier survival curves for each of the three treatment groups, each of which levels off significantly above zero and provides empirical evidence in support for a cure model for the data set.

There is some empirical evidence that the proportional hazards assumption may not be valid. This is seen by plotting the logarithm of the cumulative hazard function for the uncensored patients in each treatment group, based on the Kaplan–Meier curve, as performed by Zhang and Peng (2007) for a different data set. The uncensored subjects are assumed to be reasonably close to those that are uncured. In this instance, the logarithm of the cumulative incidence function of the uncensored subjects nearly approximates that of the uncured subjects. This is also shown in Figure 3, where treatment A does not seem parallel to the other treatments. This provides empirical evidence that the proportional hazards assumption may not hold. However, the crossing of the cumulative hazard (and hence that of survival curves) may also indicate violation of AFT model.

We fit the breast cancer data set under several settings of our proposed semiparametric Bayesian model. The logarithm of the failure times are centered by either their mean, or the mean of those observations that experience a failure. This allows the normal mixture nodes to range from -4 to 4 and easily capture the spread of

Table 3: Results of the simulation study based on a CRAFT model with bimodal error distribution (using 1000 MC replications).

Bimodal		$n = 100$				$n = 200$			
	TRUE	BIAS	MCSE	ESE	CP	BIAS	MCSE	ESE	CP
β_1	-1.0	0.04	0.20	0.22	0.97	0.01	0.11	0.12	0.97
β_2	1.0	-0.01	0.11	0.13	0.98	-0.01	0.06	0.07	0.97
γ_0	0.5	0.01	0.23	0.23	0.94	0.01	0.16	0.16	0.95
γ_1	1.0	-0.16	0.66	0.73	0.97	-0.07	0.52	0.53	0.96
γ_2	-1.0	0.06	0.43	0.45	0.95	-0.01	0.31	0.32	0.96
β_1	-1.0	0.02	0.17	0.19	0.97	0.01	0.10	0.10	0.95
β_2	1.0	0.00	0.10	0.11	0.97	0.00	0.06	0.06	0.96
γ_0	1.0	0.01	0.24	0.25	0.95	0.01	0.18	0.17	0.94
γ_1	1.0	-0.12	0.72	0.77	0.96	-0.10	0.54	0.57	0.95
γ_2	-1.0	0.04	0.47	0.48	0.95	0.01	0.36	0.34	0.95

the data. The data is fit under all combinations of $L_n = 17$ or 25 and $m = 2, 3$, or 4. The priors on w_l penalize high-dimensional models and help avoid overly complex models. Hence, provided that model complexity measures remain stable, there is no need for model selection to include a second penalization for complex models as measured by DIC. Instead model selection is based on goodness of fit, measured by deviance. In terms of prior specification, we set $\beta_1, \beta_2, \gamma_1, \gamma_2 \sim N(0, 3)$ and the incidence intercept $\gamma_0 \sim N(0, 1.5)$. Also, $a_0 = 0.1$ and $b_0 = 10(k_0 + 1)$ so $\sigma_\epsilon^2 \sim Ga(0.1, 10(k_0 + 1))$. Posterior estimation is based on 4000 burn-ins and an additional 5000 runs from three parallel chains, for a total of 15,000 Markov chain Monte Carlo samples per estimate.

Based on small levels of deviance while maintaining stable DIC values, we select the model where $L_n = 17$ and $m = 4$. The logarithm of the failure times are centered by the mean of those observations that are uncensored. The CODA package (Plummer et al. (2006)) available in R is used to perform Markov chain Monte Carlo diagnostics for the chosen model. The Gelman–Rubin 97.5% shrink factors for all statistics are ≤ 1.03, indicating good mixing and good convergence to the appropriate distributions. Appendix D provides additional diagnostic information including the trace and posterior density plots for the incidence and latency parameters and a plot of the posterior means of the mixing coefficients. The former indicates that the chains have thoroughly mixed while the later indicates that the number and location of nodes used in the normal mixture seem to be capturing the spread of the data well.

The results based on the above model parameters are summarized in Table 4. Descriptive statistics based on posterior distributions for the covariates in both the incidence and latency portions of the model include the posterior median and the posterior 2.5% and 97.5% percentiles.

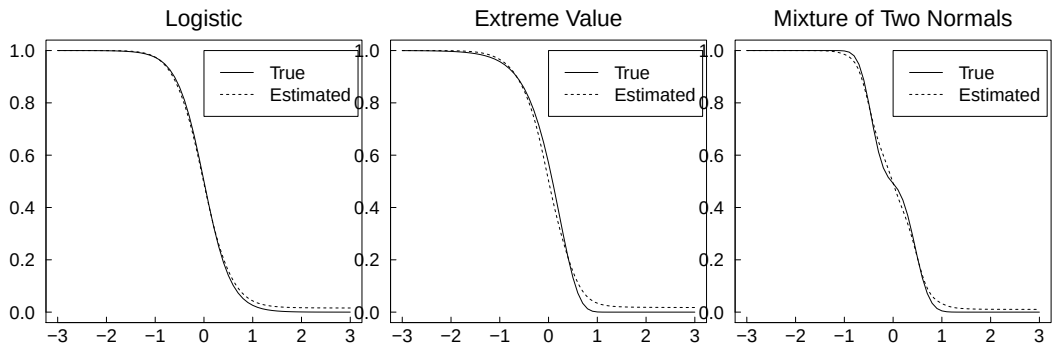

Figure 2: True and estimated (based on CRAFT model) survival probabilities for data generated from symmetric (logistic), skewed (extreme value), and bimodal (mixture of two normals) distributions when $n = 200$ and $\gamma_0 = 1$ for Monte Carlo simulations of size 1000.

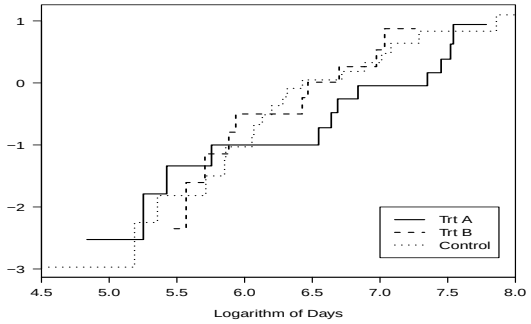

Figure 3: Breast cancer data set: The logarithm of the cumulative hazard function curves for each of the three treatments

Point estimates will not be compared to previous authors because the interpretation is different, as we are using the accelerated failure time model. A formal Bayesian testing for significance can be performed using Bayes factor, but conclusions may also be based on the concentration of posterior distribution around zero, namely, between the 2.5% and 97.5% percentiles. These results can be compared to p-values obtained from previous results. In general, statistical significance tends to be in agreement with those found by Peng and Dear (2000) and Lu and Ying (2004), as described below.

In terms of latency parameters, treatment A and clinical stage indication both have positive significant effects on short term survival times. Peng and Dear (2000) found these same results. For the incidence parameters based on a 0.05 cut-off criterion, both lymph nodes and clinical stage are significant with negative and positive effects; respectively, on long-term survival. Also, zero is just inside the tail cut-off value of 97.5% for the posterior distribution of treatment B, showing evidence of a positive effect on long-term survival. These three

Table 4: Breast cancer data: Bayesian posterior summaries for both the latency and incidence statistics associated with treatment A or B, clinical stage indication, and number of lymph nodes in the breast cancer data set using the CRAFT model

Latency: β	2.5%	50%	97.5%
Trt A	0.16	0.98	1.70
Trt B	−0.46	0.14	0.87
Clinical Stage I	0.01	0.77	1.39
Lymph Nodes	−0.88	−0.34	0.35

Incidence: γ	2.5%	50%	97.5%
Trt A	−1.57	−0.46	0.66
Trt B	−1.99	−0.97	0.03
Clinical Stage I	−1.79	−0.89	−0.01
Lymph Nodes	0.38	1.41	2.62
Intercept	−0.59	0.24	1.13

statistics were all found to be significant or nearly significant for both Peng and Dear (2000) and Lu and Ying (2004).

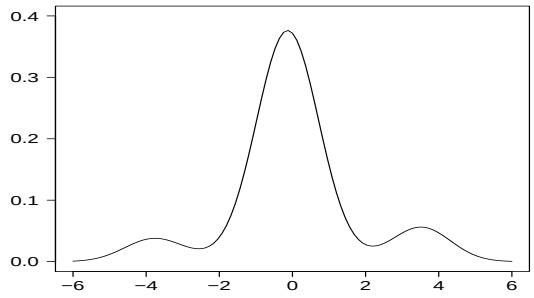

Figure 4: Breast cancer data: posterior median of the error density estimates from the CRAFT model using a mixture of 17 normal distributions

In addition to descriptive statistics, the estimated survival distributions based on posterior medians are computed for each treatment, weighting by lymph nodes and clinical stage. Weighting is performed based on the corrected group prognosis method (Chang et al., 1982). For a particular treatment, four survival curves are computed corresponding to each level of lymph nodes and clinical stage. The estimated survival curve is then constructed as the weighted average of these four curves, where the weights are proportional to the number of subjects at each level of lymph nodes and clinical stage. Figure 1 provides this plot, overlayed with the Kaplan-Meier estimates (KMEs) for each treatment, illustrating that the estimated survival curves appear to be capturing the survival distribution relatively well in the long-run. Estimated curves deviates from the empirical KMEs at shorter survival times (as the KMEs are not adjusted for baseline covariates), but the estimated curves still seem to capture the crossing survival curves feature. Figure 4 is a plot of the estimated error density based

on posterior medians, indicating that the number and location of normal mixtures seem to be capturing the error density well. Clearly, a specific parametric family would not be able to adapt to such trimodal features unless specifically known beforehand.

5 Conclusion

We have introduced a semiparametric accelerated failure time cure model within a Bayesian framework. Modeling the error term as a mixture of normal densities provides an intuitive and useful means for semiparametric estimation. However, several mixtures densities may have been considered instead of the normal. Specifically, combining our work with that of Ghosh and Ghosal (2006), posterior consistency based on a mixture of Weibull densities holds and could have sufficed in the cure rate accelerated failure time model. Also, we assume that the same vector of covariates z is present in the incidence and latency components. In practice, it is possible to partition the covariates into these two model components (Li and Taylor, 2002) or use variable selection methods (Mitchell and Beauchamp, 1988; Kinney and Dunson, 2007) to choose appropriate subsets of covariates for the incidence and latency. The penalized prior has been adapted from Komárek et al. (2005) to control smoothing but with the smoothing parameter fixed at a reasonable level. Future work may investigate selection of an optimal smoothing mechanism and/or an optimal number of normal densities used in the mixture. And lastly, although the model has been developed for a univariate response with right-censored data, it may be extended to multivariate survival analysis or to other various censoring scenarios, such as interval-censored data.

Acknowledgement

The first author would like to thank Dr. Abdus S. Wahed, Professor in the Department of Biostatistics at University of Pittsburgh for the kind invitation to submit this manuscript for 50-th year of Bangladesh's independence, and the 50th year of JSR's publication.

References

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (2003), "Posterior Consistency for Semi-parametric Regression Problems," *Bernoulli*, 9, 291–312.

Barron, A., Schervish, M., and Wasserman, L. (1999), "The Consistency of Posterior Distributions in Nonparametric Problems," *Annals of Statistics*, 27, 536–561.

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

Burnett, N. G., Benson, R. J., Williams, M. V., and Peacock, J. H. (2000), "Improving Cancer Outcomes Through Radiotherapy," *British Medical Journal*, 320, 198–199.

Campolieti, M. (2001), "Bayesian Semiparametric Estimation of Discrete Duration Models: An Application of the Dirichlet Process Prior," *Journal of Applied Econometrics*, 16, 1–22.

Chang, I., Gelman, R., and Pagano, M. (1982), "Corrected Group Prognostic Curves and Summary Statistics," *Journal of Chronic Diseases*, 35, 669–674.

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.

Chi, Y. Y. and Ibrahim, J. G. (2007), "Bayesian Approaches to Joint Longitudinal and Survival Models accommodating both Zero and Nonzero Cure Fractions," *Statistica Sinica*, 17, 445–462.

Cooner, F., Banerjee, S., Carlin, B. P., and Sinha, D. (2007), "Flexible Cure Rate Modeling Under Latent Activation Schemes," *Journal of the American Statistical Association*, 102, 560–572.

Cox, D. R. (1972), "Regression Models and Life Tables (with Discussion)," *Journal of the Royal Statistical Society, Series B*, 34, 187–220.

Diaconis, P. and Freedman, D. (1986), "On the Consistency of Bayes Estimates (with Discussion)," *Annals of Statistics*, 14, 1–26.

Dickman, P. W. and Adami, H.-O. (2006), "Interpreting Trends in Cancer Patient Survival," *Journal of Internal Medicine*, 260, 103–117.

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?" *The Canadian Journal of Statistics*, 14, 257–262.

Ghosal, S. (2000), "Dirichlet Process, Related Priors, and Posterior Asymptotics," Tech. rep., Department of Statistics, North Carolina State University, Raleigh, North Carolina.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1999), "Posterior Consistency of Dirichlet Mixtures in Density Esimation," *The Annals of Statistics*, 27, 143–158.

Ghosh, J. K. and Ramamoorthi, R. V. (2003), *Bayesian Nonparametrics*, New York: Springer-Verlag.

Ghosh, S. K. and Ghosal, S. (2006), "Semiparametric Accelerated Failure Time Models for Censored Data," *Bayesian Statistics and its Applications*, 15, 213–229.

Hansen, T. and Johnson, W. (2004), "A Bayesian Semiparametric AFT Model for Interval-Censored Data," *Journal of Computational & Graphical Statistics*, 13, 341–361.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001), "Bayesian Semiparametric Models for Survival Data with a Cure Fraction," *Biometrics*, 57, 383–388.

Ishwaran, H. and Zarepour, M. (2002), "Dirichlet Prior Sieves in Finite Normal Mixtures," *Statistica Sinica*, 12, 941–963.

Kim, S., Chen, M.-H., Dey, D. K., and Gamerman, D. (2007), "Bayesian Dynamic Models for Survival Data with a Cure Fraction," *Lifetime Data Analysis*, 13, 17–35.

Kinney, S. and Dunson, D. (2007), "Fixed and Random Effects Selection in Linear and Logistic Models," *Biometrics*, 63, 690–698.

Komárek, A., Lesaffre, E., and Hilton, J. F. (2005), "Accelerated Failure Time Model for Arbitrarily Censored Data with Smoothed Error Distribution," *Journal of Computational & Graphical Statistics*, 14, 726–745.

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

Kuo, L. and Mallick, B. (1997), "Bayesian Semiparametric Inference for the Accelerated Failure-time Model," *The Canadian Journal of Statisrics*, 25, 457–472.

Li, C.-S. and Taylor, J. M. G. (2002), "A Semi-parametric Accelerated Failure Time Cure Model," *Statistics in Medicine*, 21, 3235–3247.

Li, J. Q. and Barron, A. R. (2000), "Mixture Density Estimation," Tech. rep., Department of Statistics, Yale University, New Haven, Connecticut.

Lu, W. and Ying, Z. (2004), "On Semiparametric Transformation Cure Models," *Biometrika*, 91, 331–343.

Lunn, D. J., Best, N., and Whittaker, J. (2008), "Generic Reversible Jump MCMC Using Graphical Models," *Statistics and Computing*, DOI: 10.1007/s11222-008-9100-0.

Mitchell, T. J. and Beauchamp, J. J. (1988), "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, 83, 1023–1032.

Nieto-Barajas, L. E. and Yin, G. S. (2008), "Bayesian Semiparametric Cure Rate Model with an Unknown Threshold," *Scandinavian Journal of Statistics*, 35, 540–556.

Osman, M. and Ghosh, S. K. (2012), "Nonparametric regression models for right-censored data using Bernstein polynomials," *Computational Statistics and Data Analysis*, 56, 559–73.

Peng, Y. and Dear, K. B. G. (2000), "A Nonparametric Mixture Model for Cure Rate Estimation," *Biometrics*, 56, 237–243.

Plummer, M. (2003), "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, https://sourceforge.net/projects/mcmc-jags," .

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 6, 7–11.

Rao, C. R. (1973), "Unified Theory of Least Squares," *Communications in Statistics - Simulation and Computation*, 1, 1–8.

Roberts, G. O. and Sahu, S. H. (2001), "Approximate Predetermined Convergence Properties of the Gibbs Sampler," *Journal of Computational & Graphical Statistics*, 10, 216–229.

Schwartz, L. (1965), "On Bayes Procedures," *Probability Theory and Related Fields*, 4, 10–26.

Secasan, I., Pop, D. I., and Secasan, C. C. (2005), "Potentially New And Innovative Treatments For Superficial, Muscle-Invasive, And Metastatic Transitional Cell Carcinoma (TCC) Of The Bladder," *The Internet Journal of Oncology*, 2.

Sheng, A. and Ghosh, S. K. (2019), "Effects of Proportional Hazard Assumption on Variable Selection Methods for Censored Data," *STATISTICS IN BIOPHARMACEUTICAL RESEARCH*, 12, 199–209.

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.

Walker, S. and Mallick, B. K. (1999), "A Bayesian Semiparametric Accelerated Failure Time Model," *Biometrics*, 55, 477–483.

Wu, Y. and Ghosal, S. (2008), "Kullback Leibler Property of Kernel Mixture Priors in Bayesian Density Estimation," *Electronic Journal of Statistics*, 2, 298–331.

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

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

Yin, G. (2008), "Bayesian Transformation Cure Frailty Models with Multivariate Failure Time Data," *Statistics in Medicine*, 27, 5929–5940.

Zeng, D., Yin, G., and Ibrahim, J. G. (2006), "Semiparametric Transformation Models for Survival Data With a Cure Fraction," *Journal of the American Statistical Association*, 101, 670–684.

Zhang, J. and Peng, Y. (2007), "A new Estimation Method for the Semiparametric Accelerated Failure Time Mixture Cure Model," *Statistics in Medicine*, 26, 3157–3171.

A Proof of Theorem 1

Lemma A.1. *The model given in Equation (2.9) of our paper is identifiable*

Proof of Lemma A.1

Let $h_{\boldsymbol{\beta}_1,\boldsymbol{\gamma}_1,\sigma_{\epsilon_1},H_1}(x,\delta,\boldsymbol{z}) = h_{\boldsymbol{\beta}_2,\boldsymbol{\gamma}_2,\sigma_{\epsilon_2},H_2}(x,\delta,\boldsymbol{z})$ for all $(x,\delta,\boldsymbol{z})$, as defined in Equation (2.9). We will work with each component of $h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}(x,\delta,\boldsymbol{z})$ separately. Let $\sigma_j^2 = \sigma_{\epsilon j}^2/(k_0+1)$, $j=1,2$, so we may reparameterize the mixing distribution in terms of σ_j^2. First, let $\delta = 0$ so we work with the second component. Then for all $(x,\boldsymbol{z})$,

$$p(\boldsymbol{\gamma}_1,\boldsymbol{z})\int \Phi\left[\frac{\log x - \boldsymbol{\beta}_1'\boldsymbol{z}-\mu}{\sigma_1}\right]dH_1(\mu) = p(\boldsymbol{\gamma}_2,\boldsymbol{z})\int \Phi\left[\frac{\log x - \boldsymbol{\beta}_2'\boldsymbol{z}-\mu}{\sigma_2}\right]dH_2(\mu).$$

Now, letting $x\to\infty$ and applying Monotone Convergence Theorem (MCT), the above equation reduces to

$$p(\boldsymbol{\gamma}_1,\boldsymbol{z})\int dH_1(\mu) = p(\boldsymbol{\gamma}_1,\boldsymbol{z})\int dH_2(\mu).$$

But $\int_0^\infty dH(\mu) = 1$ for any mixing distribution. Hence, $p(\boldsymbol{\gamma}_1,\boldsymbol{z}) = p(\boldsymbol{\gamma}_2,\boldsymbol{z})$ implies that $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$ assuming that $F_o(\cdot)$ is a strictly increasing function.

Next, let $\delta = 1$ so we work with the second component. Because $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$, for all $(x,\boldsymbol{z})$, we have

$$\int \frac{1}{\sigma_1 x}\phi\left[\frac{\log x - \boldsymbol{\beta}_1'\boldsymbol{z}-\mu}{\sigma_1}\right]dH_1(\mu) = \int \frac{1}{\sigma_2 x}\phi\left[\frac{\log x - \boldsymbol{\beta}_2'\boldsymbol{z}-\mu}{\sigma_2}\right]dH_2(\mu). \tag{A.1}$$

Setting $\boldsymbol{z}=0$, the above equation reduces to

$$\int \frac{1}{\sigma_1 x}\phi\left[\frac{\log x - \mu}{\sigma_1}\right]dH_1(\mu) = \int \frac{1}{\sigma_2 x}\phi\left[\frac{\log x - \mu}{\sigma_2}\right]dH_2(\mu) \tag{A.2}$$

for all $x\in(0,\infty)$. Multiplying both sides of A.2 by e^{iwx} and integrating with respect to x we obtain

$$\int_0^\infty\int e^{iwx}\frac{1}{\sigma_1 x}\phi\left[\frac{\log x-\mu}{\sigma_1}\right]dH_1(\mu)dx = \int_0^\infty\int e^{iwx}\frac{1}{\sigma_2 x}\phi\left[\frac{\log x-\mu}{\sigma_2}\right]dH_2(\mu)dx,$$

where $i=\sqrt{-1}$ is the imaginary unit. By change of variables, $u_1 = (\log x-\mu)/\sigma_1$ and $u_2 = (\log x-\mu)/\sigma_2$ and using Fubini's Theorem we have

$$\int\int e^{iw(\mu+\sigma_1 u_1)}\phi(u_1)du_1 dH_1(\mu) = \int\int e^{iw(\mu+\sigma_2 u_2)}\phi(u_2)du_2 dH_2(\mu).$$

After re-arranging we have

$$\int e^{iw\mu}\int e^{iw\sigma_1 u_1}\phi(u_1)du_1 dH_1(\mu) = \int e^{iw\mu}\int e^{iw\sigma_2 u_2}\phi(u_2)du_2 dH_2(\mu).$$

The inside integrals are the characteristic functions for the standard normal distributions and therefore we have,

$$\int e^{iw\mu}e^{-\frac{1}{2}\sigma_1^2 w^2}dH_1(\mu) = \int e^{iw\mu}e^{-\frac{1}{2}\sigma_2^2 w^2}dH_2(\mu).$$

Let $\Psi_H(w) = \int e^{iw\mu}dH(\mu)$ be the characteristic function for $H(\cdot)$. Then the above equation simplifies to

$$e^{-\frac{1}{2}\sigma_1^2 w^2}\Psi_{H_1}(w) = e^{-\frac{1}{2}\sigma_2^2 w^2}\Psi_{H_2}(w).$$

Taking the logarithm on both sides of the equation we have

$$\sigma_1^2 w^2 - 2\log \Psi_{H_1}(w) = \sigma_2^2 w^2 - 2\log \Psi_{H_2}(w), \tag{A.3}$$

for all $w \in \mathbb{R}$. Differentiating both sides of the above equation with respect to w twice results in

$$2\sigma_1^2 - 2\frac{d^2}{dw^2}\log \Psi_{H_1}(w) = 2\sigma_2^2 - 2\frac{d^2}{dw^2}\log \Psi_{H_2}(w).$$

Note that

$$\begin{aligned}\frac{d^2}{dw^2}\log \Psi_H(w)\Big|_{w=0} &= \frac{\Psi''_H(w)\Psi_H(w) - \Psi'_H(w)^2}{\Psi_H^2(w)}\Big|_{w=0} \\ &= \Psi''_H(0) - \Psi'_H(0)^2 \\ &= -\text{Var}_H[\mu],\end{aligned}$$

since $\Psi_H(0) = 1$, $\Psi'_H(0) = -iE_H[\mu]$, and $\Psi''_H(0) = -E_H[\mu^2]$. Hence we have

$$\sigma_1^2 + \text{Var}_{H_1}[\mu] = \sigma_2^2 + \text{Var}_{H_2}[\mu]$$

Recall that $\text{Var}_H[\mu] = \frac{k_0}{k_0+1}\sigma_\epsilon^2$ and $\sigma_{\epsilon_j}^2 = \sigma_j^2(k_0+1)$. Thus, it follows that $\sigma_1^2 = \sigma_2^2$, which in turn implies that $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2$. By Equation (A.3) this also means that $\Psi_{H_1}(w) = \Psi_{H_2}(w)$ for all $w \in \mathbb{R}$. By uniqueness of characteristic functions, $H_1(\cdot) = H_2(\cdot)$. Returning to Equation (A.1) and using DCT, it follows that $\int \frac{1}{\sigma}\phi((\mu + \boldsymbol{\beta}'\boldsymbol{z} - \log(x))/\sigma)dH(\mu) \to dH(\log(x) - \boldsymbol{\beta}'\boldsymbol{z})$ as $\sigma \to 0$ for any $x > 0$ and thus we obtain for all $\boldsymbol{z}$, $\boldsymbol{\beta}_1'\boldsymbol{z}$=$\boldsymbol{\beta}_2'\boldsymbol{z}$ and hence, $\boldsymbol{\beta_1} = \boldsymbol{\beta_2}$. Alternatively, we can also multiply the equation (A.1) by $\log x$ and then integrating both sides with respect to x we can obtain for all $\boldsymbol{z}$, $\boldsymbol{\beta}_1'\boldsymbol{z}$=$\boldsymbol{\beta}_2'\boldsymbol{z}$. This completes the proof of identifiability.

Lemma A.2. *Consider the product topology on $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\epsilon, H)$, where $\boldsymbol{\beta}, \boldsymbol{\gamma}$, and σ_ϵ, are given the usual Euclidean topology and H the weak topology. On densities $h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}(x,\delta,\boldsymbol{z})$, put the total variation (or the L_1) distance defined as*

$$||h_{\boldsymbol{\beta}_1,\boldsymbol{\gamma}_1,\sigma_{\epsilon_1},H_1} - h_{\boldsymbol{\beta}_2,\boldsymbol{\gamma}_2,\sigma_{\epsilon 2},H_2}|| = \int\int_0^\infty |h_{\boldsymbol{\beta}_1,\boldsymbol{\gamma}_1,\sigma_{\epsilon_1},H_1}(x,\delta,\boldsymbol{z}) - h_{\boldsymbol{\beta}_2,\boldsymbol{\gamma}_2,\sigma_{\epsilon_2},H_2}(x,\delta,\boldsymbol{z})|dxd\boldsymbol{z}.$$

Then

$$||h_{\boldsymbol{\beta}_\nu,\boldsymbol{\gamma}_\nu,\sigma_{\epsilon_\nu},H_\nu} - h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}|| \to 0$$

if and only if $(\boldsymbol{\beta}_\nu, \boldsymbol{\gamma}_\nu, \sigma_{\epsilon_\nu}, H_\nu) \to (\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\epsilon, H)$. In other words, the variation topology on the densities is equivalent to the product topology on the indexing parameters.

Proof of Lemma A.2

The proof follows similarly as Lemma 2 in Ghosh and Ghosal (2006). In fact, the 'only if' portion follows directly as in Ghosh and Ghosal (2006) and using the identifiability property verified in Lemma 1. We provide details to the 'if' part, as the cure fraction and normal mixture provide a few differences in proof.

It suffices to show that the densities converge pointwise and then apply Scheffe's theorem. Fix $(x, \delta, \boldsymbol{z})$. We show the proof for $\delta = 1$.

Because μ has a compact range, the integrand $\phi\left(\frac{\log x-\boldsymbol{\beta}'\boldsymbol{z}-\mu}{\sigma_\epsilon/(k_0+1)}\right)\frac{\sqrt{k_0+1}}{\sigma_\epsilon}$ of $h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}$ as a family of functions of $(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon)$ indexed by μ is equicontinuous. Also $p(\boldsymbol{\gamma},\boldsymbol{\beta})$ is fixed and does not depend on μ. For a given $\varepsilon>0$, find ν large enough so that the integrands are ε-close for all μ. For such a ν,

$$|h_{\boldsymbol{\beta}_\nu,\boldsymbol{\gamma}_\nu,\sigma_{\epsilon_\nu},H_\nu}-h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}|\le|h_{\boldsymbol{\beta}_\nu,\boldsymbol{\gamma}_\nu,\sigma_{\epsilon_\nu},H_\nu}-h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H_\nu}|+|h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H_\nu}-h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}|.$$

For $n\in\mathcal{N}$, let $\{p_n\}$ and $\{A_n\}$ be sequences of real numbers. Then the following inequality holds:

$$\begin{aligned}|p_nA_n-pA| &\le |p_n-p||A|+|p_n||A_n-A|\\ &\le |p_n-p||A|+|A_n-A| \qquad \text{if } |p_n|\le 1\end{aligned} \tag{A.4}$$

Let $p_n=p(\boldsymbol{\gamma}_n,\boldsymbol{z})$ and A_n represent the density function of the error distribution, which is the same density as Ghosh and Ghosal (2006). Then $|h_{\boldsymbol{\beta}_\nu,\boldsymbol{\gamma}_\nu,\sigma_{\epsilon_\nu},H_\nu}-h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}|$ is equivalent to $|p_nA_n-pA|$. Because $p_\nu(\boldsymbol{\gamma},\boldsymbol{z})$ converges pointwise to $p(\boldsymbol{\gamma},\boldsymbol{z})$, there exists an $n>n_1$ for which $|p_n-p|<\frac{\epsilon}{2|A|}$ in the first term on the RHS of Equation (A.4). Since $|p(\boldsymbol{\gamma}_n,\boldsymbol{z})|\le 1$, and as shown in the proof of Lemma 2 in Ghosh and Ghosal (2006), there also exists an $n>n_2$ for which $|A_n-A|<\frac{\epsilon}{2}$ in the second term on the RHS of Equation (A.4). Therefore the RHS is less than $\frac{\epsilon}{2|A|}|A|+\frac{\epsilon}{2}=\epsilon$. Hence, $|h_{\boldsymbol{\beta}_\nu,\boldsymbol{\gamma}_\nu,\sigma_{\epsilon_\nu},H_\nu}-h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}|<\epsilon$. Applying Scheffe's theorem proves the 'if' part of the theorem.

Lemma A.3. *For all $\varepsilon>0$,*

$$\Pi\left\{(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H):\int\int h_{\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\sigma_{\epsilon_0},H_0}\log\frac{h_{\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\sigma_{\epsilon_0},H_0}}{h_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}}dxdz<\varepsilon\right\}>0.$$

Proof of Lemma A.3

The log likelihood ratio is given by

$$\Delta(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H) = \begin{cases}\log\frac{{}^f\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H^{(x,\delta,\boldsymbol{z})}}{{}^f\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\sigma_{\epsilon_0},H_0{}^{(x,\delta,\boldsymbol{z})}} & \text{if } \delta=0\\ \log\frac{{}^S\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H^{(x,\delta,\boldsymbol{z})}}{{}^S\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\sigma_\epsilon,H_0{}^{(x,\delta,\boldsymbol{z})}} & \text{if } \delta=1.\end{cases}$$

We shall prove the $\delta=1$ case, while the $\delta=0$ case is straighforward. Notice that $\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,\boldsymbol{z}$ are all bounded. Thus the integrand within $f_{\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_\epsilon,H}$ is bounded above and below by functions of the form $k_1e^{c_1x}e^{-x^2/c_2}$. Also, $p(\boldsymbol{\gamma},\boldsymbol{z})$ is bounded between 0 and 1. Taking the ratio and then logarithm, Δ in the tails (in x) is bounded by a multiple of a power of x. The rest of the proof follows as that in Lemma 3 of Ghosh and Ghosal (2006) and Theorem 3 of Ghosal et al. (1999).

Proof of Theorem 1 Lemma A.2 shows that it suffices to consider neighborhoods with respect to the L_1-distance. Also the space is compact. The condition of prior positivity has been verified in Lemma A.3. Hence, the proof is complete.

B Explanation of the Difference Operator Matrix

We provide an explanation of the difference operator matrix for the $m=1$ and $m=2$ cases.

First, for when $m = 1$, the penalty term includes:

$$\sum_{l=2}^{L_n}\{\Delta^1\alpha_l\}^2 = \sum_{l=2}^{L_n}(\alpha_l - \alpha_{l-1})^2 = (D_1\boldsymbol{\alpha})'(D_1\boldsymbol{\alpha}) = \boldsymbol{\alpha}' D_1' D_1 \boldsymbol{\alpha}$$

where

$$D_1\boldsymbol{\alpha} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}_{(L_n-1)\times L_n} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_{L_n} \end{pmatrix} = \begin{pmatrix} \alpha_2 - \alpha_1 \\ \alpha_3 - \alpha_2 \\ \alpha_4 - \alpha_3 \\ \vdots \\ \alpha_{L_n} - \alpha_{L_n-1} \end{pmatrix}$$

Secondly, when $m = 2$, the penalty term includes

$$\Delta^2\alpha_l = (\alpha_l - \alpha_{l-1}) - (\alpha_{l-1} - \alpha_{l-2}) = \alpha_l - 2\alpha_{l-1} + \alpha_{l-2} \qquad l = 3, \dots, L_n$$

Hence,

$$\sum_{l=3}^{L_n}\{\Delta^2\alpha_l\}^2 = (D_2\boldsymbol{\alpha})'(D_2\boldsymbol{\alpha}) = \boldsymbol{\alpha}' D_2' D_2 \boldsymbol{\alpha}$$

where

$$\begin{aligned} D_2\boldsymbol{\alpha} &= \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}_{(L_n-2)\times L_n} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_{L_n} \end{pmatrix} \\ &= \begin{pmatrix} \alpha_3 - 2\alpha_2 + \alpha_1 \\ \alpha_4 - 2\alpha_3 + \alpha_2 \\ \alpha_5 - 2\alpha_4 + \alpha_3 \\ \vdots \\ \alpha_{L_n} - 2\alpha_{L_n-1} + \alpha_{L_n-2} \end{pmatrix} \end{aligned}$$

C JAGS Code for the CRAFT Model Under Right Censoring

```
for (i in 1:n.obs){
  cen[i] ~ dbern(pi[i])
  logit(pi[i]) <- inprod(gamma[1:p],x[i,1:p])
  logtime[i] ~ dnorm(logmean[i],omega0)
```

```
  logmean[i] <- eta[latent[i]]+mu[i]
  latent[i] ~ dcat(prob[])
  mu[i]<-inprod(beta[1:p-1],x[i,2:p])}
for (i in (n.obs+1):n){
  cen[i] ~ dbern(pistar[i])
  logit(pi[i]) <- inprod(gamma[1:p],x[i,1:p])
  mu[i]<-inprod(beta[1:p-1],x[i,2:p])
  pistar[i] <- pi[i]*max(inprod(prob[],phistar[i,]),
    step(logtime[i]-maxlogtime))
  for(l in 1:N){
    phistar[i,l]<-phi((logtime[i]-eta[l]-mu[i])*sqrt(omega0))}}

inversevariance<-1/3
inversevariancesmall<-2/3
gamma[1] ~ dnorm(0,inversevariancesmall)
for (j in 2:p){
  gamma[j] ~ dnorm(0,inversevariance)}
for (j in 1:p-1){
  beta[j] ~ dnorm(0,inversevariance)}
omega0 ~ dgamma(.1,.1)
sigma0 <- 1/sqrt(omega0)

for (k in 1:(N-1)){
  expalpha[k]<-exp(alpha[k])
  prob[k] <-exp(alpha[k])/(1+sum(expalpha[]))}
prob[N]<-1-sum(prob[1:(N-1)])
for (k in 1:(N-1-m)){
  delta[k] ~ dnorm(0,lambda1)}
for (k in 1:(N-1)){
  alpha[k] <- inprod(Dcoef[k,],delta[])}
```

D Additional Figures

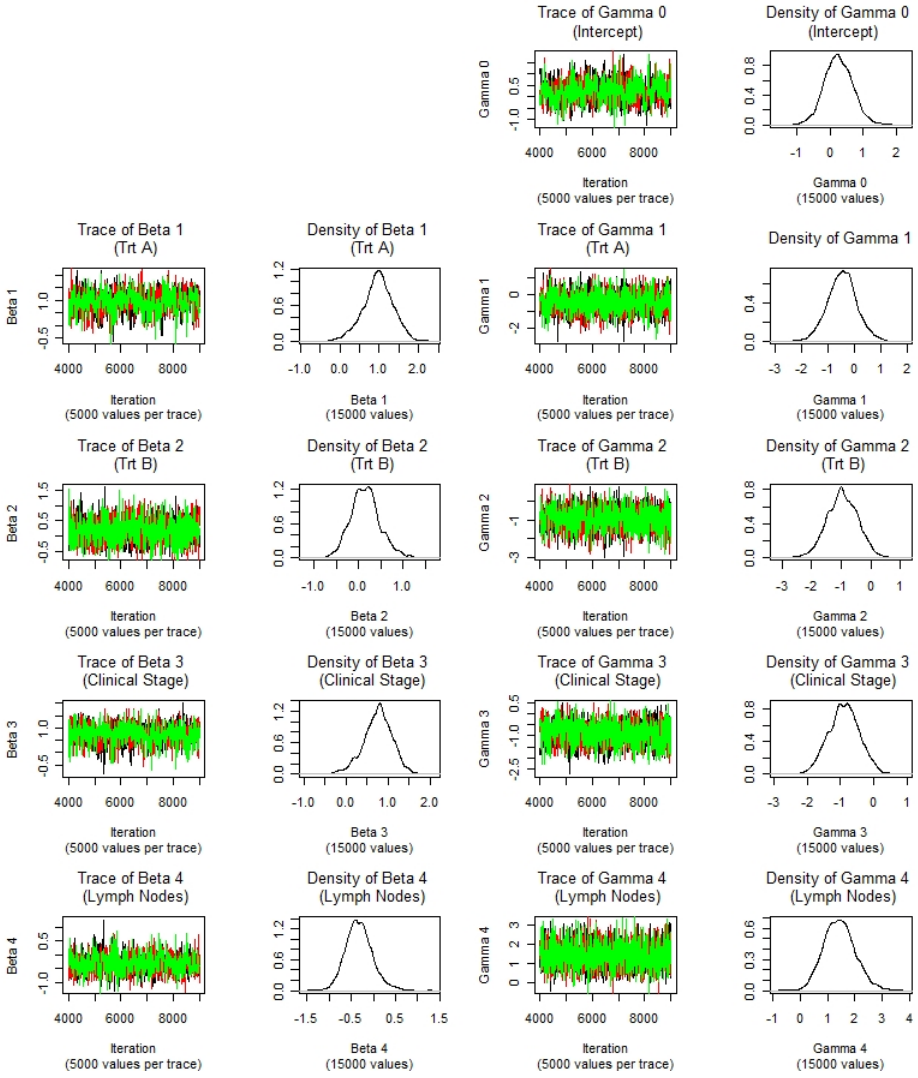

Figure 5: Breast cancer data: trace and posterior density curves for both the latency and incidence statistics associated with treatment A or B, clinical stage indication, and number of lymph nodes using the CRAFT model

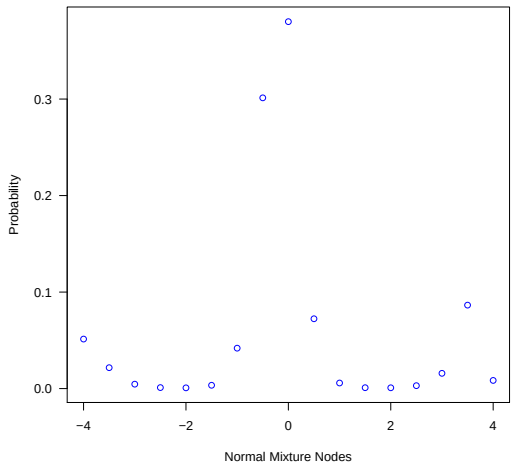

Figure 6: Breast cancer data: posterior means of the mixing weights using the CRAFT model

Received: February 25, 2021

Accepted: June 2, 2021