A BAYESIAN WEIBULL SURVIVAL MODEL

by

Jiang Qian

Institute of Statistics and Decision Sciences
Duke University

Date: _____
Approved:

_____
Donald A. Berry, Advisor

_____
Donald S. Burdick

_____
Saurabh Mukhopadhyay

_____
Dalene K. Stangl

_____
Mike West

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Institute of Statistics and Decision Sciences
in the Graduate School of
Duke University

1994

ABSTRACT

(Statistics)

A BAYESIAN WEIBULL SURVIVAL MODEL

by

Jiang Qian

Institute of Statistics and Decision Sciences
Duke University

Date: _____
Approved:

_____
Donald A. Berry, Advisor

_____
Donald S. Burdick

_____
Saurabh Mukhopadhyay

_____
Dalene K. Stangl

_____
Mike West

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Institute of Statistics and Decision Sciences
in the Graduate School of
Duke University

1994

Abstract

This dissertation explores a Bayesian Weibull survival model used to analyze data from clinical trials, and examines the Gibbs sampling scheme employed to estimate posterior distributions. I review the definition and some properties of a Weibull survival distribution, introduce two parameterizations for comparing treatments in a clinical trial, and discuss various treatment comparison measures. It is difficult to use the Weibull model to analyze and evaluate the posterior distributions of the parameters analytically. Therefore, I use an approximation based on Monte Carlo integration to obtain the posterior distributions of the parameters and predictive measures to compare the treatments.

A mixture model with a Weibull component and a surviving fraction is introduced. I employ a Gibbs sampling procedure incorporating the Adaptive Rejection Sampling Method, to obtain the posterior distribution of parameters and the predictive measure of vaccine efficacy.

This dissertation shows the practicality of the Bayesian Weibull survival modeling. I use the Weibull model to analyze data collected at several interim analysis times during a Non-Small Cell Lung Cancer (NSCLC) clinical trial. I also provide posterior distributions of parameters, give the numerical results about the relative efficacy of treatments, and address the sensitivity issue about

the prior choice.

I employ the mixture model to analyze data from a *Haemophilus Influenzae type B* vaccine trial, and to explore various methodological and computational aspects. I use a Gibbs sampling scheme to calculate the predictive protective efficacy (PE) of the vaccine. I also address the sensitivity of the choice of prior distribution.

Acknowledgements

I would like to express my loving gratitude to my husband, Guoliang Cao, for his understanding and unwavering support. He was always there in the bad times and the good times, making everything that much nicer. I also thank my parents, Ruiping Feng and Jinguo Qian, for their continuing encouragement and unselfish sacrifice during my study. Without such support, this effort would hardly have been possible.

I would like to express my sincere appreciation to Professor Donald A. Berry. Apart from being an excellent advisor, he is such a nice professor. Certainly his patience and understanding far exceed reasonable expectations. His inspiration and guidance influenced every aspect of this dissertation. He provided valuable direction and editorial suggestions that proved most helpful to me in writting this dissertation, and he reassured me during times of self-doubt.

I would like to thank the members of my dissertation committee, Professors Donald S. Burdick, Dalene K. Stangl, Mike West and Saurabh Mukhopadhy, for their advice and encouragement, comments, and criticism. A special note of thanks is due to Professor Dalene K. Stangl, for all the interest and enthusiasm that she showed, the ideas that she shared, and the suggestions that she made. Professor Steven George of the Cancer Center of Duke University deserves my

special thanks for his suggestions and comments which not only improved the quality of my dissertation, but will also be useful in my future research.

I would like to thank the faculty, staff, and graduate students of the Institute of Statistics and Decision Sciences, for providing such a good environment for the development of my dissertation. Professors Mike West, Donald S. Burdick, Saurabh Mukhopadhy, Valen Johnson, and Robert Wolpert were particularly helpful to me. These groups have instructed, inspired, and befriended me during my doctoral studies at Duke University.

In addition, I would like to thank the Cancer and Leukemia Group B (CALGB) for letting me use their NSCLC clinical trial data in my dissertation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Statement of the Problem

A clinical trial or clinical study is an experiment designed to assess the safety and efficacy of one or more medical treatments, such as new drugs, using human beings as the patients or subjects. Such a trial is an exercise in applied science; and its design and execution, as well as the analysis of the resulting data, must therefore conform as closely as possible to the principles and procedures of good scientific experimentation. Events of interest include illness, death or a similarly serious event. Frequently, clinical trials are designed to compare the efficacy of two different treatments for a disease. Usually, researchers compare data from patients treated with a standard therapy to data from patients treated with an experimental therapy, or researchers compare the effects of a new drug to those of the standard drug, or they compare the data from patients treated with a combined therapy to the data from patients treated with a standard single therapy.

When analyzing clinical data, we must decide how to deal with censored data. Sometimes, the event of interest does not occur in every one on the

treatment arm. Observations on those patients who do not experience the event of interest are called censored. Often, censoring occurs because some of the individuals under observation have not experienced the event by the end of the study when the statistical analysis is done, because some individuals withdraw from the study, or because individuals move and are lost to follow up. Because these censored observations also provide information about the treatment, they should be included in our analysis. The need to accommodate censored observations led to the development of the survival analysis.

During our analysis of clinical data, we also must decide how to analyze interim data. In many clinical trials, the primary endpoint is the time elapsed until the event of interest occurs. For example, in some clinical trials, the primary endpoint is the patient's survival time or the duration of remission or recovery. However, sometimes the clinical trial may be stopped early. Whether or not the trial may be stopped early is another major concern. Stopping a trial early has a major impact on the classical inferences that may be drawn from the accumulating death/survival data and the survival time data, but it does not affect Bayesian inferences (Berry, 1989). More statistical methods are needed to analyze those interim data.

Many different fields, including medicine, biology, and engineering use survival analysis. More recently survival models have been considered in newer application areas in financial and socio-economic studies and this trend is sure to continue (Gamerman and West, 1987). But applications for analyzing interim data within a Bayesian framework in clinical trials are just beginning to be used (Berry, 1989). In this dissertation, I propose a fully Bayesian Weibull survival model to analyze both interim and final clinical trial data. The goals of such an analysis are to provide posterior estimates for parameters of interest

and predictive measures of treatment comparisons, and also to compare results from different endpoints. I discuss the choice of parameter transformation, and employ a hierarchical structure to establish a fully Bayesian model.

1.2 Literature Review

Numerous books are available presenting survival analysis in various fields. Books such as Mann, Schafer and Singpurwalla (1974) and Nelson (1982) are oriented more towards engineering applications, while others such as Lawless (1982) and Cox and Oakes (1984) are more in the medical data. Kalbfleisch and Prentice (1980) is an advanced text with particular emphasis on proportional hazards methods, while Martz and Waller (1982) focused on the Bayesian approach. Recently, Crowder, Kimber, Smith and Sweeting (1991) discussed both the probability modeling and the statistical aspects of reliability/survival analysis. Klein and Goel (1991) presented some Bayesian methods for survival analysis.

Historically, survival analysis have generally employed classical concepts of estimation, confidence intervals and hypothesis testing. During the last 10 years or so, however, there has been increasing interest in Bayesian methods applied to survival analysis. A number of reasons might be identified for this. First, the rapid development of statistical computing, and in particular numerical integration routines for Bayesian statistics, have allowed Bayesian methods to be applied to complicated models of the sort employed in survival analysis. The papers of Naylor and Smith (1982) and Gelfand and Smith (1990), proposing new numerical integration routines which have since been extensively implemented, were particularly important in this development. Secondly, many

survival problems arise in situations where there is a high level of uncertainty, either because of the influence of uncertain factors which are hard to quantify, or because of the survival prediction problem involves significant extrapolation from any available data. Also, in some situations the data are simply inadequate for any predictions to be made with high level of confidence. Thus it is recognized that practical judgements are very often, inevitably, strongly guided by subjective judgement. Bayesian statistics provides a means of quantifying subjective judgements and combining them in a rigorous way with information obtained from experimental data.

Researchers have recently used Bayesian models to analyze survival time data. Gamerman and West (1987) applied new dynamic Bayesian models for survival data analysis in a study of contributory factors to unemployment. Grieve (1988) used a probit model to dertermin the LD50 from acute toxicity tests. A Bayesian approach to the analysis of LD50 experiments gives a meaningful classification of the toxicity of a substance even when the traditional method of maximum likelihood gives results which are not useful for practical purposes. Gamerman (1991) and Gamerman (1992) proposed dynamic models for the study of survival data with explanatory variables whose effects change through time. The hazard function was assumed to be piecewise constant along intervals. A system equation provides the stochastic link for adjacent values. The updating equations are obtained via the dynamic generalized modeling approach of West, Harrison and Migon (1985). West (1992) used piecewise proportional hazards models to analyze nasopharynx cancer survival data.

Bayesian nonparametric survival analysis was considered by Berliner and Hill (1988). Their results are applicable to a wide variety of such problems, including reliability analysis and medical survival analysis. The analysis hinged

on three assumptions: (a) The new patients and the previous sample patients are all deemed to be exchangeable with regard to survival time. (b) The posterior prediction rule, in the case of no censoring or ties among (say n) observed survival times, assigns equal probability of $\frac{1}{n+1}$ to each of the $n+1$ open intervals determined by these values. (c) The censoring mechanisms are "noninformative". Detailed discussion of these assumptions can be found in Berliner and Hill (1988). Hill (1992) compared Bayesian nonparametric survival estimator with classical nonparametric survival estimator (Kaplan-Meier estimator) and indicated that the Bayesian estimator always gives more mass to the upper tail of the distribution than the Kaplan-Meier estimator does, sometimes substantially more.

In clinical trials, Berry (1989) employed an exponential model to analyze clinical data. He assumed that patient's survival time on each treatment arm is exponentially distributed given hyperparameters λ_A and λ_B (which are the hazard rates of the treatment arms). The posterior probability that the mortality rate on A is larger than that on B depends on the initial prior distribution (λ_A and λ_B respectively). Given certain priors on λ_A and λ_B, the posterior probability of $\{\lambda_A > \lambda_B\}$ can be calculated at each interim analysis and used as guidance in deciding whether and when to stop a clinical trial.

Later on, Berry, Wolff, and Sack (1992) used an exponential distribution to model the survival times in a sequential vaccine trial. Using an exponential model to minimize the expected number of cases of *Haemophilus influenzae type B*, Berry, Wolff, and Sack (1992) suggested to close the study just three months before the actual stopping of the trial.

An exponential distribution is a special case of a Weibull distribution. That

is, the hazard function of an exponential distribution is a constant whereas the hazard function of a Weibull distribution is a function of time. Using a Weibull distribution provides a wide range of applications in a clinical study. A Weibull survival model was not used so easily until the Gibbs sampling technique was invented. This is because features such as data censoring, ordered parameters, assumed convexity or concavity of distributions, all conspire to produce complicatedly constrained regions over which numerical integrations are required. Not surprisingly, the literature therefore contains very few instances of fully Bayesian analyses in survival contexts. Recently, Gelfand and Smith (1991) have shown that the Gibbs sampler approach to Bayesian computation (for example, Gelfand and Smith, 1990) effectively side-steps the seeming problems of awkwardly defined integration regions in truncated data and constrained parameter problems, and provides an easily implemented computational procedure.

Gibbs sampling is an iterative resampling approach. Bayesian inferences can easily be made by generating posterior samples from full conditional distributions. Kuo and Smith (1992) discussed computational problems by using Gibbs sampler for censored data set or constraints in the form of ordered parameters. An adaptive rejection sampling method has been incorporated to a Gibbs sampling method for a family of log-concave density functions. Dellaportas and Smith (1993) showed that Gibbs sampling, making systematic use of an adaptive rejection algorithm proposed by Gilks and Wild (1992), provides a straightforward computational procedure for Bayesian inferences on proportional hazard models.

Escobar and West (1991) used a Gibbs sampling method to obtain Bayesian density estimation. Cao (1993) used a Gibbs sampler in Bayesian nonparamet-

ric mixture model.

Recently, Stangl (1991) used Bayesian hierarchical exponential, mixture and changepoint models to model the heterogeneity of treatment effects among medical centers in multicenter clinical trials (Stangl, 1991). She explored the heterogeneity of the relative treatment effect among medical centers in terms of the posterior distributions of first- and second-stage parameters, as well as the posterior predictive distributions of survival time differences between treatments. Stangl (1991) used the Laplace method and the Gibbs Sampling algorithm to approximate posterior distributions of parameters and predictive distributions of future observations.

A mixture model for the analysis of survival data from cancer was first proposed by Berkson and Gage (1952). Chen, Hill, Greenhouse and Fayos (1985) generalized the mixture model of Berkson and Gage (1952) and used for the analysis of survival data from cancer. An important feature of mixture analysis is that by use of the likelihood principle we can factor out certain proportionality constants. Another feature of mixture model is that it is extremely effective, even for very small samples, with only a handful of deaths. This is of extreme importance for making early decisions as to the efficacy of treatments, since inferior treatments can be rejected without the necessity of waiting to observe a substantial number of deaths. Stangle (1991) used such a mixture model to analyze a NIMH collaborative study using a Gibbs sampling approach.

There are many softwares available to perform survival analysis. Most classical inferences about survival analysis can be made by using SAS (1986). For example, The Kaplan-Meier estimator is available from SAS (1986). Gamer-

man, West and Pole (1987) developed a software to perform dynamic Bayesian modeling for survival data. Now the adaptive rejection sampling and Gibbs sampling are implemented in BUGS (1993) which is a program that carries out Bayesian inference on statistical problems using a simulation technique known as Gibbs sampling.

1.3 Outline of Dissertation

In this dissertation, I shall employ the Importance Sampling (IS) and the Gibbs sampling technique to analyze two Weibull survival models; and I shall use the Adaptive Rejection Sampling method to obtain samples from the conditional distributions of the shape parameters in the mixture model.

Chapter 2 introduces a Bayesian Weibull model with two parameterizations for comparing two treatments in a clinical trial. The first parameterization assumes that the parameters on one treatment arm are independent of those on the other treatment arm. The second parameterization assumes dependence between the scale parameters on both treatment arms. In a special case, when the shape parameters on both arms are equal to 1, the relationship between the scale parameters is modeled as log-hazard ratio in the exponential model. A numerical illustration of using the first parameterization is given using an interim data set from Berry (1989).

Chapter 3 applies the model proposed in Chapter 2 to a Non-Small Cell Lung Cancer (NSCLC) clinical trial sponsored by the Cancer and Leukemia Group B. I present numerical results of the posterior distribution and the predictive measures of treatment comparisons, and discuss the sensitivity of the results to prior distributions.

Chapter 4 introduces a mixture model, with a Weibull component and a surviving fraction. The survival times on each treatment arm are assumed to have a Weibull distribution with parameters θ_i and a_i, and with probability $1 - \pi_i$. The remaining proportion of the observations will not experience the event within the trial duration. In this mixture model, the survival function consists of two terms, one corresponding to each component of the mixture. I then describe the Gibbs sampling procedures, incorporating Adaptive Rejection Method, that were used to obtain the posterior distributions.

Chapter 5 applies the model proposed in Chapter 4 to the *Haemophilus type B* vaccine trial data (Berry, Wolff, and Sack, 1992). I present the numerical results of the posterior distributions of some interesting parameters, and I calculate the predictive protective efficacy (PE) under several prior selections and address the sensitivity issue.

In Chapter 6, I outline areas of future research related to the Bayesian Weibull survival analysis.

Chapter 2

Bayesian Weibull Model In Survival Analysis

2.1 Introduction

This chapter describes a Bayesian Weibull survival model for analyzing survival data from clinical trials. Survival time is the time until an event occurs. For example, survival time may be the lifetime of a patient or time until recurrence of some disease of the patient. Commonly, lifetime or survival data includes censored observations due to the withdrawal of experimental subjects or the termination of the experiment. For each censored observation, we only know that the patient's lifetime exceeded a given value. The exact lifetime remains unknown. Censored observations should not be ignored when analyzing survival data, because, among other considerations, the longer-lived subjects are generally more likely to be censored. The analysis methodology must correctly use the censored observations, as well as the uncensored observations.

Usually, a first step in the analysis of survival data is the estimation of the distribution of the failure times. The survival function is used to describe the

lifetimes of the population of interest. The survival function evaluated at t is the probability that an experimental subject from the population will have a lifetime exceeding t, that is

$$S(t) = Pr(T > t) \tag{2.1}$$

where $S(t)$ is the survival function and T is the lifetime of a randomly selected experimental subject. The usual estimation method is the product-limit method (also called the Kaplan-Meier method) or the life table method.

Some functions closely related to the survival function are the cumulative distribution function, $F(t)$; the probability function, $f(t)$; and the hazard function, $h(t)$. $S(t)$ is defined as $1 - F(t)$ and is the probability that a lifetime exceeds t. $f(t)$ is the derivative of $F(t)$, and the hazard function $h(t)$ is the instantaneous rate of failure given survival up to time t or

$$lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t | T > t)}{\Delta} \tag{2.2}$$

which turns out to be $f(t)/S(t)$.

The likelihood function in survival analysis is more complicated. If there are no censored data, the likelihood function is only related to the density function, $f(t)$; otherwise, the likelihood function depends on the density function and the survival function. For example, suppose that censoring is random and improper, and that the survival data is $\mathbf{t} = \{t_i; i = 1, 2, \cdots, n\}$. Then the likelihood of the data is proportional to the product of the event time density and the survival functions.

$$L(\mathbf{t}|\beta) \propto \prod_{i=1}^{n} f(t_i|\beta)^{\delta_i} S(t_i|\beta)^{1-\delta_i} \tag{2.3}$$

where β is the unknown parameters associated with the density function. δ_i is defined as follows:

$$\delta_i = \begin{cases} 0 & \text{if the observation is censored.} \\ 1 & \text{if the observation experiences the event.} \end{cases}$$

Berry (1989), Stangl (1991), Berry et al. (1992) and George et al. (1993) discussed the cases when data are assumed from an exponential distribution. Grieve (1985) applied Gaussian-Hermit quadrature method (Naylor and Smith, 1982) to a proportional hazard model. We propose here to use a Weibull distribution to model the survival time on both treatment arms.

2.2 Weibull Distribution

Survival time T has a Weibull distribution of $W(t|\theta, a)$, if its hazard function is:

$$h(t|\theta, a) = \theta a t^{a-1} \tag{2.4}$$

or its survival function is:

$$S(t|\theta, a) = exp(-\theta t^a) \tag{2.5}$$

where θ is the scale parameter, a is the shape parameter, and $\theta > 0$ and $a > 0$. One important feature of Weibull distribution $W(t|\theta, a)$ is that when $a < 1$, the failure rate of the subject decreases; when $a > 1$, the failure rate of the subject increases; when $a = 1$, the failure rate of the subject is constant, which indicates an exponential distribution. For the Weibull distribution $W(t|\theta, a)$, the probability density function is

$$f(t|\theta, a) = \theta a t^{a-1} exp(-\theta t^a) \tag{2.6}$$

and its cumulative distribution is

$$F(t|\theta, a) = 1 - exp(-\theta t^a) \tag{2.7}$$

The expectation and variance of the Weibull distribution are:

$$E(t) = ?(1 + a^{-1})/\theta^{\frac{1}{a}} \tag{2.8}$$

$$Var(t) = [?(1 + 2a^{-1}) - (?(1 + a^{-1}))^2]/\theta^{\frac{2}{a}} \tag{2.9}$$

The likelihood function for Weibull distributed data, $\mathbf{t} = \{t_i; i = 1, 2, \cdots, n\}$ with random censoring $\delta = \{\delta_i; i = 1, 2, \cdots, n\}$, is

$$L(\mathbf{t}|\theta, a) \propto \prod_{i=1}^{n} (\theta a t_i^{a-1} e^{-\theta t_i^a})^{\delta_i} (e^{-\theta t_i^a})^{1-\delta_i} \tag{2.10}$$

2.3 Weibull Model with Independent Parameters

2.3.1 The Model

We assume that the survival times of patients on each treatment arm in one clinical trial are *i.i.d.* observations from a Weibull distribution, and that our main goal is to compare the relative efficacy of two or more treatments. Each treatment corresponds to a pair of Weibull parameters, say, $(\theta_i, a_i), i = 1, 2$.

We use $i = 1$ to denote the standard treatment and $i = 2$ to denote the experimental treatment, and t_{ij} to denote the survival time of the j-th patient on the i-th treatment that we observed at the time of our analysis. Thus, our observations are actually $(t_{ij}, \delta_{ij}), j = 1, 2, \cdots, n_i, i = 1, 2$, where t_{ij} is the survival time and δ_{ij} is the status of the j-th patient on the i-th treatment. That is:

$$\delta_{ij} = \begin{cases} 0 & \text{if } t_{ij} \text{ is censored.} \\ 1 & \text{otherwise.} \end{cases}$$

Then the distribution of t_{ij} is the Weibull distribution $W(t_{ij}|\theta_i, a_i)$. n_i represents the number of total patients in treatment i, and d_i represents the number of deaths in treatment i at the time of our analysis (for example, at interim analysis time point). $\{t_{ij}, j = 1, 2, \cdots, d_i, i = 1, 2\}$ represents the uncensored observations, and $\{t_{ij}, j = d_i + 1, \cdots, n_i, i = 1, 2\}$ represents the censored observations. Thus, we can specify a Bayesian Weibull hierarchical model within the Bayesian paradigm. At the first stage of the hierarchy, we assume that observations t_{ij}, $j = 1, 2, \cdots, n_i, i = 1, 2$, conditional on (θ_i, a_i), are independent and identically distributed with $W(t_{ij}|\theta_i, a_i)$. This stage provides a parametric model for lifetime t_{ij}. At the second stage, we give prior distributions for unknown parameters (θ_i, a_i) $i = 1, 2$. Under the stated assumptions, the Weibull model takes the form:

Stage I:

$$(t_{ij}|\theta_i, a_i) \sim W(t_{ij}|\theta_i, a_i) \qquad j = 1, 2, \cdots, n_i, \quad i = 1, 2. \tag{2.11}$$

Stage II:

$$\begin{aligned} (\theta_i|u_\theta, v_\theta) &\overset{i.i.d.}{\sim} f(\theta_i|u_\theta, v_\theta) \\ (a_i|u_a, v_a) &\overset{i.i.d.}{\sim} g(a_i|u_a, v_a) \\ i &= 1, 2. \end{aligned} \tag{2.12}$$

In this model, inference may initially focus on the posterior distribution of the shape parameters a_i, $i = 1, 2$. But because they are not directly related

to the hazard rate or the survival function by themselves, we will mainly focus on the predictive difference between the survival function at some time point of interest, and on the predictive probability that a future patient will survive longer on one treatment than on another until some time point of interest, or on the predictive probability that the mean survival time for one treatment is greater than that for the other.

If t_{ij} is a censored time, then its contribution to the likelihood is the survival function $S(t_{ij}|\theta_i, a_i) = e^{-\theta_i t_{ij}^{a_i}}$. Thus, the likelihood function for $(\theta_1, a_1, \theta_2, a_2)$ is:

$$\begin{aligned} L(\theta_1, a_1, \theta_2, a_2) &= \prod_{i=1}^{2} \{ \prod_{j=1}^{d_i} f(t_{ij}|\theta_i, a_i) \prod_{j=d_i+1}^{n_i} S(t_{ij}|\theta_i, a_i) \} \\ &= \prod_{i=1}^{2} \{ (\theta_i a_i)^{d_i} (\prod_{j=1}^{d_i} t_{ij})^{a_i - 1} exp(-\theta_i \sum_{j=1}^{n_i} {t_{ij}}^{a_i}) \} \end{aligned} \tag{2.13}$$

2.3.2 The Prior for $(\theta_1, a_1, \theta_2, a_2)$

We assume that the priors on (θ_1, a_1) and (θ_2, a_2) are independent. Since, in a Weibull distribution, the shape parameter $a_i < 1$ corresponds to decreasing failure rate and vice versa, and $a_i = 1, i = 1, 2$ is an interesting point to be considered in choosing the prior distribution for $a_i, i = 1, 2$. If we want to determine whether or not an exponential model is reasonable for the data set, we may chose a gamma prior for both a_1 and a_2 with a mode equal to 1 (in which case the Weibull model collapses and becomes an exponential model). That is:

$$a_i \sim f(a_i|u_a, v_a) = G(a_i|u_a, v_a), \quad i = 1, 2 \tag{2.14}$$

where $G(a_i|u_a, v_a)$ denotes a gamma distribution with shape parameter u_a and scale parameter v_a. (u_a, v_a) are chosen such that the mode of the gamma distribution is 1 and the distribution has some reasonable variance. By reasonable, we mean that if we have very vague prior information about whether the disease has an increasing or a decreasing death rate then we should choose u_a, v_a so that the variance of the gamma distribution is large, otherwise, we may choose (u_a, v_a) which gives smaller variance for the gamma distribution.

Then, we can check the posterior distribution of the $a_i, i = 1, 2$ to see if the mode of the posterior density of $a_i, i = 1, 2$ is still about 1 or if the 95% highest posterior density (HPD) contains $a_i = 1, i = 1, 2$.

For the scale parameter $\theta_i, i = 1, 2$, since if $a_i = 1$ for $i = 1, 2$, the Weibull distribution collapses and becomes an exponential distribution with parameter $\theta_i, i = 1, 2$ as its hazard rate. We shall choose a gamma prior for both θ_1 and θ_2 with a mode equal to estimated hazard rate using information from historical studies or training data set under the *exponential* survival time assumption. That is:

$$\theta_i \sim g(\theta_i|u_\theta, v_\theta) = G(\theta_i|u_\theta, v_\theta), \quad i = 1, 2 \tag{2.15}$$

We chose u_θ and v_θ so that the mode of the gamma distribution is the estimated hazard rate derived from our prior knowledge about the disease, and its variance reflects the uncertainty about the hazard rate of the disease.

Thus, the prior for $(\theta_1, a_1, \theta_2, a_2)$ is:

$$\pi(\theta_1, a_1, \theta_2, a_2) = G(\theta_1|u_\theta, v_\theta) \times G(a_1|u_a, v_a) \times G(\theta_2|u_\theta, v_\theta) \times G(a_2|u_a, v_a) \tag{2.16}$$

2.3.3 The Posterior Distributions

Under the model and the prior $\pi(\theta_1, a_1, \theta_2, a_2)$, the posterior density for $(\theta_1, a_1, \theta_2, a_2)$ is:

$$\begin{aligned} p(\theta_1, a_1, \theta_2, a_2|data) &\propto L(\theta_1, a_1, \theta_2, a_2)\pi(\theta_1, a_1, \theta_2, a_2) \\ &\propto \prod_{i=1}^{2}\{(\theta_i a_i)^{d_i}(\prod_{j=1}^{d_i} t_{ij})^{a_i-1} exp(-\theta_i \sum_{j=1}^{n_i} {t_{ij}}^{a_i}) \\ &\quad \mathrm{G}(\theta_i|u_\theta, v_\theta)\mathrm{G}(a_i|u_a, v_a)\} \\ &\propto \prod_{i=1}^{2}\{\theta_i^{d_i+u_\theta-1} e^{-\theta_i(v_\theta+\sum_{j=1}^{n_i} t_{ij}^{a_i})} \\ &\quad a_i^{d_i+u_a-1}(\prod_{j=1}^{d_i} t_{ij})^{a_i-1} e^{-v_a a_i}\} \end{aligned} \tag{2.17}$$

2.4 Weibull Model with Dependent Parameters

2.4.1 The Model

In the Weibull model with independent parameters, let $v = log(\frac{\theta_2}{\theta_1})$. Thus, when $a_1 = a_2 = 1$, v stands for the log-hazard ratio of Treatment 2 (the experimental treatment) to Treatment 1 (the standard treatment) and it is positive or negative depending upon which treatment is better. The absolute value of the v represents the degree of the difference between the two treatments; and the larger the $|v|$, the more significant the difference between the two treatments. When $a_i, i = 1, 2$ is not exactly equal to 1, the meaning of the v may be a little bit different, but we can still use this parameterization to compare the relative efficacy of the treatments. When we use this parameterization, the

Weibull model takes the form:

Stage I:

$$(t_{ij}|\theta_i, a_i) \sim W(t_{ij}|\theta_i, a_i); \qquad j = 1, 2, \cdots, n_i, \quad i = 1, 2. \tag{2.18}$$

Stage II:

$$\begin{aligned} (\theta_1|u_\theta, v_\theta) &\sim f(\theta_1|u_\theta, v_\theta) \\ (v|\sigma^2) &\sim l(v|\sigma^2) \\ (a_i|u_a, v_a) &\overset{i.i.d.}{\sim} g(a_i|u_a, v_a), \quad i = 1, 2 \end{aligned}$$

2.4.2 The Prior for (θ_1, v, a_1, a_2)

As in Section 2.3.2, we will take the prior distributions for (a_1, a_2) to be $i.i.d.$ gamma distribution with shape parameter u_a and scale parameter v_a. For θ_1, we will take the prior distribution to be $G(\theta_1|u_\theta, v_\theta)$. The hyperparameters u_a, v_a, u_θ, v_θ were chosen using the criteria described in Section 2.3.2 and some historical information about the standard treatment. Since v represents the log- hazard ratio when $a_1 = a_2 = 1$, we will choose a normal distribution with mean 0 and variance σ^2 as the prior distribution. This prior gives the same prior probability for $v < 0$ (Treatment 2 is better than Treatment 1) and $v > 0$ (Treatment 1 is better than Treatment 2). σ's value reflects the degree of our prior belief about v, i.e., the smaller the σ, the stronger our prior belief about the relative efficacy of the two treatments.

2.4.3 The Posterior Distributions

Laplace Method

Calculating posterior distributions requires integration. When integrands are not in a tractable form approximations methods are needed. One useful integral approximation method is the Laplace method adapted by Tierney and Kadane (1986). The Laplace method is suited for integration problems when the integrand is smooth and has a single dominant peak in a region that is roughly known. Partition the $d \times 1$ vector β as (β_1, β_2), where β_1 is a scalar and β_2 is $(d-1)\times 1$. For a given value of β_1, let $\hat{\beta}_2^* = \hat{\beta}_2^*(\beta_1)$ maximize $p(\beta_1, \beta_2)$ with β_1 fixed and let:

$$\hat{\Sigma}^* = [\frac{\partial^2 log p(\beta_1, \beta_2|Y)}{\partial \beta_2^2}|_{\hat{\beta}_2^*}]^{-1} \tag{2.19}$$

where $\hat{\Sigma^*}$ is a $(d-1)\times(d-1)$ matrix. Tierney and Kadane (1986) approximate $p(\beta|Y)$ with

$$\hat{p}(\beta_1|Y) \propto (det\hat{\Sigma}^*)^{\frac{1}{2}} p(\beta_1, \beta_2^*|Y) = (det\hat{\Sigma}^*)^{\frac{1}{2}} \text{profile posterior of } \beta_1). \tag{2.20}$$

To obtain the marginal of an arbitrary function $g(\beta)$, Tierney, Kass and Kadane (1988) approximate $\hat{p}(\gamma|Y)$ with

$$\hat{p}(\gamma|Y) \propto [\frac{d}{e} t\{\Sigma(\gamma)\} det\{(D_g)^T \Sigma(\gamma)(D_g)\}]^{\frac{1}{2}} p\{\hat{\beta}(\gamma)|Y\} \tag{2.21}$$

$\hat{\beta}(\gamma)$ maximizes the posterior $p(\beta|Y)$ subject to the constraint that $g(\beta) = \gamma$, $\Sigma(\gamma)$ is the inverse Hessian of the log posterior evaluated at $\hat{\beta}(\gamma)$ and D_g is the gradient or Jacobian evaluated at $\hat{\beta}(\gamma)$.

The Posterior Distributions

Given the prior $\pi(\theta_1, v, a_1, a_2)$ and data $\{t_{ij}; j = 1, 2, \cdots, n_i, \quad i = 1, 2\}$ at any particular time point of interest, the posterior density $p(\theta_1, v, a_1, a_2|data)$ is as follows:

$$\begin{aligned} p(\theta_1, v, a_1, a_2|data) &\propto L(\theta_1, v, a_1, a_2) \times \pi(\theta_1, v, a_1, a_2) \\ &\propto L(\theta_1, v, a_1, a_2) \times \theta_1^{u_\theta - 1} e^{-v_\theta \theta_1} \times e^{\frac{-v^2}{2\sigma^2}} \times a_1^{u_a - 1} e^{-v_a a_1} \times a_2^{u_a - 1} e^{-v_a a_2} \\ &\propto (\theta_1 a_1)^{d_1} (\prod_{j=1}^{d_1} t_{1j})^{a_1 - 1} exp(-\theta_1 \sum_{j=1}^{n_1} {t_{1j}}^{a_1}) \\ &\quad (\theta_1 e^v a_2)^{d_2} (\prod_{j=1}^{d_2} t_{2j})^{a_2 - 1} exp(-\theta_1 e^v \sum_{j=1}^{n_2} {t_{2j}}^{a_2}) \\ &\quad \times \theta_1^{u_\theta - 1} e^{-v_\theta \theta_1} \times e^{\frac{-v^2}{2\sigma^2}} \times a_1^{u_a - 1} e^{-v_a a_1} \times a_2^{u_a - 1} e^{-v_a a_2} \end{aligned} \tag{2.22}$$

where $data = \{t_{ij}; j = 1, 2, \cdots, n_i, \quad i = 1, 2\}$.

The Laplace method based software XlispStat may be used to obtain all the marginal posterior distributions: $p(a_1|data), p(a_2|data), p(\theta_1|data)$, and $p(v|data)$.

2.5 Comparing Two Treatments

There are many ways to compare the effects of different treatments. Four of them are:

1. $\mathrm{D}(t|data) = Pr(S_1(t) < S_2(t)|data)$

   $\mathrm{D}(t|data)$ is the predictive probability that survival probability after time t on Treatment 1 is smaller than that on Treatment 2. Thus it is a function of time t. Here $S_1(t)$ and $S_2(t)$ are the survival functions under

Treatment 1 and Treatment 2, respectively. Note that if an exponential model is used, then D(t) will be constant over the time t.

2. $\mathrm{d}(t|data) = E(S_1(t) - S_2(t)|data)$

   $\mathrm{d}(t|data)$ is the predictive mean difference between the survival probability after time t on Treatment 1 and that on Treatment 2. Like D(t), d(t) will be constant over the time t if an exponential model is used.

3. $Pr(m_1 < m_2|data)$

   where m_i is the mean survival time of patients on treatment i. $Pr(m_1 < m_2|data)$ is the predictive probability that the mean survival time of patients on Treatment 2 will be longer than that for those on Treatment 1.

4. $Pr(T_1 < T_2|data)$

   where $Pr(T_1 < T_2|data)$ is the predictive probability that any future patient will live longer on Treatment 2 than on Treatment 1.

2.6 Numerical Method for Computing $\mathbf{D}(t|data)$, $\mathbf{d}(t|data)$

Although we may obtain the marginal posterior distribution of interested parameters via XlispStat, we need numerical methods to compute the treatment comparison measures. In this section, we will discuss the Importance Sampling technique to compute the treatment comparison measures. We may also use other methods (such as Gibbs sampling plus a rejection method) to perform these calculations.

2.6.1 The Importance Sampling Method

Consider the problem of calculating the integral

$$p(y) = \int f(y|x)g(x)dx \tag{2.23}$$

If one can sample directly from $g(x)$, then $p(y)$ can be easily got from

$$\begin{aligned} p(y) &= \int f(y|x)g(x)dx \\ &= \frac{1}{N}\sum_{i=1}^{N} f(y|x_i) \end{aligned} \tag{2.24}$$

where N is the Monte Carlo sample size and $x_1, x_2, \cdots, x_N$ are $i.i.d.$ samples from $g(x)$.

In cases where one can not directly sample from $g(x)$, Importance Sampling Method can be used. Let $J(x)$ be an easy to sample from density which approximates $g(x)$. The Method of Importance Sampling approximates $p(y)$ as:

1. Draw $x_1, x_2, \cdots, x_N \overset{i.i.d.}{\sim} J(x)$
2. $\tilde{p(y)} = \sum_{i=1}^{N} w_i f(y|x_i) / \sum_{i=1}^{N} w_i, \qquad w_i = \frac{g(x_i)}{J(x_i)}$

Importance sampling gives more weight to regions where $J(x) < g(x)$ and downweights where $J(x) > g(x)$ to correctly calculate $p(y)$ given a sample from $J(x)$. Geweke (1989) has shown that if the support of $J(x)$ includes the support of $g(x)$, the N i.i.d. sample from $J(x)$, and $p(y)$ exists and is finite, then

$$\tilde{p}(y) \xrightarrow{a.s} p(y) \tag{2.25}$$

the first condition is sensible, for if the support of $J(x)$ is strictly contained in the support of $g(x)$, then there would be no hope of generating deviates in

the complement of support of $J(x)$. The rate of convergence depends on how closely $J(x)$ mimics $g(x)$. As in Geweke (1989), it is important that the tails of $J(x)$ do not decay faster than the tail of $g(x)$.

2.6.2 Under the Independent Model

Using the Rejection Sampling Method

It is easy to see that the marginal posterior distributions of the a_is are as follows:

$$p(a_i|data) \propto a_i^{d_i+u_a-1}(\prod_{j=1}^{d_i} t_{ij})^{a_i-1} e^{v_a a_i}/(v_\theta + \sum_{j=1}^{n_i} {t_{ij}}_i^a)^{d_i+u_\theta} \tag{2.26}$$

and the conditional distributions of the θ_is given a_is are:

$$\begin{aligned} p(\theta_i|a_i, data) &\propto \theta_i^{d_i+u_\theta-1} exp(-\theta_i(v_\theta + \sum_{j=1}^{n_i} {t_{ij}}^{a_i})) \\ &\propto G(u_\theta + d_i, v_\theta + \sum_{j=1}^{n_i} t_{ij}^{a_i}) \end{aligned} \tag{2.27}$$

Thus we can calculate the predictive measures, for example $\mathrm{D}(t|data)$, through following steps:

1. Draw sample ${a_i}^{(k)}, i = 1, 2$ from $p(a_i|data)$ using rejection sampling method.

2. Draw sample ${\theta_i}^{(k)}, i = 1, 2$ from $p(\theta_i|a_i^{(k)}, data)$, which is a gamma distribution.

Thus where $k = 1, 2, \cdots, N$ and N is the Monte-Carlo sample size.

$$\mathrm{D}(t|data) = \iiiint I(e^{-\theta_1 t^{a_1}} < e^{-\theta_2 t^{a_2}}) p(\theta_1, a_1, \theta_2, a_2|data) d_{\theta_1} d_{a_1} d_{\theta_2} d_{a_2}$$

$$= \frac{1}{N}\sum_{k=1}^{N} I(e^{-\theta_1 t^{a_1^{(k)}}} < e^{-\theta_2 t^{a_2^{(k)}}})$$

$$= \frac{1}{N}\sum_{k=1}^{N} I(\theta_1 t^{a_1^{(k)}} > \theta_2 t^{a_2^{(k)}}) \tag{2.28}$$

The marginal posterior distributions of a_is can be easily plotted using S-plus software, and the marginal posterior distributions of θ_is can be calculated using Monte-Carlo method also.

Using the Importance Sampling Method

If we have obtained the posterior mean m and covariance matrix Σ for the parameters $(\theta_1, a_1, \theta_2, a_2)$ via XlispStat using Laplace approximation, then we may choose the multivariate T distribution with the same mean and covariance matrix and a degree of freedom equal to 3 as the Importance Sampling function that we will use to compute the predictive characteristics. For example, we use the following steps to calculate $\mathrm{D}(t|data)$:

$$\begin{aligned}
\mathrm{D}(t|data) &= \int\int\int\int I(e^{-\theta_1 t^{a_1}} < e^{-\theta_2 t^{a_2}}) p(\theta_1, a_1, \theta_2, a_2|data) d_{\theta_1} d_{a_1} d_{\theta_2} d_{a_2} \\
&= \int\int\int\int \frac{I(e^{-\theta_1 t^{a_1}} < e^{-\theta_2 t^{a_2}}) p(\theta_1, a_1, \theta_2, a_2|data)}{T_3(\theta_1, a_1, \theta_2, a_2|m, \Sigma)} \\
&\quad T_3(\theta_1, a_1, \theta_2, a_2|m, \Sigma) d_{\theta_1} d_{a_1} d_{\theta_2} d_{a_2} \\
&= E_{T_3}[\frac{I(e^{-\theta_1 t^{a_1}} < e^{-\theta_2 t^{a_2}}) p(\theta_1, a_1, \theta_2, a_2|data)}{T_3(\theta_1, a_1, a_2|m, \Sigma)}] \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{I(e^{-\theta_1^{(i)} t^{a_1^{(i)}}} < e^{-\theta_2^{(i)} t^{a_2^{(i)}}}) p(\theta_1^{(i)}, a_1^{(i)}, \theta_2^{(i)}, a_2^{(i)}|data)}{T_3(\theta_1^{(i)}, a_1^{(i)}, \theta_2^{(i)}, a_2^{(i)}|m, \Sigma)}
\end{aligned} \tag{2.29}$$

where N is the Monte Carlo sample size, and $(\theta_1^{(i)}, a_1^{(i)}, \theta_2^{(i)}, a_2^{(i)})$ are the i-th sample from the multivariate $T_3(\theta_1, a_1, \theta_2, a_2|m, \Sigma)$ distribution. $I(.)$ is an

indicator function, which means:

$$I(x) = \begin{cases} 1 & \text{if } x \text{ is true.} \\ 0 & \text{if } x \text{ is false.} \end{cases}$$

$p(.|data)$ is the posterior distribution.

2.6.3 Under the Dependent Model

Under this model, suppose we have got the posterior mean m and covariance matrix Σ for the parameters (θ_1, v, a_1, a_2) via XlispStat using the Laplace approximation. Then we may choose the multivariate T distribution with the same mean and covariance matrix and a degree of freedom equal to 3 as the Importance Sampling function that we will use to compute the predictive characteristics. For example, we use the following steps to calculate $\mathrm{D}(t|data)$:

$$\begin{aligned}
\mathrm{D}(t|data) &= \int\int\int\int I(e^{-\theta_1 t^{a_1}} < e^{-\theta_1 v t^{a_2}}) p(\theta_1, v, a_1, a_2|data) d_{\theta_1} d_v d_{a_1} d_{a_2} \\
&= \int\int\int\int \frac{I(e^{-\theta_1 t^{a_1}} < e^{-\theta_1 v t^{a_2}}) p(\theta_1, v, a_1, a_2|data)}{T_3(\theta_1, v, a_1, a_2|m, \Sigma)} \\
&\quad T_3(\theta_1, v, a_1, a_2|m, \Sigma) d_{\theta_1} d_v d_{a_1} d_{a_2} \\
&= E_{T_3}[\frac{I(e^{-\theta_1 t^{a_1}} < e^{-\theta_1 v t^{a_2}}) p(\theta_1, v, a_1, a_2|data)}{T_3(\theta_1, v, a_1, a_2|m, \Sigma)}] \\
&= \frac{1}{N} \sum_{i=1}^{N} \frac{I(e^{-\theta_1^{(i)} t^{a_1^{(i)}}} < e^{-\theta_1^{(i)} v^{(i)} t^{a_2^{(i)}}}) p(\theta_1^{(i)}, v^{(i)}, a_1^{(i)}, a_2^{(i)}|data)}{T_3(\theta_1^{(i)}, v^{(i)}, a_1^{(i)}, a_2^{(i)}|m, \Sigma)} \qquad (2.30)
\end{aligned}$$

where N is the Monte Carlo sample size, and $(\theta_1^{(i)}, v^{(i)}, a_1^{(i)}, a_2^{(i)})$ are the i-th sample from the multivariate $T_3(\theta_1, v, a_1, a_2|m, \Sigma)$ distribution and $I(.)$ is as defined as in last subsection.

Patient number	Treatment	Entry day	Withdrawal (or death) day	Days in study	Death?
1	A	0	8	8	yes
2	B	2	92	90	no
3	B	10	100	90	no
4	A	8	98	90	no
5	A	30	37	7	no
6	A	45	60	15	yes
7	B	50	100	50	yes
8	A	60	none	45	no
9	A	79	80	1	yes
10	B	75	none	45	no
11	B	100	none	20	no
12	A	110	none	10	no

Table 2.1: Interim data for 12 patients; current day is day 120

2.7 An Example Using the Independent Weibull Model

In this section, I will illustrate an example from Berry (1989) using the Weibull model with independent parameters. Table 2.1 shows an interim data set for 12 patients; the current day is day 120.

The total exposure time of Treatment A was $T_A = 191$ days, and there were 3 deaths among the 7 patients. The total exposure time of Treatment B was $T_B = 295$ days, and there was 1 death among the 5 patients.

We assume a Weibull model with independent parameters and let $(\theta_1, a_1, \theta_2, a_2)$ denote the parameter set in which 1 represents Treatment A and 2 represents Treatment B.

According to some training data set, we will use $G(\theta_1|1.0157, 1.0)$ as the

prior for θ_1, and $G(\theta_2|1.0034, 1.0)$ for θ_2, which will give the modes of the gammas' equal to the estimated hazard rates of the training data set. Later on, we will extend this prior to extreme case (uniform on $(0, \infty)$) to explore the sensitivity of the prior.

The first prior we used is:

$$\begin{aligned} \theta_1 &\sim G(\theta_1|1.0157, 1.0) \\ \theta_2 &\sim G(\theta_2|1.0034, 1.0) \\ a_i &\sim U(0, \infty), \quad i = 1, 2. \end{aligned} \tag{2.31}$$

which assumes no prior knowledge about a_1 and a_2.

The second prior we used is a completely improper prior that assumes no prior knowledge about either the $\theta_i's$ or the $a_i's$.

$$\begin{aligned} \theta_i &\sim U(0, \infty) \\ a_i &\sim U(0, \infty), \quad i = 1, 2. \end{aligned} \tag{2.32}$$

The third prior we used is:

$$\begin{aligned} \theta_i &\sim U(0, \infty) \\ a_i &\sim G(a_i|11, 10) \quad i = 1, 2 \end{aligned} \tag{2.33}$$

which assumes that the modes of the $a_i's$ are equal to 1 and assumes no prior knowledge about the $\theta_i's$.

The fourth prior we used is:

$$\begin{aligned} \theta_1 &\sim G(\theta_1|1.0157, 1.0) \\ \theta_2 &\sim G(\theta_2|1.0034, 1.0) \\ a_i &\sim G(a_i|11, 10) \quad i = 1, 2 \end{aligned} \tag{2.34}$$

which assumes that the modes of the $a_i's$ are equal to 1.

I performed all the calculations by using rejection sampling and Monte-Carlo method. The results I am showing below used Monte-Carlo sample size of 2000, though I did vary the Monte-Carlo sample size from $1000, 2000, 3000$ to 4000. The results did not change much essentially. The marginal posterior distributions of a_is are from S-plus directly.

Figure 2.1 shows the posterior distributions of a_1 and a_2 for four prior choices detailed above. This figure demonstrates that under any prior, most of the density for a_1 falls between 0 and 1, which suggests a decreasing failure rate for the Treatment A. For a_2, under the improper prior, the posterior for a_2 suggests a decreasing failure rate for the Treatment B, whereas under informative prior $G(a_i|11, 10), i = 1, 2$, the posterior for a_2 suggests that it may be also reasonable to assume a constant failure rate.

Figure 2.2 displays the posterior distributions of θ_1 and θ_2 for each of the four prior choices.

Figure 2.3 shows the predictive measures $\mathrm{D}(t|data)$ and $\mathrm{d}(t|data)$ which were introduced in Section 2.5. Under the improper prior for the $a_i's$, the predictive probability that the survival on Treatment A is smaller than the survival on Treatment B $\mathrm{D}(t|data)$ decreases as a function of time t, from about 0.975 to 0.75; whereas under the informative prior $G(a_i|11, 10), i = 1, 2$, $\mathrm{D}(t|data)$ decreases slowly as a function of time t, from about 0.91 to 0.88.

Comparing the result from Berry (1989), using an exponential model and an informative prior for hazard rates $\lambda_i, i = 1, 2$ on both treatment arms, the posterior probability $Pr(\lambda_A > \lambda_B|data) = 0.91$, which is equivalent to the $\mathrm{D}(t|data)$ in our Weibull model.

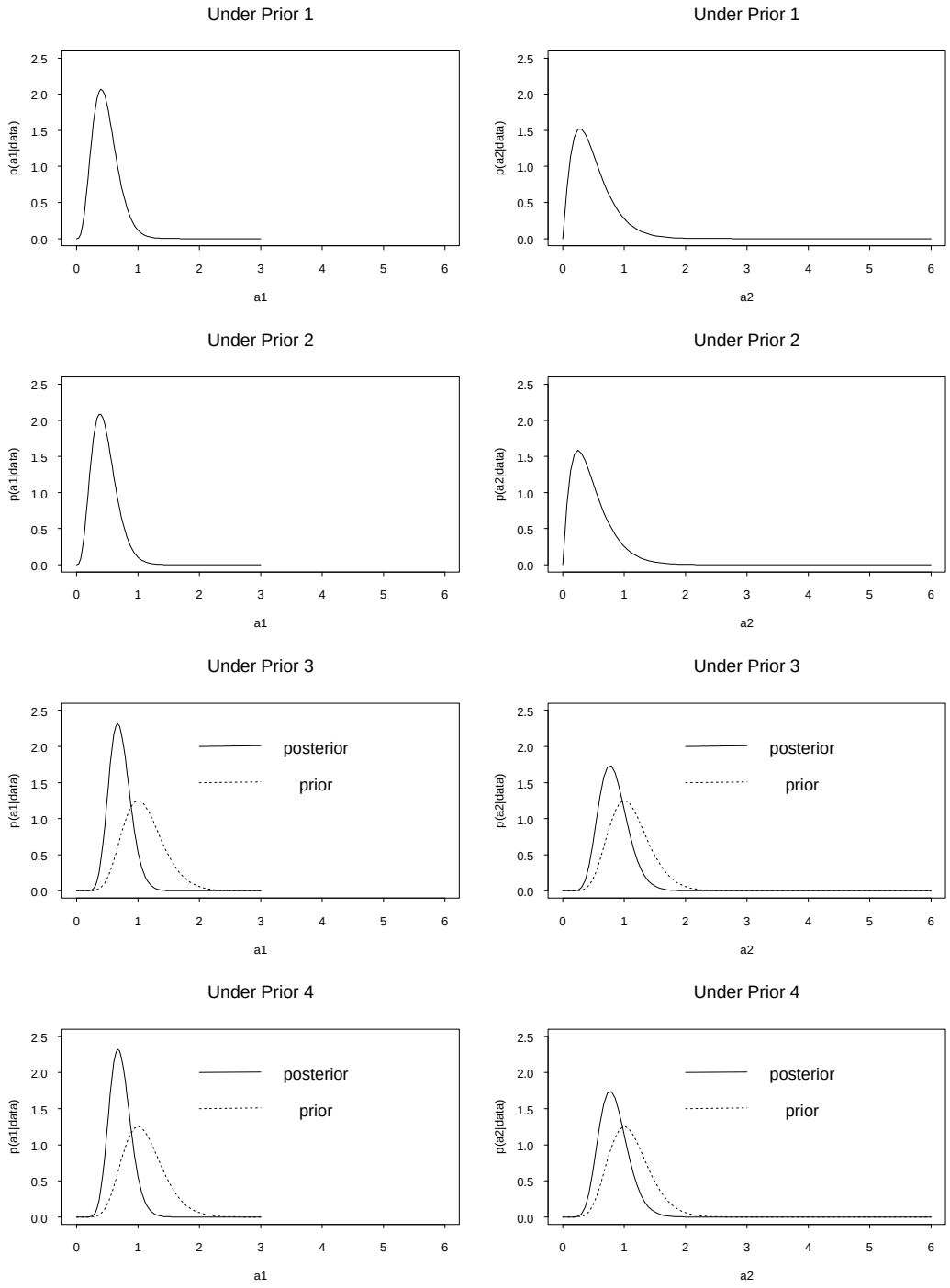

Figure 2.1: Posterior distributions of a_1 and a_2 under four different priors.

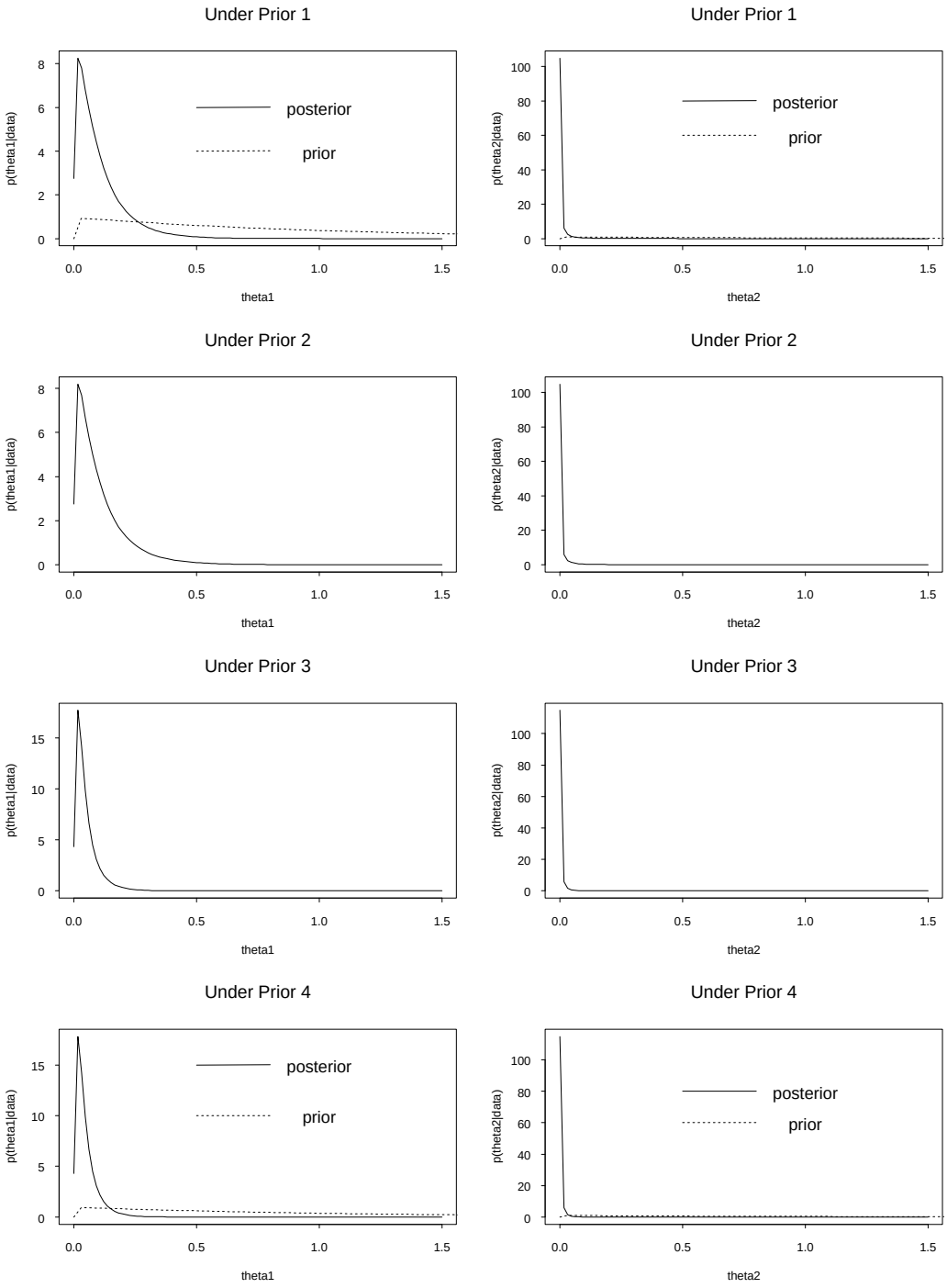

Figure 2.2: Posterior distributions of θ_1 and θ_2 under four different priors.

On the other hand, I found that under the improper prior for the $a_i's$, the mean survival difference between Treatment A and Treatment B $\mathrm{d}(t|data)$ decreased as a function of time t, reached its minimum at about $t = 46$ months, and then increased as a function of time t. In contrast, under the informative prior $G(a_i|11, 10), i = 1, 2$, $\mathrm{d}(t|data)$ decreased at first, but then began to stablize around -0.33. This is because the data will influence our prediction a lot more if an improper prior is used in stead of an informative one. Under the informative priors, the predictions tell us that any future patient will be better off on Treatment B than on A, no matter how many months he expects to survive. While under improper prior, he will not that much better off if he expects to survive till 80 months other than 46 months. In other words, our prediction favoring Treatment B will decrease as the prediction interval increases.

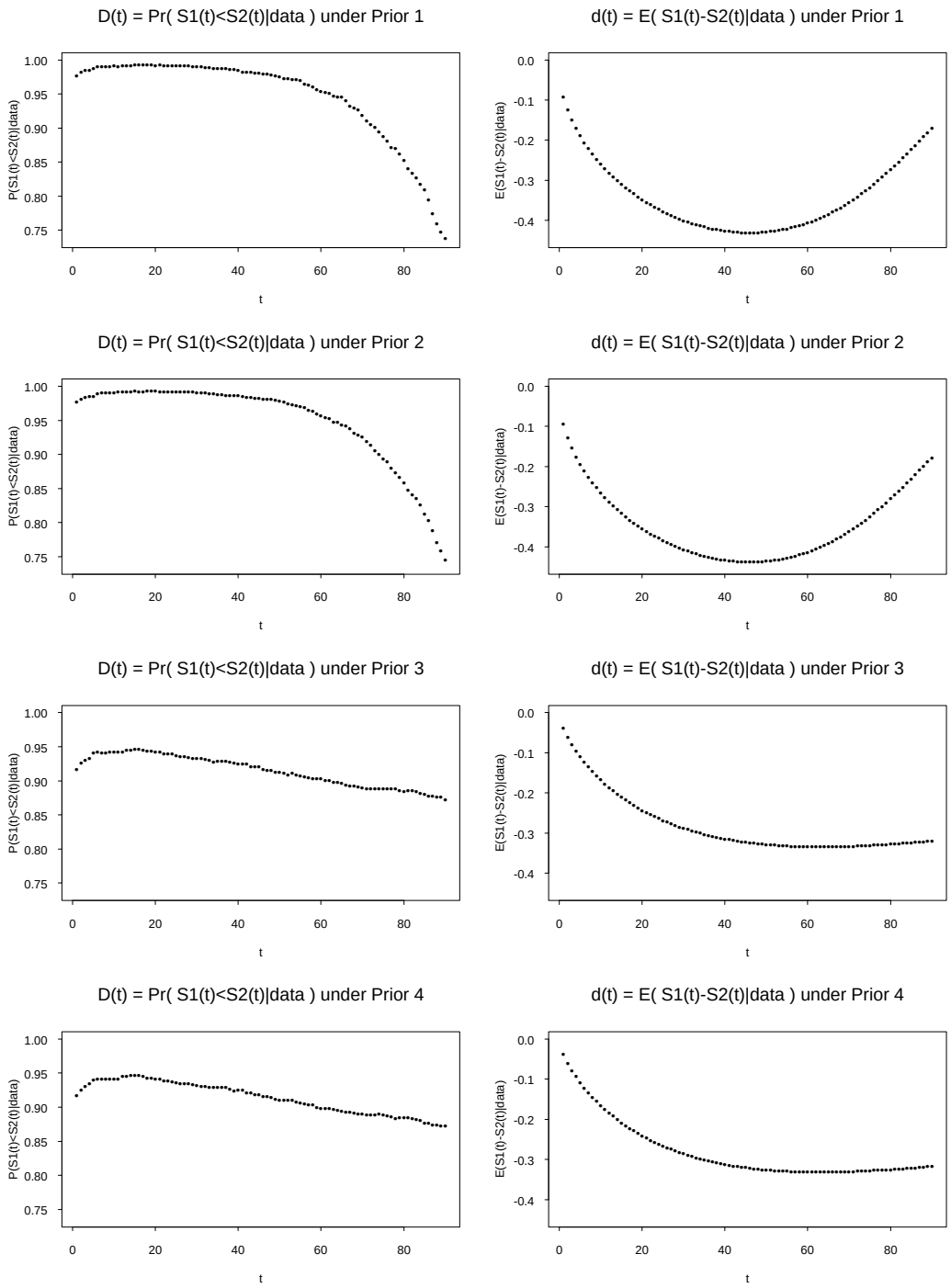

Figure 2.3: Predictive measures D($t|data$) and d($t|data$) under four prior choices.

Chapter 3

Application of the Weibull Model to the NSCLC Trial

3.1 Introduction

3.1.1 Purpose of the Study

Carcinoma of the lung is the leading cause of cancer deaths in the United States. For treatment purposes, lung cancer is generally divided into two categories, based on the cell type presenting at diagnosis. The first type, small cell (or oat cell) anaplastic carcinoma, is strongly associated with smoking and accounts for roughly 30% of all lung cancers. Most other cell types are classified as non-small cell lung cancer (NSCLC), which includes adenocarcinoma, squamous cell carcinoma, and large cell anaplastic carcinoma.

All cancers are usually classified further according to the extent or stage of disease so that therapies may be tailored to the particular disease stage. Patients with Stage III non-small cell lung cancer are those who have no demonstrable distant metastases but do have locally extensive or invasive disease or involvement of mediastinal lymph nodes. The cancer is generally consid-

ered incurable by surgery because of the association of these findings with micrometastatic disease.

For many years, radiation therapy (RT) has been the treatment of choice for such patients, although the practice has been questioned. Thoracic radiation to the primary tumor and the hilar and mediastinal lymph nodes was considered "standard" therapy based on the ability to control local symptoms with this treatment. However, the overall impact of radiation alone on survival in stage III NSCLC patients is minimal at best, producing median survivals in the range of 8–10 months with a 2-year survival of 10 to 20 percent and a three-year survival of 5 to 10 percent (Perez, Pajak, and Rubin, 1987). In attempts to improve survival in these patients, clinical researchers in the early 1980s considered the possibility that RT alone might not be sufficient to eradicate micro metastatic disease. At that time there was also some evidence that platinum-based chemotherapy (CT) was beneficial in terms of survival in more advanced disease patients. Therefore, various systematic approaches to treatment of stage III patients, which included CT in conjunction with standard RT, were proposed. This study was designed to compare the standard treatment (RT only), consisting of RT delivered over 6 weeks to the original tumor volume and involved regional lymph nodes, to an experimental treatment(CT+RT), which employed 5 weeks of cisplatin plus vinblastine prior to the RT. The RT on the experimental arm was identical in dose, schedule, and volume to that for the standard treatment.

3.1.2 Design of the Study

The study we describe here was a phase III clinical trial for patients with stage III NSCLC. It was opened in 1984 by the Cancer and Leukemia Group

B (CALGB) to compare standard treatment (RT alone) to an experimental treatment regimen of two courses of combination chemotherapy (CT) given prior to the standard radiotherapy as treatment for Stage III NSCLC. The primary objective of this study was to compare survival on the experimental treatment (CT+RT) to that on the standard treatment (RT only).

The original fixed sample size for this trial was 240 patients (120 patients per treatment arm). This sample size was calculated to provide 80% power to detect a hazard ratio of 1.5:1 assuming that the logrank test would be used at a two-sided significance level of $\alpha = 0.05$. This hazard ratio represents a 50% difference in median survival between the two treatment groups. Shortly after the trial began, it was decided to apply group sequential concepts by using a truncated O'Brien-Fleming (1979) stopping rule, implemented via a Lan-Demets (1983) α–spending function. A data monitoring committee (DMC) was established to review the analyses as they were produced.

The patient population for this study was limited to patients with documented regional stage III NSCLC. Patient eligibility criteria included no prior CT, RT, or total resection (which means the tumor was completely removed in the surgical procedure), performance status (PS) of 0 or 1 (PS is given as an integer from 0 to 4, 0 = fully active, 1 = ambulatory, capable of light work), and weight loss of less than 5% in the 3-month interval prior to study entry. Standard CALGB eligibility criteria regarding laboratory values, other diseases, and so on, were also used. All eligible patients who began treatment and follow–up data were included in the analysis. Eight percent of the patients on this study failed to meet these criteria and were excluded from the analysis.

The study was stopped at the fifth interim analysis in May 1987 after 155

eligible patients had been entered, of which follow–up data were only available for 105 patients.

3.2 Applying the Bayesian Weibull Model to the NSCLC Trial

3.2.1 Introduction

In this section, we will apply the Weibull model with the dependent parameters proposed in Section 2.3 to the NSCLC trial described in Section 3.1.

3.2.2 Preliminary Analysis of the NSCLC Trial

From May 1984 to May 1987, five interim looks were performed and presented to the DMC. Table 3.1 provides specific information about the number of eligible patients accrued and the number of deaths on each treatment arm at each interim look time. It also lists a summary of observed p-values, and boundary significance levels, and it indicates whether the DMC decided to close the trial or to keep it open. Initial accrual was slow, and the first interim analysis was not performed until the fall of 1985. The sample size at this time was too small to allow any meaningful comparisons of survival by treatment.

The O'Brian-Fleming boundary truncated at 3.0 standard deviations (a boundary significance level of 0.0013) was selected as the boundary to be used for all monitoring of the study. I have included the boundary value for the Pocock (1977) boundary to illustrate how the decision process would have differed if other boundary was used.

By the October 1986 analysis, the observed significance for the comparison of survival by treatment had decreased to $p = 0.0015$, which almost touched

Analysis	d_1	n_1	d_2	n_2	Logrank p–value	Truncated O'Brien–Fleming	Pocock	Decision
Sept. 1985	7	25	3	25	$--$	0.0013	0.0041	keep open
Mar. 1986	12	41	4	38	0.021	0.0013	0.0034	keep open
Aug. 1986	20	41	14	47	0.0071	0.0013	0.0078	keep open
Oct. 1986	24	46	18	49	0.0015	0.0013	0.0061	keep open
Mar. 1987	32	51	24	54	0.0015	0.0013	0.0081	close

Table 3.1: Summary of statistics and boundary significance levels used by DMC

the truncated O'Brian-Fleming boundary significance level of 0.0013. Despite the close proximity to the boundary, the DMC again decided that there was insufficient evidence to recommend closure of the study and that the next analysis would be performed as scheduled in March 1987.

The March 1987 interim analysis led to the decision to close the trial to further accrual. At this time, 163 out of the projected 240 patients had been accrued, and follow-up data were available for 105 patients. The observed p-values was 0.0015 with a total of 56 failures. A comprehensive analysis including use of the Cox (1972) proportional hazards model was undertaken. The prognostic factors examined were comparable in the two treatment arms. The results of the Cox analysis, which controlled for various prognostic factors known for survival in NSCLC, gave $p = 0.0008$ for the treatment comparison. This analysis reaffirmed that the observed difference represented a real treatment effect. This adjusted p-value acrossed the truncated O'Brian-Fleming boundary, and the DMC unanimously voted to close the study. After the trial stopped accruing new patients, the enrolled patients were followed up until June, 1992.

Figure 3.1 shows the Kaplan-Meier plots for each of the interim analyses

and as of June, 1992. It is easy to see that at the beginning of the trial, there appear to be huge differences between patients receiving RT only and those undergoing CT+RT. But as time goes on, the difference began to smooth out, although patients on CT+RT still had a better survival.

A standard assumption in clinical trials is that survival time is exponential. But I will assume it is Weibull here since this assumption is more general. One way to check whether lifetimes are exponential is by plotting $-log(\widehat{S(t)})$ vs. t, where $\widehat{S(t)}$ is the product limit estimate of the survival function $S(t)$ and t is lifetime (days, months, etc.). If the plot is far from a straight line, one may question the assumption of an exponential lifetime. I plot $log(-\widehat{S(t)})$ against time t in Figure 3.2 for each interim look time and as of 1992. It is easy to see that at second and fourth interim look, an exponential lifetime assumption seems to be valid for the control arm (RT only), but the exponential assumption is not true for the experimental arm (CT+RT) at any of the interim look time. If the lifetimes have the Weibull distribution $W(t|\theta, a)$, then

$$S(t) = exp(-\theta t^a) \tag{3.1}$$

Thus, the plot of $log(-log(\widehat{S(t)}))$ vs. $log(t)$ will be roughly a straight line with a slope equal to a. We then plot $log(-log(\widehat{S(t)}))$ against $log(t)$ in Figure 3.3 for each interim look time and as of current (1992). These plots suggest that the Weibull assumption is an improvement over the exponential assumption.

3.2.3 The Prior for (θ_1, v, a_1, a_2)

For the Weibull distribution, the shape parameter $a < 1$ corresponds to a decreasing failure rate, and the shape parameter $a > 1$ corresponds to an increasing failure rate. A gamma prior was chosen for both a_1 and a_2 with

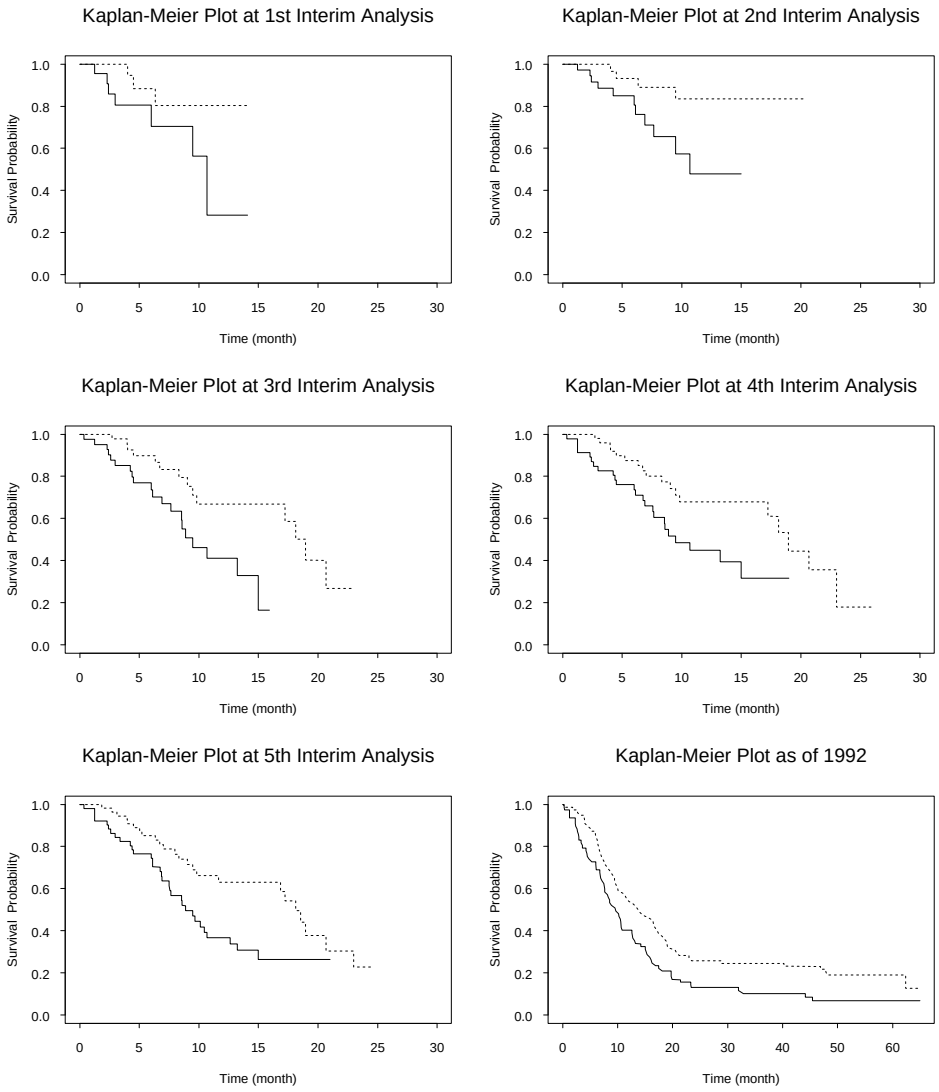

Figure 3.1: Kaplan-Meier Plots..... for RT only—— for CT+RT.

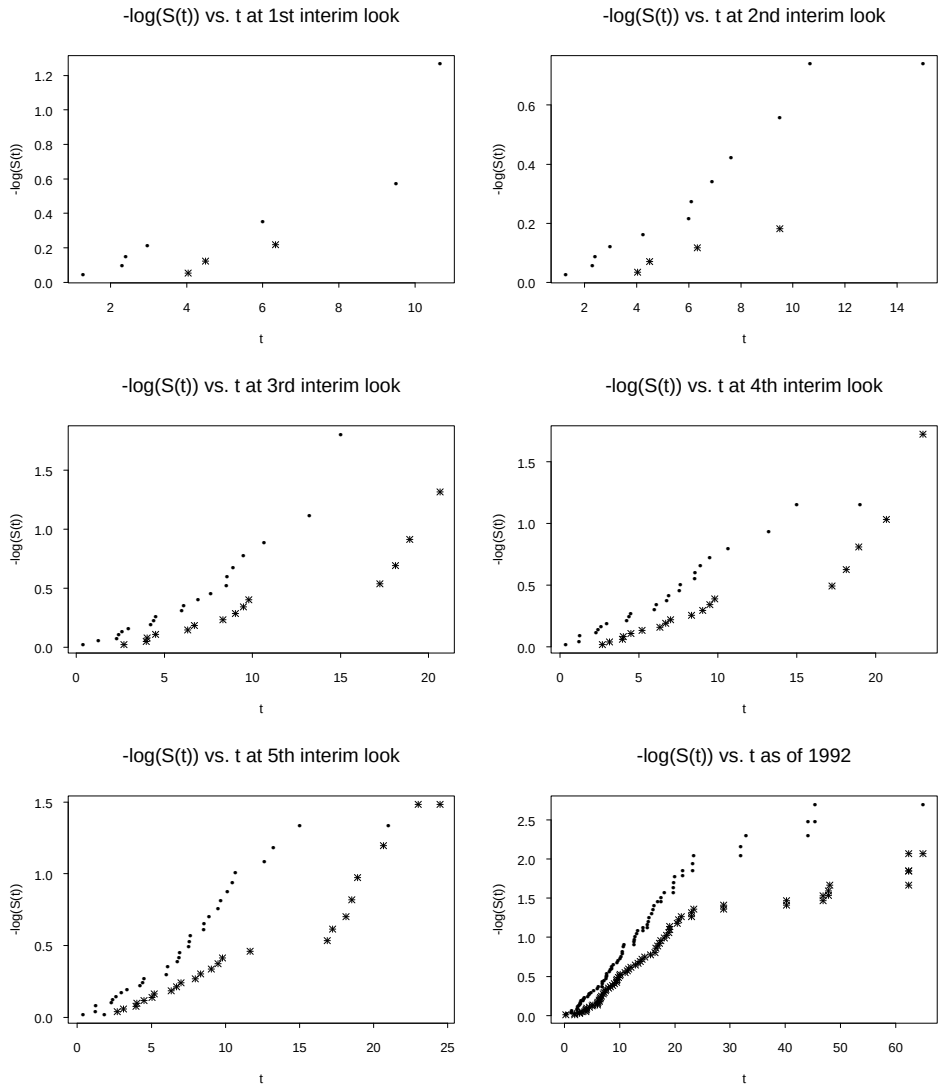

Figure 3.2: Log($\widehat{S(t)}$) vs. Time t (months)

..... for RT only
***** for CT+RT.

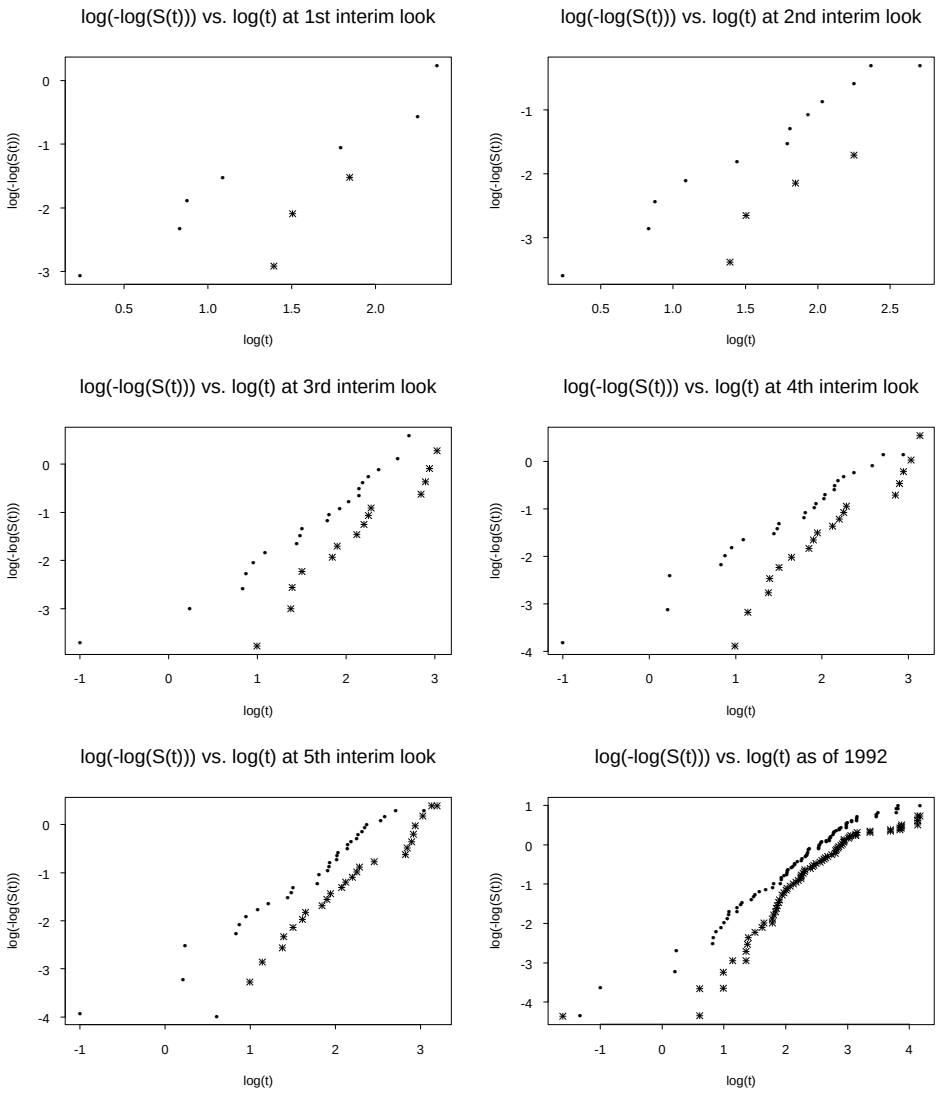

Figure 3.3: log(-log($\widehat{S(t)}$) vs. log(t)..... for RT only***** for CT+RT.

the mode equal to 1. If a_1 and a_2 are equal to 1, then the Weibull model is just an exponential model. A review of the pre-1984 literature about NSCLC shows that the median survival time for NSCLC patients is 8–10 months, with a 2-year survival of 10–20% and a 3-year survival of 5–10%. A gamma prior distribution with a shape parameter of 2 and a scale parameter of 20 approximates this information, and estimates a median survival time of 8.3 months. In an exponential model, v represents the log-hazard ratio of Treatment 2 to 1. A normal distribution with mean 0 and variance 1, was chosen for v to represent our prior belief about v. For this prior, we consider both treatments to be equally likely to be the most effective and believe that there are only small differences between the treatments.

3.2.4 The Posterior

Given the prior $\pi(\theta_1, v, a_1, a_2)$ on (θ_1, v, a_1, a_2) and data $\{t_{ij}; j = 1, 2, ..., n_i, \ \ i = 1, 2\}$ at the time of our analysis, we can obtain the posterior $p(\theta_1, v, a_1, a_2|data)$ as follows:

$$\begin{aligned} p(\theta_1, v, a_1, a_2|data) &\propto L(\theta_1, v, a_1, a_2|data) \times \pi(\theta_1, v, a_1, a_2) \\ &\propto (\theta_1 a_1)^{d_1} (\prod_{j=1}^{d_1} t_{1j})^{a_1 - 1} exp(-\theta_1 \sum_{j=1}^{n_1} {t_{1j}}^{a_1}) \\ &\quad (\theta_1 e^v a_2)^{d_2} (\prod_{j=1}^{d_2} t_{2j})^{a_2 - 1} exp(-\theta_1 e^v \sum_{j=1}^{n_2} {t_{2j}}^{a_2}) \\ &\quad \times \theta_1^{u_\theta - 1} e^{-v_\theta \theta_1} \times e^{\frac{-v^2}{2}} \times a_1^{u_a - 1} e^{-v_a a_1} \times a_2^{u_a - 1} e^{-v_a a_2} \end{aligned} \quad (3.2)$$

3.3 Comparing Treatments

There are many ways to compare the effects of the treatments; Four of which are listed in previous chapter. They are: D$(t|data)$, which is defined as $Pr(S_1(t) < S_2(t)|data)$; d$(t|data)$, which is defined as $E(S_1(t) - S_2(t)|data)$; $Pr(m_1 < m_2|data)$; and $Pr(T_1 < T_2|data)$.

3.3.1 Results of the Interim, Final, and Current Analysis

As we discussed above, a prior distribution for (θ_1, v) is chosen as $G(\theta_1|2, 20) \times N(v|0, 1)$.

The first prior chosen for the $a_i's$ is $G(a_i|10001, 10000)$. This prior distribution has small variance of $10001/10000^2 \approx 10^{-4}$(std $\approx 10^{-2}$). So under this prior distribution, the Weibull model corresponds closely to the simple exponential model (George *et al.*, 1993). By choosing this prior, we actually pushed the Weibull model close enough to the exponential model. Thus, it will be interesting to see how the results from this model differ from those of the exponential one.

The results I am showing below is based on a Monte-Carlo sample size of 2000, though I did several reanalysis to check how varying the Monte-Carlo sample size might affect the results. The MC sample size I ever tried are $1000, 2000, 3000, 4000$ and 5000. Through these reanalysis, I found that the results do not change much essentially.

Figure 3.4 shows the prior and posterior distributions of parameters and predictive characteristics D$(t|data)$ and d$(t|data)$ and various predictive probabilities of v. When we compare the posterior distributions of θ_1 and v with

those in George *et al.* (1993), we almost cannot find any difference between them. We do not find much shrinkage when comparing the posterior distribution with the prior for (a_1, a_2), because our prior is so strong (almost a point mass on 1) and allows very little movement with moderate sample size. The plot of $\mathrm{D}(t|data)$ shows us that this is a nearly exponential model. For the exponential model, $\mathrm{D}(t|data)$ should not change with time t and should be constant over time. The second prior chosen for the $a_i's$ is $G(a_i|101, 100)$. This is a gamma prior with mode 1 and standard deviation 0.1, which is a much flatter prior distribution than the $G(a_i|10001, 10000)$. Figure 3.5 shows the prior and the posterior distributions of the parameters and the predictive characteristics of $\mathrm{D}(t|data)$ and $\mathrm{d}(t|data)$ and various predictive probabilities of v. The posterior distributions are a little bit different from those under the first prior. The posterior density of the log-hazard ratio v shrinks over time. However, the posterior density of θ_1 at the second interim look has the smallest variance among all the looks. There is a lot of variation in the data at second interim look.

At the second interim look, there are 12 deaths among the patients on RT therapy and only 3 deaths among the patients on CT+RT therapy. This interim look has the biggest contrast in deaths between therapies among all the looks. As of 1992, for both a_1 and a_2, we can see that although 1 is within the 95% posterior HPD region of $a_i, i = 1, 2$, it is not the posterior mode of either a_1 or a_2. Thus, the exponential model does not capture the data very well. Most of the density of $a_i, i = 1, 2$ falls between 0 and 1. When $a_i < 1$, the failure rate decreases in the Weibull distribution. Thus, on both treatment arms, a patient who survives the first few years after therapy has a better prognosis than one who has been treated only recently. The Weibull distribution can capture this

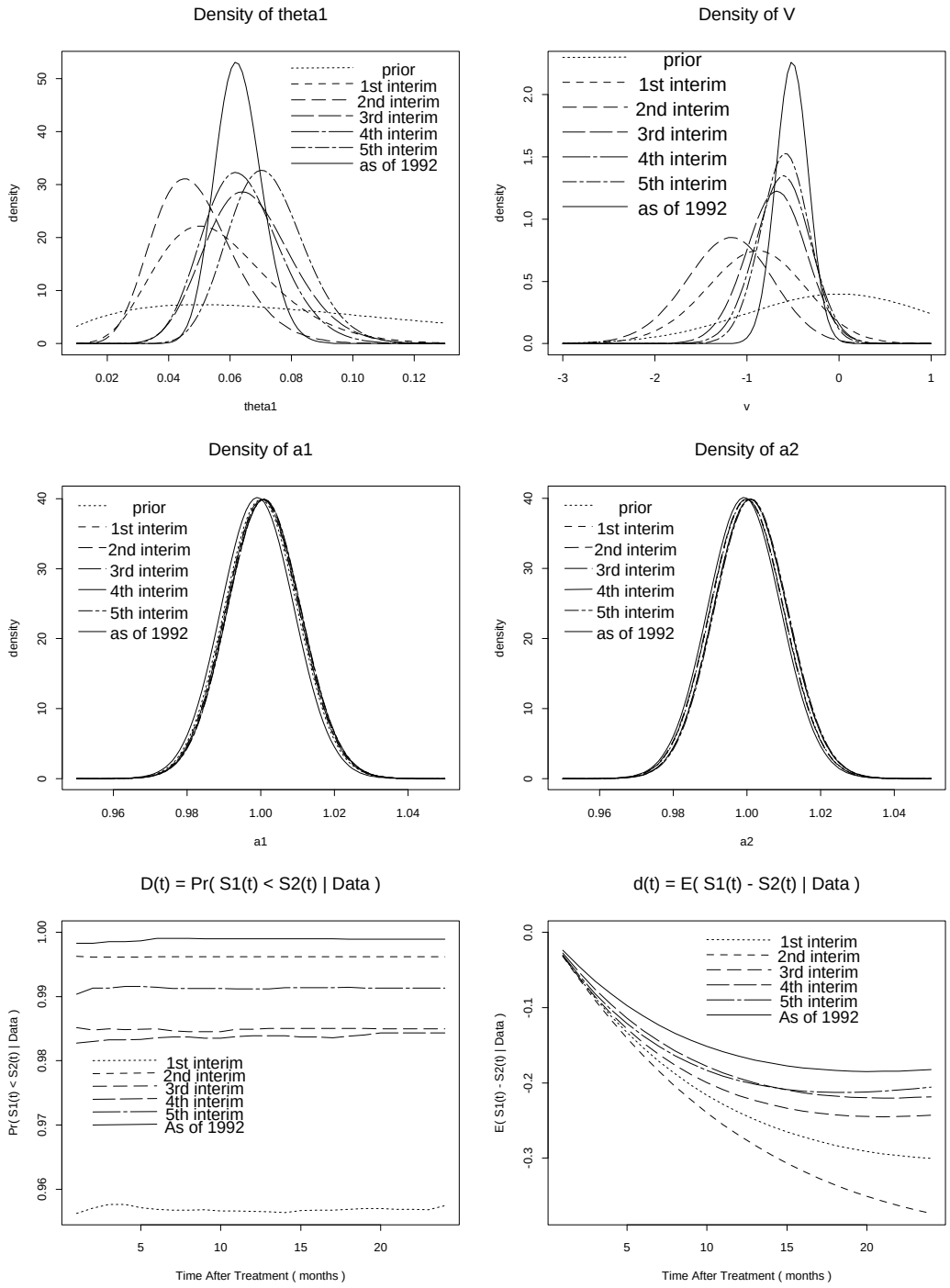

Figure 3.4: Posterior distributions of (θ_1, v, a_1, a_2) and predictive characteristics when the prior is $a_i \sim G(a_i|10001, 10000)$, $\theta \sim G(\theta|2, 20)$, and $v \sim N(v|0, 1)$.

feature.

With the second prior, the $\mathrm{D}(t|data)$ changes with the time, and our Weibull model is no longer close to the exponential one.

3.3.2 Sensitivity Analysis

To see how different priors on (θ_1, v, a_1, a_2) affect my conclusions about this trial, I tried 10 more values for σ. They are 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9.

Again the results I am showing below is based on a Monte-Carlo sample size of 2000.

Figure 3.6 shows how changing σ affects the various posterior probabilities of the log-hazard ratio v. As the information is accumulating (via the interim looks), the probabilities of $v < 0, v < -0.25$, and $v < -0.5$ change less as σ increases. Except for the third interim look, the $p(v < 0)$ is always greater than 0.5 for all the σ's. For σ smaller than 0.3, there is no compelling advantage of CT+RT over RT. But the data from the later interim looks show a convincing advantage for all $\sigma > 0.3$.

I also varied priors for the θ_is and a_is to see how my conclusions about this trial might differ. I extend the priors in last subsection for θ_is and a_is to gammas with larger variance, and further more I extend these priors into improper prior (uniform on $(0, \infty)$). I found that my conclusions about this trial essentially remain the same under different priors I ever tried.

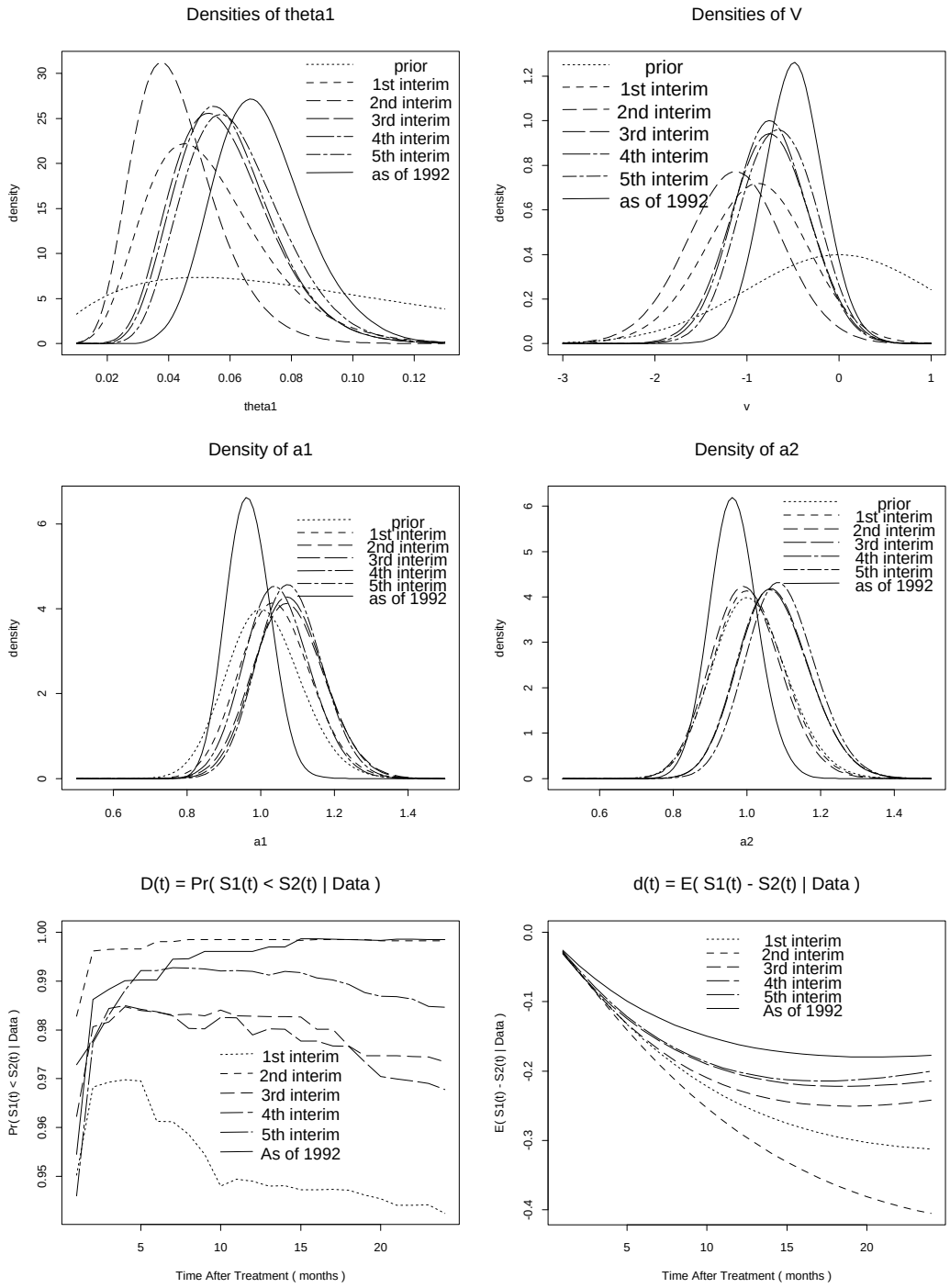

Figure 3.5: Posterior distributions of (θ_1, v, a_1, a_2) and predictive characteristics when the prior is $a_i \sim G(a_i|101, 100)$, $\theta \sim G(\theta|2, 20)$, and $v \sim N(v|0, 1)$.

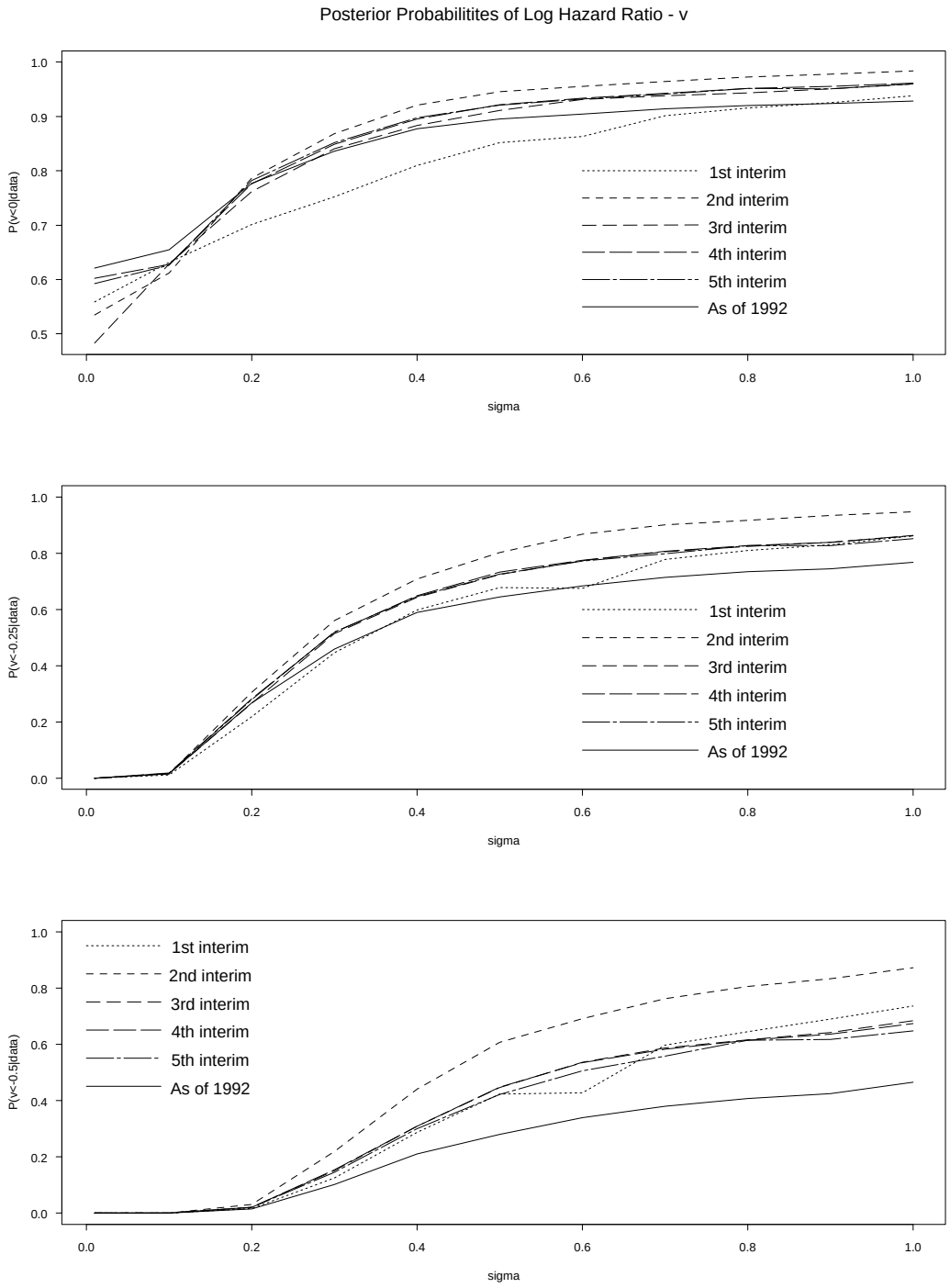

Figure 3.6: Posterior probabilities of v as a function of σ.

Analysis	d_1	n_1	d_2	n_2	$p(v<0)$	$p(v<-0.25)$	$p(v<-0.5)$
1st Interim	7	25	3	25	0.937918	0.862504	0.736147
2nd Interim	12	41	4	38	0.983479	0.946832	0.871660
3rd Interim	20	41	14	47	0.959806	0.862281	0.683985
4th Interim	24	46	18	49	0.961223	0.862627	0.674363
5th Interim	32	51	24	54	0.960187	0.851531	0.648247
As of 1992	71	77	65	78	0.928275	0.767728	0.465101

Table 3.2: Statistics in the interim analysis

3.4 Using Prediction in Stopping Decisions

If the study had stopped at interim analysis 1, with a total of 50 patients admitted, and followed to their deaths, what would have been our conclusion about this trial at the 5-th interim look? We can address this by simulating another 55 patients' lifetimes in order to get a total 105 patients used in the 5th interim analysis. We can ask this very same question at each interim analysis. Table 3.2 shows the accrual numbers and deaths on each treatment at the various interim analysis points.

Figure 3.7 is a histogram of D($t|data$) and d($t|data$) and various posterior probabilities of v based on 500 simulations which assume that the trial was stopped at the first interim analysis. The black triangles are the true values at the fifth interim analysis. Figure 3.8 shows a histogram of D($t|data$) and d($t|data$) and various posterior probabilities of v based on 500 simulations which assume that the trial was stopped at the 2nd interim analysis. Figure 3.9 shows us the histogram of D($t|data$) and d($t|data$) and various posterior probabilities of v based on 500 simulations which assume that the trial was stopped at the 3rd analysis. Figure 3.10 shows a histogram of D($t|data$) and d($t|data$) and various posterior probabilities of v based on 500 simulations

which assume the trial was stopped at the 4th interim analysis. These plots demonstrates that the variation between these simulations gets smaller as the trial continues. That is, if the trial had been stopped at the 1st interim analysis, then there would have been more uncertainties about the final conclusions that we would have drawn at the 5th analysis than if the trial had been stopped at 4th interim analysis.

3.5 What If We Had Not Stopped the Trial in 1987?

There are many disputes about the early stopping of the trial. The original plan was to accrue a total of 240 patients and randomly assign 120 patients to each treatment arm. The trial was stopped in April, 1987 because the DMC thought that the data was already very supportive concerning the CT+RT treatment. Thus, the trial was stopped before the plan was fully executed. The total patients accrued at that time was 155: 77 on RT only treatment and 78 on CT+RT treatment. Thus, another 85 patients were needed to fulfill the plan.

As of June 1992, we could look back and ask, what would have happened if we had not stopped the trial, and to what degree would these extra 85 patients have influenced our conclusion? According to the actual accrual rate from the first interim look time to the end of the trial, we accrued patients after April 1987, at the same rate (about 2 patients every 11 days). We assigned one to Treatment 1, one to Treatment 2. To obtain their lifetime based on all the information we have as of current (June 1992), I used the following procedure:

$$t_{ij} \sim W(t_{ij}|\theta_i, a_i),$$

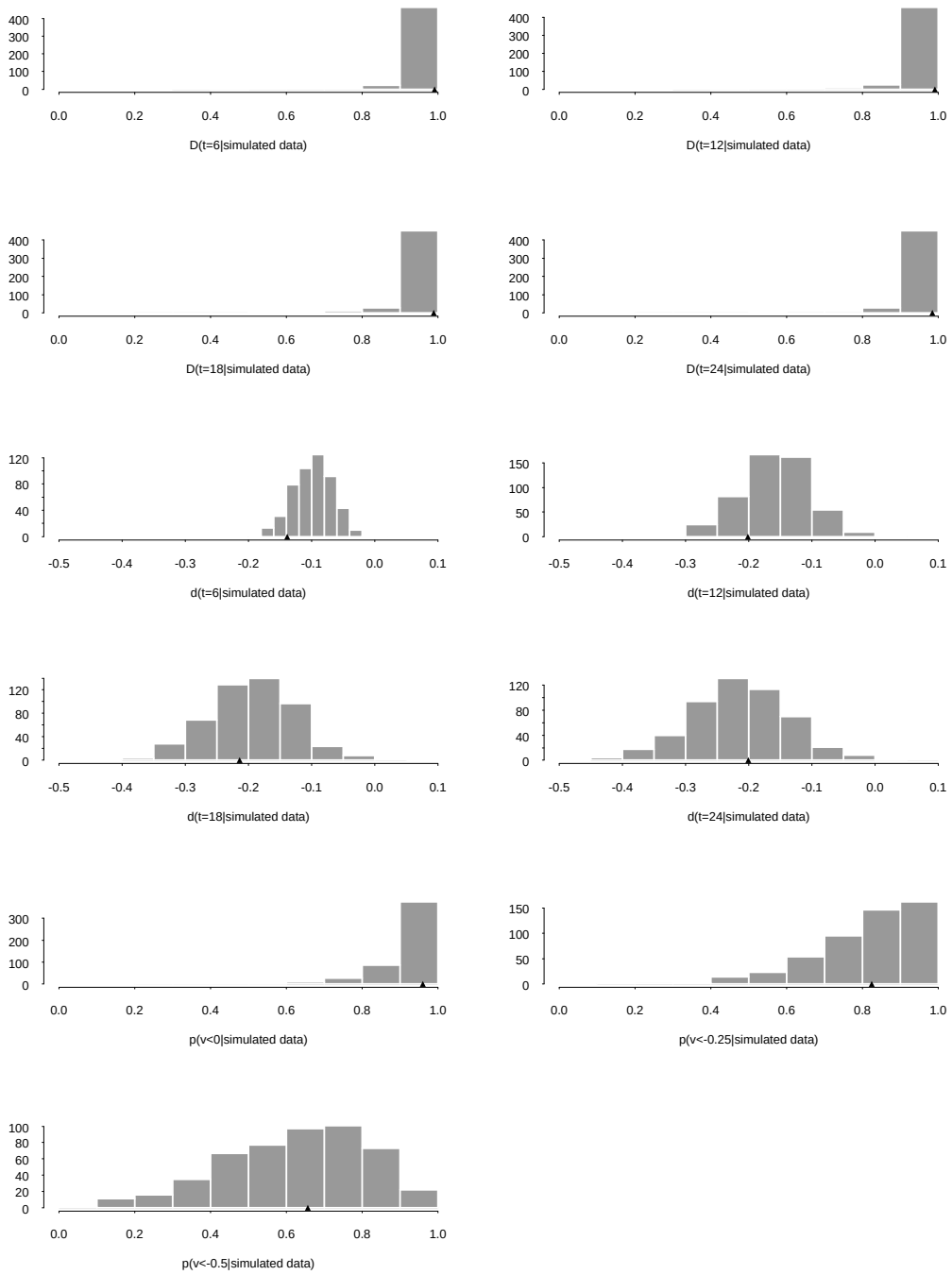

Figure 3.7: Histograms of predictive characteristics based on 500 simulations if the trial had been stopped at 1st interim look.

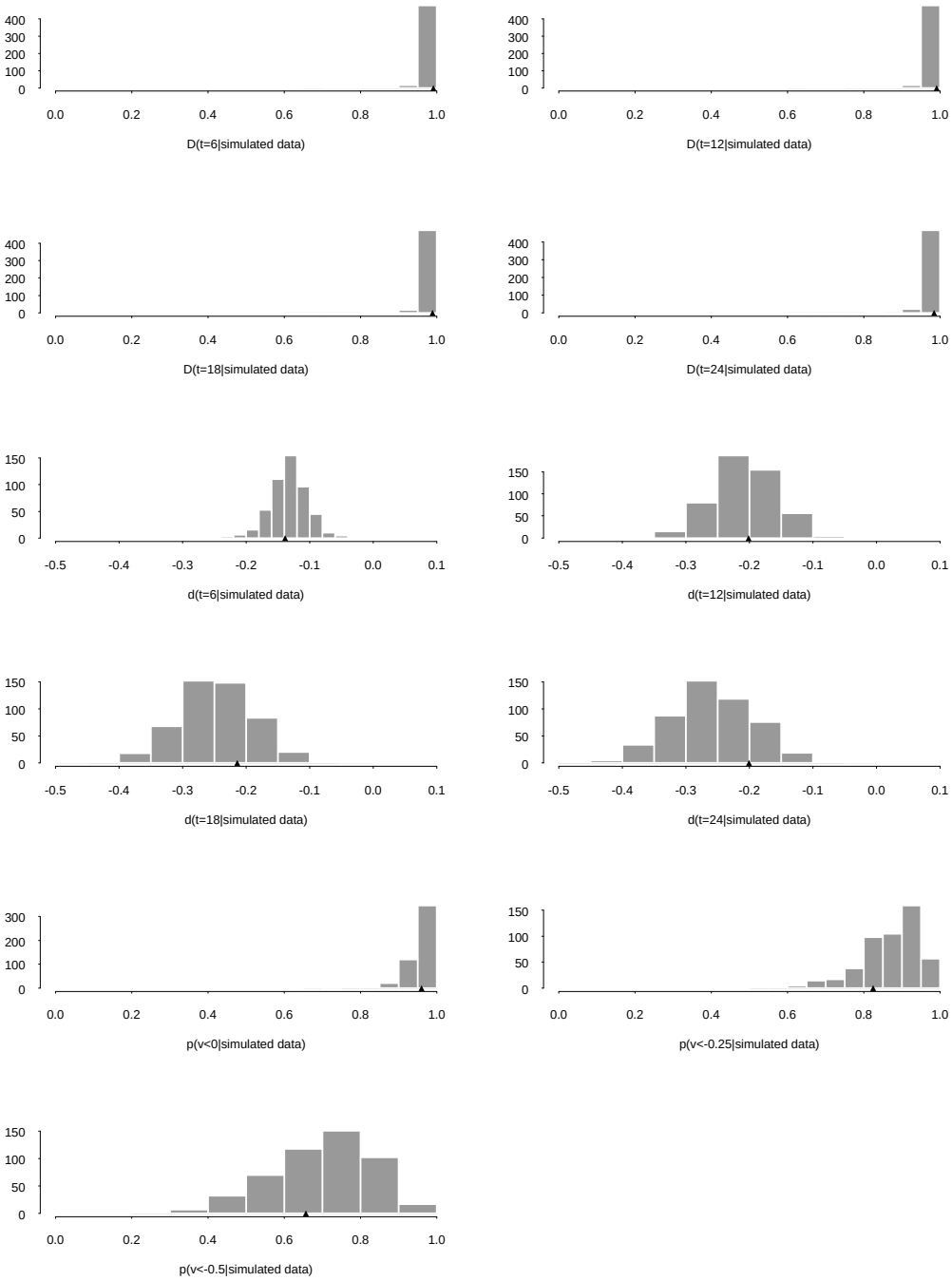

Figure 3.8: Histograms of predictive characteristics based on 500 simulations if the trial had been stopped at 2nd interim look.

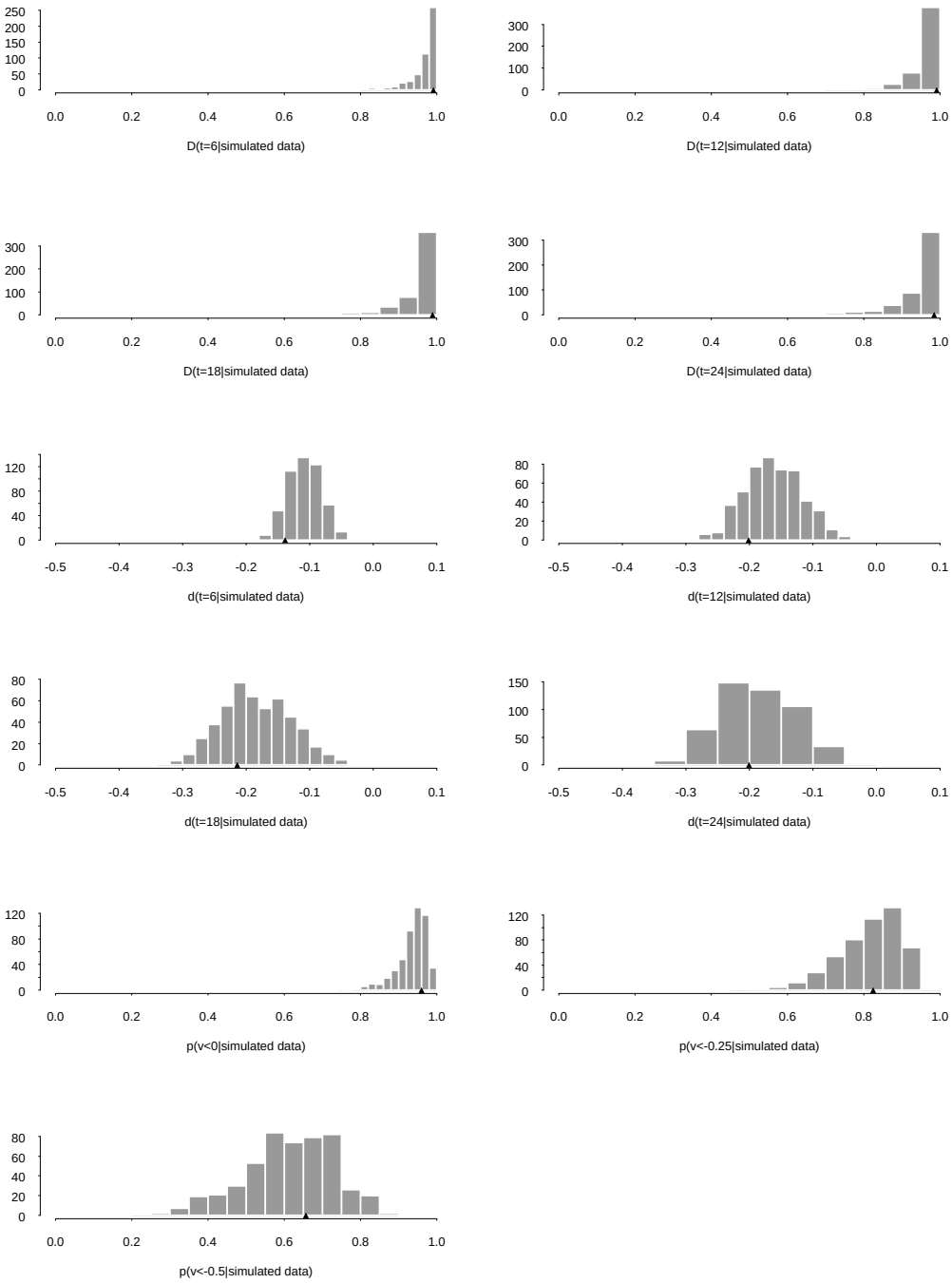

Figure 3.9: Histograms of predictive characteristics based on 500 simulations if the trial had been stopped at 3rd interim look.

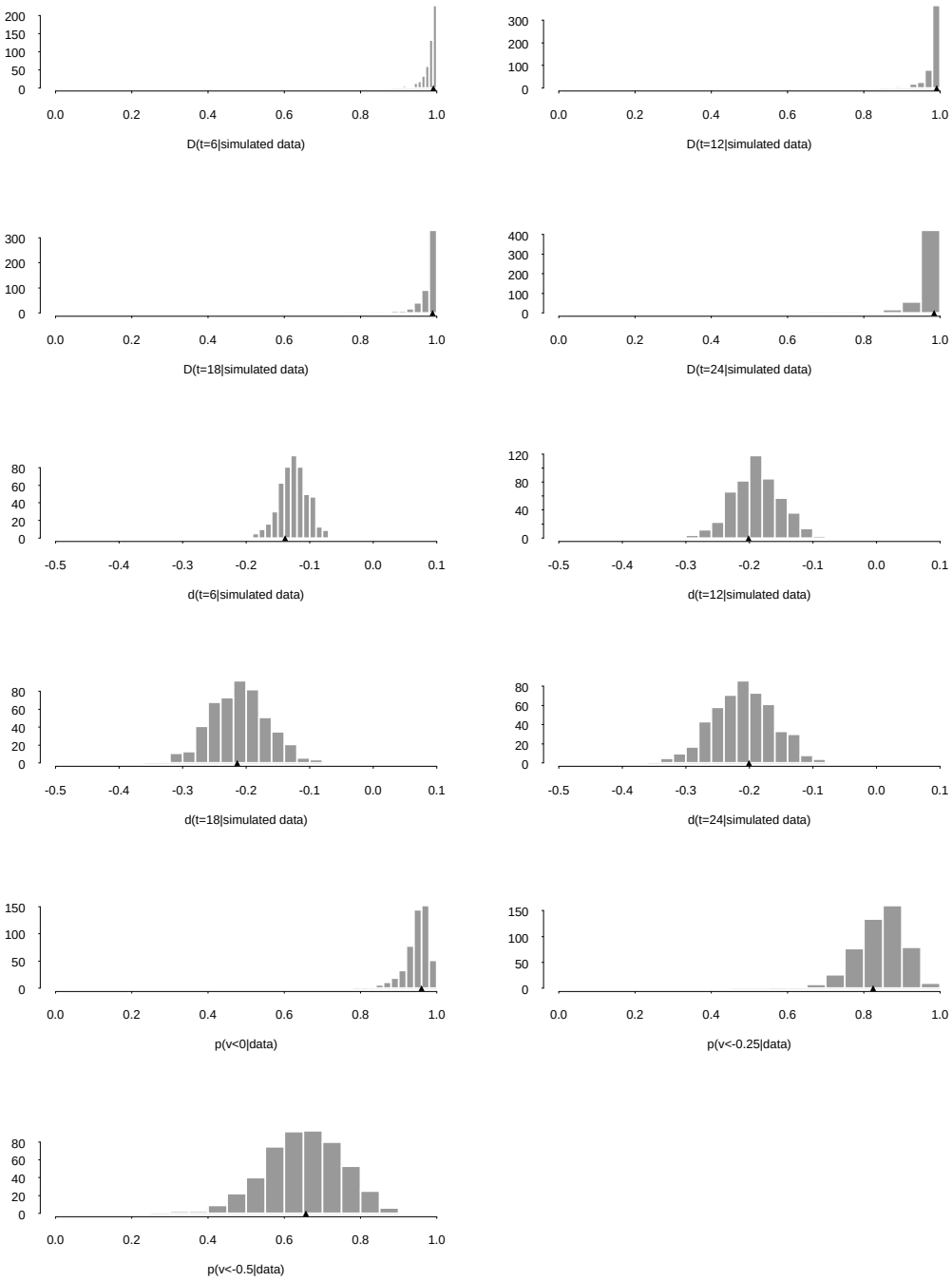

Figure 3.10: Histograms of predictive characteristics based on 500 simulations if the trial had been stopped at 4th interim look.

$$
\begin{aligned}
i &= 1 : j = 78, ..., 120; \\
i &= 2 : j = 79, ..., 120. \\
(\theta_1, v, a_1, a_2) &\sim p(\theta_1, v, a_1, a_2 | data \text{ as of } 06/92)
\end{aligned} \tag{3.3}
$$

I repeated this procedure 500 times and found that these extra 85 patients would not have affected our 1987 conclusion, even if we had accrued them after the trial was stopped. This can be seen in Figure 3.11, which compares the histogram of $\mathrm{d}(t|data)$, $t = 6, 12, 18, 24$ for these 500 simulations to the true $\mathrm{d}(t|data)$ as of June, 1992. The black triangles on the plots shows the true $\mathrm{d}(t|data)$ value as of June, 1992. Figure 3.11 also compares the histogram of $p(v < 0), p(v < -0.25)$, and $p(v < -0.5)$ for these 500 simulations to the true $p(v < 0), p(v < -0.25)$, and $p(v < -0.5)$ as of June, 1992. The black triangles on the plot shows the true $p(v < 0), p(v < -0.25)$, and $p(v < -0.5)$ value as of June, 1992. All these plots demonstrates that there is really not that much that would have changed the decision to stop the trial in 1987.

3.6 Discussion

All the results presented in this chapter is based on the Weibull model with the dependent parameterization. But I did try the Weibull model with the independent parameterization on this data set. I used same priors for a_i's and θ_i's. The posterior distributions of a_i's, θ_1, the predictive measures $\mathrm{D}(t|data)$ and $\mathrm{d}(t|data)$ are pretty much the same as those I presented before.

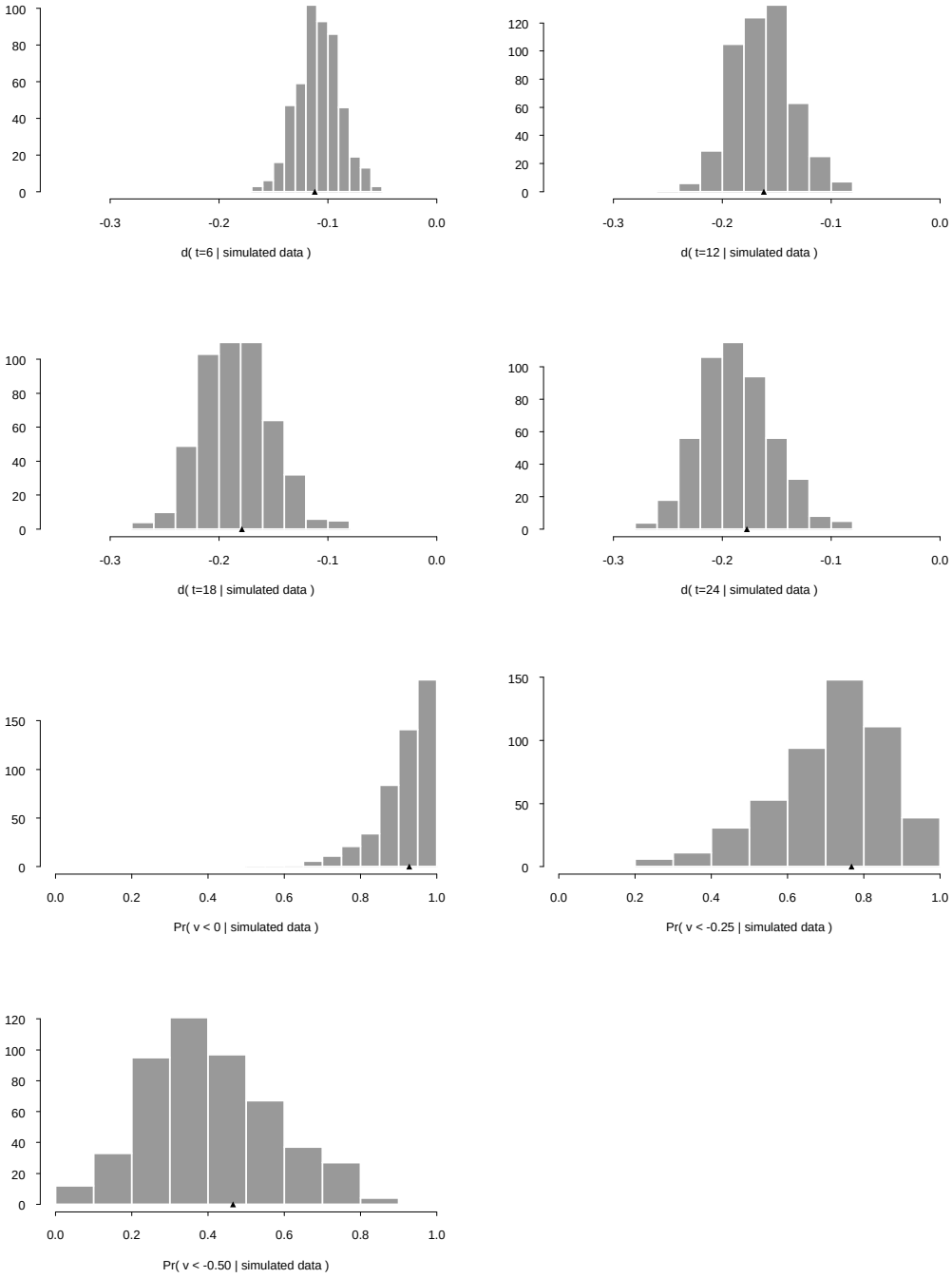

Figure 3.11: Histograms of $\mathrm{d}(t|data)$, $p(v < 0)$, $p(v < -0.25)$, and $p(v < -0.5)$ based on 500 simulations if the trial had not been stopped at 5th interim look.

Chapter 4

A Mixture Survival Model

4.1 Introduction

A mixture survival model can be found in Berkson and Gage (1952). Berkson and Gage (1952) proposed that a proportion C of patients treated for cancer might be effectively cured in the sense that their mortality experience would follow, approximately, the appropriate actuarial tables for the general population without cancer; while a proportion 1-C would be subject to additional forces of mortality. Chen, Hill, Greenhouse and Fayos (1985) generalized the mixture model of Berkson and Gage (1985) and used for the analysis of survival data from cancer. A subject Bayesian approach was taken towards inference and decision-making. Such a model are effective for very small samples, including cases where only a handful of patients have died, and a moderate number are still surviving. In some clinical trials (such as vaccine trials), more than 99 percent of the subjects may not experience the event of interest. Thus, the Weibull model introduced in Chapter 2 may not fit the data very well. In this chapter, I describe a mixture model, with a Weibull component and a surviving fraction for the purpose of analyzing such clinical trial.

4.2 The Mixture Model

Assuming conditional independence, the mixture model takes the form:

Stage I:

$$\begin{aligned}(t_{ij}|\theta_i, a_i, \pi_i) &\overset{i.i.d.}{\sim} \theta_i a_i t_{ij}^{a_i - 1} exp(-\theta_i t_{ij}^{a_i}) \text{ with probability } (1-\pi_i) \\ i &= 1, 2; \quad j = 1, 2, \cdots, n_i \end{aligned} \tag{4.1}$$

Stage II:

$$\theta_i \overset{i.i.d.}{\sim} G(\theta_i|u_\theta, v_\theta) \tag{4.2}$$

$$a_i \overset{i.i.d.}{\sim} G(a_i|u_a, v_a) \tag{4.3}$$

$$\begin{aligned}\pi_i &\overset{i.i.d.}{\sim} B(\pi_i|u_\pi, v_\pi) \\ i &= 1, 2. \end{aligned} \tag{4.4}$$

For this model, survival time has a Weibull distribution $W(t_{ij}|\theta_i, a_i)$ with probability $1 - \pi_i$. The remaining proportion of the observations, π_i, will not experience the event during the study. The survival function for this model consists of two terms, one corresponding to each component of the mixture:

$$S(t) = (1-\pi)e^{-\theta t^a} + \pi \tag{4.5}$$

During the second stage, I assume the scale parameters $\theta_i's$ are conditionally $i.i.d.$ according to a gamma distribution. I chose a gamma prior for the $\theta_i's$, because it is a conjugate prior for the $\theta_i's$.

I also assumed the shape parameters $a_i's$ are conditionally $i.i.d.$ according to a gamma distribution.

For each treatment, the surviving fraction π_i's will be assumed conditionally $i.i.d.$ according to a beta distribution. The parameters of the beta distribution,

u_π and v_π respectively represent the number of successes and failures in a sequence of Bernoulli trials. The mean of the beta distribution is $\frac{u_\pi}{u_\pi+v_\pi}$, and the variance is $\frac{u_\pi v_\pi}{(u_\pi+v_\pi)^2(u_\pi+v_\pi+1)}$. A beta distribution with $u_\pi = 1$ and $v_\pi = 1$ is a uniform distribution on $(0, 1)$.

4.3 Choose the Prior Distribution

If $a_1 = a_2 = 1$, then our model becomes a mixture model with an exponential component and a surviving proportion (Stangl, 1991). Thus, I chose the prior for $\theta_i, i = 1, 2$ the same as the prior for the hazard rate in an exponential model. Since a gamma prior is a conjugate prior for the θ_i's, for simplicity's sake, I chose a gamma prior for the $\theta_i's$. For sensitivity analysis, we may generalize this prior to extreme cases. $G(\theta_i|1, \epsilon)$ is the uniform prior as $\epsilon \to 0$.

For the shape parameter $a_i, i = 1, 2$, we will still be very interested in the special point 1, because it is the change point at which the Weibull distribution will have a constant hazard rate (the exponential distribution). At less than 1, the Weibull distribution has a decreasing hazard rate, whereas at more than 1, it has an increasing hazard rate. Thus, a gamma distribution with a mode of 1 and certain variance will be chosen as the prior for the $a_i's$. Again, the uniform prior may be approximated by $G(a_i|1, \epsilon)$ for the $a_i's$ when $\epsilon \to 0$.

Because the surviving fraction parameter π_i is between 0 and 1, and because the beta prior is a conjugate prior, we chose a beta distribution as the prior for the $\pi_i's$. A uniform prior is equivalent to $B(\pi_i|1, 1)$.

4.4 The Posterior Distributions

Using the notation in Chapter 2, d_i denotes the number of deaths in treatment i. $\{t_{ij}, j = 1, 2, \cdots, d_i, i = 1, 2\}$ is the uncensored observations, and $\{t_{ij}, j = d_i + 1, \cdots, n_i, i = 1, 2\}$ is the censored observations. Thus, given a prior distribution $p(\theta_1, a_1, \pi_1, \theta_2, a_2, \pi_2)$, the joint posterior distribution of $(\theta_1, a_1, \pi_1, \theta_2, a_2, \pi_2)$ is:

$$\begin{aligned} p(\theta_1, a_1, \pi_1, \theta_2, a_2, \pi_2|\mathbf{t}) &\propto p(\theta_1, a_1, \pi_1, \theta_2, a_2, \pi_2) L(\theta_1, a_1, \pi_1, \theta_2, a_2, \pi_2|t) \\ &\propto \prod_{i=1}^{2} \{((1-\pi_i)\theta_i a_i)^{d_i} (\prod_{j=1}^{d_i} t_{ij})^{a_i - 1} \\ &\quad \prod_{j=d_i+1}^{n_i} (\pi_i + (1-\pi_i) e^{-\theta_i {t_{ij}}^{a_i}}) \} \end{aligned} \tag{4.6}$$

where $p(\theta_1, a_1, \pi_1, \theta_2, a_2, \pi_2|\mathbf{t})$ is the joint posterior distribution for the parameter set, $L(\theta_1, a_1, \pi_1, \theta_2, a_2, \pi_2|\mathbf{t})$ is the likelihood function based on data $\mathbf{t} = \{t_{ij}, \quad j = 1, 2, \cdots, n_i, \quad i = 1, 2\}$.

There is no closed form for the above density. Thus, I will use a Gibbs sampling scheme to calculate the above density numerically.

4.5 The Gibbs Sampling Algorithm for the Mixture Model

To obtain the posterior density of interesting parameters in the mixture model, I employed the Gibbs sampling algorithm.

The Gibbs sampling algorithm is a Markovian updating scheme enabling one to obtain (in the limit) samples from a joint distribution, via iterated sampling from full conditional distributions. Given a joint posterior density $f(\beta|Y)$, functional forms of the k univariate full conditional densities can

be readily written down, at least up to proportionality. These full conditional densities are denoted by $f(\beta_1|\beta_2, \beta_3, \ldots, \beta_k, Y)$, $f(\beta_2|\beta_1, \beta_3, \ldots, \beta_k, Y)$, $\ldots$, $f(\beta_k|\beta_1, \beta_2, \ldots, \beta_{k-1}, Y)$, then the Gibbs sampling algorithm proceeds as follows:

choose initial values $\beta_2^{(0)}, \beta_3^{(0)}, \ldots, \beta_k^{(0)}$ and generate a value $\beta_1^{(1)}$ from the conditional density $f(\beta_1|\beta_2^{(0)}, \beta_3^{(0)}, \ldots, \beta_k^{(0)}, Y)$. Similarly, generate a value $\beta_2^{(1)}$ from the conditional density $f(\beta_2|\beta_1^{(1)}, \beta_3^{(0)}, \ldots, \beta_k^{(0)}, Y)$. and continue up to the value $\beta_k^{(1)}$ from the conditional density $f(\beta_k|\beta_1^{(1)}, \beta_2^{(1)}, \ldots, \beta_{k-1}^{(1)}, Y)$. Then, with the new realization of β, $\beta^{(1)}$, we can replace $\beta^{(0)}$ to repeat above process, say t times, to produce $\beta^{(t)}$. Under mild conditions, as shown by Geman and Geman (1984) for discrete distributions, we can treat $\beta_i^{(t)}$ approximately from posterior distribution $f(\beta_i|Y)$, the marginal distribution of β_i. The posterior density function of $(\beta_i|Y)$ can be obtained by

$$f(\beta_i|Y) = \frac{1}{N}\sum_{t=1}^{N} f(\beta_i|\beta_j^{(t)}, j \neq i). \tag{4.7}$$

See Gelfand and Smith (1990).

The Gibbs sampler involves drawing random samples from all full conditional densities. Often the likelihood and prior forms specified in Bayesian analysis lead to distributions in full conditional expression which are of a familiar form, such as normals and gammas. See for example Gelfand and Smith (1990), Escobar and West (1991) and Cao (1993). In these cases, standard algorithms are available to generate random variates. However, in other cases, it is very difficult to generate random samples from full conditional distributions. Depending on the nature of the distribution family, an efficient choice of a method generally requires mathematical insight on the part of the designer

of the sampling scheme, e.g. exploiting a property such as log-concavity, or the knowledge of certain density characteristics such as the supremum of the density or the explicit form of the inverse of the cumulative density function. Recently, Gilks and Wild (1992) have proposed a novel 'adaptive rejection sampling' method of sampling from any log-concave univariate probability density function. Their suggested algorithm is based on the fact that any concave function can be bounded by piecewise linear upper and lower bounds (hulls), constructed by using tangents at, and chords between, evaluated points on the function over its domain. Dellaportas and Smith (1993) used this algorithm to analyze generalized linear and proportional hazards models. The adaptive rejection sampling method based on the derivative-free version (Gilks 1992) is also implemented in *BUGS*.

Mixture models have been analyzed via a Gibbs sampling method. To use the Gibbs sampling algorithm, we must introduce an extra parameter vector of dummy variables indicating group membership. Thus, the cases (observations) are categorized according to whether or not they belong to the survival group. Although this membership is apparent for observations experiencing the event of interest, it is not apparent for censored observations. This membership vector of dummy variables influenced the priors of the scale and shape parameters of each treatment arm when they were updated by the observations from the corresponding treatment arm. Sampling occurs from this updated conditional density, and once it takes place, the posterior probabilities of group membership will be updated. Alternating between these two steps leads to a sample from the joint posterior distribution. The algorithm begins by arbitrarily assigning values to the $2+2+2+n_1+n_2 = 6+n_1+n_2$ parameters, ${\theta_i}^0, {a_i}^0, {\pi_i}^0$, and ${\lambda_{ij}}^0, j = 1, 2, \cdots, n_i$, and $i = 1, 2$. Here $\lambda = \{\lambda_{ij}, j = 1, 2, \cdots, n_i, i = 1, 2\}$

is the membership vector of dummy variables indicating group membership of each observation. If $\lambda_{ij} = 1$, then the observation belongs to the survival group; otherwise, $\lambda_{ij} = 0$. Then, we begin iterations $k = 1, 2, \cdots, icyc$ (total number of samples).

Let $\Theta = (\theta_1, a_1, \pi_1, \theta_2, a_2, \pi_2)$ and $\Theta_{(-\theta_1)}$ denote Θ without θ_1, etc.

The full conditional posterior distributions for the $\theta_i,\ i = 1, 2$ are:

$$\begin{aligned} p(\theta_i|\Theta_{(-\theta_i)}, \lambda, \mathbf{t}) &\propto {\theta_i}^{u_\theta + \sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij})-1} exp(-(v_\theta + \sum_{j=1}^{n_i} {t_{ij}}^{a_i}(1-\lambda_{ij}))\theta_i) \\ &\propto G(\theta_i|u_\theta + \sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij}), v_\theta + \sum_{j=1}^{n_i} {t_{ij}}^{a_i}(1-\lambda_{ij})) \quad (4.8) \end{aligned}$$

Thus, we can sample θ_i from a gamma distribution with prior parameters updated by $\sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij})$ and $\sum_{j=1}^{n_i} {t_{ij}}^{a_i}(1-\lambda_{ij}))$. Note that the updating uses only those observations that do not belong to the survival group. Thus, in this model, the θ_i should be interpreted as a conditional shape parameter, i.e., the shape parameter in the truncated (within the study duration) Weibull distribution.

In the second step, we need to sample the shape parameters $a_i,\ i = 1, 2$. The full conditional posterior distributions for $a_i,\ i = 1, 2$ are:

$$\begin{aligned} p(a_i|\Theta_{(-a_i)}, \lambda, \mathbf{t}) &\propto {a_i}^{u_a + \sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij})-1} \exp((a_i - 1)\sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij}) log(t_{ij}) \\ &\quad -(v_a a_i + \theta_i \sum_{j=1}^{n_i} {t_{ij}}^{a_i}(1-\lambda_{ij}))) \quad (4.9) \end{aligned}$$

There is no closed form for the above density. Thus, sampling from the above density may seem complicated. But the recently developed Adaptive Rejection Sampling Method (Wild and Gilks, 1993) can be used to accelerate the

sampling procedure. See Appendix B for the proof of the log-concavity of the conditional distribution of a_i given data and other parameters.

Next, the surviving fractions (π_i) are sampled. The full conditional distributions for the surviving fractions are:

$$\begin{aligned} p(\pi_i|\Theta_{(-\pi_i)},\lambda,\mathbf{t}) &\propto \pi_i{}^{u_\pi+\sum_{j=1}^{n_i}\lambda_{ij}}(1-\pi_i)^{v_\pi+n_i-\sum_{j=1}^{n_i}\lambda_{ij}} \\ &\propto B(\pi_i|u_\pi+\sum_{j=1}^{n_i}\lambda_{ij}, v_\pi+n_i-\sum_{j=1}^{n_i}\lambda_{ij}) \end{aligned} \tag{4.10}$$

Thus, the full conditional distribution of π_i is a beta distribution with updated parameters $u_\pi+\sum_{j=1}^{n_i}\lambda_{ij}$ and $v_\pi+n_i-\sum_{j=1}^{n_i}\lambda_{ij}$.

Finally, the dummy membership variables λ_{ij} are sampled from a Bernoulli(p_{ij}) distribution. The p_{ij} is the conditional posterior probability that the ij-th observation belongs to the surviving group (SG) and can be calculated as:

$$\begin{aligned} p_{ij} &= P(t_{ij}\in SG|\Theta,t) \\ &= \frac{L(t|t_{ij}\in SG)P(t_{ij}\in SG)}{L(t|t_{ij}\in SG)P(t_{ij}\in SG)+L(t|t_{ij}\notin SG)P(t_{ij}\notin SG)} \end{aligned} \tag{4.11}$$

L here refers to the likelihood. Since observation t_{ij} is independent of the other observations, the above equation includes all but one of the likelihood factors, we have the following:

$$p_{ij}=\frac{f(t_{ij}|t_{ij}\in SG)^{\delta_{ij}}S(t_{ij}|t_{ij}\in SG)^{1-\delta_{ij}}\pi_i}{f(t_{ij}|t_{ij}\in SG)^{\delta_{ij}}S(t_{ij}|t_{ij}\in SG)^{1-\delta_{ij}}\pi_i+f(t_{ij}|t_{ij}\notin SG)^{\delta_{ij}}S(t_{ij}|t_{ij}\notin SG)^{1-\delta_{ij}}(1-\pi_i)} \tag{4.12}$$

Since the density and survival functions depend upon whether the observation t_{ij} is in the surviving group, we need to have the following four equations:

$$f(t_{ij}|t_{ij}\in SG,\Theta) = 0 \tag{4.13}$$

$$S(t_{ij}|t_{ij} \in SG, \Theta) = 1 \tag{4.14}$$

$$f(t_{ij}|t_{ij} \notin SG, \Theta) = \theta_i a_i {t_{ij}}^{a_i - 1} e^{-\theta_i t_{ij}^{a_i}} \tag{4.15}$$

$$S(t_{ij}|t_{ij} \notin SG, \Theta) = e^{-\theta_i t_{ij}^{a_i}} \tag{4.16}$$

That is, if t_{ij} is an observation from the survival group, it will not experience the event of interest. Thus, the density for any t_{ij} within the study duration should be 0, and the survival function must be equal to 1. If t_{ij} is not from the survival group, then it will experience the event of interest according to a Weibull distribution. Finally we have:

$$p_{ij} = \begin{cases} 0; & t_{ij} \text{ uncensored i.e. } \delta_{ij} = 1. \\ \frac{\pi_i}{\pi_i + (1-\pi_i) exp(-\theta_i t_{ij}^{a_i})}; & t_{ij} \text{ censored i.e. } \delta_{ij} = 0. \end{cases}$$

Note that p_{ij} is monotonically increasing in terms of the censoring time. In fact, the probability that an observation belongs to the surviving group is simply π_i if the censoring time is near 0. As the censoring time increases, p_{ij} increases monotonically to 1, which means that censored observations are likely to be included in the survival group and will have little or no influence on the posterior density of the Weibull parameters.

Thus, the full conditional distribution for λ_{ij} has a Bernoulli(p_{ij}) distribution when the observation is censored. When the observation is not censored, λ_{ij} is set to 0 and remains at zero throughout the algorithm. That is:

$$\begin{aligned} p(\lambda_{ij}|\Theta, \lambda_{-\lambda_{ij}}, \mathbf{t}) &= p(\lambda_{ij}|\Theta, \mathbf{t}) \\ &= \lambda_{ij}^{\frac{\pi_i}{\pi_i + (1-\pi_i) exp(-\theta_i t_{ij}^{a_i})}} \times (1 - \lambda_{ij})^{1 - \frac{\pi_i}{\pi_i + (1-\pi_i) exp(-\theta_i t_{ij}^{a_i})}} \end{aligned} \tag{4.17}$$

4.6 Convergence of the Gibbs sampler

Establishing convergence is straightforward by using the theory of Harris recurrent chains, as outlined by Tierney (1991), Roberts (1992). Diagnostics for monitoring convergence will be of value. Though much has been written on this subject recently and the theory has been considerably clarified, some of the simplest aspects have remained controversial. The most basic issue in dispute is whether valid inference from Gibbs sampler results from averaging over one long run of the chain, as the name of the method and all the theory suggest, or whether multiple shorter runs are desirable or even necessary for valid inference (Geyer, 1992). Throughout this thesis, we recommend that inference be based on a single long run, and that starting values and exact form of the algorithm be chosen on the basis of experimentation. A bad starting value can lead to slow convergence. There is certainly advantage to a systematic search for good starting values.

Chapter 5

A *Haemophilus Influenzae type B* Vaccine Trial

5.1 Introduction

This chapter applies the model proposed in the last chapter to a vaccine trial (Berry, Wolff and Sack, 1992). In this trial, a vaccine for *Haemophilus influenzae type b* is compared with a placebo vaccine in a randomized trial. The authors tried to minimize the expected number of cases of *H. influenzae type b* among Navajo Indians during the next several years. The data that accumulated during the trial includes the time each subject was at risk. Berry, Wolff, and Sack (1992) found optimal designs by numerical calculations using backwards induction, and they evaluated optimal designs in an actual trial and compared them with the frequentist design that was actually used in the trial. However, in this chapter, our focus is restricted to the posterior behavior of parameters of interest and predictive protective efficacy (PE) of the vaccine under the mixture model assumption.

5.1.1 Purpose of the Study

H. influenzae type b (HIB) is a disease that usually affects infants and babies; it is a bacterium that causes severe infections affecting the blood stream (septicemia), the central nervous system (meningitis), the bones and joints (osteomyelitis or septic arthritis), the upper airways (epiglottitis), and other sites. Collectively, these are known as "invasive *H. influenzae* infections". If these infections are untreated or if treatment is delayed, they are often fatal. Many patients treated for HIB meningitis develop neurologic complications, especially loss of hearing or a seizure disorder, and thus are impaired for life.

These infections occur in all groups. However, the groups with the highest incidence are certain tribes of American Indians, including Navajos and the Apaches. Among these people, the risk that a child will develop invasive HIB infection is about 1% at some time before age 3. This rate is 10-50 times greater than that of other American populations. Thus, although a successful vaccine will be useful in the general U.S. population, it will be even more important to these tribes.

A new HIB vaccine (HIB-OMP, developed by the Merck, Sharpe and Dohme Research Lab) was tested for efficacy among Navajo infants living on the Navajo reservations in Arizona, New Mexio, and Utah. It was found to be highly efficacious.

5.1.2 Design of the Trial

The study was conducted among Navajo Indians by the Department of International Health of The Johns Hopkins School of Hygiene and Public Health. Each month there were 105 2-month-old Navajo babies allocated to each treat-

ment group. Babies in the trial were vaccinated at 2 and 4 months of age. They were followed until age 18 months. A case during the intervening 16 months is a failure of the vaccine. The study was to continue until a fixed total of 5000 subjects completing the trial. It was estimated that approximately 5600 subjects must be enrolled in the trial to reach this target and that approximately 40 months must elapse to follow the subjects to completion. There was the possibility of stopping early based on confidence intervals for efficacy. The study actually accumulated about 5250 subjects, and it was stopped when the vaccine/placebo comparison was "highly significant" and sufficient experience had accumulated in the oldest age group. As is evident from the fact that the study was not stopped even though the interim data were quite convincing, the criteria for early stopping were stringent.

5.1.3 Efficacy of a Vaccine

Therapeutic clinical trials involve patients who have a particular disease or condition. The objective of treatment in such trials is to improve some aspect of the patient's condition. But in a prophylactic clinical trial (such as vaccine trial), the subjects do not have the disease in question, and the objective of treatment is to keep them disease-free. The incidence of some important diseases is less than one percent, and so success rates of a placebo vaccine may exceed 99%. Therefore, vaccine trials usually require large numbers of subjects, so that researchers may draw conclusive inferences concerning the vaccine's efficacy.

The value of a vaccine can be expressed as its protective efficacy (PE), which

is defined as the reduction in incidence of disease for vaccinated individuals:

$$\text{PE} = 1 - \frac{\text{incidence in the vaccine group}}{\text{incidence in the nonvaccinated group}} \tag{5.1}$$

Protective efficacy is usually measured in placebo-controlled, randomized trials; and the incidence in the vaccine group is compared with the incidence in the placebo group. When calculated during such a trial, PE provides an estimate of the biological power of the vaccine to protect the individual.

5.2 Preliminary Analysis

Appendix A displays the HIB-OMP vaccine trial data. There were a total 22 HIB cases accumulated in the placebo group and 1 HIB case in the vaccine group when the trial was stopped. The numbers of subjects accrued in each treatment group are 2625. Figure 5.1 shows the Product–Limit (PL) estimate of the survivor function of the placebo group and vaccine group, where the black triangles designates the cases.

5.3 Applying the Mixture Model to the HIB Vaccine Trial

The mixture model for the HIB vaccine data, thus we have:

$$\begin{aligned}
(t_{ij}|\theta_i, a_i, \pi_i) &\overset{iid}{\sim} \theta_i a_i t_{ij}^{a_i - 1} exp(-\theta_i t_{ij}^{a_i}) \text{ with probability } (1-\pi_i) \\
\theta_i &\overset{iid}{\sim} G(\theta_i|u_\theta, v_\theta) \\
a_i &\overset{iid}{\sim} G(a_i|u_a, v_a) \\
\pi_i &\overset{iid}{\sim} B(u_\pi, v_\pi) \\
i &= 1, 2; \quad j = 1, 2, \cdots, 2625.
\end{aligned} \tag{5.2}$$

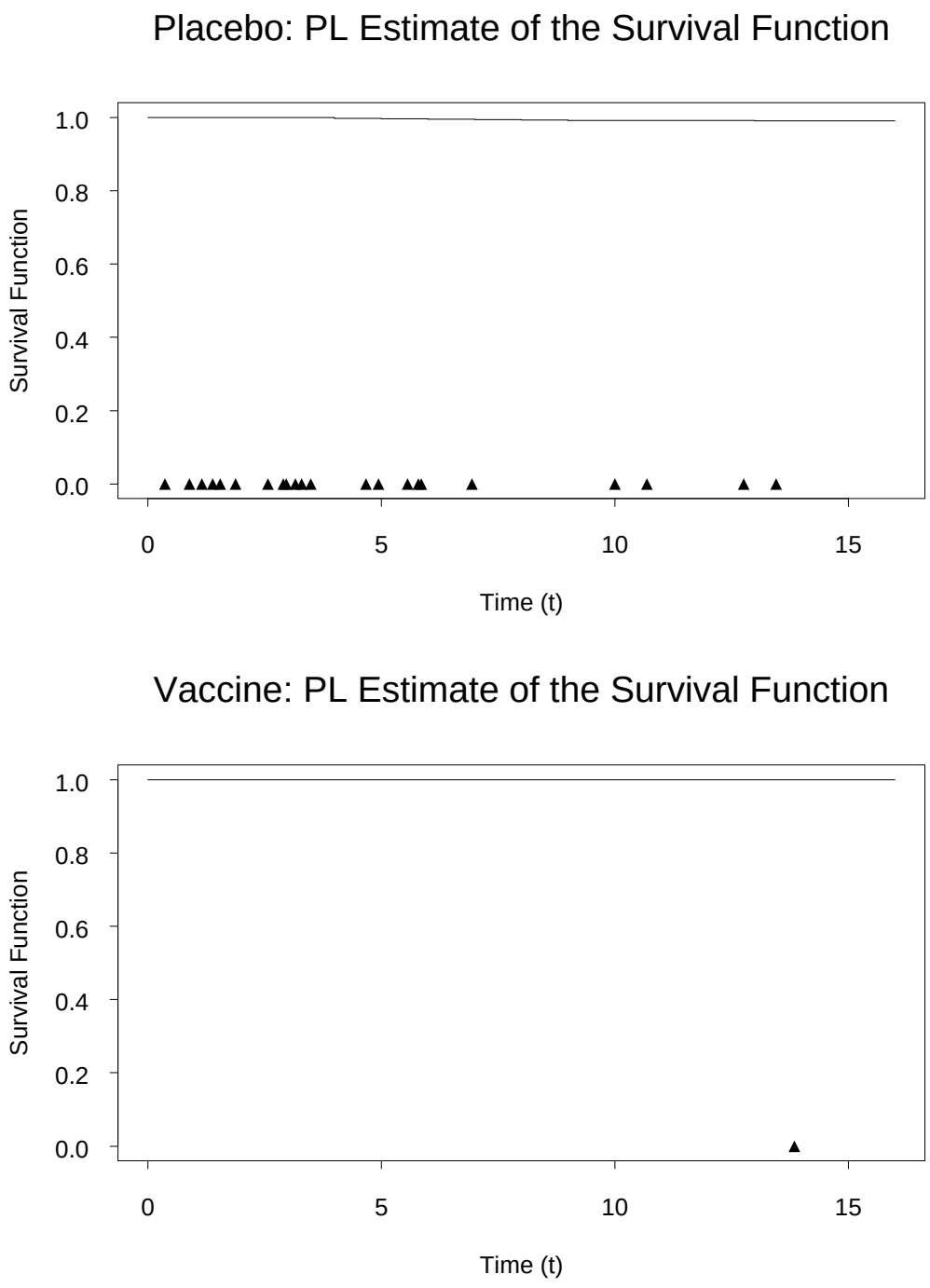

Figure 5.1: Product–Limit (PL) estimate of the survival function of the placebo group.

5.3.1 Prior for $(\theta_i, \pi_i, a_i), i = 1, 2$

According to Berry, Wolff, and Sack (1992), when an exponential model was used, their prior choices for the θ_1 (corresponding to the placebo) and θ_2 (corresponding to the vaccine) are $G(\theta_1|1, 3200) \times G(\theta_2|5, 10700)$ based on various considerations including the historical information, the relative risk (RR), and the prior experience of experts in the vaccine field, etc. Thus we will choose $G(\theta_1|1, 3200) \times G(\theta_2|5, 10700)$ as the prior for θ_1 and θ_2.

For the shape parameter $a_i, i = 1, 2$, we will first choose $G(a_i|11, 10)$ as the prior.

For the surviving fraction $\pi_i, i = 1, 2$, according to some previous studies of HIB, I will choose a $B(\pi_i|50, 2)$ as the prior for $\pi_i, i = 1, 2$, which gives the mode equal to $(50 - 1)/(50 + 2 - 2) = 0.98$ and a standard deviation equal to $\sqrt{50 \times 2/(50 + 2)^2(50 + 2 + 1)} = 0.0264$.

5.3.2 The Posterior Distributions and Predictive PE

Our goals here are to estimate the posterior distributions of a_1, a_2, θ_1 and θ_2 as well as predictive PE. The Monte Carlo sample size used here was 3,000 after an initial 3,000 'burn-in' cycles. Since the starting value can have an important effect on the performance of the Gibbs sampler, there is certainly advantage to a systematic search for good starting values. For this analysis, we choose initial values for $a_i = 1$ for $i = 1, 2$. The starting values of θ_1 and θ_2 are $1/3200$ and $5/10700$ respectively. We monitor all convergence features for a_1, a_2, θ_1 and θ_2. After Gibbs samples converge, we use Monte Carlo (sample size of 3000) averages of conditional distribution to obtain posterior distributions of interesting parameters and predictive PE.

To illustrate the convergence of the Gibbs sampler and how the starting values affect the convergence, I use the algorithm and sample the first 3000 iterations and exibit them in the following four pictures. In stead of displaying all features of the model, I only select four intersting parameters to show the convergence of the Gibbs sampler here. I also monitor all other convergence features for other parameters which are not displayed here.

Figure 5.2 show the convergence of the Gibbs sampler by choosing starting values $\theta_1 = \frac{1}{3}200, \theta_2 = \frac{5}{1}0700, a_1 = a_2 = 1, \pi_1 = \pi_2 = 0.99$. After several Gibbs runs, these parameters reach their stable distribution.

Figure 5.3 shows the convergence of the Gibbs sampler by choosing the same starting values except $a_1 = a_2 = 2, \pi_1 = \pi_2 = 0.98$. After serveral Gibbs iterations, Figure 5.3 displays the same feature as Figure 5.2.

Figure 5.4 displays the features of the Gibbs sampler by choosing the same starting values as in Figure 5.3 except $\theta_1 = \theta_2 = \frac{4}{1}0700$. Again, this figure displays the same feature as Figure 5.2, 5.3.

Figure 5.5 displays the features of the Gibbs sampler by choosing the same starting values as in Figure 5.4 except $a_1 = a_2 = 0.5$. From above four figures, we are pretty sure that the Gibbs sampler converges after 3000 burn-in.

Figure 5.6 shows the posterior distributions of parameters under the prior choice above, and

Figure 5.7 shows the predictive density of the "surviving fraction" $\pi_i{}^*, i = 1, 2$, which is

$$\pi_i{}^* = \pi_i + \int_{16}^{\infty} f(t|\theta_i, a_i, \pi_i) d_t, \quad i = 1, 2 \tag{5.3}$$

and the predictive density of PE.

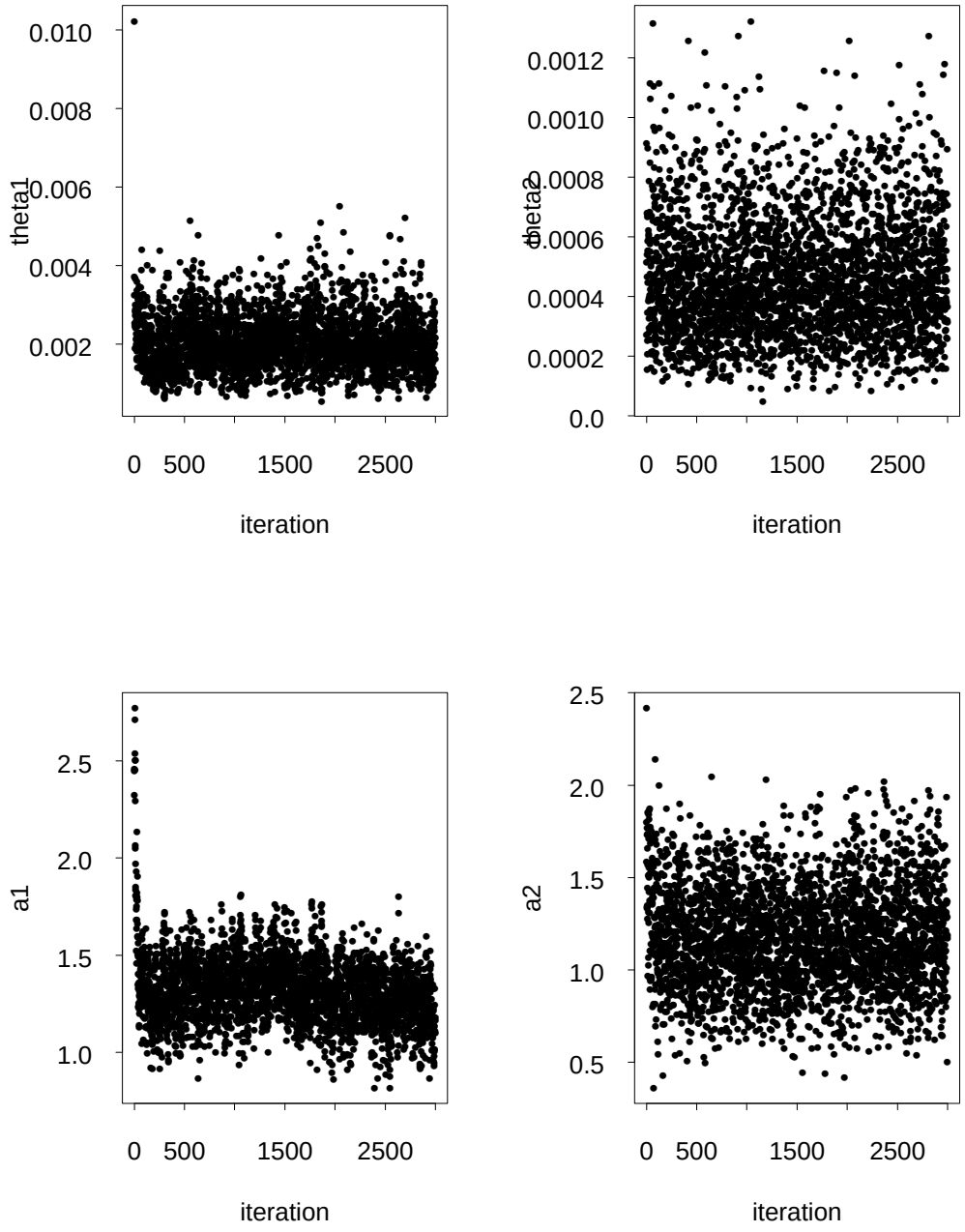

Figure 5.2: Convergence of the Gibbs Sampler (I).

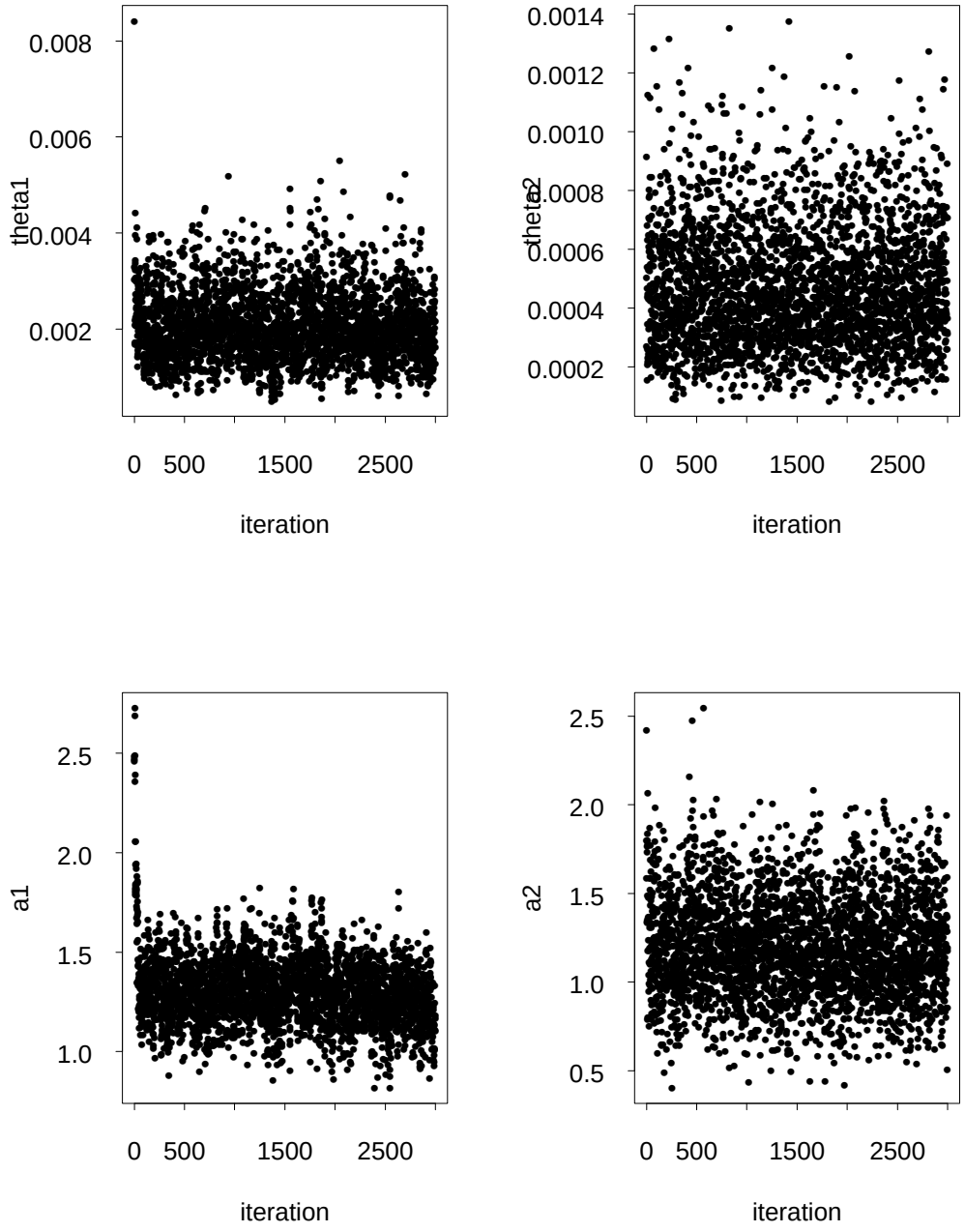

Figure 5.3: Convergence of the Gibbs Sampler (II).

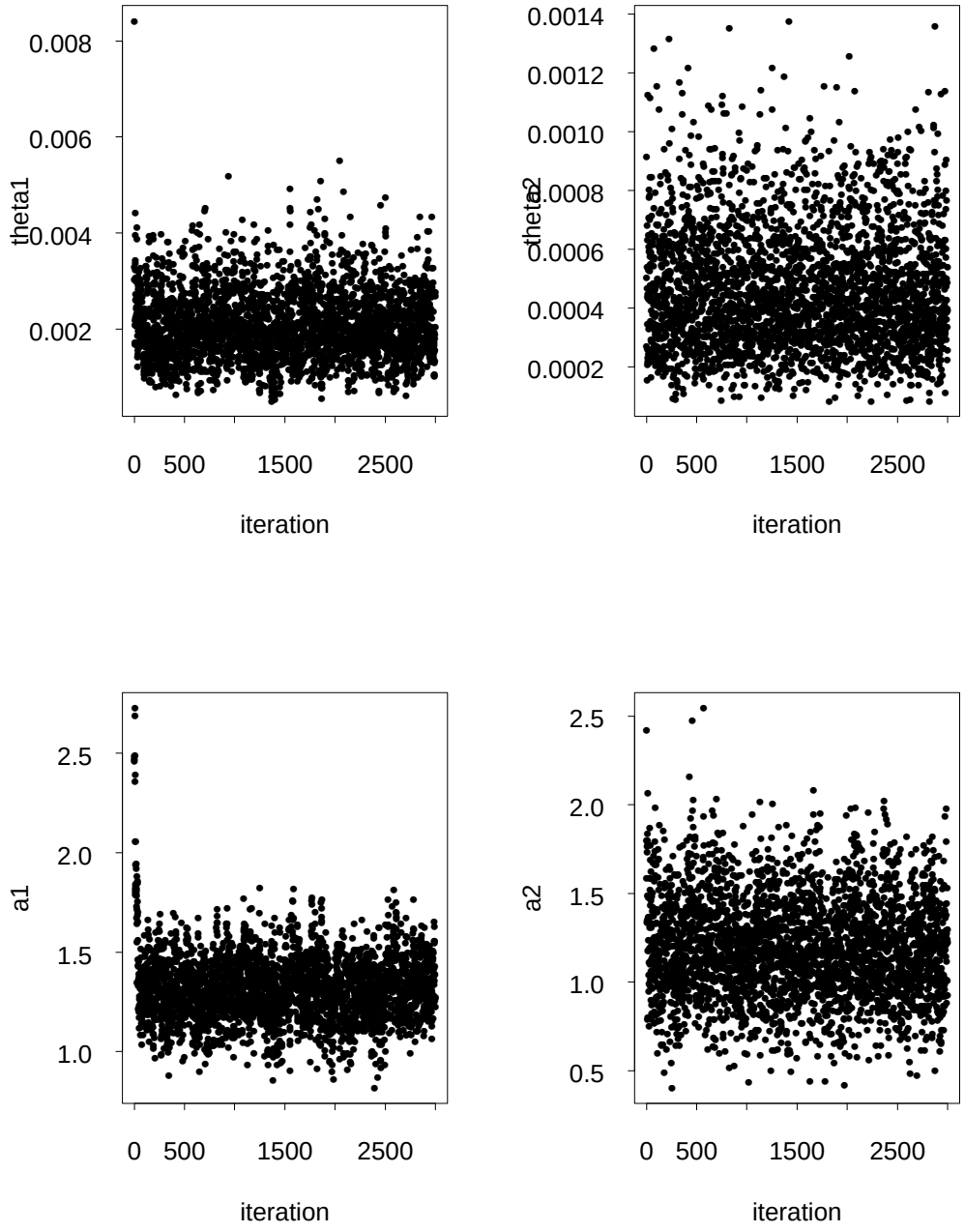

Figure 5.4: Convergence of the Gibbs Sampler (III).

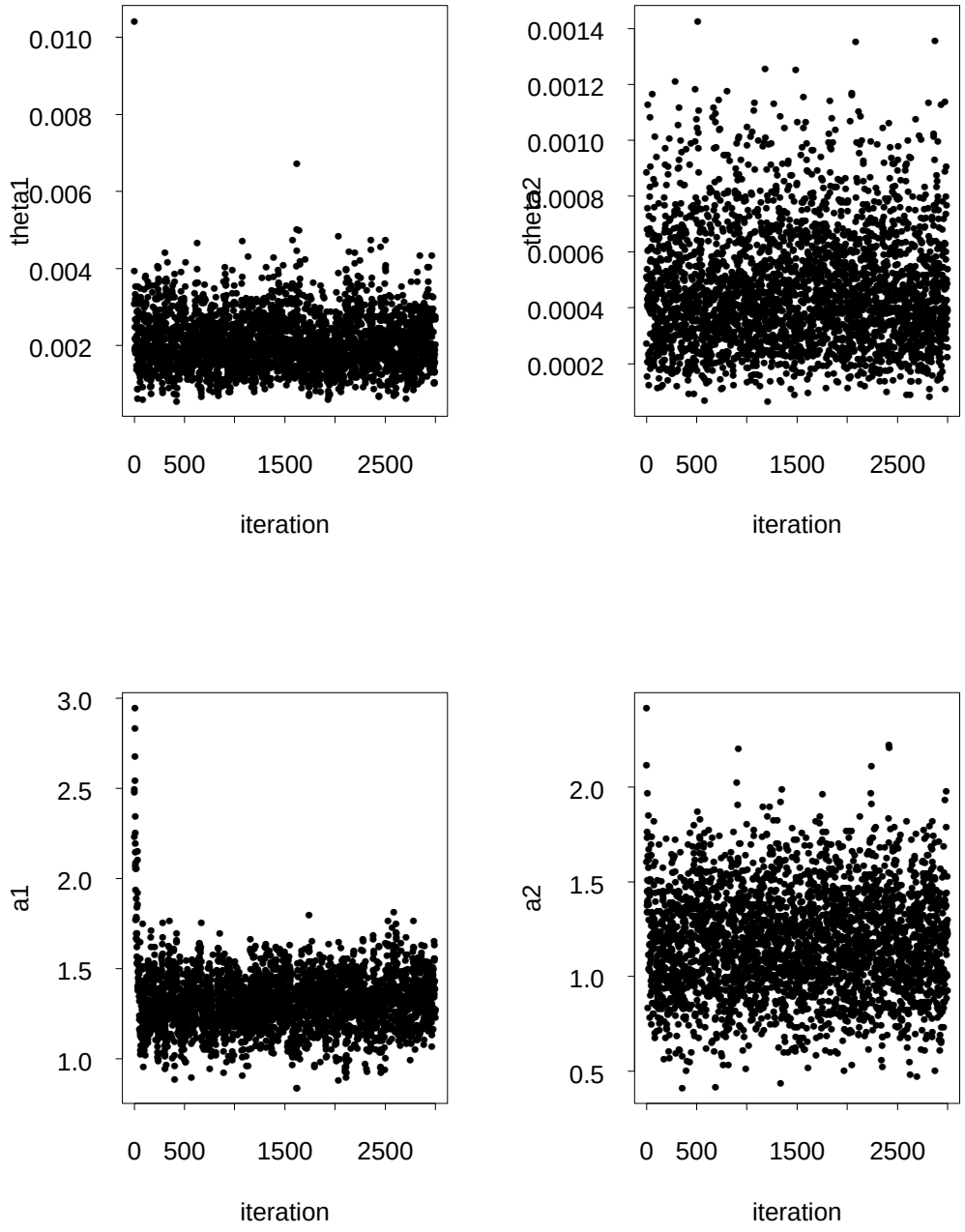

Figure 5.5: Convergence of the Gibbs Sampler (IV).

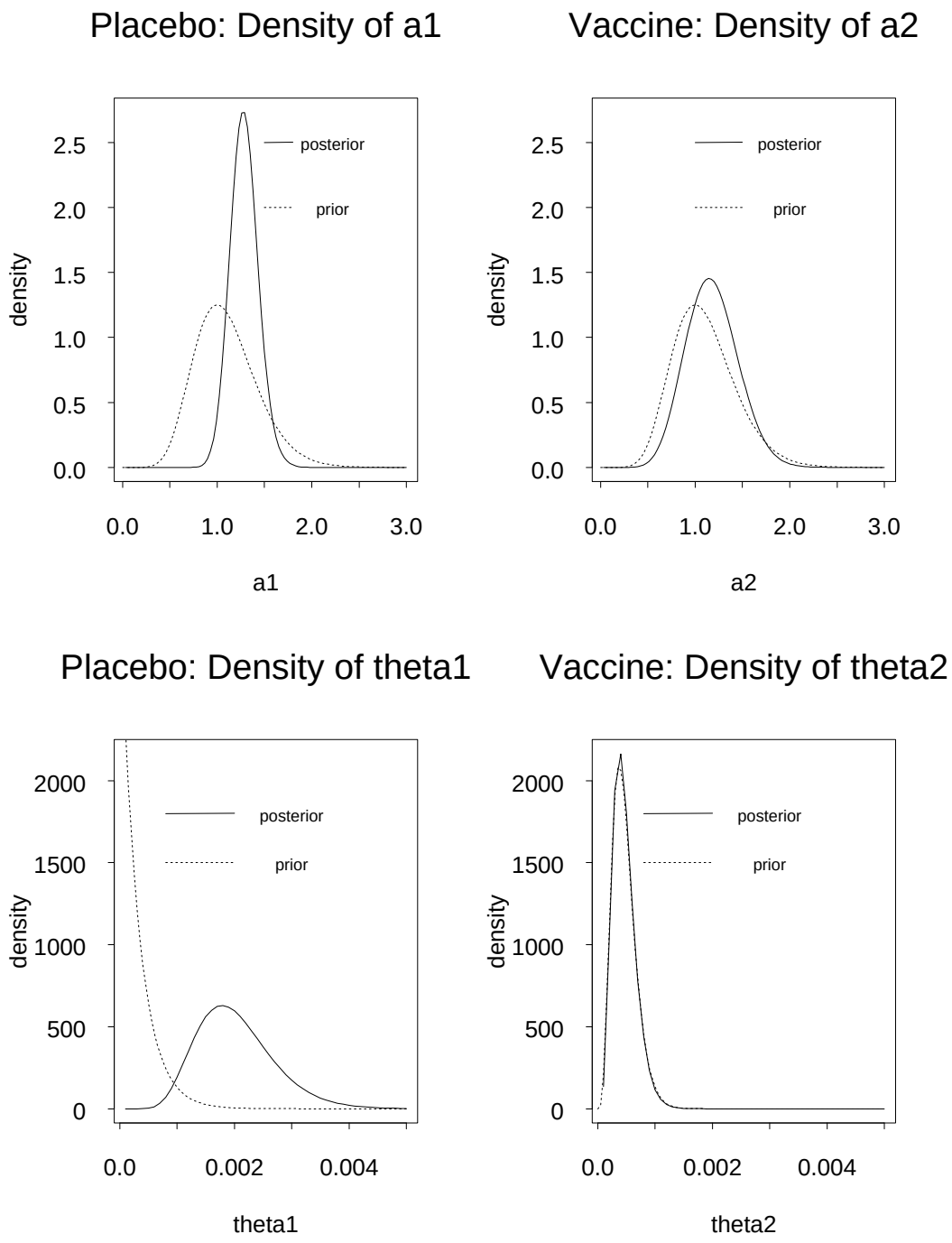

Figure 5.6: Posterior distributions of $(\theta_i, a_i), i = 1, 2$.

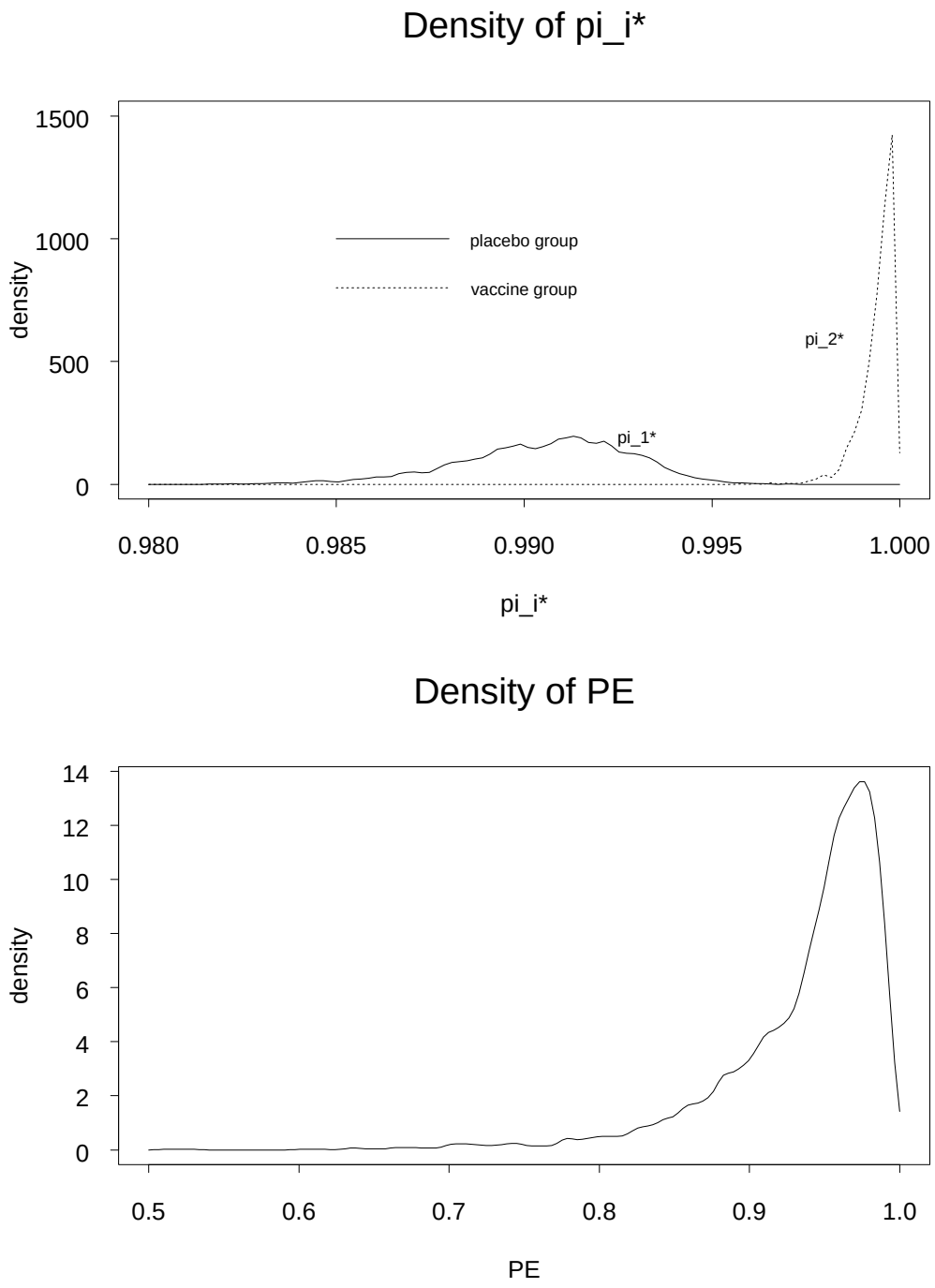

Figure 5.7: Predictive distributions of $\pi_i{}^*, i = 1, 2$ and PE.

From Figure 5.7, it is easy to see that the model fits the "surviving fraction" π_i^* very well. The empirical estimates of the "surviving fraction" are 0.991619 in the placebo group and 0.999619 in the vaccine group. These fractions are both in the HPD region of the predictive distributions. The mean of the predictive distribution is 0.9905195 for π_1^* and 0.9994364 for π_2^*.

From Figure 5.7, we may conclude that the vaccine is highly effective with a mean predictive PE= 0.9368128.

5.3.3 Sensitivity Analysis

The results shown in the previous section are based on a single prior. To see how much the results may differ, I chose four different priors for the parameters.

The first prior that I chose for investigating sensitivity is:

$$\begin{aligned} \theta_1 &\sim G(1, 3200) \\ \theta_2 &\sim G(5, 10700) \\ a_i &\sim G(11, 10) \\ \pi_i &\sim U(0,1) \quad i = 1, 2 \end{aligned} \tag{5.4}$$

This prior is the same as the prior that we used in last section, except that I changed the prior for $\pi_i, i = 1, 2$ to a uniform one.

The second prior that I chose is:

$$\begin{aligned} \theta_1 &\sim G(1, 3200) \\ \theta_2 &\sim G(5, 10700) \\ a_i &\sim U(0, \infty) \\ \pi_i &\sim B(50, 2) \quad i = 1, 2 \end{aligned} \tag{5.5}$$

This prior is the same as the prior that we used in last section, except that I changed the prior for $a_i, i = 1, 2$ to an uniform one.

The third prior that I chose is:

$$\begin{aligned} \theta_i &\sim U(0, \infty) \\ a_i &\sim G(11, 10) \\ \pi_i &\sim B(50, 2) \quad i = 1, 2 \end{aligned} \tag{5.6}$$

This prior is the same as the prior we used in last section, except that I changed the prior for $\theta_i, i = 1, 2$ to an uniform one.

The fourth prior that I chose is:

$$\begin{aligned} \theta_1 &\sim G(1, 3200) \\ \theta_2 &\sim G(5, 10700) \\ a_i &\sim G(5, 4) \\ \pi_i &\sim B(50, 2) \quad i = 1, 2 \end{aligned} \tag{5.7}$$

This prior is the same as the prior that we used in last section, except that I changed the prior for $\theta_i, i = 1, 2$ to flatter one.

Figure 5.8 shows the posterior distributions of the shape parameters $a_i, i = 1, 2$ under various prior choices.

Figure 5.9 shows the posterior distributions of the scale parameters $\theta_i, i = 1, 2$ under various prior choices above;

Figure 5.10 shows the predictive distributions of the "surviving fraction" $\pi_i^*, i = 1, 2$ and the protective efficacy PE.

Although there are some variations among the posterior distributions of the shape and scale parameters under different prior choices, the predictive

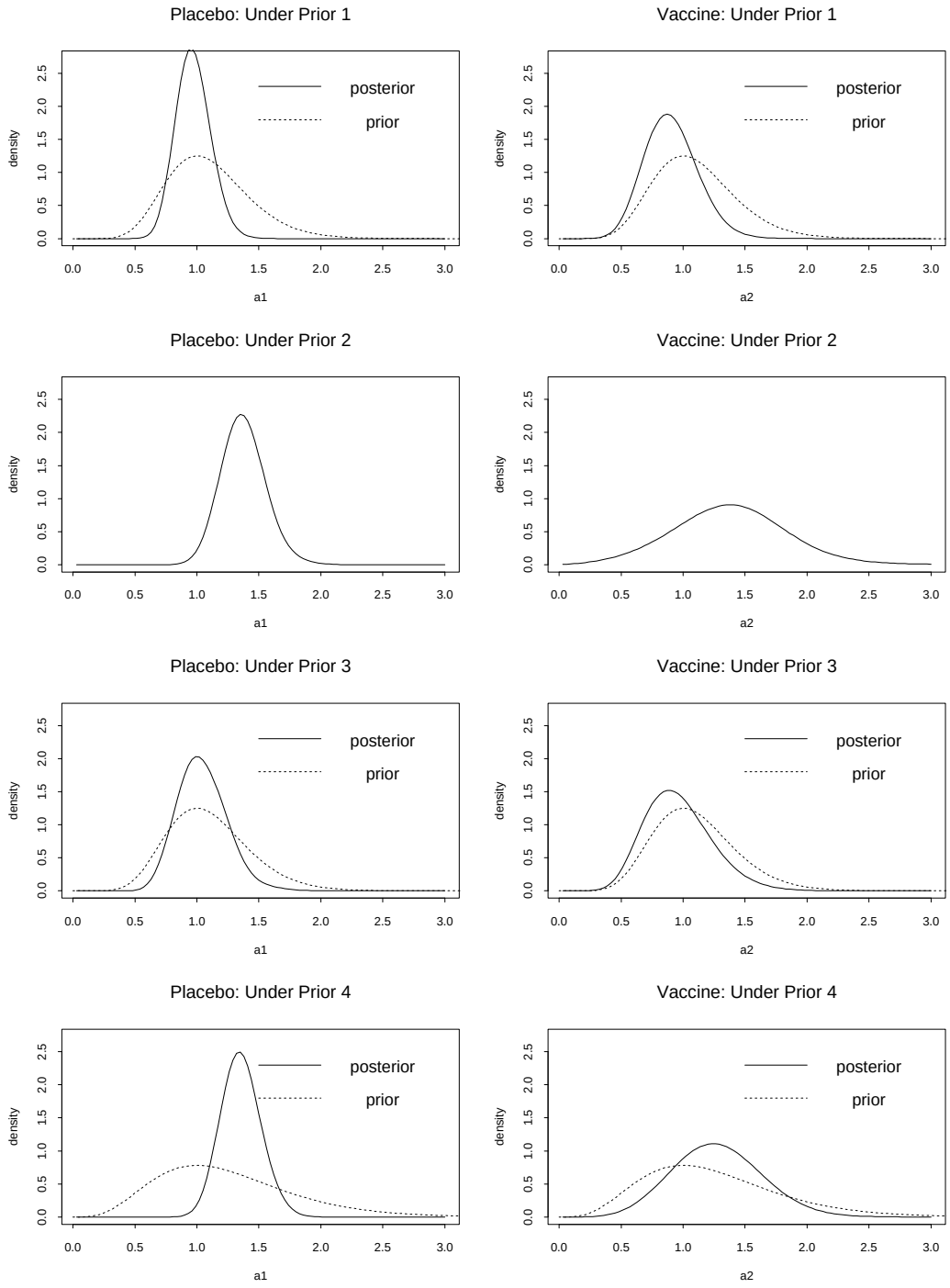

Figure 5.8: Posterior distributions of $a_i, i = 1, 2$ under various priors.

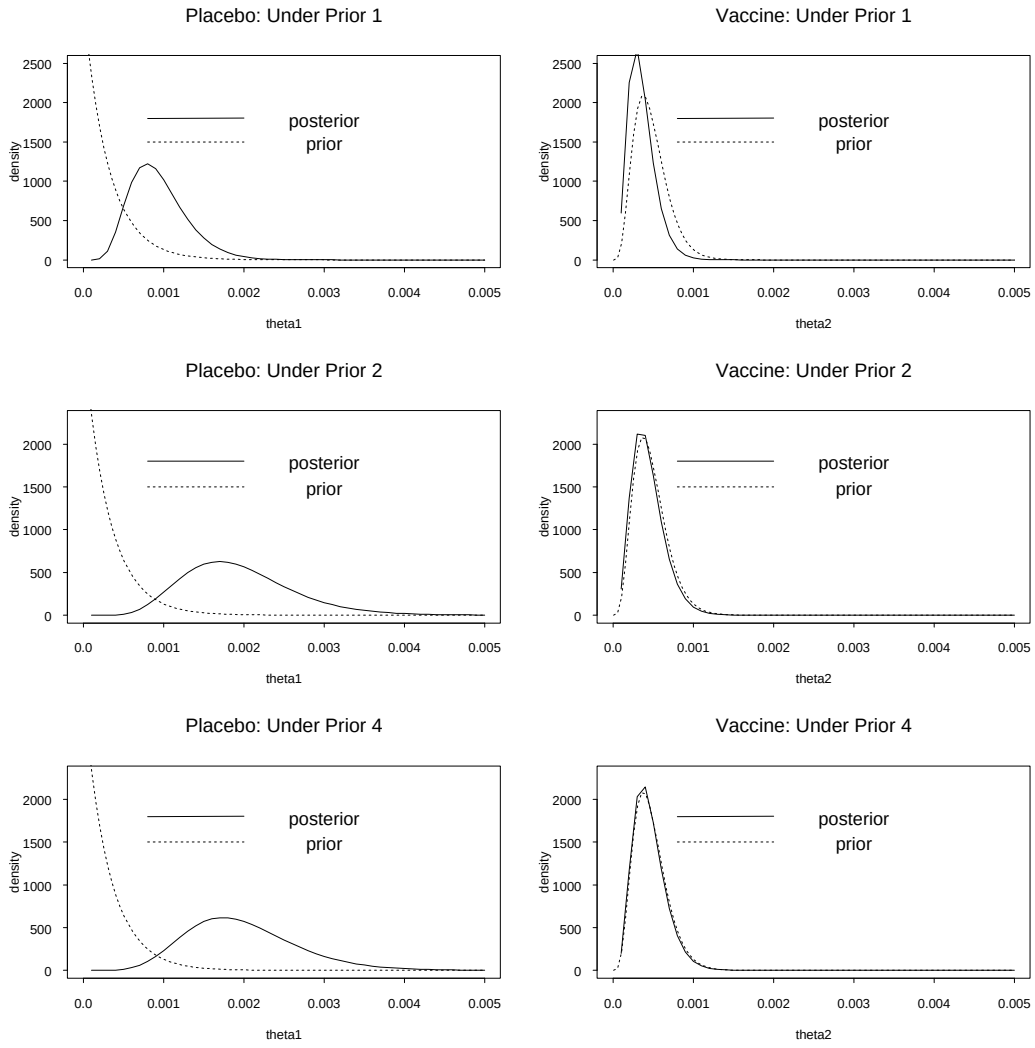

Figure 5.9: Posterior distributions of $\theta_i, i = 1, 2$ under various priors.

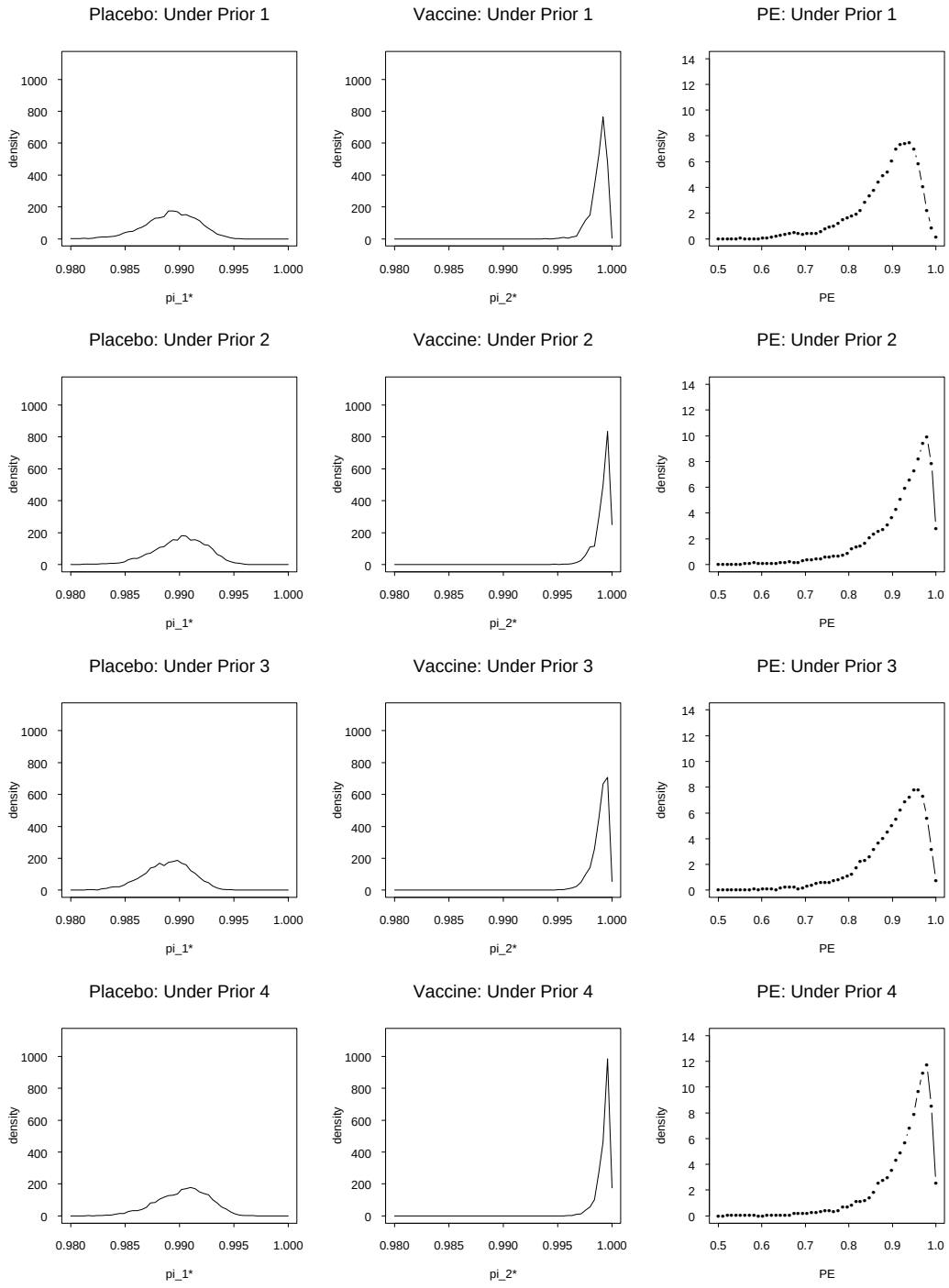

Figure 5.10: The predictive distributions of $\pi_i^*, i = 1, 2$ and PE under various priors.

distributions do not change much. The Monte Carlo estimates of the PE are all well above 90% under any prior choice considered.

Thus, the efficacy of the vaccine seems not very sensitive to the choice of the prior.

5.4 Discussion

The hazard rate for the placebo group increases at first, then decreases to 0, then increases again, and finally decreases again to 0. This is an artifact of small sample. Babies under 10 months old are more susceptible to the disease than babies over 10 months old. Thus, we should not expect that a simple increasing or decreasing rate model will fit the data very well.

In our mixture model case, the results from last section suggest an increasing hazard rate for the placebo group. This rate is determined by the model setting. Thus we may need to analyze the data by using a piecewise Weibull model (assuming one Weibull model for one interval and another Weibull model for another interval).

Chapter 6

Conclusions and Further Work

This dissertation examined a Bayesian Weibull survival analysis for clinical trials and the use of the Gibbs sampling scheme to approximate posterior estimates. Chapter 2 defines and describes some characteristics of the Weibull survival distribution. It is extremely difficult to obtain analytic results of posterior distributions for parameters of Weibull distributions and the predictive measures of treatment comparison, so I made numerical approximations. I used the Gibbs sampling method to solve this problem.

In this dissertation, I introduced a Bayesian Weibull survival model for comparing two treatments in a clinical trial, and performed a numerical analysis of an interim data set. I also introduced a Non-Small Cell Lung Cancer (NSCLC) clinical trial, presented the numerical results of posterior distributions and predictive measures and discussed the sensitivity analysis for the prior distributions.

I developed a mixture model, with a Weibull component and a surviving fraction; and I explored various computational aspects of this model. I then used this model to analyze the data from a *Haemophilus type B* vaccine trial

and to calculate predictive efficacy (PE). A Gibbs sampling scheme, incorporating with the Adaptive Rejection Method, easily overcame some computational difficulties.

I plan to conduct further research in several areas. As I mentioned in Chapter 5, we should not be restricted to using a fixed model when we are dealing with practical problems. It will be interesting to analyze the *Haemophilus type B* vaccine trial data using a piece-wise Weibull model. We have reason to expect that it might fit the hazard rate better than the mixture model. It certainly will be one of my future research topics.

Though I focus on survival data ignoring prognosis of individual patients, the model can be generalized to incorporating the prognosis of patients.

The model can also be extended to deal with multicenter clinical trial data. Although the numbers of parameters will increase a lot, the calculations should be not essentially more difficult.

Moving in a more computational direction, I also shall study how fast and when the Gibbs sampler converges.

DOB	DT_DX	VAC_PLAC	DAYINTOSTD	MNINTOSTD	AGE_DAYS
08/07/88	11/03/88	P	125	4.11	88
07/09/88	12/23/88	P	175	5.75	167
05/15/88	12/31/88	P	183	6.01	230
10/14/88	01/30/89	P	213	7.00	108
01/03/89	05/01/89	P	304	9.99	118
10/04/88	05/03/89	P	306	10.05	211
06/22/88	06/22/89	P	356	11.70	365
03/23/89	10/12/89	P	468	15.38	203
06/29/89	11/25/89	P	512	16.82	149
06/30/89	12/04/89	P	521	17.12	157
08/29/88	12/24/89	V	541	17.77	482
05/17/89	01/09/90	P	557	18.30	237
11/06/89	01/17/90	P	565	18.56	72
11/11/89	02/22/90	P	601	19.75	103
06/11/89	03/10/90	P	617	20.27	272
12/17/88	03/11/90	P	618	20.30	449
04/03/89	04/24/90	P	662	21.75	386
11/25/89	04/25/90	P	663	21.78	151
10/10/89	06/06/90	P	705	23.16	239
02/26/89	06/11/90	P	710	23.33	470
03/08/90	06/12/90	P	711	23.36	96
02/15/90	07/04/90	P	733	24.08	139
02/17/90	07/28/90	P	757	24.87	161

Appendix A

Data from the H.I.B. Vaccine Trial

where "DT_DX" means the date when the baby got the shot, "DOB" means the birth date of the baby, and "V" means vaccine and "P" means placebo. "MNINTOSTD" means months in the study when the baby became a case,

and "AGE_DAYS" means the age (in days) of the baby when he/she became a case.

Appendix B

Proof of Log-Concavity of the Full Conditional $p(a_i|\Theta_{(-a_i)}, \lambda, \mathbf{t})$, $i = 1, 2$

The full conditional distribution of the a_i's in the mixture model is:

$$\begin{aligned} p(a_i|\Theta_{(-a_i)}, \lambda, \mathbf{t}) &\propto {a_i}^{u_a + \sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij})-1} \\ &\quad \exp\left((a_i - 1)\sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij}) log(t_{ij}) - (v_a a_i + \theta_i \sum_{j=1}^{n_i} {t_{ij}}^{a_i}(1-\lambda_{ij})\right) \\ i &= 1, 2. \end{aligned} \tag{B.1}$$

Calculating the log of the above density gives us:

$$\begin{aligned} log\left(p(a_i|\Theta_{(-a_i)}, \lambda, \mathbf{t})\right) &= \left(u_a + \sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij}) - 1\right) * log(a_i) \\ &\quad + (a_i - 1)\sum_{j=1}^{n_i} \delta_{ij}(1-\lambda_{ij}) log(t_{ij}) \\ &\quad - \left(v_a a_i + \theta_i \sum_{j=1}^{n_i} {t_{ij}}^{a_i}(1-\lambda_{ij})\right) \end{aligned} \tag{B.2}$$

Let $h(a_i) = log\left(p(a_i|\Theta_{(-a_i)}, \lambda, \mathbf{t})\right)$, and $i = 1, 2$, then taking the second derivative of $h(a_i)$ with respect to a_i gives us:

$$\begin{aligned}\frac{d^2{}_{h(a_i)}}{d_{a_i^2}} &= -\left(u_a + \sum_{j=1}^{n_i}\delta_{ij}(1-\lambda_{ij}) - 1\right)/a_i^2 - \theta_i \sum_{j=1}^{n_i}(1-\lambda_{ij}){t_{ij}}^{a_i} log(t_{ij})^2 \\ &= -(u_a + d_i - 1)/a_i^2 - \theta_i \sum_{j=1}^{d_i}{t_{ij}}^{a_i} log(t_{ij})^2 \qquad \text{(B.3)}\end{aligned}$$

where $u_a + d_i - 1 > 0$ for sure if we choose prior parameter $u_a > 0$ and $\theta_i > 0$, and there is at least one case (death) on the i-th treatment arm (otherwise, the posterior for a_i should be the same as the prior, because censored observations will not provide any information for estimating the scale parameter a_i). Also $log(t_{ij})^2 > 0$. Thus, if $u_a > 0$, then $\frac{d^2{}_{h(a_i)}}{d_{a_i^2}} < 0$. That means that the $p(a_i|\Theta_{(-a_i)}, \lambda, \mathbf{t})$ is log-concave, and the Adaptive Rejection Method can be used for sampling from $p(a_i|\Theta_{(-a_i)}, \lambda, \mathbf{t})$.

Appendix C

Program for the Mixture Model

```
/*****************************************************************
 *****************************************************************
 *      This is modified on Mar. 18, 1994 for the mixture model, *
 *      the prior is set as:                                     *
 *                                                               *
 *              theta1 ~ Gamma(u_th1, v_th1)                     *
 *              theta2 ~ Gamma(u_th2, v_th2)                     *
 *              a1 = a2 ~ Gamma(u_a2, v_a2)                      *
 *              pi1 = pi2 ~ Beta(alpha, beta)                    *
 *                                        ^^^^^^^^^^             *
 *                                        JIANG QIAN             *
 *                                        ^^^^^^^^^^             *
 *****************************************************************
 *****************************************************************/
#include <math.h>
#include <stdio.h>

#define sq(x)   (x)*(x)
#define MAX     5250
#define NMC     3000    /* number of Monte Carlo samples        */
#define NIN     3000    /* Number of initial cycles             */
#define NZth    50      /* # of grid points for plot of theta[i]*/
```

```
#define NZa     100     /* # of grid points for plot of a[i]    */
#define NZpi    50      /* # of grid points for plot of pi[i]   */
#define ath     0.0     /* plot of theta[i] over (ath, bth)     */
#define bth     0.0050
#define aa      0.0     /* plot of a[i] over (aa, ba)           */
#define ba      3.0
#define api     0.80    /* plot of pi[i] over (api, bpi )       */
#define bpi     1.0

double  pow(double , double );
double  *dvector(int, int);
int     *ivector(int, int);
double  **dmatrix(int, int, int , int);
double  ran1(int *);
double  gasdev(int * );
double  gamdev(double *, int * );
double  betadev(double *, double *);
void    free_dvector(double *, int ,int);
void    free_ivector(int *, int ,int);
void    free_dmatrix(double **, int, int, int , int);
void    nrerror(char *);
double  gammln(double );
void    sort1(int , double *);
void    sort2(int , double *, double *);

void    gibbs(double **, double **, double **, double **,
                double *, double *, double *, double *,
                double , double , double, double, double ,
                double , double , double, double, double, int);
void    sampleA(double *, double *,
                double **, double *, double *, double ,
                double *, double **, double **, int *, int *);
double  h(double , double **, double *, double *,
```

```
                double *, double **, double **, double , int *);
double  hh(double , double **, double *, double *,
                double *, double **, double **, double , int *);
double  h_l(double , double *, double *, int );
double  h_u(double , double *, double *, double *, double *, int );

main()
{
        int     i, j, k, icyc, l, jj, m, mm, flag;
        double  **D, **t, **P, **lambda, **delta,
                *theta, *a, *pi, *x, *Pi1, *Pi2, *PE,
                u_a1, v_a1, u_th1, v_th1,
                u_a2, v_a2, u_th2, v_th2, alpha, beta,
                temp, temp1, temp2, sumdellam1, sumdellam2;
        double  *xth1, *xth2, *pth1, *pth2,
                *xa1, *xa2, *pa1, *pa2, *temppa1, *temppa2,
                *xpi1, *xpi2, *ppi1, *ppi2;
        double  sum1, sum2, a1, a2, b1, b2, h[26], tt[26];

        FILE    *fp;

        mm = 3;
        D       = dmatrix(1, MAX, 1, 3);
        t       = dmatrix(1, 2, 1, MAX/2);
        P       = dmatrix(1, 2, 1, MAX/2);
        lambda  = dmatrix(1, 2, 1, MAX/2);
        delta   = dmatrix(1, 2, 1, MAX/2);
        x       = dvector(1, 3);
        theta   = dvector(1, 2);
        a       = dvector(1, 2);
        pi      = dvector(1, 2);
        Pi1     = dvector(1, NMC);
        Pi2     = dvector(1, NMC);
        PE      = dvector(1, NMC);
```

```
if ( (fp = fopen("cur.dat", "r")) == NULL )
        nrerror("Data file cur.dat not found\n");
for( i=1; i<=MAX; i++ ) {
        fscanf(fp, "%lf\%lf\%lf\n", &D[i][1], &D[i][2], &D[i][3]);
}
/***    Read in the initial 3 points for the Adaptive   ***
 ***    Rejection Sampling drawing here                 ***/
fscanf(fp, "%lf\t%lf\t%lf\n", &x[1], &x[2], &x[3]);
/***    Read in the prior parameters                    ***
 ***    for "(pi_i, a_i, theta_i)" here                 ***/
fscanf(fp, "%lf\t%lf\n", &u_th1, &v_th1);
fscanf(fp, "%lf\t%lf\n", &u_th2, &v_th2);
fscanf(fp, "%lf\t%lf\n", &u_a1, &v_a1);
fscanf(fp, "%lf\t%lf\n", &u_a2, &v_a2);
fscanf(fp, "%lf\t%lf\n", &alpha, &beta);
fclose(fp);
fflush(stdout);

for(i=0; i<26; i++)     {
        tt[i] = i;
        h[i] = 0.0;
}
tt[0] = 0.005;

/***    Set up "delta[i][j] and                 ***
 ***    intial values for "lambda[i][j]"        ***/
for( i=1; i<=MAX/2; i++ )       {
        t[1][i] = D[i][1]-2.;
        if( D[i][3] == 1.0 )    {
                lambda[1][i] = 0;
                delta[1][i] = 1;
        }
```

```
        else    {
                lambda[1][i] = 1;
                delta[1][i] = 0;
        }
        t[2][i] = D[i+MAX/2][1]-2.;
        if( D[i+MAX/2][3]==1 )  {
                lambda[2][i] = 0;
                delta[2][i] = 1;
        }
        else    {
                lambda[2][i] = 1;
                delta[2][i] = 0;
        }
}

/***    set up initial value for Gibbs draw     ***/
theta[1] = 1./3200.;
theta[2] = 5./10700.;
a[1]    = 1.0;
a[2]    = 1.0;
pi[1]   = 0.99;
pi[2]   = 0.99;

printf("NIN=%d\tNMC=%d\n", NIN, NMC);
/*** Now burn-in cycles ***/
for (icyc=1; icyc<=NIN; icyc++){
        gibbs(t,P,lambda, delta, theta, a, pi, x, u_th1, v_th1,
                u_a1, v_a1, u_th2, v_th2, u_a2, v_a2,
                alpha, beta, mm);
}

/*** Now MC cycles      ***/
xth1    = dvector(1, NZth);
xth2    = dvector(1, NZth);
```

```
pth1    = dvector(1, NZth);
pth2    = dvector(1, NZth);
xa1     = dvector(1, NZa);
xa2     = dvector(1, NZa);
pa1     = dvector(1, NZa);
pa2     = dvector(1, NZa);
temppa1 = dvector(1, NZa);
temppa2 = dvector(1, NZa);
xpi1    = dvector(1, NZpi);
xpi2    = dvector(1, NZpi);
ppi1    = dvector(1, NZpi);
ppi2    = dvector(1, NZpi);
for (j=1; j<=NZth; j++){
        xth1[j] = ath + (bth-ath)* (double) (j)/ (double )(NZth);
        xth2[j] = xth1[j];
        pth1[j] = 0.0;          /* pdf for theta[1] */
        pth2[j] = 0.0;          /* pdf for theta[2] */
}
for (j=1; j<=NZa; j++){
        xa1[j] = aa+(ba-aa)* (double) (j)/ (double) (NZa);
        xa2[j] = xa1[j];
        pa1[j] = pa2[j] = 0.0;          /* pdf for a[i] */
}
for (j=1; j<=NZpi; j++){
        xpi1[j] = api +(bpi-api)* (double) (j)/(double) (NZpi);
        xpi2[j] = xpi1[j];
        ppi1[j] = ppi2[j] = 0.0;        /* pdf for pi[i] */
}

for (icyc=1; icyc<=NMC; icyc++){
        gibbs(t,P,lambda, delta, theta, a, pi, x, u_th1, v_th1,
                u_a1, v_a1, u_th2, v_th2, u_a2, v_a2,
```

```
        alpha, beta, mm);

/***    Calculating "pi1, pi2"                ***/
Pi1[icyc] = pi[1] + (1. - pi[1])*
                exp(-theta[1]*pow(16., a[1]));
Pi2[icyc] = pi[2] + (1. - pi[2])*
                exp(-theta[2]*pow(16., a[2]));

/***    Calculating Protective Efficacy PE    ***/
PE[icyc] = (Pi2[icyc] - Pi1[icyc])/(1. - Pi1[icyc]);
printf("%lf\t%lf\t%lf\n", Pi1[icyc], Pi2[icyc], PE[icyc]);

/***    Calculating hazard rate               ***
 ***    for t=1, 2, 3, ..., 26                ***/
for (i=0; i<26; i++)
        h[i] += (1.-pi[1])*theta[1]*a[1]*pow(tt[i],
                a[1]-1.)*exp(-theta[1]*pow(tt[i], a[1])) /
               (pi[1]+(1.-pi[1])
                *exp(-theta[1]*pow(tt[i],a[1])));

/************************************************
 ***    Now posterior distribution of theta[i]  ***
 ***    theta[1] ~Gamma(a1, b1)                 ***
 ***    theta[2] ~Gamma(a2, b2)                 ***
 ************************************************/
/***    calculating constant                    ***
 ***    sum_j delta[i][j]*(1-lambda[i][j])      ***/
sumdellam1 =0.0;
sumdellam2 =0.0;
for(j=1; j<=MAX/2; j++) {
        sumdellam1 += delta[1][j]*(1.-lambda[1][j]);
        sumdellam2 += delta[2][j]*(1.-lambda[2][j]);
}
a1= sumdellam1 + u_th1;
```

```
a2= sumdellam2 + u_th2;
b1=0.0;
b2=0.0;
for(j=1; j<=MAX/2; j++) {
        b1 += pow(t[1][j], a[1])*(1.-lambda[1][j]);
        b2 += pow(t[2][j], a[2])*(1.-lambda[2][j]);
}
b1 +=v_th1;
b2 +=v_th2;
b1 +=v_th1;
b2 +=v_th2;
temp1 = a1*log(b1)-gammln(a1);
temp2 = a2*log(b2)-gammln(a2);
for (j=1; j<=NZth; j++){
        temp = temp1 + (a1-1.0)*log(xth1[j])
               - b1*xth1[j];
        pth1[j] += exp(temp);
        temp = temp2 + (a2-1.0)*log(xth2[j])
               - b2*xth2[j];
        pth2[j] += exp(temp);
}

/*********************************************************
 *   Now posterior distribution of a[i]                  *
 *********************************************************/
a1= sumdellam1+u_a1-1.0;
a2= sumdellam2+u_a2-1.0;
b1=0.0;
b2=0.0;
for(j=1; j<=MAX/2; j++) {
        b1 += log(t[1][j])*delta[1][j]*(1.-lambda[1][j]);
        b2 += log(t[2][j])*delta[2][j]*(1.-lambda[2][j]);
}
```

```
sum1=sum2=0.0;
for (j=1; j<=NZa; j++){
        temp1 = a1*log(xa1[j])+b1*(xa1[j]-1.0);
        temp= v_a1*xa1[j];
        for (i=1; i<=MAX/2; i++){
                temp +=theta[1]*pow(t[1][i], xa1[j])
                        *(1.0-lambda[1][i]);
        }
        temp =temp1 -temp;
        temppa1[j] = exp(temp);
        sum1 += temppa1[j]*(ba-aa)/(NZa+1.0);

        temp2 = a2*log(xa2[j])+b2*(xa2[j]-1.0);
        temp = v_a2*xa2[j];
        for (i=1; i<=MAX/2; i++){
                temp +=theta[2]*pow(t[2][i], xa2[j])
                        *(1.0-lambda[2][i]);
        }
        temp =temp2 -temp;
        temppa2[j] = exp(temp);
        sum2 += temppa2[j]*(ba-aa)/(NZa+1.0);

}
for (j=1; j<=NZa; j++){
        pa1[j] += temppa1[j]/sum1;
        pa2[j] += temppa2[j]/sum2;
}

/*********************************************************
 * Now posterior distribution of pi[i]                   *
 * pi[1] ~ Beta(a1, b1)                                  *
 * pi[2] ~ Beta(a2, b2)                                  *
 *********************************************************/
```

```
        temp1 = 0.0;
        temp2 = 0.0;
        for(j=1; j<=MAX/2; j++) {
                temp1 +=  lambda[1][j];
                temp2 +=  lambda[2][j];
        }
        a1= alpha+temp1;
        b1= beta+MAX/2-temp1;
        a2= alpha+temp2;
        b2 = beta+MAX/2-temp2;
        temp=a1+b1;
        temp1 = gammln(temp)-gammln(a1)-gammln(b1);
        temp= a2+b2;
        temp2= gammln(temp)-gammln(a2)-gammln(b2);
        for (j=1; j<=NZpi; j++){
                temp = (a1-1.0)*log(xpi1[j])+
                        (b1-1.0)*log(1.0-xpi1[j]);
                ppi1[j] += exp(temp+temp1);
                temp = (a2-1.0)*log(xpi2[j])+
                        (b2-1.0)*log(1.0-xpi2[j]);
                ppi2[j] += exp(temp+temp2);
        }
}
printf("\n\n");

/**** Now output ****/
printf ("xth1, pth1, xth2, pth2,for plot of theta[i]\n");
for (j=1; j<=NZth; j++){
        pth1[j] /= NMC;
        pth2[j] /= NMC;
        printf("\%lf\t %lf\t %lf\t %lf\n", xth1[j], pth1[j],
                xth2[j], pth2[j]);
```

```
        }

        printf ("xa1, pa1, xa2, pa2, for plot of a[i]\n");
        for (j=1; j<=NZa; j++){
                pa1[j]  /= NMC;
                pa2[j]  /= NMC;
                printf("\%lf\t %lf\t %lf\t %lf\n", xa1[j], pa1[j],
                        xa2[j], pa2[j]);
                printf("\%lf\t %lf\t %lf\t %lf\n", xpi1[j], ppi1[j],
                        xpi2[j], ppi2[j]);
        }
        printf("\n\n");
        printf("t\t\t h(t)\n");
        for(i=0;i<26;i++){
                h[i] /=NMC;
                printf("%lf\t%lf\n", tt[i], h[i]);
        }
}

void gibbs(double **t, double **P, double **lambda, double **delta,
        double *theta, double *a, double *pi, double *x,
        double u_th1, double v_th1, double u_a1, double v_a1,
        double u_th2, double v_th2, double u_a2, double v_a2,
        double alpha, double beta, int mm)
{
        int     i, j, k, l, jj, m, flag;
        double
                uu1, vv1, uu2, vv2, u, v, w, p, nin1, nin2, PE,
                temp, temp1, temp2;
                /*********************************/
                /***    sample "theta_i" now    ***/
                /*********************************/
                uu1 = u_th1;
```

```
uu2 = u_th2;
temp1 = v_th1;
temp2 = v_th2;
for(j=1; j<=MAX/2; j++) {
        uu1 += delta[1][j]*(1.-lambda[1][j]);
        uu2 += delta[2][j]*(1.-lambda[2][j]);
        temp1 += pow(t[1][j], a[1])*(1.-lambda[1][j]);
        temp2 += pow(t[2][j], a[2])*(1.-lambda[2][j]);
}
jj = 1;
theta[1] = gamdev(&uu1, &jj)/temp1;
jj += 1;
theta[2] = gamdev(&uu2, &jj)/temp2;

/*********************************/
/***    sample "a_i" now        ***/
/*********************************/
uu1 = uu1 - u_th1 + u_a1;
uu2 = uu2 - u_th2 + u_a2;
flag = 1;
sampleA(&a[1], x, t, &u_a1, &v_a1, uu1,
        &theta[1], delta, lambda, &flag, &mm);
flag = 2;
sampleA(&a[2], x, t, &u_a2, &v_a2, uu2,
        &theta[2], delta, lambda, &flag, &mm);

/*********************************/
/***    sample "pi" now         ***/
/*********************************/
temp1 = 0;
temp2 = 0;
for(j=1; j<=MAX/2; j++) {
        temp1 +=  lambda[1][j];
```

```
        temp2 +=  lambda[2][j];
}
uu1 = alpha+temp1;
vv1 = beta+MAX/2-temp1;
jj += 1;
u = gamdev(&uu1, &jj);
jj += 1;
v = gamdev(&vv1, &jj);
pi[1] = u/(u+v);

uu2 = alpha+temp2;
vv2 = beta+MAX/2-temp2;
jj += 1;
u = gamdev(&uu2, &jj);
jj += 1;
v = gamdev(&vv2, &jj);
pi[2] = u/(u+v);

/************************************************/
/***    sample "lambda_[ij]" now                ***/
/************************************************/
jj = 1;
for (i=1; i<=2; i++)    {
        for(j=1; j<=MAX/2; j++) {
                if( delta[i][j]==1 )
                        P[i][j]=0.;
                else    {
                        p = pi[i];
                        P[i][j] = p/(p+(1.-p)*
                           exp(-theta[i]*pow(t[i][j],a[i])));

                }
                lambda[i][j] = 0.0;
```

```
                                u = ran1(&jj);
                                if ( u<=P[i][j] )       lambda[i][j] += 1;
                                jj += 1;
                        }
                }

}

void sampleA(double *a, double *x,
        double **t, double *u, double *v, double uu,
        double *theta, double **delta,
                double **lambda, int *flag, int *mm)
{
        int     i, jj, k, m, MAXm;
        double  *xsort, *ysort, y1, y2, *w, *z, *yz,
                U, Cu, X, temp1, temp2, temp;

        m = *mm;
        MAXm=50;                /* memory allocation for xsort, etc */
        xsort   = dvector(1, MAXm); /* give more spaces at one time */
        ysort   = dvector(1, MAXm); /* give more spaces at one time */
        z       = dvector(0, MAXm); /* more spaces at one time */
        yz      = dvector(0, MAXm); /* more spaces at one time */
        w       = dvector(1, MAXm); /* more spaces at one time */

        /***    allocate dummy vector   ***/
        for(i=1; i<=m; i++)     {
                xsort[i] = x[i];
        }
hate:   sort1(m, xsort);
        for(i=1; i<=m; i++)             {
                ysort[i] = h(xsort[i], t, u, v, theta, delta, lambda, uu,
                                flag);
                w[i] = hh(xsort[i], t, u, v, theta, delta, lambda, uu,
```

```
                        flag);
}

z[0] = 0.00001;
z[m] = 15.;
yz[0] = h(z[0], t, u, v, theta, delta, lambda, uu, flag);
yz[m] = h(z[m], t, u, v, theta, delta, lambda, uu, flag);
for(i=1; i<m; i++)      {
        y1 = h(xsort[i], t, u, v, theta, delta, lambda, uu, flag);
        y2 = h(xsort[i+1], t, u, v, theta, delta, lambda, uu,
                         flag);
        z[i] = xsort[i] + (y1 - y2 + w[i+1]*(xsort[i+1]-xsort[i]))
               /(w[i+1]-w[i]);
        yz[i] = w[i]*(z[i] - xsort[i]) + y1;
}

Cu = 0.0;
for (i=0; i<m; i++)     {
        Cu += (exp(yz[i+1]) - exp(yz[i]))/w[i+1];
}
/***    Find the LARGEST z(i) S.T. "temp<U"     ***/
jj = 1;
U = ran1(&jj);
temp = 0.0;
i = 0;
do      {
        temp += (exp(yz[i+1]) - exp(yz[i])) / w[i+1] ;
        i += 1;
}       while( temp/Cu < U);
i -= 1;
temp -= (exp(yz[i+1]) - exp(yz[i])) / w[i+1] ;
temp /= Cu;

/***    Sample X from upper hull s(x)   ***/
```

```
        X = z[i] + log(1. + w[i+1] * Cu * (U - temp)/exp(yz[i]) )/w[i+1];
        if( X < 0.0 ) goto hate;
        jj += 1;
        U = ran1(&jj);

        temp1 = h_l(X, xsort, ysort, m);
        temp2 = h_u(X, xsort, w, z, yz, m);
        if ( U < exp(temp1 - temp2) && X > 0.0 )        *a = X;
        else    {
                temp1 = h(X, t, u, v, theta, delta, lambda, uu, flag);
                if ( U < exp(temp1 - temp2) && X > 0.0 )    *a = X;
                else    {
                        xsort[m+1] = X;
                        ysort[m+1] =
                                h(X, t, u, v, theta, delta, lambda, uu,
                                        flag);
                        m += 1;
                        if ( m > MAXm ) puts("warning m > MAXm");
                        goto hate;
              }
        }
        free_dvector(xsort, 1, MAXm);
        free_dvector(ysort, 1, MAXm);
        free_dvector(z, 0, MAXm);
        free_dvector(yz, 0, MAXm);
        free_dvector(w, 1, MAXm);
}

double  h(double a, double **t, double *u, double *v,
                double *theta, double **delta,
                double **lambda, double uu, int *flag)
{
        int     i, j;
```

```
        double  temp1, prod, temp;

        if( *flag==1 )  i=1;
        else    i=2;

        temp1 = *v * a;
        prod = 0.0;
        for( j=1; j<=MAX/2; j++ )       {
                temp1 += *theta * (1. - lambda[i][j]) * pow(t[i][j], a);
                if( lambda[i][j] == 0 ) prod += delta[i][j] * (1. -
                                        lambda[i][j]) * log(t[i][j]);
        }
        temp = (uu-1.) * log(a) + (a-1) * prod - temp1;
        return temp;
}

double hh(double a, double **t, double *u, double *v,
                double *theta, double **delta,
                double **lambda, double uu, int *flag)
{
        int i, j;
        double temp1, prod, temp;

        if (*flag==1)   i=1;
        else    i=2;

        temp1 = *v;
        prod = 0.0;
        for( j=1; j<=MAX/2; j++ )       {
                if( lambda[i][j] != 1.0 )       {
                        temp1 += *theta * (1.0 - lambda[i][j])
                                * pow(t[i][j], a)
                                * log(t[i][j]);
                        prod += delta[i][j] * (1. -
```

```
                                        lambda[i][j]) * log(t[i][j]);
                }
        }
        temp = (uu-1.)/a + prod - temp1;
        return temp;
}

double h_l(double X, double *xx, double *yy, int mm)
{
        int i, m, k;
        double temp;

        m = mm;
        for (i=1; i<m; i++)     {
                if( xx[i] < X && X <= xx[i+1] )         {
                        temp = yy[i] + (X - xx[i]) *
                                (yy[i+1] - yy[i])/(xx[i+1] - xx[i]);
                }
        }
        if( X < xx[1] ) temp = yy[1] + (X - xx[1]) *
                                (yy[2]-yy[1])/(xx[2]-xx[1]);
        if( X > xx[m] ) temp = yy[m-1] +  (X - xx[m-1]) *
                                (yy[m]-yy[m-1])/(xx[m]-xx[m-1]);
        return temp;
}

double h_u(double X, double *xsort, double *w, double *z, double *yz,
        int mm)
{
        int i, m, k;
        double temp;

        m = mm;
        for(i=0; i<=m; i++)     {
                if (z[i] < X && X <= z[i+1]){
```

```
                    temp = yz[i] + w[i+1]*(X-z[i]);
                }
        }
        return temp;
}
```

Bibliography

[1] James O. Berger (1985), *Statistical Decision Theory and Bayesian Analysis,* New York: Springer-Verlag.

[2] L. M. Berliner and B. M. Hill (1988), Bayesian nonparametric survival analysis, *Journal of the American Statistical Association* **83**, 772-781.

[3] D. A. Berry (1985), Interim Analysis in Clinical Trials: Classical vs. Bayesian Approaches, *Statistics in Medicine* **4**, 521-526.

[4] D. A. Berry (1987), Interim Analysis in Clinical Trials: The Role of the Likelihood Principle, *American Statistician* **41**, 117-122.

[5] D. A. Berry (1988), Interim Analysis in Clinical Research, *Cancer Investigation* **5**, 469-477.

[6] D. A. Berry (1989), Monitoring Accumulating Data in a Clinical Trial, *Biometrics* **45**, 1197-1211.

[7] D. A. Berry (1989b), Ethics and ECMO: Comment on a Paper by Ware, *Statistical Science* **4**, 306-310.

[8] D. A. Berry, Mark C. Wolff, and David Sack (1992), Public Health Decision Making: A Sequential Vaccine Trial, *Bayesian Statistics* **4**, 79-96.

[9] D. A. Berry, S. M. Berry, and K. Gillingham (1993), Bayesian Meta-nalysis for Treatment Comparisons: Dichotomous Responses, *in process*.

[10] BUGS (1993), Bayesian Inference Using Gibbs Sampling Version 1.0 (ed. by D. Spiegelhalter, A. Thomas and W. Gilks).

[11] G. Cao (1993) Bayesian Nonparametric Mixture Modeling, *Ph.D. Thesis*, ISDS, Duke University.

[12] W. C. Chen, B. M. Hill, J. B. Greenhouse and J. V. Fayos (1985) Bayesian Analysis of Survival Curves for Cancer Patients Following Treatment, *Bayesian Statistics 2* (ED. by J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), 299-328.

[13] D. G. Clayton (1991), A Monte Carlo Method for Bayesian Inference in Frailty Models, *Biometrics* **47**, 467-485.

[14] D. R. Cox and D. Oakes (1984), *Analysis of Survival Data*, London: Chapman and Hall.

[15] M. J. Crowder, A. C. Kimber, R. L. Smith and T. J. Sweeting (1991), *Statistical Analysis of Reliability Data*, Chapman and Hall.

[16] P. Dellaportas and A. F. M. Smith (1993) Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling, *Applied Statistics* **42**, 443-459.

[17] L. Devroye (1986), *Non-uniform Random Variate Generation,* New York: Springer-Verlag.

[18] R. O. Dillman, S. L. Seagren, K. J. Propert, J. Guerra, W. L. Eaton, M. C. Perry, R. W. Carey, E. F. Frei, III, and M. R. Green (1990), A Randomized Trial of Induction Chemotherapy Plus High-dose Radiation Versus Radiation Alone in Stage III Non-Small-Cell Lung Cancer, *The New England Journal of Medicine* **323**, 940-945.

[19] Elandt-Johnson, R. C. and Johnson, N. L. (1980), *Survival Models and Data Analysis*, New York: Wiley.

[20] M. D. Escobar and M. West (1991) Bayesian Density Estimation and Inference Using Mixtures. *To appear Journal of American Statistical Association.*

[21] D. Gamerman (1987), Dynamic inference on survival function, *Probability and Bayesian Statistics* (ed. R. Viertl), 183-192. New York: Plenum.

[22] D. Gamerman (1991), Dynamic Bayesian models for survival data, *Applied Statistics* **40**, 63-79.

[23] D. Gamerman and M. West (1987), An application of dynamic survival models in unemployment studies, *The Statistician* **36**, 269-274.

[24] D. Gamerman, M. West and A. Pole (1987), A guide to SURVIVAL. *Warwick Research Report* **121**, University of Warwick, Coventry.

[25] A. E. Gelfand and A. F. M. Smith (1990), Sampling Based Approaches to Calculate Marginal Densities, *Journal of the American Statistical Association* **85**, 398-409.

[26] S. Geman and D. Geman (1984), Stochastic Relation, Gibbs Distributions

and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.

[27] S. George, C. Li, and D. A. Berry (1993) Stopping a Clinical Trial Early: Frequentist and Bayesian Approaches Applied to a CALGB Trial in NON-Small Cell Lung Cancer, to appear in *Statistics in Medicine.*

[28] W. R. Gilks (1992), Derivative-free adaptive rejection sampling for Gibbs sampling, *Bayesian Statistics 4*, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), 641-665. Clarendon Press, Oxford, UK.

[29] W. R. Gilks and P. Wild (1992), Adaptive Rejection Sampling for Gibbs Sampling, *Applied Statistics* **41**, 337-348.

[30] J. Geweke (1989) Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrika* **57**, 97-109.

[31] A. P. Grieve (1987) Applications of Bayesian software: two examples, *The Statistician* **36** 283-288.

[32] A. P. Grieve (1988) A Bayesian approach to the analysis of LD50 experiments, *Bayesian Statistics* **3** (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), 631-640.

[33] B. M. Hill (1992) Bayesian nonparametric survival analysis: a comparison of the Kaplan-Meier and Berliner-Hill estimators, *Survival Analysis: State of the Art* (ed. J. P. Klein and P. K. Goel), 25-46.

[34] K. K. G. Lan and D. L. Demets (1983), Discrete Sequential Boundaries for Clinical Trials, *Biometrika* **70**, 659-663.

[35] J. D. Kalbfleisch and R. L. Prentice (1980) *The Statistical Analysis of Failure Time Data*, New York : John Wiley & Sons.

[36] E. L. Kaplan and P. Meier (1958), Nonparametric Estimation from Incomplete Observations, *Journals of American Statistical Association* **53**, 457-481.

[37] L. Kuo and A. F. M. Smith (1992), Bayesian computations in survival models via the Gibbs sampler, *Survival Analysis: State of the Art*,(ed. J. P. Klein and P. K. Goel), 11-24.

[38] M. Lavine and M. West (1992) A Bayesian Method for Classification and Discrimination, *The Canadian Journal of Statistics* **20**, 451-461.

[39] J. F. Lawless (1982), *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

[40] Zhengning Lin (1993), Statistical Method in Combining Historical and Current Clinical Trial Data, *Ph.D. Thesis*, Duke University.

[41] D. V. Lindley and A. F. M. Smith (1972), Bayes Estimates for the Linear Model, *Journal of the Royal Statistical Society, B* **34**, 1-41.

[42] N. R. Mann, R. E. Schafer and N. D. Singpurwalla (1974), *Methods for Statistical Analysis of Reliability and Lifetime Data*, Wiley, New York.

[43] N. Mantel (1966), Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration, *Cancer Chemother. Rep.*, **50**, 163-170.

[44] N. Mantel and W. Haenszel (1959), Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease, *J. Nat. Cancer Inst.,* **22**, 719-748.

[45] H. F. Martz and B. S. Duran (1985), *Bayesian Reliability Analysis,* Wiley, New York.

[46] P. Müller (1991) A Generic Approach to Posterior Integration and Gibbs Sampling, preprint, Duke University.

[47] J. C. Naylor and A. F. M. Smith (1982), Applications of a Method for the Efficient Computation of Posterior Distributions, *Applied Statistics* **31**, 214-225.

[48] W. B. Nelson (1982) *Applied Life Data Analysis.* Wiley, New York.

[49] P. C. O'Brien and T. R. Fleming (1979) A Multiple Testing Procedure for Clinical Trials, *Biometrics* **35**, 549-556.

[50] C. A. Perez, T. F. Pajak, P. Rubin, et al. (1987), Long-term Observations of the Patterns of Failure in Patients with Unresectable Non-oat Cell Carcinoma of the Lung Treated with Definitive Radiotherapy, *Report by the Radiation Therapy Oncology Group*, *Cancer*, **59**, 1874-1881.

[51] K. J. Propert and K. Kim (1992), Group Sequential Methods in Multi-institutional Cancer Clinical Trials, *Mercel Dekker, Inc.*, 133-153.

[52] G. O. Roberts (1992), Convergence Diagnostics of the Gibbs Sampler. *Bayesian Statistics*, **4**, 775-782.

[53] D. K. Stangl (1991), Modeling Heterogeneity in Multi-center Clini.al Trials Using Bayesian Hierarchical Survival Models, *Ph.D. Thesis*, Carnegie Mellon University.

[54] L. Tierney, R. Kass, and J. B. Kadane (1986), Approximation of Expectations and Variances Using Laplace's Method. Technical Report, Department of Statistics, Carnegie Mellon University.

[55] L. Tierney (1990), *LISP-STAT*, New York: John Wiley & Sons.

[56] L. Tierney (1991), Markov Chains for Exploring Posterior Distributions, *Technical Report* **560**, University of Minnesota.

[57] J. H. Ware (1989), Investigating Therapies of Potentially Great Benefit: ECMO (with Discussion), *Statistical Science*, Vol. 4, No. 4, 298-340.

[58] M. West (1992), Modelling Time-Varying Hazards and Covariate Effects (with discussion) , *Survival Analysis: State of the Art*(ed. by J. P. Klein and P. K. Goel) 47-60.

[59] M. West (1992), Modelling with mixtures (with discussion), In *Bayesian Statistics* **4** (J. O. Berger, J. M. Bernardo, A. P. David and A. F. M. Smith Eds) Oxford University Press, Oxford.

[60] M. West, P. J. Harrison and H. S. Migon (1985), Dynamic Generalized Linear Models and Bayesian Forecasting (with discussion), Journal of the American Statistical Association

[61] N. Wetmore, D. McEwen, M. O'Connor, and R. H. Bartlett (1979), Defining Indications for Artificial Organ Support in Respiratory Failure, *Trans. Am. Soc. Artif. Intern. Organs* **25**, 459-461.

[62] P. Wild and W. R. Gilks (1993) Adaptive Rejection Sampling from Log-concave Density Functions, *Applied Statistics* **42**, 701-709.

[63] *SAS/STAT User's Guide, Version 6,* Fourth Edition, Cary, N.C.: SAS Institute Inc.

Biography

Jiang Qian was born on September 13, 1966, in Jiangsu, China. In 1985, she received her Bachelor of Science degree in Applied Mathematics from Southeast University (formerly Nanjing Institute of Technology) in Nanjing, China; and in 1988 she earned her Master of Science degree in Statistics from the same university. From 1988 to 1990 Jiang Qian taught Statistics in the Department of Mathematics, Southeast University. In January, 1991, she entered the Graduate School of Duke University in Durham, North Carolina, to pursue a Doctor of Philosophy degree in Statistics.