Session 2: Bayesian Cure Rate Models

2.1 Introduction

2.2 Parametric Cure Rate Models

2.3 Semiparametric Cure Rate Models

2.4 Multivariate Cure Rate Models
Overview

In this session, we examine the cure rate models in detail. We derive the cure rate model, review parametric and semiparametric cure rate models, examine the properties of the models, and discuss the construction of the likelihood function, prior elicitation strategies, and computational methods. We also present multivariate extensions of the cure rate model. In addition, we will discuss informative prior elicitation for this model based on the power prior, and discuss its properties. Several case studies will be presented to illustrate these models.
Objectives

The main objectives of this session are to introduce and examine

- Cure rate models and their properties;
- Computational algorithms; and
- Issues involved in data analysis using cure rate models.
2.1 Introduction

- **When Do We Need A Cure Rate Model?**

  The cure rate model is needed for modelling time-to-event data for various types of cancers, including breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma, and head and neck cancer, where for these diseases, a significant proportion of patients are “cured”.

- **Example:**

  To demonstrate such a phenomenon, we consider a recent phase III clinical trial in malignant melanoma (E1684) undertaken by the Eastern Cooperative Oncology Group (ECOG). The graph in Figure 2.1 gives the Kaplan-Meier survival curve for \( n = 286 \) patients in E1684, with the survival time given in years.
Figure 2.1. Kaplan-Meier plot for E1684 data.

We see from Figure 2.1 that a plateau in the curve occurs at approximately 0.36, suggesting that 36% fraction of patients are "cured" after sufficient follow-up.
**Standard Cure Rate Model**

Berkson and Gage (1952) first introduced the following cure rate model. In this model, it is assumed that a certain fraction $\pi$ of the population is “cured”, and the remaining $1 - \pi$ are not cured. The survivor function for the entire population, denoted by $S_1(y)$, for this model is given by

$$S_1(y) = \pi + (1 - \pi)S^*(y),$$

where $S^*(y)$ denotes the survivor function for the non-cured group in the population. Common choices for $S^*(y)$ are the exponential and Weibull distributions.
• Disadvantages of the Standard Cure Rate Model

(i) It does not have a proportional hazards structure, which is a desirable property for survival models.

(ii) When including covariates through the parameter $\pi$ via a standard binomial regression model, the standard cure rate model yields improper posterior distributions for many types of noninformative improper priors, including the uniform prior for the regression coefficients.

(iii) There does not appear to be a natural multivariate extension of the standard cure rate model.
2.2 Parametric Cure Rate (PCR) Models

♦ The PCR Models

- \( N \): the number of carcinogenic cells for an individual, which has a Poisson distribution.
- \( Z_i \): the random time for the \( i^{th} \) clonalogenic tumor cell to produce a detectable tumor, which can also be viewed as an incubation time. Assume that \( Z_i, i = 1, 2, \ldots \) are independent and identically distributed with a common distribution function \( F(y) = 1 - S(y) \) and independent of \( N \).
- \( Y = \min \{ Z_i, 0 \leq i \leq N \} \): the tumor latency time.
♠ The Survival Function

\[
S_p(y) = P(\text{no cancer by time } y) \\
= P(N = 0) + P(Z_1 > y, \ldots, Z_N > y, N \geq 1) \\
= \exp(-\theta) + \sum_{k=1}^{\infty} S(y)^k \frac{\theta^k}{k!} \exp(-\theta) \\
= \exp(-\theta + \theta S(y)) \\
= \exp(-\theta F(y)).
\]

The cure fraction (i.e., cure rate) is given by

\[
P(N = 0) = \exp(-\theta) = \lim_{y \to \infty} S_p(y).
\]

Note that \( S_p(y) \) is not a proper survival function as

\[
\lim_{y \to \infty} S_p(y) = \exp(-\theta) \neq 0.
\]
Some Properties

- As $\theta \to \infty$, the cure fraction tends to 0, whereas as $\theta \to 0$, the cure fraction tends to 1.

- The corresponding density is given by

$$f_p(y) = \theta f(y) \exp(-\theta F(y)),$$

and the hazard function is given by

$$h_p(y) = \theta f(y),$$

where $f(y) = \frac{d}{dy} F(y)$.

- $h_p(y)$ is multiplicative in $\theta$ and $y$, and thus has the proportional hazards structure, with the covariates modelled through $\theta$.

- $S_p(y)$ explicitly implies the contribution to the relapse-free time of two distinct characteristics of tumor growth: the initial number of clonogens and the rate of their progression. Thus the model incorporates parameters bearing clear biological meaning.
• The survivor function for the “non-cured” population is given by

\[ S^*(y) = P(Y > y \mid N \geq 1) = \frac{\exp(-\theta F(y)) - \exp(-\theta)}{1 - \exp(-\theta)}. \]

We note that \( S^*(0) = 1 \) and \( S^*(\infty) = 0 \) so that \( S^*(y) \) is a proper survivor function.

• The survival density for the non-cured population is given by

\[ f^*(y) = -\frac{d}{dy} S^*(y) = \left( \frac{\exp(-\theta F(y))}{1 - \exp(-\theta)} \right) \theta f(y). \]

• The hazard function for the non-cured population is given by

\[ h^*(y) = \frac{f^*(y)}{S^*(y)} = \frac{\theta f(y)}{1 - \exp(-\theta F(y))}. \]
Relationship Between the PCR and Standard Models

\[ S_p(y) = \exp(-\theta F(y)) \]
\[ = \exp(-\theta) + (1 - \exp(-\theta)) \left( \frac{\exp(-\theta F(y)) - \exp(-\theta)}{1 - \exp(-\theta)} \right) \]
\[ = \exp(-\theta) + (1 - \exp(-\theta)) S_\theta(y), \]

where \( S_\theta(y) = \frac{\exp(-\theta F(y)) - \exp(-\theta)}{1 - \exp(-\theta)} \).

It is easily shown that \( S_\theta(y) \) defines a proper survivor function. Thus the new model is a standard cure rate model with cure rate equal to \( \pi = \exp(-\theta) \) and survivor function for the non-cured population given by \( S_\theta(y) \).
Parametric Form of the Cure Rate Model

We specify a parametric form for $F(\cdot)$, such as a Weibull or gamma distribution. We denote the indexing parameter (possibly vector valued) by $\psi$, and thus write $F(\cdot|\psi)$ and $S(\cdot|\psi)$. For example, if $F(\cdot|\psi)$ corresponds to a Weibull distribution, then $\psi = (\alpha, \lambda)'$, where $\alpha$ is the shape parameter and $\lambda$ is the scale parameter.
♠ The Likelihood Function

◊ Notations

- $y_i$: survival time (failure time or right-censored)
- $\nu_i$: the censoring indicator; $\nu_i = 1$ if $y_i$ is the failure time and 0 if $y_i$ is right censored.
- $N_i$: the number of carcinogenic cells
- $x'_i = (x_{i1}, \ldots, x_{ip})$: the $p \times 1$ vector of covariates
- $n$: the sample size
- $\beta = (\beta_1, \ldots, \beta_p)'$: the vector of regression coefficients
- We model $\theta_i = \exp(x'_i\beta)$
- The observed data: $D_{obs} = (n, y, \nu)$, where $y = (y_1, y_2, \ldots, y_n)'$, and $\nu = (\nu_1, \nu_2, \ldots, \nu_n)'$.
- The complete data: $D = (n, y, \nu, N)$, where $N = (N_1, N_2, \ldots, N_n)'$. 

2-14
◊ The Complete Data Likelihood Function

\[
L(\beta, \psi|D) = \left( \prod_{i=1}^{n} S(y_i|\psi)^{N_i-\nu_i} (N_i f(y_i|\psi))^{\nu_i} \right) \\
\times \exp \left\{ \sum_{i=1}^{n} [N_i x_i' \beta - \log(N_i!) - \exp(x_i' \beta)] \right\},
\]

where we assume a Weibull density for \( f(y_i|\psi) \), so that

\[
f(y|\psi) = \alpha y^{\alpha-1} \exp \{ \lambda - y^\alpha \exp(\lambda) \}.
\]

◊ The Likelihood Function Based on the Observed Data

After summing out \( N \), we obtain

\[
L(\beta, \psi|D_{obs}) = \prod_{i=1}^{n} \left( \theta_i f(y_i|\psi) \right)^{\nu_i} \exp\{-\theta_i (1 - S(y_i|\psi))\},
\]

where \( \theta_i = \exp(x_i' \beta) \).
The Prior and Posterior Distributions

Noninformative Prior

Prior

We take

\[ \pi(\beta, \psi) \propto \pi(\psi), \]
\[ \pi(\psi) = \pi(\alpha|\delta_0, \tau_0)\pi(\lambda), \]

where \( \pi(\psi) = \pi(\alpha|\delta_0, \tau_0)\pi(\lambda), \)

\[ \pi(\alpha|\delta_0, \tau_0) \propto \alpha^{\delta_0-1} \exp(-\tau_0\alpha), \]

and \( \delta_0 \) and \( \tau_0 \) are two specified hyperparameters.

Posterior

With these specifications, the posterior distribution of \((\beta, \psi)\) based on the observed data \(D_{obs} = (n, y, X, \nu)\) is given by

\[ \pi(\beta, \psi|D_{obs}) \propto L(\beta, \psi|D_{obs})\pi(\alpha|\delta_0, \tau_0)\pi(\lambda). \]
• **Propriety of the Posterior**

The following theorem, due to Chen, Ibrahim, and Sinha (1999), gives conditions concerning the propriety of the posterior distribution $\pi(\beta, \psi | D_{obs})$ using the noninformative prior $\pi(\beta, \psi) \propto \pi(\psi)$.

**Theorem 2.1** Let $d = \sum_{i=1}^{n} \nu_i$ and $X^*$ be an $n \times p$ matrix with rows $\nu_i x'_i$. Then if (i) $X^*$ is of full rank, (ii) $\pi(\lambda)$ is proper, and (iii) $\tau_0 > 0$ and $\delta_0 > -d$, the posterior $\pi(\beta, \psi | D_{obs})$ is proper.
• **Note:**

- The conditions given in Theorem 2.1 are sufficient but *not* necessary for the propriety of the posterior distribution.

- The conditions stated in the theorem are quite general and are typically satisfied for most datasets.

- A proper prior for \( \alpha \) is not required in order to obtain a proper posterior. This can be observed from condition (iii) since \( \pi(\alpha|\delta_0, \tau_0) \) is no longer proper when \( \delta_0 < 0 \).

- Based on condition (ii), \( \pi(\lambda) \) is required to be proper. Although several choices can be made, we will use a normal density for \( \pi(\lambda) \).

- Theorem 2.1 guarantees propriety of the posterior distribution of \( \beta \) using an improper uniform prior. This enables us to carry out Bayesian inference with improper priors for the regression coefficients and facilitates comparisons with maximum likelihood.
The Standard Cure Rate Model

Under the improper priors $\pi(\beta, \psi) \propto \pi(\psi)$, the standard cure rate model always leads to an improper posterior distribution for $\beta$. This result is stated in the following theorem.

**Theorem 2.2** For the standard cure rate model, suppose we relate the cure fraction $\pi$ to the covariates via a standard binomial regression

$$\pi_i = G(x'_i \beta),$$

where $G(\cdot)$ is a continuous cumulative distribution function (cdf). Assume that the survival function $S^*(\cdot)$ for the non-cured group depends on the parameter $\psi^*$. Let $L_1(\beta, \psi^* | D_{obs})$ denote the resulting likelihood function based on the observed data. Then, if we take an improper uniform prior for $\beta$ (i.e., $\pi(\beta) \propto 1$), the posterior distribution

$$\pi_1(\beta, \psi^* | D_{obs}) \propto L_1(\beta, \psi^* | D_{obs}) \pi(\psi^*)$$

is always improper regardless of the propriety of $\pi(\psi^*)$. 
◊ Informative Prior: Power Prior

• A Brief Introduction

– Our prior construction is based on the notion of the existence of a previous similar study that measures the same response variable and covariates as the current study.

– Prior elicitation using historical data has been discussed by the authors for several different types of models for different contexts, including GLMs, GLMMs, the Cox model, and time series models. For details and motivation on prior constructions from historical data, see, for example, Ibrahim, Ryan, and Chen (1997), Chen, Ibrahim, and Shao (2000), Ibrahim, Chen, and MacEachern (1997), Chen, Ibrahim, and Yiannoutsos (1999), Ibrahim and Chen (1998, 2000) and Chen, Ibrahim, Shao, and Weiss (2002).

– The justification of the power prior has been elaborated and discussed in detail in Ibrahim, Chen, and Sinha (2003) and Chen and Ibrahim (2005).
• Notation

Let $n_0$ denote sample size for the historical data, $y_0$ be an $n_0 \times 1$ of right censored event times for the historical data with censoring indicators $\nu_0$, $N_0$ is the unobserved vector of latent counts of carcinogenic cells, and $X_0$ is an $n_0 \times p$ matrix of covariates corresponding to $y_0$. Let $D_0 = (n_0, y_0, X_0, \nu_0, N_0)$ and $D_{0,obs} = (n_0, y_0, X_0, \nu_0)$ denote the complete and observed historical data, respectively.

• The Formulation

Let $\pi_0(\beta, \psi)$ denote the initial prior distribution for $(\beta, \psi)$. The power prior for $(\beta, \psi)$ takes the form

$$
\pi(\beta, \psi|D_{0,obs}, a_0) \propto \left[ \sum_{N_0} L(\beta, \psi|D_0) \right]^{a_0} \pi_0(\beta, \psi)
$$

$$
= [L(\beta, \psi|D_{0,obs})]^{a_0} \pi_0(\beta, \psi),
$$

where $L(\beta, \psi|D_0)$ is the complete data likelihood based on $D_0$. 

2-21

M.-H. Chen
Session 2

• **Interpretation of** $\pi(\beta, \psi|D_{0,\text{obs}}, a_0)$

  - $a_0$: a dispersion parameter for the historical data; $0 \leq a_0 \leq 1$.
  
  - Role of $a_0$: controls the heaviness of the tails of the prior for $(\beta, \psi)$. As $a_0$ becomes smaller, the tails of $\pi(\beta, \psi|D_{0,\text{obs}}, a_0)$ become heavier, thus resulting in a less informative prior.

  - Properties: Setting $a_0 = 1$, $\pi(\beta, \psi|D_{0,\text{obs}}, a_0)$ corresponds to the usual Bayesian update of $\pi_0(\beta, \psi)$ via Bayes theorem, weighting the historical data and the current data equally.

  When $a = 0$, then the prior does not depend on the historical data, and in this case, $\pi(\beta, \psi \mid D_{0,\text{obs}}, a_0) \equiv \pi_0(\beta, \psi)$.

  - The parameter $a_0$ allows the investigator to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the previous and current study, or when the sample sizes of the two studies are quite different.
• Joint Prior

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

$$
\pi(\beta, \psi, a_0 \mid D_{0,obs}) \propto \left[ \sum_{N_0} L(\beta, \psi \mid D_0) \right]^{a_0} \pi_0(\beta, \psi)
\times a_0^{\gamma_0 - 1} (1 - a_0)^{\lambda_0 - 1},
$$

where $(\gamma_0, \lambda_0)$ are specified prior parameters.
• Propriety of the Power Prior

The joint prior \( \pi(\beta, \psi, a_0 | D_{0,obs}) \) is guaranteed to be proper if \( \pi_0(\beta, \psi) \) is proper. Further, \( \pi(\beta, \psi, a_0 | D_{0,obs}) \) can be proper even if \( \pi_0(\beta, \psi) \) is improper.

**Theorem 2.3** Assume that

\[
\pi_0(\beta, \psi) \propto \pi_0(\psi) = \pi_0(\alpha | \delta_0, \tau_0) \pi_0(\lambda) \propto \alpha^{\delta_0-1} \exp(-\tau_0 \alpha) \pi_0(\lambda),
\]

where \( \delta_0 \) and \( \tau_0 \) are specified hyperparameters. Let \( d_0 = \sum_{i=1}^{n_0} \nu_0i \) and \( X_0^* \) be an \( n_0 \times p \) matrix with rows \( \nu_0i \mathbf{x}'_{0i} \). If (i) \( X_0^* \) is of full rank, (ii) \( \delta_0 > 0 \) and \( \tau_0 > 0 \), (iii) \( \pi_0(\lambda) \) is proper, and (iv) \( \gamma_0 > p \) and \( \lambda_0 > 0 \), then the joint prior \( \pi(\beta, \psi, a_0 | D_{0,obs}) \) is proper.
• **Note:** For the standard cure rate model, suppose we relate the cure
fraction $\pi$ to the covariates via a standard binomial regression

$$\pi_i = G(x_i' \beta),$$

where $G(\cdot)$ is a continuous cumulative distribution function (cdf).
Assume that the survival function for the non-cured group $S^*(.)$ depends
on the parameter $\psi^*$. Let $L_1(\beta, \psi^* | D_{0, obs})$ and $L_1(\beta, \psi^* | D_{obs})$ denote
the likelihood functions based on the observed historical and current
data. Suppose we use an improper uniform initial prior for $\beta$ (i.e.,
$\pi_0(\beta) \propto 1$) to construct the joint prior as

$$\pi_1(\beta, \psi^*, a_0 | D_{0, obs}) \propto [L_1(\beta, \psi^* | D_{0, obs})]^{a_0} \pi_0(\gamma^*) a_0^{\gamma_0-1} (1 - a_0)^{\lambda_0-1},$$

where $\gamma_0$ and $\lambda_0$ are specified hyperparameters. Then,
$\pi_1(\beta, \psi^*, a_0 | D_{0, obs})$ is always improper regardless of the propriety of
$\pi_0(\psi^*)$. In addition, if we use $\pi_1(\beta, \psi^*, a_0 | D_{0, obs})$ as a prior, the
resulting posterior, given by

$$\pi_1(\beta, \psi^*, a_0 | D_{obs}) \propto L_1(\beta, \psi^* | D_{obs}) \pi_1(\beta, \psi^*, a_0 | D_{0, obs})$$

is also improper.
♣ Posterior Computation

• Objective

We aim to sample from the joint posterior density of \((\beta, \psi, a_0)\). The joint posterior of \((\beta, \psi, a_0)\) based on the observed data \(D_{obs}\) is given by

\[
\pi(\beta, \psi, a_0|D_{obs}) \propto \sum_{N} L(\beta, \psi|D) \left[ \sum_{N_0} L(\beta, \psi|D_0) \right]^{a_0} \\
\times \pi_0(\beta, \psi)a_0^{\gamma_0-1}(1 - a_0)^{\lambda_0-1},
\]

where \(L(\beta, \psi|D)\) and \(L(\beta, \psi|D_0)\) are the complete data likelihoods with data \(D\) and \(D_0\), respectively.

• An Assumption on Initial Prior

We assume that \(\pi_0(\beta, \psi)\) is log-concave in each component of \((\beta, \psi)\).

• WinBUGS CODES

The WinBUGS codes for the parametric cure rate model with the power prior is given in the website

“http://www.stat.uconn.edu/~mhchen/survbook.”
- Two Key Formulas Used in WinBUGS Codes

\[
\sum_{N} L(\beta, \psi | D) = \prod_{i=1}^{n} \left( \theta_i f(y_i | \psi) \right)^{\nu_i} \exp\{-\theta_i(1 - S(y_i | \psi))\},
\]

\[
\sum_{N_0} L(\beta, \psi | D_0) = \prod_{i=1}^{n_0} \left( \theta_{0i} f(y_{0i} | \psi) \right)^{\nu_{0i}} \exp\{-\theta_{0i}(1 - S(y_{0i} | \psi))\},
\]

where \( \theta_i = \exp(x_i' \beta) \), \( \theta_{0i} = \exp(x_{0i} \beta) \),

\[
f(y | \psi) = \alpha y^{\alpha-1} \exp\{\lambda - y^\alpha \exp(\lambda)\},
\]

and

\[
S(y | \psi) = \exp\{-y^\alpha \exp(\lambda)\}.
\]
♠ Example: Melanoma Data

- **Issues Involved in Data Analysis**
  - Why the cure rate models?
  - Why the PCR model?
  - Why the incorporation of historical data is important?
  - Sensitivity of Bayesian analysis.
• The Current Data

We consider the E1684 data, which was obtained from a recent Eastern Cooperative Oncology Group (ECOG) phase III clinical trial. The response variable is overall survival, which is defined as the time from randomization until death. One of our main goals in this example is to compare inferences between the standard cure rate model and the PCR model. First, we compare the maximum likelihood estimates of the cure rates between the two models. Second, using the cure rate model, we carry out a Bayesian analysis with covariates using the proposed power priors, and compare the results to the estimates based on the standard cure rate model. Three covariates and an intercept are included in the analyses. The covariates are age \((x_1)\), gender \((x_2)\) (male, female), and performance status \((x_3)\) (fully active, other).

After deleting missing observations, a total of \(n = 284\) observations are used in the analysis. Table 2.1 provides a summary of the E1684 data. For the survival time summary in Table 2.1, the Kaplan-Meier estimate of the median survival and its corresponding 95% confidence interval
(CI) are given. In all of the analyses, we standardized the age covariate to stabilize the posterior computations.

**Table 2.1. Summary of E1684 Data (current)**

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival Time (y) (years)</td>
<td></td>
<td>((2.34, 4.33))</td>
</tr>
<tr>
<td>Status (frequency))</td>
<td></td>
<td>censored</td>
</tr>
<tr>
<td></td>
<td></td>
<td>death</td>
</tr>
<tr>
<td>Age (x_1) (years)</td>
<td>Mean</td>
<td>47.0</td>
</tr>
<tr>
<td></td>
<td>Std Dev</td>
<td>13.0</td>
</tr>
<tr>
<td>Gender (x_2) (frequency)</td>
<td>Male</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>113</td>
</tr>
<tr>
<td>PS (x_3) (frequency)</td>
<td>Fully Active</td>
<td>253</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>31</td>
</tr>
</tbody>
</table>
• Kaplan-Meier Plots

Figure 2.2 shows three superimposed plots of the survival curve based on the Kaplan-Meier method (dashed line), the standard cure rate model (dotted line), and the parametric cure rate model (solid line). We see that the three plots are nearly identical, giving essentially the same results. Figure 2.2 also shows that a plateau in the curve occurs at approximately 0.36, suggesting that 36% fraction of patients are “cured” after sufficient follow-up.
Figure 2.2. Superimposed survival curves for the E1684 data.
• Cure Rates

We now consider several analyses with the covariates included. Figure 2.3 shows a box-plot of the maximum likelihood estimates (MLE’s) of the cure rates for all patients for the two models. We see that the two box-plots are very similar. The first, second, and third quartiles for the two box-plots are 0.32, 0.36, and 0.40 for the standard cure rate model and 0.33, 0.36, and 0.39 for the parametric cure rate model. We see from Figure 2.3 that the variation in the cure rate estimates from the standard model is greater than that of the parametric cure rate model. In fact, the standard deviations of the cure rate estimates are 0.06 for the standard cure rate model and 0.05 for the parametric cure rate model.
Figure 2.3. Box plots of the MLE’s of the Cure Rates for All Patients where the label ‘1’ denotes the standard cure rate model and the label ‘2’ denotes the PCR model.
• The Historical Data

Several years earlier, a similar melanoma study with the same patient population was conducted by ECOG. This study, denoted by E1673, serves as the historical data for our Bayesian analysis of E1684. A total of $n_0 = 650$ patients are used in the historical data. Table 2.2 summarizes the historical data E1673.
### Table 2.2. Summary of E1673 Data

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survival Time ($y_0$) (years)</strong></td>
<td>8.80</td>
<td>(5.44, 11.62)</td>
</tr>
<tr>
<td><strong>Status (frequency)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Censored</td>
<td>257</td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>393</td>
<td></td>
</tr>
<tr>
<td><strong>Age ($x_{01}$) (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>48.0</td>
<td></td>
</tr>
<tr>
<td>Std Dev</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td><strong>Gender ($x_{02}$) (frequency)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>375</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>275</td>
<td></td>
</tr>
<tr>
<td><strong>PS ($x_{03}$) (frequency)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully Active</td>
<td>561</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>89</td>
<td></td>
</tr>
</tbody>
</table>
• Initial Prior

For the initial prior for \( \beta \), we take an improper uniform prior, and for \( \pi_0(\alpha|\nu_0, \tau_0) \), we take \( \nu_0 = 1 \) and \( \tau_0 = 0.01 \) to ensure a proper prior. We note that this choice for \( \pi_0(\alpha|\nu_0, \tau_0) \) also guarantees log-concavity. The parameter \( \lambda \) is taken to have a normal distribution with mean 0 and variance 10,000.
### Table 2.3. Posterior Estimates of the Model Parameters

| $E(a_0|D_{obs})$ | Variable | Posterior Mean | Posterior Std Dev | 95% HPD Interval |
|------------------|----------|----------------|-------------------|------------------|
| 0 (w.p.1)        | intercept | 0.09           | 0.11              | (−0.12, 0.30)    |
|                  | age      | 0.09           | 0.07              | (−0.05, 0.23)    |
|                  | gender   | −0.12          | 0.16              | (−0.44, 0.19)    |
|                  | ps       | −0.23          | 0.26              | (−0.73, 0.28)    |
|                  | α        | 1.31           | 0.09              | (1.15, 1.48)     |
|                  | λ        | −1.36          | 0.12              | (−1.60, −1.11)   |
| 0.14             | intercept | 0.25           | 0.10              | (0.05, 0.45)     |
|                  | age      | 0.12           | 0.06              | (−0.00, 0.24)    |
|                  | gender   | −0.20          | 0.14              | (−0.47, 0.07)    |
|                  | ps       | −0.09          | 0.22              | (−0.53, 0.31)    |
|                  | α        | 1.06           | 0.06              | (0.95, 1.17)     |
|                  | λ        | −1.62          | 0.12              | (−1.85, −1.39)   |
| 0.29             | intercept | 0.26           | 0.09              | (0.08, 0.43)     |
|                  | age      | 0.13           | 0.06              | (0.02, 0.24)     |
|                  | gender   | −0.24          | 0.12              | (−0.48, 0.00)    |
|                  | ps       | −0.01          | 0.19              | (−0.38, 0.35)    |
|                  | α        | 1.03           | 0.05              | (0.93, 1.13)     |
|                  | λ        | −1.70          | 0.11              | (−1.91, −1.50)   |
| 1 (w.p.1)        | intercept | 0.22           | 0.06              | (0.11, 0.35)     |
|                  | age      | 0.16           | 0.04              | (0.08, 0.24)     |
|                  | gender   | −0.32          | 0.09              | (−0.50, −0.15)   |
|                  | ps       | 0.14           | 0.13              | (−0.11, 0.39)    |
|                  | α        | 1.00           | 0.04              | (0.93, 1.07)     |
|                  | λ        | −1.82          | 0.08              | (−1.97, −1.67)   |
Posterior Estimates

Table 2.3 indicates a fairly robust pattern of behavior. The estimates of the posterior mean, standard deviation, or highest posterior density (HPD) intervals of $\beta$ do not change a great deal if a low or moderate weight is given to the historical data. However, if a higher than moderate weight is given to the historical data, these posterior summaries can change a lot. For example, when the posterior mean of $a_0$ is less than 0.14, we see that all of the HPD intervals for $\beta$ include 0, and when the posterior mean of $a_0$ is greater than or equal to 0.14, some HPD intervals for $\beta$ do not include 0. The HPD interval for age does not include 0 when the posterior mean of $a_0$ is 0.29, and it includes 0 when less weight is given to the historical data. This finding is interesting, since it indicates that age is a potentially important prognostic factor for predicting survival in melanoma. Such a conclusion is not possible based on a frequentist or Bayesian analysis of the current data alone.
• The Posterior Estimates of the Cure Rates

**Figure 2.4.** Box plots of the Posterior Means of the Cure Rates for All Patients where “1” denotes no incorporation of historical data and “2” corresponds to $E(a_0|D_{obs}) = 0.29$
Session 2

- **Interpretation of Figure 2.4**

The mean and standard deviations are 0.36 and 0.05 for box-plot 1 ($a_0 = 0$) and 0.31 and 0.06 for box-plot 2 ($E(a_0|D_{obs}) = 0.29$). Thus we see that the mean cure rate drops from 0.36 to 0.31 when the historical data are incorporated. A partial explanation of this result is due to the fact that the historical data are much more mature than the current data, with nearly 20 years of follow-up and a smaller fraction of censored cases. Thus, the results in Figure 2.4 are not surprising, and in fact appealing, since they give us a better estimate of the cure rate compared to an estimate based on the current data alone. Such a conclusion cannot be reached by a frequentist or Bayesian analysis using only the E1684 data.
• Sensitivity Analysis

A detailed sensitivity analysis for the regression coefficients was conducted by varying the hyperparameters for \( a_0 \) (i.e., \( (\gamma_0, \lambda_0) \)) and varying the hyperparameters for \( \psi = (\alpha, \lambda) \). The posterior estimates of the parameters are fairly robust as the hyperparameters \( (\gamma_0, \lambda_0) \) are varied. When we vary the hyperparameters for \( \psi \), the posterior estimates of \( \beta \) are also robust for a wide range of hyperparameter values. For example, when fixing the hyperparameters for \( a_0 \) so that \( E(a_0|D_{obs}) = 0.29 \) and taking \( \alpha \sim \mathcal{G}(1, 1) \) and \( \lambda \sim \mathcal{N}(0, 10) \), we obtain the posterior estimates shown in Table 2.4. We see that these priors for \( (\alpha, \lambda) \) are fairly informative relative to those of Table 2.3. Other moderate to informative choices of hyperparameters for \( (\alpha, \lambda) \) also led to fairly robust posterior estimates of \( \beta \).
Table 2.4. Posterior Estimates of the Model Parameters with $E(a_0|D_{obs}) = 0.29$, $\alpha \sim G(1, 1)$ and $\lambda \sim N(0, 10)$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Posterior Mean</th>
<th>Posterior Std Dev</th>
<th>95% HPD</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.26</td>
<td>0.09</td>
<td>(0.07, 0.42)</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>0.13</td>
<td>0.06</td>
<td>(0.02, 0.24)</td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>-0.24</td>
<td>0.12</td>
<td>(-0.48, 0.00)</td>
<td></td>
</tr>
<tr>
<td>ps</td>
<td>-0.01</td>
<td>0.19</td>
<td>(-0.38, 0.35)</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1.02</td>
<td>0.05</td>
<td>(0.93, 1.12)</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-1.69</td>
<td>0.11</td>
<td>(-1.90, -1.48)</td>
<td></td>
</tr>
</tbody>
</table>
2.3 Semiparametric Cure Rate Models

◊ Version I

• Piecewise Constant Hazard Model for $F(y)$

We specify a semiparametric form for $F(\cdot)$. To do this, we first construct a finite partition of the time axis, $0 < s_1 < \ldots < s_J$, with $s_J > y_i$ for all $i = 1, 2, \ldots, n$. Thus, we have the $J$ intervals $(0, s_1], (s_1, s_2], \ldots, (s_{J-1}, s_J]$. We then assume that the hazard for $F(y)$ is equal to $\lambda_j$ for the $j^{th}$ interval, $j = 1, 2, \ldots, J$, leading to

$$F(y) = 1 - \exp \left\{ -\lambda_j (y - s_{j-1}) - \sum_{g=1}^{j-1} \lambda_g (s_g - s_{g-1}) \right\}.$$ 

We note that when $J = 1$, $F(y)$ reduces to the parametric exponential model.
• Likelihood Function

The complete data likelihood can be written as

\[
L(\beta, \lambda|D) = \prod_{i=1}^{n} \prod_{j=1}^{J} \exp \left\{ - (N_i - \nu_i) \nu_{ij} \left[ \lambda_j (y_i - s_{j-1}) + \sum_{g=1}^{j-1} \lambda_g (s_g - s_{g-1}) \right] \right\}
\times \prod_{i=1}^{n} \prod_{j=1}^{J} (N_i \lambda_j)^{\nu_{ij}} \nu_i \exp \left\{ - \nu_i \nu_{ij} \left[ \lambda_j (y_i - s_{j-1}) + \sum_{g=1}^{j-1} \lambda_g (s_g - s_{g-1}) \right] \right\}
\times \exp \left\{ \sum_{i=1}^{n} \left[ N_i \mathbf{x}_i' \beta - \log(N_i!) - \exp(\mathbf{x}_i' \beta) \right] \right\},
\]

where \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_J)' \) and \( \nu_{ij} = 1 \) if the \( i^{th} \) subject failed or was censored in the \( j^{th} \) interval, and 0 otherwise.
- **Attractive Features**

First, we note the degree of the nonparametricity is controlled by $J$. The larger the $J$, the more nonparametric the model is. However, by picking a small to moderate $J$, we get more of a parametric shape for $F(y)$. This is an important aspect for the cure rate model, since the estimation of the cure rate parameter $\theta$ could be highly affected by the nonparametric nature of $F(y)$. For this reason, it may be desirable to choose small to moderate values of $J$ for cure rate modeling. In practice, we recommend doing analyses for several values of $J$ to see the sensitivity of the posterior estimates of the regression coefficients. This semiparametric cure rate model is quite flexible, as it allows us to model general shapes of the hazard function, as well as choose the degree of parametricity in $F(y)$ through suitable choices of $J$. 
• **Initial Prior**

The initial prior $\pi_0(\beta, \lambda)$ is taken to be

$$
\pi_0(\beta, \lambda) \propto \prod_{j=1}^{J} \lambda_j^{\zeta_0 - 1} \exp(-\tau_0 \lambda_j),
$$

so that $\beta$ has an improper uniform initial prior and the $\lambda_j$'s have i.i.d. $\mathcal{G}(\zeta_0, \tau_0)$ distributions.

• **Power Prior**

The joint power prior for $(\beta, \lambda, a_0)$ is given by

$$
\pi(\beta, \lambda, a_0|D_{0,obs}) \propto \left( \sum_{N_0} \mathcal{L}(\beta, \lambda|D_0) \right)^{a_0} \pi_0(\beta, \lambda) a_0^{\gamma_0 - 1} (1 - a_0)^{\lambda_0 - 1},
$$

where $(\gamma_0, \lambda_0)$ are specified hyperparameters for the prior distribution of $a_0$. 

2-47

M.-H. Chen
• Choices of Partitions

We propose three different constructions of \((s_{j-1}, s_j]\). That is, we choose the subintervals \((s_{j-1}, s_j]\) with

(i) equal numbers of failures or censored observations;

(ii) approximately equal lengths subject to the restriction that at least one failure occurs in each interval;

(iii) decreasing numbers of failures or censored observations.

In case (iii), \(s_j\) may be taken to be the \(((1 - e^{(-j/J)})/(1 - e^{-1}))^{th}\) quantile of the \(y_j\)'s.
Version II

• Rationale

A crucial issue with cure rate modeling, and semiparametric survival models in general, is the behavior of the model in the right tail of the survival distribution. In these models, there are typically few subjects at risk in the tail of the survival curve after sufficient follow-up, and therefore estimation of the cure rate can be quite sensitive to the choice of the semiparametric model. Thus there is a need to carefully model the right tail of the survival curve, and allow the model to be more parametric in the tail, while also allowing the model to be nonparametric in other parts of the curve.
• Finite Partition

We construct a finite partition of the time axis, \(0 < s_1 < \ldots < s_J\), with \(s_J > y_i\) for all \(i = 1, 2, \ldots, n\). Thus, we have the \(J\) intervals \((0, s_1], (s_1, s_2], \ldots, (s_{J-1}, s_J]\). We then assume that the hazard for \(F(y)\) is equal to \(\lambda_j\) for the \(j^{th}\) interval, \(j = 1, 2, \ldots, J\).

• Smooth Parameter \(\kappa\)

Let \(F_0(y|\psi_0)\) denote the parametric survival model we wish to choose for the right tail of the survival curve, and let \(H_0(y)\) denote the corresponding cumulative baseline hazard function. Now we take the \(\lambda_j\)'s to be independent a priori, each having a gamma prior distribution with mean

\[\mu_j = E(\lambda_j|\psi_0) = \frac{H_0(s_j) - H_0(s_{j-1})}{s_j - s_{j-1}},\]

and variance

\[\sigma_j^2 = \text{Var}(\lambda_j|\psi_0, \kappa) = \mu_j \kappa^j,\]

where \(0 < \kappa < 1\) is the smoothing parameter.
• Attractive Features

- As $\kappa \to 0$, $\sigma_j^2 \to 0$, so that small values of $\kappa$ imply a more parametric model in the right tail.

- As $j \to \infty$, $\sigma_j^2 \to 0$, implying that the degree of parametricity is increased at a rate governed by $\kappa$ as the number of intervals increases. This property also implies that as $j \to \infty$, the survival distribution in the right tail becomes more parametric regardless of any fixed value of $\kappa$.

- If $F_0(.|\psi_0)$ is an exponential distribution, then
  
  $$F_0(y|\psi_0) = 1 - \exp(-\psi_0 y),$$
  
  so that $\mu_j = \psi_0$ and $\sigma_j^2 = \psi_0 \kappa^j$.

- If $F_0(.|\psi_0)$ is a Weibull distribution, then
  
  $$F_0(y|\psi_0) = 1 - \exp(-\gamma_0 y^{\alpha_0}),$$
  
  $\psi_0 = (\alpha_0, \gamma_0)'$, so that
  
  $$\mu_j = \gamma_0 \frac{(s_j^{\alpha_0} - s_{j-1}^{\alpha_0})}{s_j - s_{j-1}}$$
  
  and $\sigma_j^2 = \gamma_0 \frac{(s_j^{\alpha_0} - s_{j-1}^{\alpha_0})}{s_j - s_{j-1}} \kappa^j$. 

• Properties

Let

\[ F^*(y|\lambda) = 1 - \exp \left\{ -\lambda_j (y - s_{j-1}) - \sum_{g=1}^{j-1} \lambda_g (s_g - s_{g-1}) \right\} . \]

**Property 2.1** Assume that \( \frac{s_j + s_{j-1}}{2} \to y \) as \( s_j - s_{j-1} \to 0 \). Then for any \( j \), according to this prior process, \( E(\lambda_j|\psi_0) \to h_0(y) \) as \( s_j - s_{j-1} \to 0 \), where \( h_0(y) = \frac{d}{dt} H_0(y) \).

For example, when \( F_0(y|\psi_0) = 1 - \exp(-\psi_0 y) \), then \( E(\lambda_j|\psi_0) = \psi_0 \) regardless of our choice of \( s_1, s_2, \ldots, s_J \). When \( F_0(y|\psi_0) = 1 - \exp(-\gamma_0 y^{\alpha_0}) \), then \( E(\lambda_j|\psi_0) \to \gamma_0 \alpha_0 t^{\alpha_0-1} \) as \( s_j - s_{j-1} \to 0 \). This assures that as \( j \) becomes large and \( s_j - s_{j-1} \to 0 \), then this prior process approximates any prior process with prior mean \( h_0(y) \) defined on the promotion time hazard \( h^*(y|\lambda) \) corresponding to \( F^*(y|\lambda) \).
Property 2.2 Let $S_p^*(y|\lambda) = \exp(-\theta F^*(y|\lambda))$. Then, $S_p^*(y|\lambda) \to S_p(y|\psi_0)$ as $\kappa \to 0$, where $S_p(y|\psi_0) = \exp(-\theta F_0(y|\psi_0))$.

Property 2.3 Let $f^*(y|\lambda) = \frac{d}{dy} F^*(y|\lambda)$, and $h_p^*(y|\lambda) = \theta f^*(y|\lambda)$ denote the corresponding hazard function. Then $h_p^*(y|\lambda) \to \theta f_0(y|\psi_0)$ as $\kappa \to 0$, where $f_0(y|\psi_0) = \frac{d}{dy} F_0(y|\psi_0)$.
• Prior Distribution

We specify a hierarchical model and consider a joint (improper) noninformative prior distribution for \((\beta, \lambda, \psi_0)\). We specify the joint prior of these parameters as

\[
\pi(\beta, \lambda, \psi_0) = \pi(\beta)\pi(\lambda|\psi_0)\pi(\psi_0) \propto \pi(\beta) \left[ \prod_{j=1}^{J} \pi(\lambda_j|\psi_0) \right] \pi(\psi_0).
\]

As noted earlier, we take each \(\pi(\lambda_j|\psi_0)\) to be independent gamma densities with mean \(\mu_j\) and variance \(\sigma_j^2\). If \(F_0(.|\psi_0)\) is an exponential distribution, then \(\psi_0\) is a scalar, and we specify a gamma prior for it, i.e.,

\[
\pi(\psi_0) \propto \psi_0^{\zeta_0-1} \exp(-\tau_0 \psi_0),
\]

where \(\zeta_0\) and \(\tau_0\) are specified hyperparameters. If \(F_0(.|\psi_0)\) is a Weibull distribution, then \(\psi_0 = (\alpha_0, \gamma_0)'\). In this case, we take a prior of the form

\[
\pi(\psi_0) = \pi(\alpha_0, \gamma_0) \propto \alpha_0^{\zeta_0} e^{-\tau_0 \alpha_0} \gamma_0^{\zeta_0} e^{-\tau_0 \gamma_0},
\]

where \(\zeta_{\alpha_0}, \tau_{\alpha_0}, \zeta_{\gamma_0}\) and \(\tau_{\gamma_0}\) are specified hyperparameters. For \(\beta\), we consider a uniform improper prior.
Propriety of the Posterior

Suppose (i) when \( \nu_i = 1, y_i > 0 \), (ii) there exists \( i_1, i_2, \ldots, i_J \) such that \( \nu_{i_j} = 1 \), and \( s_{j-1} < y_{i_j} \leq s_j, j = 1, 2, \ldots, J \), (iii) the design matrix \( X^* \) with \( i^{th} \) row equal to \( \nu_i x_i' \) is of full rank, iv) if \( F_0(.|\psi_0) \) is an exponential distribution, \( \zeta_0 > 0 \) and

\[
\tau_0 > \sum_{j=1}^{J} \frac{1}{\kappa_j} \log[(1/\kappa_j)/((y_{i_j} - s_{j-1})/2 + 1/\kappa_j)],
\]

and if \( F_0(.|\psi_0) \) is a Weibull distribution, \( \zeta_{\gamma_0} > 0, \tau_{\gamma_0} \geq 0, \zeta_{\alpha_0} > 0, \) and \( \tau_{\alpha_0} > -\zeta_{\gamma_0} \log(s_J) \). Then the posterior distribution of \( (\beta, \lambda, \psi_0) \) is proper, i.e.,

\[
\int L(\beta, \lambda|D)\pi(\beta, \lambda, \psi_0) \, d\beta d\lambda d\psi_0 < \infty,
\]

where \( L(\beta, \lambda|D) \) is the likelihood function based on the observed data \( D \).
Example: Melanoma Data

- We revisit the E1684 and E1690 trials discussed earlier.
- One purpose in this example is to examine the tail behavior of the model as $\kappa$, $a_0$, $F_0$, and $J$ are varied.
- Of particular interest is the sensitivity of the posterior estimates of $\beta$, $\lambda$, and $S^*(t|\lambda) = 1 - F^*(t|\lambda)$, as these parameters are varied.
- The E1690 study is quite suitable for our purposes here since the median follow-up for E1690 (4.33 years) is considerably smaller than E1684 (6.9 years). Thus, cure rate estimation based on the E1690 study alone, i.e., $a_0 = 0$, may be more sensitive than that of an analysis which incorporates the historical data E1684.
- The covariates included in the model are treatment (IFN, OBS), age, which is continuous, and gender (male, female).
### Table 2.5. Posterior Estimates of $\beta$

<table>
<thead>
<tr>
<th>$F_0$</th>
<th>$\kappa$</th>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exponential</strong></td>
<td>0.05</td>
<td>intercept</td>
<td>0.183</td>
<td>0.096</td>
<td>(−0.009, 0.367)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>treatment</td>
<td>−0.242</td>
<td>0.115</td>
<td>(−0.469, −0.018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>age</td>
<td>0.099</td>
<td>0.058</td>
<td>(−0.011, 0.214)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gender</td>
<td>−0.118</td>
<td>0.120</td>
<td>(−0.361, 0.110)</td>
</tr>
<tr>
<td>0.95</td>
<td></td>
<td>intercept</td>
<td>0.157</td>
<td>0.098</td>
<td>(−0.038, 0.345)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>treatment</td>
<td>−0.242</td>
<td>0.116</td>
<td>(−0.472, −0.017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>age</td>
<td>0.097</td>
<td>0.058</td>
<td>(−0.015, 0.211)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gender</td>
<td>−0.113</td>
<td>0.121</td>
<td>(−0.348, 0.128)</td>
</tr>
<tr>
<td><strong>Weibull</strong></td>
<td>0.05</td>
<td>intercept</td>
<td>0.202</td>
<td>0.102</td>
<td>(0.002, 0.404)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>treatment</td>
<td>−0.242</td>
<td>0.115</td>
<td>(−0.471, −0.019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>age</td>
<td>0.100</td>
<td>0.057</td>
<td>(−0.012, 0.213)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gender</td>
<td>−0.118</td>
<td>0.121</td>
<td>(−0.358, 0.114)</td>
</tr>
<tr>
<td>0.95</td>
<td></td>
<td>intercept</td>
<td>0.160</td>
<td>0.097</td>
<td>(−0.034, 0.345)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>treatment</td>
<td>−0.244</td>
<td>0.116</td>
<td>(−0.473, −0.022)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>age</td>
<td>0.097</td>
<td>0.058</td>
<td>(−0.013, 0.213)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gender</td>
<td>−0.115</td>
<td>0.120</td>
<td>(−0.351, 0.119)</td>
</tr>
</tbody>
</table>
• Table 2.5 gives posterior means, standard deviations, and 95% HPD intervals of $\beta$ for several values of $\kappa$ using the exponential and Weibull models for $F_0$ with $J = 10$ intervals, and when $E(a_0|D) = 0.33$.

• As $\kappa$ is varied for a given $a_0$ using an exponential or Weibull $F_0$, we see small to moderate changes in the posterior estimates of $\beta$.

• As $a_0$ is varied, more substantial changes occur in the posterior estimates of $\beta$ across values of $a_0$. For example, using an exponential $F_0$ and $\kappa = 0.05$, the posterior means, standard deviations, and 95% HPD intervals for the treatment coefficient, i.e., $\beta_2$, are $-0.209$, $0.130$, and $(-0.461, 0.050)$ when $a_0 = 0$ with probability 1; $-0.242$, $0.115$, and $(-0.469, -0.018)$ when $E(a_0|D) = 0.33$; and $-0.277$, $0.079$, and $(-0.462, -0.087)$ when $a_0 = 1$ with probability 1.

• Overall, we conclude that the estimates of $\beta$ are reasonably robust as $\kappa$ is varied for given $a_0$, but change substantially as $a_0$ is varied.
Figure 2.5. Plots of survival function $S^*(t|\lambda)$ for non-cured patients with $J = 10$ and $E(a_0|D) = 0.33$. 
• Figure 2.5 shows the posterior estimates of $S^*(t|\lambda)$ for $E(a_0|D) = 0.33$ using several values of $\kappa$.

• In Figure 2.5, (a) $F_0(\cdot|\psi_0)$ is an exponential distribution; (b) $F_0(\cdot|\psi_0)$ is a Weibull distribution; and the solid, dotted, dashed, and dot-dashed curves correspond to $\kappa = 0.05, 0.30, 0.60, 0.95$, respectively.

• We see from Figure 2.5 that small to moderate changes in the survival estimates occur as $\kappa$ is varied. The biggest changes occur in the interval $1 \leq t \leq 5$.

• Thus, $S^*(t|\lambda)$ can be moderately sensitive to the choice of $\kappa$. 
2.4 Multivariate Cure Rate Models

♣ The Model

• $Y = (Y_1, Y_2)$: a bivariate failure time (such as $Y_1 =$ time to cancer relapse and $Y_2 =$ time from relapse to death).

• $N = (N_1, N_2)$: the unobserved number of latent risk factors for that individual with respect to $(Y_1, Y_2)$.

• Assume that $N_k$ has a Poisson distribution with mean $\theta_k \omega$, $k = 1, 2$ and $(N_1, N_2)$ are independent, and unobserved. $N_1$ can denote the unobserved number of clonogens (carcinogenic cells) remaining in the patient’s body following a treatment, and $N_2$ can denote the unobserved number of complications due to the adverse effects of treatment as well as the patient’s physical condition, which may cause death.
• The quantity $w$ is a frailty component in the model which induces a correlation between $(N_1, N_2)$.

• We take $w \sim S_\alpha(1, 1, 0)$ (a positive stable distribution indexed by the parameter $\alpha$), where $0 < \alpha < 1$.

• $Z_i = (Z_{1i}, Z_{2i})$: the bivariate promotion time for the $i^{th}$ metastasis-competent tumor cell.

• Assume $Z_{ki}$ has a cdf $F_k(y)$ and a survival function $S_k(y) = 1 - F_k(y)$, and $Z_{ki}$'s are independent of $N$.

• $Y_k = \min \{Z_{ki}, 0 \leq i \leq N_k\}$: the observed survival time for $k = 1, 2$. 
♦ The Survival Function

- The survival function for the population given $w$ is given by

$$S_{pop}(y_1, y_2 | w)$$

$$= \prod_{k=1}^{2} \left( P(N_k = 0) + P(Z_{k1} > y_k, \ldots, Z_{kN_k} > y_k, N_k \geq 1) \right)$$

$$= \prod_{k=1}^{2} \exp(-w \theta_k) + \prod_{k=1}^{2} \left( \sum_{r=1}^{\infty} S_k(y_k)^r \frac{(w \theta_k)^r}{r!} \exp(-w \theta_k) \right)$$

$$= \prod_{k=1}^{2} \left( \exp(-w \theta_k + \theta_k w S_k(y_k)) \right)$$

$$= \exp(-w [\theta_1 F_1(y_1) + \theta_2 F_2(y_2)]) .$$

- If we take $w \sim S_{\alpha}(1, 1, 0)$, using the Laplace transform of $w$ (i.e.,

$$E(\exp(-sw)) = \exp(-s^\alpha))$$

we obtain the unconditional survival function:

$$S_{pop}(y_1, y_2) = \exp \left\{ -\left[ \theta_1 F_1(y_1) + \theta_2 F_2(y_2) \right]^\alpha \right\} .$$
Some Properties

- The marginal survival functions are
  \[ S_k(y) = \exp(-\theta_k^\alpha (F_k(y))^\alpha), \quad k = 1, 2. \]

- The marginal hazard function is given by
  \[ \alpha \theta_k^\alpha f_k(y)(F_k(y))^{\alpha-1}. \]

- We can write the marginal survival functions in terms of standard cure rate models:
  \[
  S_k(y) &= \exp(-\theta_k^\alpha (F_k(y))^\alpha) \\
  &= \exp(-\theta_k^\alpha) + (1 - \exp(-\theta_k^\alpha)) \left( \frac{\exp(-\theta_k^\alpha (F_k(y))^\alpha) - \exp(-\theta_k^\alpha)}{1 - \exp(-\theta_k^\alpha)} \right) \\
  &= \exp(-\theta_k^\alpha) + (1 - \exp(-\theta_k^\alpha)) S_k^*(y),
  \]
  where
  \[
  S_k^*(y) = \frac{\exp(-\theta_k^\alpha (F_k(y))^\alpha) - \exp(-\theta_k^\alpha)}{1 - \exp(-\theta_k^\alpha)}.
  \]
 Meaning of $\alpha$

- Following Clayton (1978) and Oakes (1989), we can compute a local measure of association, denoted, $\theta^*(y_1, y_2)$, as a function of $\alpha$. This measure of association is defined as

$$\theta^*(y_1, y_2) = \frac{S_{pop}(y_1, y_2)\frac{\partial^2}{\partial y_1 \partial y_2} S_{pop}(y_1, y_2)}{(\frac{\partial}{\partial y_1} S_{pop}(y_1, y_2))(\frac{\partial}{\partial y_2} S_{pop}(y_1, y_2))}.$$ 

- For the multivariate cure rate model, we have

$$\theta^*(y_1, y_2) = \alpha^{-1}(1 - \alpha)(\theta_1 F_1(y_1) + \theta_2 F_2(y_2))^{-\alpha} + 1.$$ 

We see that $\theta^*(y_1, y_2)$ decreases in $(y_1, y_2)$. That is, the association between $(Y_1, Y_2)$ is greater when $(Y_1, Y_2)$ are small and the association decreases over time.

- The parameter $\alpha$ ($0 < \alpha < 1$) is a scalar parameter that is a measure of association between $(Y_1, Y_2)$. Small values of $\alpha$ indicate high association between $(Y_1, Y_2)$. As $\alpha \to 1$, $\theta^*(y_1, y_2) \to 1$, which implies less association between $(Y_1, Y_2)$. 

2-65 M.-H. Chen
Session 2

♦ Note

A global measure of dependence such as Kendall’s \( \tau \) or the Pearson correlation coefficient is not well defined for the multivariate cure rate model since no moments for cure rate models exist due to the improper survival function.

♦ The Likelihood Function

• Notation

- \( y_{ki} \): survival time (failure time or right-censored)
- \( \nu_{ki} \): the censoring indicator; \( \nu_{ki} = 1 \) if \( y_{ki} \) is the failure time and 0 if \( y_{ki} \) is right censored.
- \( N_{ki} \): the number of carcinogenic cells
- \( \mathbf{x}'_i = (x_{i1}, \ldots, x_{ip}) \): the \( p \times 1 \) vector of covariates
- \( n \): the sample size
- \( \boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp})' \): the vector of regression coefficients
- We model \( \theta_{ki} = \exp(\mathbf{x}'_i \boldsymbol{\beta}_k) \)
Session 2

- Complete data: $D = (n, y_1, y_2, \nu_1, \nu_2, X, N_1, N_2, w)$, where $y_k = (y_{k1}, \ldots, y_{kn})'$, $\nu_k = (\nu_{k1}, \ldots, \nu_{kn})'$, $w = (w_1, \ldots, w_n)'$, $X$ is the $n \times p$ matrix of covariates, and $N_k = (N_{k1}, \ldots, N_{kn})'$.

- Observed data: $D_{obs} = (n, y_1, y_2, \nu_1, \nu_2, X)$.

- Assume a Weibull distribution for cdf $F_k(y_{ki}|\psi_k)$ and pdf $f_k(y_{ki}|\psi_k)$ so that

$$f_k(y_{ki}|\psi_k) = \xi_k y_i^{\xi_k-1} \exp \left\{ \lambda_k - y_i^{\xi_k} \exp(\lambda_k) \right\},$$

and $\psi_k = (\xi_k, \lambda_k)'$. 

The Complete Data Likelihood Function

Let $\theta = (\theta_1, \theta_2)'$ and $\psi = (\psi_1, \psi_2)'$. Then, the likelihood function of $(\theta, \psi)$ based on the complete data $D$ is given by

$$L(\theta, \psi | D) = \left( \prod_{k=1}^{2} \prod_{i=1}^{n} S_k(y_{ki} | \psi_k)^{N_{ki} - \nu_{ki}} (N_{ki} f_k(y_{ki} | \psi_k))^\nu_{ki} \right)$$

$$\times \exp \left\{ \sum_{i=1}^{n} (N_{ki} \log(w_i \theta_k) - \log(N_{ki}!) - w_i \theta_k) \right\}.$$
teams: The Observed Data Likelihood Function

Let \( f_s(w_i|\alpha) \) denote the \( S_\alpha(1, 1, 0) \) density for each \( w_i \). The likelihood function based on the observed data, denoted \( L(\theta, \psi, \alpha|D_{obs}) \), is given by

\[
L(\theta, \psi, \alpha|D_{obs}) \equiv \int_{\mathbb{R}^n} \sum_{\mathcal{N}} L(\theta, \psi|D) \times \left[ \prod_{i=1}^n f_s(w_i|\alpha) \right] dw
\]

\[
= \theta_1^{d_1} \theta_2^{d_2} \alpha^{d_1 + d_2} \left[ \prod_{k=1}^{2} \prod_{i=1}^{n} f_k(y_{ki}|\psi_k)^{\nu_{ki}} \right]
\]

\[
\times \prod_{i=1}^{n} \left\{ [\theta_1 F_1(y_{1i}|\psi_1) + \theta_2 F_2(y_{2i}|\psi_2)]^{(\alpha-1)(\nu_{1i} + \nu_{2i})} \right\}
\]

\[
\times \prod_{i=1}^{n} \left[ \alpha^{-1} (1 - \alpha)(\theta_1 F_1(y_{1i}|\psi_1) + \theta_2 F_2(y_{2i}|\psi_2))^{-\alpha} + 1 \right]^{\nu_{1i} \nu_{2i}}
\]

\[
\times \prod_{i=1}^{n} \exp \left\{ -(\theta_1 F_1(y_{1i}|\psi_1) + \theta_2 F_2(y_{2i}|\psi_2))^{\alpha} \right\},
\]

where \( d_k = \sum_{i=1}^{n} \nu_{ki} \) for \( k = 1, 2 \), \( R^{+n} = R^+ \times R^+ \times \cdots \times R^+ \), and \( R^+ = (0, \infty) \).
We relate $\theta$ to the covariates by

$$
\theta_{ki} \equiv \theta(x_i'\beta_k) = \exp(x_i'\beta_k),
$$

so that the cure rate for subject $i$ is

$$
\exp(-\theta_{ki}) = \exp(-\exp(x_i'\beta_k)),
$$

for $i = 1, 2, \ldots, n$ and $k = 1, 2$. Letting $\beta = (\beta_1', \beta_2')'$, we can write the observed data likelihood of $(\beta, \psi, \alpha)$ as

$$
L(\beta, \psi, \alpha|D_{obs}) = \left(\alpha^{d_1+d_2} \prod_{k=1}^{2} \prod_{i \in D_k} e^{x_i'\beta_k} \right) \left[ \prod_{k=1}^{2} \prod_{i=1}^{n} f_k(y_{ki}|\psi_k)^{\nu_{ki}} \right] \\
\times \prod_{i=1}^{n} \left\{ \left[ e^{x_i'\beta_1} F_1(y_{1i}|\psi_1) + e^{x_i'\beta_2} F_2(y_{2i}|\psi_2) \right]^{(\alpha-1)(\nu_1+\nu_2)} \right\} \\
\times \prod_{i=1}^{n} \left\{ \frac{1-\alpha}{\alpha} \left[ e^{x_i'\beta_1} F_1(y_{1i}|\psi_1) + e^{x_i'\beta_2} F_2(y_{2i}|\psi_2) \right]^{-\alpha} + 1 \right\}^{\nu_1\nu_2} \\
\times \prod_{i=1}^{n} \exp \left\{ -\left( e^{x_i'\beta_1} F_1(y_{1i}|\psi_1) + e^{x_i'\beta_2} F_2(y_{2i}|\psi_2) \right)^{\alpha} \right\},
$$

where $D_k$ consists of those patients who failed according to $Y_k$, $k = 1, 2$, and $F_k(y_{ki}|\psi_k) = 1 - \exp(-y_{ki}^{\xi_k} \exp(\lambda_k))$. 

2-70  

M.-H. Chen
The Prior and Posterior Distributions

We consider a joint improper prior for \((\beta, \psi, \alpha) = (\beta_1, \beta_2, \psi_1, \psi_2, \alpha)\) of the form

\[
\pi(\beta, \psi, \alpha) = \pi(\beta_1, \beta_2, \psi_1, \psi_2, \alpha) \propto \pi(\psi_1)\pi(\psi_2)I(0 < \alpha < 1) = \prod_{k=1}^{2} \pi(\xi_k, \lambda_k)I(0 < \alpha < 1),
\]

where \(I(0 < \alpha < 1) = 1\) if \(0 < \alpha < 1\), and 0 otherwise,

\[
\pi(\psi_k) = \pi(\xi_k, \lambda_k) = \pi(\xi_k \vert \nu_0, \tau_0)\pi(\lambda_k),
\]

\[
\pi(\xi_k \vert \delta_0, \tau_0) \propto \xi_k^{\delta_0-1} \exp\{-\tau_0\xi_k\}, \quad \pi(\lambda_k) \propto \exp\{-c_0\lambda_k^2\},
\]

and \(\delta_0, \tau_0,\) and \(c_0\) are specified hyperparameters.

**Theorem 2.4** Let \(X_k^*\) be an \(n \times p\) matrix with rows \(\nu_k x_{ki}'\) for \(k = 1, 2\). Then if (i) \(X_k^*\) is of full rank for \(k = 1, 2\), (ii) \(\pi(\lambda_k)\) is proper, and (iii) \(\tau_0 > 0\) and \(\delta_0 > -\min\{d_1, d_2\}\), then the posterior,

\[
\pi(\beta, \psi, \alpha \vert D_{obs}) \propto L(\beta, \psi, \alpha \vert D_{obs}) \prod_{k=1}^{2} \pi(\xi_k \vert \delta_0, \tau_0)\pi(\lambda_k),
\]

is proper.
Computational Implementation

Following Ibragimov and Chernin (1959), the positive stable \( \text{Stable}(\alpha) \) density can be expressed in the form

\[
f_s(w|\alpha) = aw^{-(a+1)} \int_0^1 s(u) \exp \left\{-\frac{s(u)}{w^a}\right\} du, \quad w > 0,
\]

where

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

It can be shown that \( f_s(w|\alpha) \) is the marginal distribution of the joint density

\[
f(w, u|\alpha) = aw^{-(a+1)} s(u) \exp \left\{-\frac{s(u)}{w^a}\right\},
\]

\( w > 0, \quad 0 < u < 1 \). This relationship plays an important role in the implementation of the Gibbs sampler.
To facilitate the Gibbs sampler, we introduce several auxiliary (latent) variables. These include $\mathbf{N} = (\mathbf{N}_1, \mathbf{N}_2)$, where $\mathbf{N}_k = (N_{k1}, \ldots, N_{kn})'$ for $k = 1, 2$, $\mathbf{w} = (w_1, w_2, \ldots, w_n)$, and $\mathbf{u} = (u_1, u_2, \ldots, u_n)$. The joint posterior distribution of $(\beta, \psi, \alpha, \mathbf{N}, \mathbf{w}, \mathbf{u}|D_{obs})$ is given by

$$
\pi(\beta, \psi, \alpha, \mathbf{N}, \mathbf{w}, \mathbf{u}|D_{obs}) \\
\propto \left( \prod_{k=1}^{2} \prod_{i=1}^{n} S_k(y_{ki} | \psi_k)^{N_{ki} - \nu_{ki}} \ (N_{ki}f_k(y_{ki} | \psi_k))^{\nu_{ki}} \right) \\
\times \exp \left\{ \sum_{i=1}^{n} (N_{ki} \log(w_i \theta_{ki}) - \log(N_{ki}!)) - w_i \theta_{ki} \right\} \\
\times \prod_{i=1}^{n} \left[ a w_i^{-(\alpha+1)} s(u_i) \exp \left\{ - \frac{s(u_i)}{w_i^\alpha} \right\} \right] \\
\times \prod_{k=1}^{2} \left( \pi(\xi_k|\nu_0, \tau_0)\pi(\lambda_k) \right),
$$

where $\theta_{ki} = \exp(\mathbf{x}_i' \beta_k)$.
◊ Outline of the MCMC Sampling Algorithm

- We use the collapsed Gibbs procedure of Liu (1994).
- We standardize all covariates.
- We need to sample from the following conditional distributions:
  
  (i) \([\psi | \beta, \alpha, N, w, u, D_{obs}]\) (logconcave, exact sampling);

  (ii) \([\beta | \alpha, w, u, \psi, D_{obs}]\) (collapsing, logconcave, exact sampling);

  (iii) \([\alpha | \beta, \psi, D_{obs}]\) (collapsing, Metropolis-Hastings);

  (iv) \([w, u | \alpha, \beta, \psi, D_{obs}]\) (ROU, exact sampling); and

  (v) \([N | \beta, \alpha, w, u, \psi, D_{obs}]\) (collapsing, exact sampling).
Example: E1684 Melanoma Data

- **Response Variables**: $Y_1$ = time to relapse from randomization and $Y_2$ = relapse to death.

- **Covariates**: age ($x_1$), gender ($x_2$) (male, female), and performance status ($x_3$) (fully active, other).

- **Sample size**: $n = 274$.

- **Priors**: We use the noninformative improper prior with $\pi(\beta) \propto 1$, $\lambda_k \sim N(0, 10, 000)$, $\xi_k \sim G(1, 0.01)$, and independent for each $k = 1, 2$. Also, we take a uniform prior for $\alpha$ on the interval (0, 1).

- **Posterior Estimate of $\alpha$**: The posterior mean of $\alpha$ is 0.709, with a 95% HPD interval of (0.585, 0.840). This indicates a moderate association between time to relapse and relapse to death for these data, as was expected.
- **Estimates of Cure Rates**: The median cure rate for time to relapse is 0.285 and the median cure rate for relapse to death is 0.103.

**Figure 2.6.** Box-plots of the posterior means of the cure rates for all patients.
- **Superimposed Marginal Survival Curves**: Figure 2.7 shows two superimposed plots, where plot (a) represents time to relapse and plot (b) represents relapse to death. The covariates are not used in constructing plots (a) and (b). In plot (a), the two superimposed plots correspond to the Kaplan-Meier estimate of survival, and the maximum likelihood estimate of the marginal survival function based on the multivariate cure rate model. We see that the two curves in plot (a) are nearly identical, and appear to plateau after approximately six years of follow-up. In plot (b), the relapse to death variable appears to plateau after approximately four years of follow-up.
Figure 2.7. Superimposed survival curves for a subset of E1684.
• **Joint Survival Surface**: Figure 2.8 shows a three-dimensional plot of the posterior mean survival surface based on average age for males with fully active performance status. We see in this plot how the survival curve plateaus for each failure time variable. The joint survival function approaches a joint cure fraction, and the marginal survival functions each approach a cure fraction. From this figure, it is clear that the estimated cure rate for the time to relapse variable is larger than the estimated cure rate for the relapse to death variable.
Figure 2.8. The bivariate posterior survival surface.