Comparing Survival Data For Two Therapies: Nonhierarchical And Hierarchical Bayesian Approaches

Article · March 1996

Source: CiteSeer

6 authors, including:

Donald A Berry
University of Texas MD Anderson Cancer Center
392 PUBLICATIONS **31,860** CITATIONS
SEE PROFILE

Robert L. Wolpert
Duke University
157 PUBLICATIONS **4,195** CITATIONS
SEE PROFILE

Dalene Stangl
Duke University
95 PUBLICATIONS **5,695** CITATIONS
SEE PROFILE

COMPARING SURVIVAL DATA FOR TWO THERAPIES: NONHIERARCHICAL AND HIERARCHICAL BAYESIAN APPROACHES

by

Chengchang Li

Department of Statistics
Duke University

Date: _____
Approved:

_____
Donald A. Berry, Supervisor

_____
Robert L. Wolpert

_____
Donald S. Burdick

_____
Giovanni Parmigiani

_____
Dalene Stangl

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Statistics
in the Graduate School of
Duke University

1994

ABSTRACT

(Statistics)

COMPARING SURVIVAL DATA FOR TWO THERAPIES: NONHIERARCHICAL AND HIERARCHICAL BAYESIAN APPROACHES

by

Chengchang Li

Department of Statistics
Duke University

Date: _____
Approved:

_____
Donald A. Berry, Supervisor

_____
Robert L. Wolpert

_____
Donald S. Burdick

_____
Giovanni Parmigiani

_____
Dalene Stangl

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

1994

Abstract

The problem of comparing two therapies with survival data is considered from a Bayesian point of view. Survival times on each therapy are assumed to have an exponential distribution. The posterior distribution of the log hazard ratio of the experimental therapy to the standard therapy is the basis of inference. Two models are proposed in this dissertation. The first assumes center homogeneity and the second uses a Bayesian hierarchical model for heterogeneity of therapy effects among different centers in a multicenter trial. Center heterogeneity in a multicenter trial is explored in terms of the posterior distributions of the first and second stage parameters as well as the predictive distributions of survival time on each therapy at every center. Posterior distributions of parameters in the first model are derived by doing numerical integration on univariate functions. Posterior distributions of parameters in the second model are derived by using a sampling based algorithm, called Gibbs sampling.

Sensitivity of results to the prior belief is examined by doing the analysis on some different prior distributions. In the first model, stress is given to the prior variance of the log hazard ratio. In the second model, stress is given to the prior belief of the center heterogeneity.

Two clinical trials are analyzed in this dissertation as examples. One is a

phase III clinical trial conducted by Cancer and Leukemia Group B to test two therapies for treatment of patients with stage III non-small cell lung cancer. The other is a NIMH-PRB Collaborative Study of Long-Term Maintance Drug Therapy in Recurrent Affective Illness.

Acknowledgements

To my advisor, Professor Donald A. Berry, I would like to express my deepest gratitude and appreciation for his guidance, advice and support that I have received throughout. His advice and encouragement was very important in my development as a statistician. He provided direction and many editorial suggestions in writing my dissertation, and reassured my progress during times of self-doubt. I am very grateful for having the opportunity to work with him.

I would like to thank the members of my committee, Professor Robert L. Wolpert, Professor Donald S. Burdick, Professor Giovanni Parmigiani and Professor Dalene Stangl, for their substantive advice and encouragement, and for their comments and criticism of my dissertation, and for their help in my development as a statistician. Special note of thanks are due to Professor Robert L. Wolpert for all his guidance and support in my research at Duke. Professor Dalene Stangl deserves special thanks for her suggestions and comments that not only improved the quality of my dissertation, but also helpful in my future research. I am thankful to Professor Steve MacEachern, who is visiting from Ohio State University, for his helpful suggestions and comments on my dissertation.

I would like to extend my thanks to all the faculty and staff of the Insti-

tute of Statistics and Decision Sciences. Their support made my stay here so pleasant.

Professor Stephen L. George at Duke University Medical Center deserves my thanks. He provided access to the data set of the Cancer and Leukemia Group B trial in this dissertation. I also wish to acknowledge the National Institute of Mental Health for providing access to the data set of their trial that serves as an example in this dissertation. I also want to thank all my fellow graduate students and friends at Duke University, for their friendship, and for their much valued assistance and encouragement.

Finally, my special gratitude goes to my wife, Yunfang Gao, for her love and support during the whole process of my dissertation research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Statement of the Problem

Clinical trials are designed to assess the efficacy of therapies. Usually, data from patients treated with experimental therapies are compared to data from patients treated with standard therapies or placebos. In many clinical trials, the primary endpoint occurs when a patient dies or when some other specified event happens. In this dissertation, we are interested in the length of time that it takes for a particular event to occur in the individual patient. If the event is death, then it is the patient's survival time that interests us. In these kinds of clinical trials, some individuals do not experience the event during their observation period, so we can only observe them during the time that they stay in the trial. The data on such individuals are said to be right censored. Right censoring occurs when the individual under observation has not experienced the event at the end of the study when the statistical analysis is performed, when the individual withdraws from the study, or when the individual moves and is lost to follow up.

Clinical trials may be conducted in one or more medical centers. Although

it may be reasonable to assume that all the patients in a single center are exchangeable, patients at different centers may not be.

In clinical trials conducted at different medical centers, there may be differences in inclusion and exclusion criteria or in the way such criteria are applied. For example, a patient with a large tumor may be admitted to a cancer trial at one medical center, but not at another.

Many clinical trials are multicenter trials. Conducting these trials at more than one center allows researchers to evaluate the efficacy of a therapy in a variety of patients and centers. Moreover, it might be difficult or even impossible for a single center to recruit the required number of patients in a given time period, especially for some rare diseases.

It is important for us to recognize the differences between centers when we analyze data from a multicenter trial, especially when the results vary substantially from center to center. In modeling this kind of data, we can not assume that the observations from all the centers are exchangeable. Fleiss (1986) discussed the controversy about the appropriate analysis of data from a multicenter clinical trial. He posited that pooling in the sense of averaging within-clinic differences is almost always justified, and pooling in the sense of throwing together all the data is only rarely justified.

One way to pool the data from all these different centers is to use a hierarchical model, with the idea that these centers are a random sample from a population of centers. (This is a random-effects as opposed to a fixed-effects model.) Such a hierarchical structure takes advantage of the information from all centers to estimate the efficacy of therapy in individual centers and the center population. In this dissertation, I will propose a full Bayesian model for

analyzing a clinical trial without center differences and a full Bayesian hierarchical model for analyzing a multicenter clinical trial with center differences. The models will be applied to some clinical trial data.

1.2 Literature Review

For comparing survival distribution functions, Mantel (1966) first proposed the log rank test based on the work of Mantel and Haenszel (1959). The test is a nonparametric method of comparing survival distribution functions, it has been a commonly used classical method in survival analysis.

One of the earliest works with Bayesian hierarchical models is that of Lindley and Smith (1972). They analyzed the usual linear regression model, using a Bayesian hierarchical model. In the first stage of the hierarchy, observations are assumed to have independent normal distribution, given first stage parameters. Moreover there is a linear regression relationship between the observations and the first stage parameters. In the second stage of the hierarchy, the first stage parameters are assumed to have independent normal distributions, given second stage parameters; and there is a linear regression relationship between the second-stage parameters and the first-stage parameters. In the third and final stage of the hierarchy, the relationship between the second-stage parameters and the third-stage parameters is also a linear regression. Lindley and Smith derived the posterior distribution of the first-stage parameters analytically, and they used the model in some examples.

Chakrovorti and Grizzle (1975) proposed a mixed-effects model to analyze data from multicenter experiments for a continuous treatment response. Using a normal linear model, they treated treatment effects as fixed factors; they

assumed that clinic effects and clinic by treatment effects were random. Maximum likelihood estimates and likelihood ratio statistics were used as the basis of inference.

Hierarchical models have been used to analyze clinical trial data for many years. Skene and Wakefield (1990) presented an application for discrete data, and DuMouchel (1990) presented an application for normal data. Berry and Berry (1993) used a Bayesian hierarchical model in a multi-study of binary data. In the first stage, they assumed that the distribution of the observations in each study was binomial. In the second stage, they assumed that the success probabilities in each study were exchangeable and had a beta distribution with unknown parameters. In the third stage, the parameters in the beta distribution were assumed to have a known prior distribution. Berry and Berry derived the posterior distributions of the success probabilities in each study and the parameters in the beta distribution. These distributions provided information about each individual study and the heterogeneity between the studies.

In recent years, researchers have developed applications for continuous but not normal distributed data. Perhaps the first work applying hierarchical models to continuous time survival data is that of Clayton (1991). He used a Bayesian hierarchical model termed a "Bayesian Frailty Model" to model the heterogeneity between subgroups in the proportional hazards model. In the proportional hazards model, the hazard function for an individual characterized by covariate vector z is denoted

$$\lambda(t|z) = \lambda_0(t)exp(\beta^T z).$$

The baseline hazard function, $\lambda_0(t)$, is modeled nonparametrically, whereas the second factor on the right-hand side is modeled parametrically.

Stangl (1991) used an exponential-gamma hierarchical model to analyze continuous time survival data in a multicenter study aimed at comparing two interventions. In the first stage, she assumed that survival time under one intervention in each center had an exponential distribution. In the second stage, Stangl assumed that the hazard rates from all centers in each intervention were exchangeable gamma random variables, and that the hazard rates for different treatments were independent. She applied both empirical and nonempirical Bayesian methods to the model. Stangl extended the general exponential distribution to mixture exponential and changepoint exponential distributions for survival time in the first stage in her other models, and she compared those different models by applying them to a multicenter trial data set.

Robert Gray (1993) used a Bayesian method to investigate the amount of center variation in a multicenter clinical trial with a censored failure time endpoint. He used a hierarchical structure to model the center effects in a proportional hazards model. In the first stage, he modeled failure time as a piecewise exponential distribution. In the second stage, he assumed that the constant hazard rates in each interval were lognormally distributed, and he took some covariates into account by using the Cox proportional hazard model. The jumps of log hazard rates in each interval are assumed to be exchangeable and normally distributed, and so are the log hazard ratios for the treatments and the coefficients of the covariates in the Cox proportional model. Gray then derived the numerical results of the posterior distributions.

Another Bayesian hierarchical model for a multicenter trial is proposed in this dissertation. In the first stage, exponential distributions are assumed for survival times. In the second stage, the hazard rate of the standard therapy and the log hazard ratio of the experimental to the standard therapy at differ-

ent centers are assumed to be samples from a center population. This model connects different centers and the two therapies together, the inference on each therapy will use the information on the therapy as well as the information on the other therapy at all centers. Modelling the log hazard ratios in the second stage of the hierarchy makes the inference about them in the center population very straightforward.

1.3 Outline of Dissertation

In Chapter 2 of this dissertation, I will introduce a general Bayesian model for comparing two therapies in a clinical trial that was conducted in one medical center, or that was conducted in multicenters but ignored possible heterogeneity between different centers. Survival time on each therapy will be modeled with an exponential distribution. I will model the log hazard ratio directly. I will present general procedures for deriving posterior distribution and predictive distribution.

In Chapter 3, a phase III clinical trial of non-small cell lung cancer (NSCLC) will be introduced, this study was conducted by the Cancer and Leukemia Group B (CALGB). I will apply the model proposed in Chapter 2 to this trial, present the numerical results of the posterior distributions and the predictive distributions, and discuss the sensitivity to prior distributions.

In Chapter 4, a general Bayesian hierarchical model for modeling heterogeneity between centers in a multicenter clinical will be introduced. I assume that the survival time on each therapy at every center has an exponential distribution, and that the hazard rate of the standard therapy and the log hazard ratio of the experimental therapy to the standard therapy in each center are

a sample from a larger population. I will present the general procedures of deriving posterior distribution of the parameters in each individual center and the population. The Gibbs sampling technique will be applied for getting the posterior distributions.

In Chapter 5, the hierarchical model proposed in chapter 4 will be applied to the data from the NSCLC trial sponsored by CALGB. I will model the heterogeneity between different centers and estimate the parameters in each individual center and the center population. I will present the results derived from different prior distributions to explore the impact of prior distribution on posterior distribution.

In Chapter 6, the model in chapter 4 will be applied to a NIMH-PRB Collaborative Study of Long-Term Maintenance Drug Therapy in a Recurrent Affective Illness. I will present the numerical results of the posterior distributions of the parameters in each individual center and the center population. I will also present the predictive survival functions for each therapy in all individual centers. Finally, to check the sensitivity to prior distributions, I will present the results of some different prior distributions.

Chapter 2

A Bayesian Model for Comparing Two Therapies

2.1 Introduction

This chapter consists of a Bayesian model designed without concern for differences between centers. This model can be applied to a single-center clinical trial or a multicenter clinical trial in which the number of patients in the individual centers is small. How long a patient survives on a therapy or how long it takes for a disease to recur is of major interest in this kind of trial. Quite often, some observations will be right censored. The main goal of the trial is to compare the efficacy of two therapies, recognizing that often the efficacy of a given therapy is measured by how long a patient can survive on that therapy.

2.2 Statistical Model

It is assumed that survival times on each therapy are exchangeable with an exponential distribution, and that each therapy has a constant hazard rate. Let λ_1 and λ_2 be the hazard rates for the standard therapy and the experimental

therapy respectively. The hazard rate completely determines the exponential distribution; a greater hazard rate means a shorter survival time. To compare the efficacy of the two therapies is to compare λ_1 and λ_2.

Let t_{jk} denote the length of time until the event or censoring occurs on the kth patient assigned therapy j in the trial. The density of t_{jk} is:

$$f(t_{jk}|\lambda_j) = \lambda_j e^{-\lambda_j t_{jk}}, \quad j = 1, 2$$
$$k = 1, \cdots, n_j$$

where $j = 1$ corresponds to the standard therapy, and $j = 2$ corresponds to the experimental therapy.

Let v stand for the log hazard ratio of experimental therapy to standard therapy: $v = ln(\lambda_2/\lambda_1)$. Parameter v is positive or negative depending on whether the standard or the experimental therapy is better (in the sense of having smaller hazard rate), and the value of v represents the degree of difference between the two therapies. The larger the $|v|$, the greater the difference between the two therapies. For example, $|v| = ln2 = 0.7$ implies that the mean survival time on one therapy is twice that on the other therapy.

λ_2 is determined by λ_1 and v through relation $\lambda_2 = \lambda_1 e^v$, so ($\lambda_1$, v) forms a full parameter space for this model.

Assume our observations are (t_{jk}, δ_{jk}), $k = 1, ..., n_j$, $j = 1, 2$, where δ_{jk} is the status at the observing.

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

If t_{jk} is a censored point, then the corresponding observation's contribution to the likelihood function is the survival function $S(t_{jk}) = 1 - F(t_{jk}) = e^{-\lambda_j t_{jk}}$.

The likelihood function of λ_1 and v is

$$\begin{aligned} L(\lambda_1, v|data) &= \Pi_{j=1}^{2}\Pi_{k=1}^{n_j} f(t_{jk}, \delta_{jk}|\lambda_j) \\ &= \Pi_{j=1}^{2}\Pi_{k=1}^{n_j} \lambda_j^{\delta_{jk}} e^{-\lambda_j t_{jk}} \\ &= \Pi_{j=1}^{2} \lambda_j^{\sum_{k=1}^{n_j} \delta_{jk}} e^{-\lambda_j \sum_{k=1}^{n_j} t_{jk}} \\ &= \lambda_1^{d_1} e^{-\lambda_1 T_1} \lambda_2^{d_2} e^{-\lambda_2 T_2} \\ &= \lambda_1^{d_1+d_2} e^{d_2 v} exp[-\lambda_1(T_1 + T_2 e^{v})] \end{aligned}$$

where $d_j = \sum_{k=1}^{n_j} \delta_{jk}, \quad j = 1, 2$ is the total number of uncensored observations on therapy j, and $T_j = \sum_{k=1}^{n_j} t_{jk}, \ j = 1, 2$ is the total exposure time observed on therapy j. Sufficient statistics for (λ_1, v) are $d_j, T_j, j = 1, 2$.

2.3 Prior Distributions

Prior probability distributions of λ_1 and v represent our knowledge about λ_1 and v before the clinical trial. When much knowledge about λ_1 and v is available, the prior distributions are quite concentrated. If our knowledge about λ_1 and v is slight, then the prior distributions are disperse.

In prior distributions, λ_1 and v are assumed to be independent. A conjugate prior distribution for λ_1 is a gamma distribution:

$$f(\lambda_1|a, b) \propto \lambda_1^{a-1} e^{-b\lambda_1}$$

Usually, there is historical information about the standard therapy. For example, we might know the median or mean survival time, as well as survival rate up to a certain time for the patients treated with the standard therapy in previous studies. It is possible that historical and current patients are not

exchangeable (Lin, 1993), so we only use the information from historical patients to build our prior distribution, we do not use historical patients in our likelihood function. Parameters a and b are chosen by assessing the historical information, so that Gamma(a,b) distribution represents prior belief about λ_1.

In this model, prior distribution of v is $N(0, \sigma^2)$, and $v = 0$ corresponds to $\lambda_1 = \lambda_2$. Prior belief about the relative efficacy of the two therapies is symmetric about zero. The value of σ reflects the degree of our prior knowledge about v; a small σ indicates a firm belief in the relative efficiency of the two therapies. When σ is varied to address sensitivity, a larger σ represents more open-mindedness concerning the relative efficiency of the experimental therapy to the standard therapy in the sense that the data then have a greater impact on the posterior distribution.

2.4 Posterior Distributions

The joint posterior density of λ_1 and v is

$$\begin{aligned} f(\lambda_1, v|data) &\propto f(\lambda_1, v)L(\lambda_1, v|data) \\ &\propto f(\lambda_1)f(v)L(\lambda_1, v|data) \\ &\propto \lambda_1^{a-1}e^{-b\lambda_1}e^{-\frac{v^2}{2\sigma^2}}\lambda_1^{d_1+d_2}e^{d_2 v}exp[-\lambda_1(T_1 + T_2e^v)] \\ &\propto \lambda_1^{a+d_1+d_2-1}e^{d_2 v}e^{-\frac{v^2}{2\sigma^2}}exp[-\lambda_1(b + T_1 + T_2e^v)] \end{aligned}$$

So the posterior density of v is

$$\begin{aligned} f(v|data) &= \int_0^\infty f(\lambda_1, v|data)d\lambda_1 \\ &\propto e^{d_2 v}e^{-\frac{v^2}{2\sigma^2}}\int_0^\infty \lambda_1^{a+d_1+d_2-1}exp[-\lambda_1(b + T_1 + T_2e^v)]d\lambda_1 \end{aligned}$$

$$\propto \frac{e^{d_2 v - \frac{v^2}{2\sigma^2}}}{(b + T_1 + T_2 e^v)^{a+d_1+d_2}}$$

$$f(v|data) = c\frac{e^{d_2 v - \frac{v^2}{2\sigma^2}}}{(b + T_1 + T_2 e^v)^{a+d_1+d_2}}$$

where c is the normalizing constant.

Noticing

$$\begin{aligned} f(\lambda_1|v, data) &\propto f(\lambda_1, v) f(data|\lambda_1, v) \\ &\propto f(\lambda_1) f(data|\lambda_1, v) \\ &\propto \lambda_1^{a-1} e^{-b\lambda_1} \lambda_1^{d_1+d_2} exp[-\lambda_1(T_1 + T_2 e^v)] \\ &\propto \lambda_1^{a+d_1+d_2-1} exp[-\lambda_1(b + T_1 + T_2 e^v)] \end{aligned}$$

so,

$$f(\lambda_1|v, data) = \frac{(b + T_1 + T_2 e^v)^{a+d_1+d_2}}{\Gamma(a + d_1 + d_2)} \lambda_1^{a+d_1+d_2-1} exp[-\lambda_1(b + T_1 + T_2 e^v)] \ ,$$

a $Gamma(a + d_1 + d_2, b + T_1 + T_2 e^v)$ distribution.

The posterior density of λ_1 is

$$\begin{aligned} f(\lambda_1|data) &= \int_{-\infty}^{\infty} f(\lambda_1, v|data) dv \\ &= \int_{-\infty}^{\infty} f(\lambda_1|v, data) f(v|data) dv \\ &= \frac{c}{\Gamma(a + d_1 + d_2)} \lambda_1^{a+d_1+d_2-1} e^{-\lambda_1(b+T_1)} \int_{-\infty}^{\infty} exp[-(\frac{v^2}{2\sigma^2} + \lambda_1 T_2 e^v)] dv \end{aligned}$$

Numerical integration in one dimension is required for calculating $f(v|data)$ and $f(\lambda_1|data)$. The integrating functions are smooth and unimodal. Simp-

son's rule will be used to do the numerical integration in the example in the next chapter.

2.5 Predictive Survival Functions

The predicted survival function for each therapy is of particular interest to new patients who need to choose a therapy. Let X_1 and X_2 denote the survival time of new patients treated with therapy 1 and therapy 2, respectively. The predicted survival functions for the two therapies are

$$\begin{aligned} S_1(t) &= P(X_1 > t|data) \\ &= \int_0^\infty P(X_1 > t|\lambda_1) f(\lambda_1|data) d\lambda_1 \\ &= \int_0^\infty e^{-\lambda_1 t} f(\lambda_1|data) d\lambda_1 \end{aligned}$$

$$\begin{aligned} S_2(t) &= P(X_2 > t|data) \\ &= \int_{-\infty}^\infty \int_0^\infty P(X_2 > t|\lambda_1, v) f(v, \lambda_1|data) d\lambda_1 dv \\ &= \int_{-\infty}^\infty \int_0^\infty e^{-\lambda_1 e^v t} f(v, \lambda_1|data) d\lambda_1 dv \end{aligned}$$

Chapter 3

Study of a NSCLC Trial

3.1 Introduction

3.1.1 Purpose of the Study

In this chapter, we present a case study of a phase III clinical trial conducted by Cancer and Leukemia Group B (CALGB) to test two therapies for treatment of patients with stage III non-small cell lung cancer (NSCLC). Lung cancer is generally subdivided into two categories, based on the cell type determined at diagnosis. The first type, small cell (or oat cell) anaplastic carcinoma, accounts for roughly 30% of all lung cancers. Most other cell types are classified as 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 NSCLC who have extensive disease in the chest but no demonstrable distant metastases are defined as having stage III NSCLC. Such patients are not generally considered curable by surgery alone.

For many years, radiotherapy (RT) alone had been the standard treatment

of choice for these patients. In attempting to improve survival time in these patients, clinical researchers in the early 1980s considered the possibility that RT alone might not be sufficient to eradicate micrometastatic disease, and they accumulated some considerable evidence that indicated that platinum-based chemotherapy (CT) increased the survival time of patients who had more advanced disease. Therefore, these researchers proposed that CT be administered in conjunction with standard RT.

The study presented in this chapter was designed to compare the standard treatment (RT only), which consists of RT delivered over six weeks to the original tumor volume and involved regional lymph nodes, to an experimental treatment (CT+RT), which employs five weeks of cisplatin plus vinblastine prior to the RT.

3.1.2 Design of the Study

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; performance status of 0 or 1; and weight loss of less than 5% in the three-month interval prior to study entry. It also imposed standard CALGB eligibility criteria for laboratory values, other diseases, and so on. Details of the eligibility and other clinical aspects of this study are outlined in Dillman et al. (1990).

The trial was a prospective, randomized, nonblinded study. The central office of the CALGB stratified patients according to historic type to ensure a balanced distribution between therapy groups, and then randomly assigned the patients to receive either RT only or CT+RT. For the RT only group, radiation

therapy was started within five days after entry, and the first day of therapy was defined as day 1. For the CT+RT group, chemotherapy was started on day 1 and stopped on either day 29 or day 36; and radiation therapy was started on day 50.

The primary study objective was to compare the overall survival time for the two therapy groups. The original fixed sample size for this trial was 240 patients (120 patients in each therapy group). 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-side significance level of $\alpha = 0.05$. If survival times are assumed to be distributed exponentially, a 1.5:1 hazard ratio represents a 50% increase in median survival time in one therapy group over another.

3.1.3 Analysis of Survival Time

Patients were enrolled from May 1984 to May 1987. Although the trial was designed as a fixed sample size study, several interim analyses were performed and the trial was terminated early in response to a therapy difference emerging over time. When the first interim analysis was performed in the fall of 1985, the sample size (10 deaths in 50 eligible patients) was too small and the following time was too short to allow any meaningful comparison of survival time by therapy. Four more interim analyses were performed in March 1986, August 1986, October 1986 and March 1987. In these analyses, the logrank test was used to compare the two survival curves, and truncated O'Brien-Fleming boundary level for each p-value was adopted in decision making. The Kaplan-Meier plot for the two therapy groups at each interim analysis is presented in Figure 3.1, and Table 3.1 lists the observed p-values and boundary significance

Analysis	Logrank p-value	Boundary	Decision
1st Interim	–	0.0013	Keep open
2nd Interim	0.021	0.0013	Keep open
3rd Interim	0.0071	0.0013	Keep open
4th Interim	0.0015	0.0013	Keep open
5th Interim	0.0015*	0.0013	Close

$\star$ After adjusting covariates with Cox model, p-value is 0.0008

Table 3.1: Observed p-values and the boundary at the interim analyses

levels used in the monitoring process.

Results of the March 1987 interim analysis convinced the CALGB researchers to close the trial to further accrual. At that time, 155 eligible patients had been accrued and follow-up data were available for 105 patients. After the trial stopped enrolling new patients, the enrolled patients were followed up until the summer of 1992. We will analyze the data observed in 1992 as the most recent analysis.

3.2 Bayesian Analysis of the NSCLC Study

3.2.1 The Model

In this section, I will apply the model proposed in Chapter 1 to this trial, and perform interim analyses and an analysis of the information gathered in 1992. The primary interest is to compare efficacies of the two therapies.

It is assumed that survival times for each therapy are distributed exponentially with hazard rates λ_j, $j = 1, 2$. Here, $j = 1$ corresponds to the RT only therapy, and $j = 2$ corresponds to the CT+RT therapy. Parameter v denotes the log hazard ratio: $v = ln(\lambda_2/\lambda_1)$. Posterior distributions of λ_1 and v are derived at each interim analysis and as of 1992. The prior distributions of λ_1

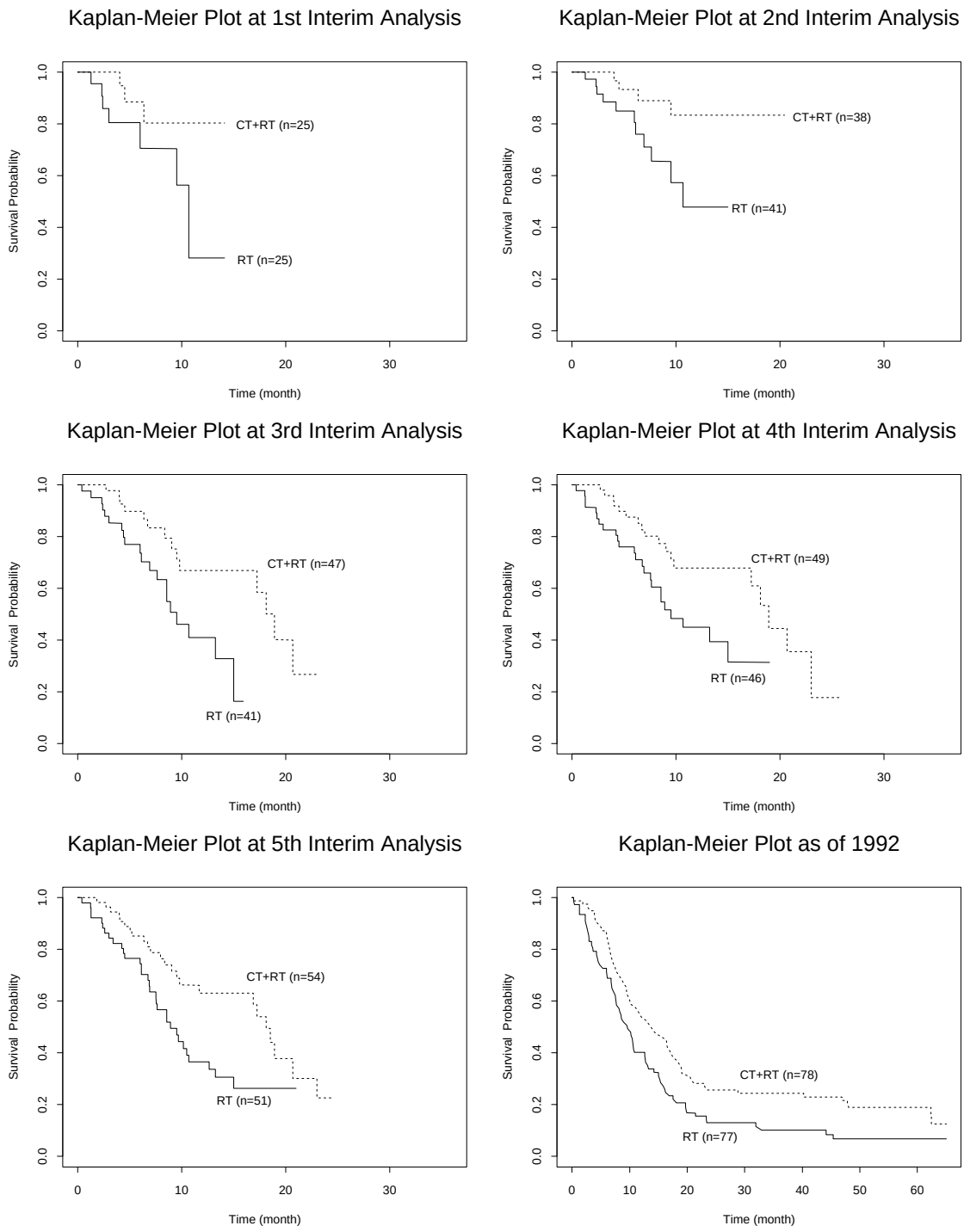

Figure 3.1: Kaplan-Meier plot at interim analyses and as of 1992

	On RT			On CT+RT		
Analysis Time	n_1	d_1	T_1	n_2	d_2	T_2
1st Interim	25	7	122.03	25	3	164.43
2nd Interim	41	12	240.63	38	4	341.07
3rd Interim	41	20	298.53	47	14	432.77
4th Interim	46	24	375.97	49	18	532.67
5th Interim	51	32	441.83	54	24	611.13
as of 1992	77	71	1135.70	78	65	1737.60

Table 3.2: Available information at time of analysis

and v are Gamma(a,b) and $N(0, \sigma^2)$, respectively.

3.2.2 Data

The data from the study, with information up to 1992, are presented in Appendix A. In the model, sufficient statistics of $(\lambda_1,\ v)$ are $d_j, T_j,\ j = 1, 2$, which, respectively, represent the total number of deaths and the total exposure time observed on each therapy. Table 3.2 displays the sufficient statistics and the total number of patients enrolled on each therapy (n_i) at the times of the five interim analyses and as of June 1992. The time unit is one month.

3.2.3 Choice of Prior Distributions

Parameters a, b and σ^2 in prior distributions need to be specified. The prior distribution reflects our knowledge of the parameters λ_1 and v prior to the trial. Available historical information about λ_1 and v are used to determine their prior distribution.

Note that the prior distribution function of survival time on RT therapy is

$$f_1(t \mid a, b) \ = \ \int_0^\infty f(t \mid \lambda_1) f(\lambda_1 \mid a, b)\, d\lambda_1$$

$$
\begin{aligned}
&= \int_0^\infty \lambda_1 e^{-\lambda_1 t} \frac{b^a}{\Gamma(a)} \lambda_1^{a-1} e^{-\lambda_1 b} d\lambda_1 \\
&= \frac{b^a}{\Gamma(a)} \int_0^\infty \lambda_1^a e^{-\lambda_1 (b+t)} d\lambda_1 \\
&= \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a+1)}{(b+t)^{a+1}} \\
&= \frac{a b^a}{(b+t)^{a+1}}
\end{aligned}
$$

Prior survival function on RT therapy is

$$
\begin{aligned}
S_1(t \mid a, b) &= \int_t^\infty f_1(x \mid a, b) dx \\
&= \int_t^\infty \frac{a b^a}{(b+x)^{a+1}} dx \\
&= \left(\frac{b}{b+t}\right)^a
\end{aligned}
$$

Evidence available before this trial suggests that the median survival time on RT is 8 to 10 months, with a two-year survival rate of 10 to 20 percent and a three-year survival rate of 5 to 10 percent (Perez et al., 1987). In this chapter, parameter values $a = 2$ and $b = 20$ are chosen to approximate the historical information. These parameter values set the prior median survival time to 8.3 months, two-year survival rate to 20 percent and three-year survival rate to 13 percent. So, the prior distribution of λ_1 is Gamma(2,20).

σ^2, the prior variance of v, reflects the variation of our prior belief of v. A large σ reflects a prior belief of a possible large difference between the two therapies, and it represents open-mindedness concerning the effect of adding CT to RT in the sense that the data then have a greater impact on the posterior distribution, A small σ reflects a prior belief of only a small difference between

	MLE		v		λ_1	
Analysis Time	v_{MLE}	λ_{1MLE}	$\hat{v}$	$\hat{\sigma}_v^2$	$\hat{\lambda}_1$	$\hat{\sigma}_{\lambda_1}^2$
1st Interim	-1.146	0.057	-0.883	0.288	0.050	0.00036
2nd Interim	-1.447	0.050	-1.176	0.221	0.045	0.00018
3rd Interim	-0.728	0.067	-0.680	0.106	0.064	0.00020
4th Interim	-0.636	0.064	-0.607	0.087	0.062	0.00016
5th Interim	-0.612	0.072	-0.586	0.067	0.070	0.00015
as of 1992	-0.514	0.063	-0.509	0.028	0.062	0.00005

Table 3.3: MLE, posterior modes and variances of v and λ_1 at times of analyses

the two therapies. Since our prior belief is that adding CT to RT might be profitable, σ is set to 1 first. Different values of σ will be considered to see the sensitivity. $N(0,1)$ distribution for v is a quite disperse prior distribution and a rather open-minded choice. Data will strongly influence the posterior distribution. For example, the trial was designed to detect a log hazard ratio of $-ln(1.5) = -0.41$, and -0.41 is close to the middle of a standard normal distribution. Choosing a value of σ larger than 1 leaves the conclusions essentially unchanged from assuming $\sigma = 1$.

3.2.4 Posterior Distributions

Results of analysis are presented in the following tables and figure. Table 3.3 displays the maximum likelihood estimators (MLE), posterior modes ($\hat{v}$ and $\hat{\lambda}_1$) and variance ($\hat{\sigma}_v^2$ and $\hat{\sigma}_{\lambda_1}^2$) of v and λ_1 at the times of the analyses. Table 3.4 displays posterior probabilities of $v < 0$, $v < -0.25$ and $v < -0.5$ at times of the analysis. Figure 3.2 shows the marginal posterior densities of v and λ_1 at the five interim analyses and as of 1992.

From those tables and the figure, one can see how our beliefs about the parameters changed as the trial went on. They clearly shown that the posterior

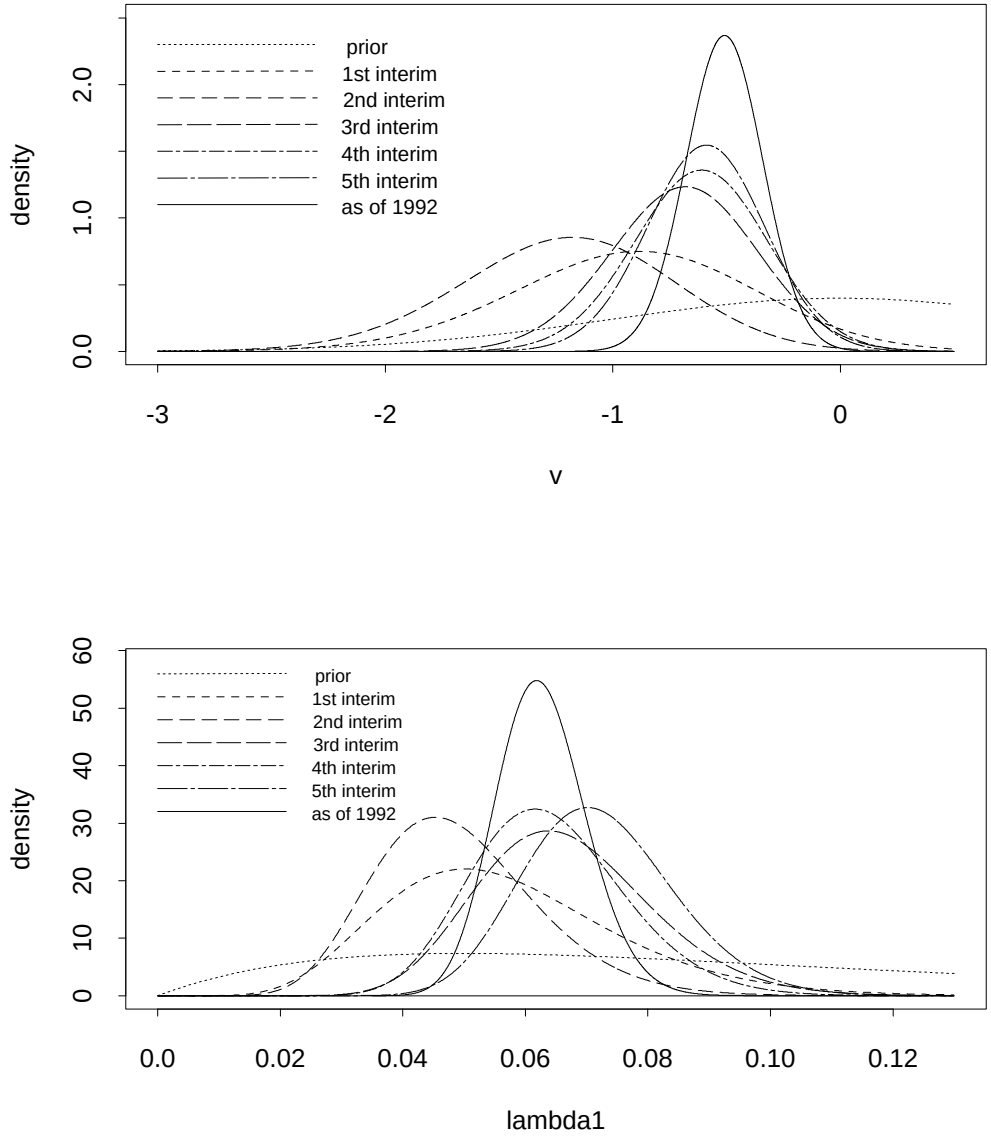

Figure 3.2: Posterior densities at time of analysis

	Some Probabilities of v		
Analysis Time	$P(v < 0 \mid D)$	$P(v < -0.25 \mid D)$	$P(v < -0.5 \mid D)$
Prior	0.500	0.401	0.309
1st Interim	0.976	0.911	0.794
2nd Interim	0.997	0.984	0.940
3rd Interim	0.987	0.916	0.720
4th Interim	0.985	0.895	0.650
5th Interim	0.990	0.909	0.637
as of 1992	0.999	0.939	0.523

Table 3.4: Some posterior probabilities of v

variance of v decreased as the trial went on, and that the posterior variance of v in the 1992 analysis is about 10 times smaller than that in the first interim analysis. As the trial went on, more and more information was obtained, the posterior density of v got more and more concentrated. The posterior density of v is mainly concentrated on the left side of $v = 0$ at any time of analysis, so the information about v consistently indicates that v is very likely negative.

The posterior densities of λ_1 in the interim analyses have fluctuated somewhat. This is apparent in the plottings of the densities of λ_1 and the posterior variances of λ_1; and it suggests that the new information in the interim analyses might not very consistent with previous information, because in the interim analyses, only a small number of observations were uncensored, and not much information was available. The posterior density of λ_1 in the analysis of 1992 is much more concentrated than those in the interim analyses because much more complete information was obtained in 1992.

There were many more early deaths in the RT group than in the CT+RT group. In the first interim analysis, seven out of 25 patients in the RT only group had died, whereas three out of 25 patients in the CT+RT group had

died. In the second interim analysis, 12 out of 41 patients receiving RT only had died, whereas only 4 out of 38 patients receiving CT+RT had died. This suggests that a patient's chances of surviving are a lot better with the CT+RT than with the RT alone. This can be seen from the posterior densities of v in these interim analyses.

The posterior modes of v at the first and second interim analysis times are small, especially at the second interim analysis time when $\hat{v} = -1.176$. In the model, $v = -1.176$ means a 224% increase in median or mean survival time with CT+RT over RT only. Because there was less information in the first and second interim analyses, the densities of v were relatively flat. The mode of v increased in the later analyses because more deaths were observed in the CT+RT group. Evidence at any given time suggested that patients would be better off on the CT+RT regimen than on the RT only regimen because the posterior probability of $v < 0$ is very close to 1 at any time of analysis.

3.2.5 Predictive Survival Functions

Patients with this disease would like to know how well they will respond on each therapy and which therapy will help them survive longer. Table 3.5 displays the predictive mean survival times on the two therapies at the times of analyses. Figure 3.3 presents the predictive survival functions on each therapy at the times of analyses.

Because of the discrepancy between the results from the two therapies at the early interim analysis times, the difference between the predictive mean survival time on the two therapies is large at early interim analyses. The difference was greatest (57.6 months) at the time of the second interim analysis because the

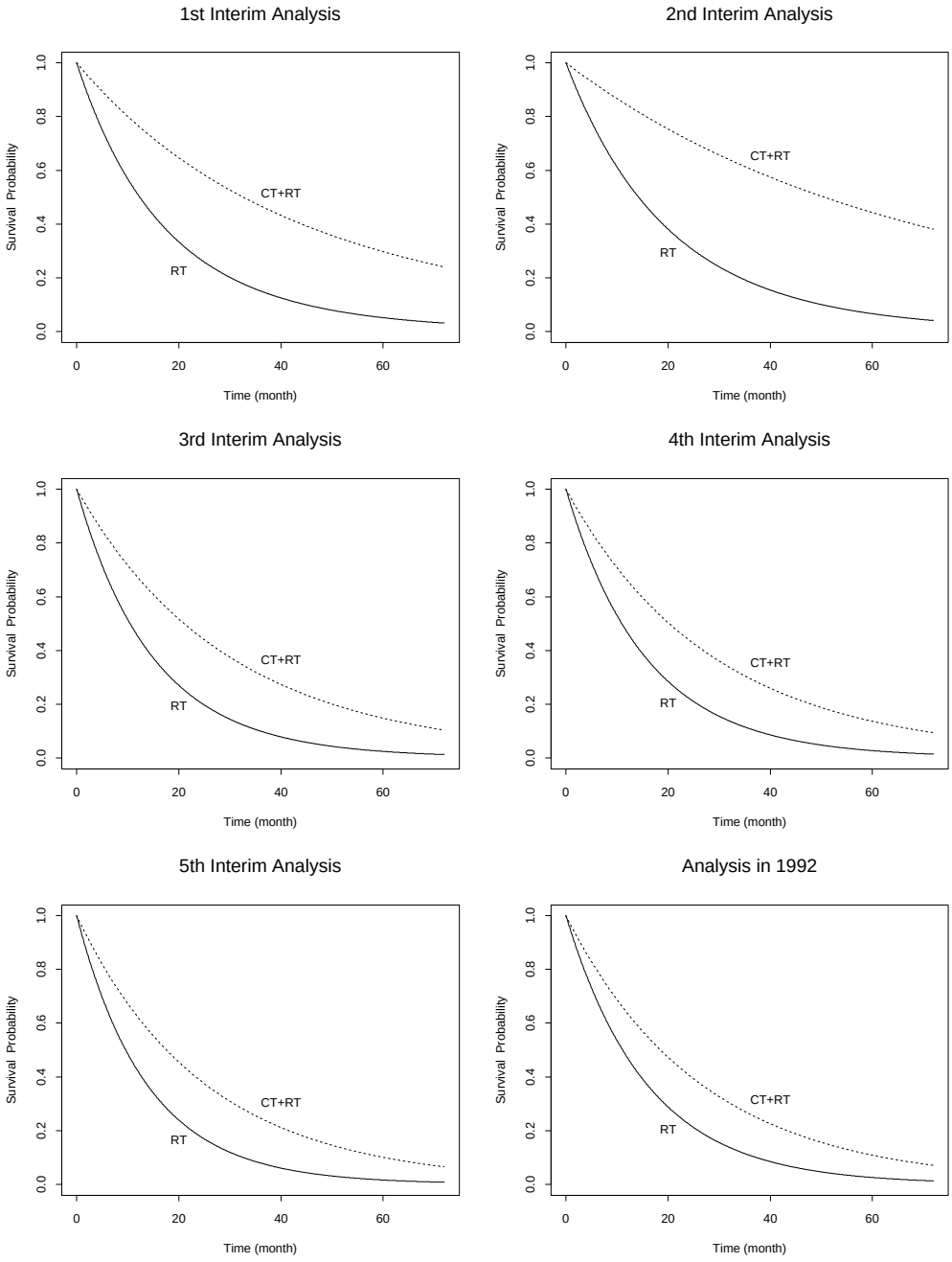

Figure 3.3: Predictive survival function at time of Analysis

	Mean Survival (month)	
Analysis Time	RT	CT+RT
1st Interim	17.9	52.9
2nd Interim	20.1	77.7
3rd Interim	15.6	31.5
4th Interim	16.2	30.2
5th Interim	14.2	25.9
as of 1992	16.2	27.0

Table 3.5: Predictive mean survival time at time of analysis

largest difference between the data on the two therapy groups occurred then. In Figure 3.3, the predictive survival function for the CT+RT group is always above that for the RT only group. These plots indicate that at any time point, the probability of a patient in the CT+RT group would survive to that time point is always greater than that of a patient in the RT only group. The difference between the two survival functions was large in the early interim analyses, especially in the second interim analysis. The difference between the two survival functions and the two predictive mean survival times decreased in later analyses and became stable.

3.2.6 Sensitivity to Prior Distributions

Since the log hazard ratio is unlikely to have a large absolute value, the $N(0,1)$ prior distribution of v is quite disperse. The choice of $\sigma = 1$ in the $N(0,\sigma^2)$ prior distribution of v is quite open-minded; that means that the data play a crucial role in making an inference about v. Choosing the $N(0,\sigma^2)$ prior distribution for v shows that we are neutral concerning the effect of adding CT to RT, and the value of σ reflects our degree of confidence about this concern. A larger σ represents more open-mindedness in the sense that the data then

have a greater impact on the posterior distribution.

All priors in the $N(0, \sigma^2)$ family have the same probability of $v < 0$ (CT is a beneficial add-on) as of $v > 0$ (CT detrimental). This symmetry may not be appropriate, because CT was not regarded a priori as likely to be detrimental. However, the likelihood function concentrates on $v < 0$. So, the fact that half the prior probability is associated with $v > 0$ turns out to be unimportant. An open-minded prior means that if early data point to CT being effective, the posterior probability of $v<0$, say, may be quite large. For someone with an open-minded prior, this is appropriate; but it is not appropriate if the effectiveness of CT is questionable. If σ is very small, then our prior belief that CT+RT is as good as RT alone is strong, and early data would not change this belief much.

Figure 3.4 shows how $P(v < 0 \,|\, D)$, $P(v < -0.25 \,|\, D)$ and $P(v < -0.5 \,|\, D)$ change when σ varies. We can see from the figure that the posterior probabilities have a larger variation for a smaller σ, and the variation decreases as σ increases. The variation is larger at the earlier analysis times than at the later analysis times, because when more data is available, the prior distribution becomes less impact on the posterior distribution.

3.2.7 What If the Trial Had Not Been Stopped Early

The trial became controversial because it was stopped early. A very important question raised is, what would the trial have concluded if it had not been stopped in 1987? Obviously, no one can be certain. An advantage of the Bayesian approach is that all uncertainties have probabilities. One can find the distribution of lifetimes that would have been concluded had the trial

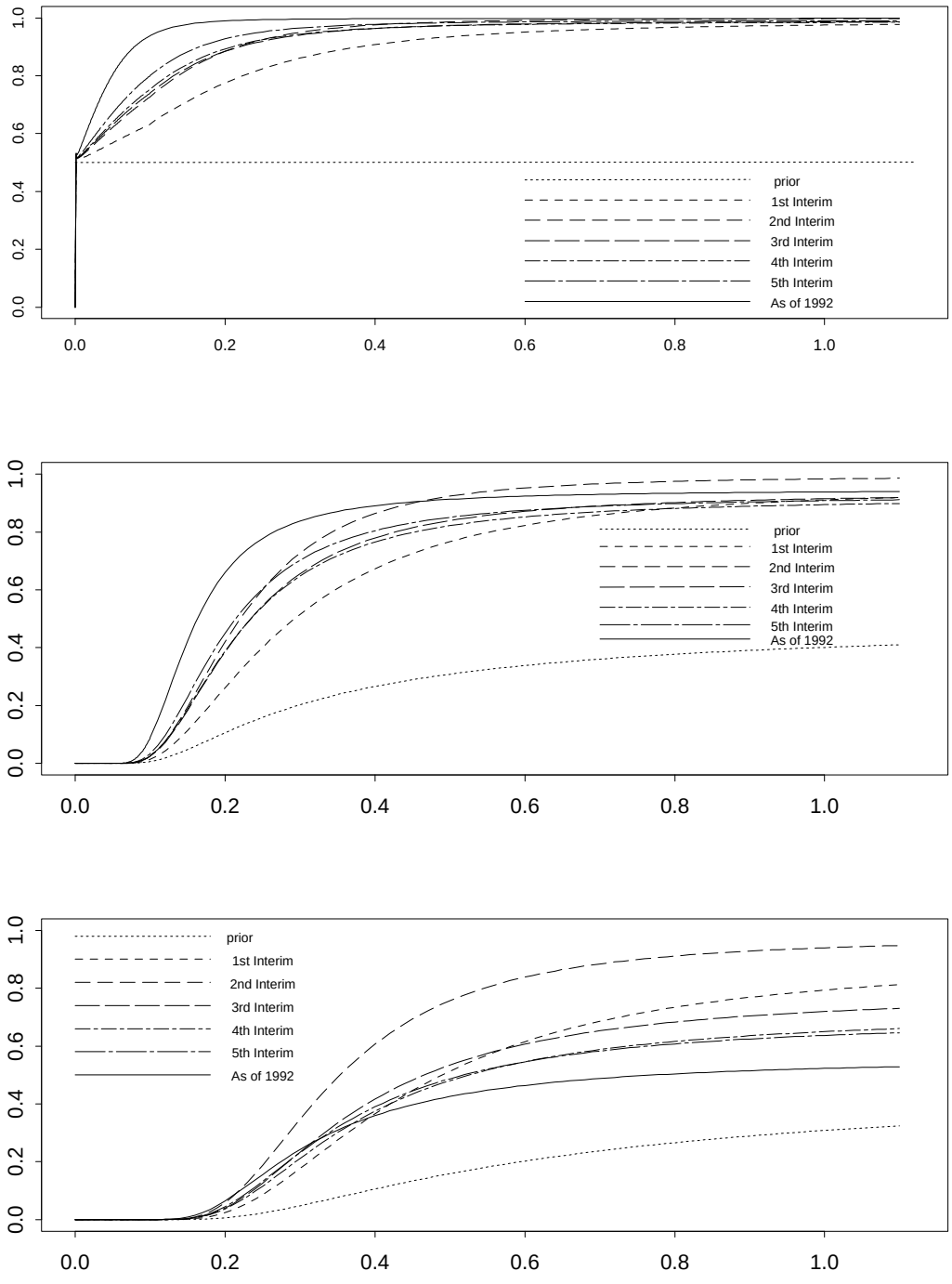

Figure 3.4: The effect on the posterior probabilities of changing σ

continued to its originally planned size. This distribution can be found at any time. The most interesting times during the conduct of a trial occur at the interim analyses. From our perspective, the present is the most interesting time. So, the predictive distributions conditioned on information available today are calculated.

According to its design, the trial was to stop once 240 patients had been accrued or a total of 190 deaths had occurred. In the simulations to be described, the 240th patient was always admitted before the 190th death occurred, so the former criterion is really the only effective one. In 1987, when the trial stopped, 155 patients had been admitted. These patients are included in the "as of 1992" analysis. According to the predictive distribution based on the information from the 155 patients in 1992, I simulated the information that would be available in June 1992 had an additional 85 patients accrued (uniformly over the 17 months between April 1987 and September 1988), for a total of 240.

Figure 3.5 displays simulated values of $P(v < 0|240\ pts)$, $P(v < -0.25|240\ pts)$ and $P(v < -0.50|240\ pts)$. In each case, the mean is the actual current value of the corresponding probability, which is based on the data for the 155 patients in the "as of 1992" analysis of Table 3.4. The histogram for $P(v < 0|240\ pts)$ makes it clear that the strong conclusion that CT is beneficial is unlikely to have changed even with an increase in sample size of 55 percent. The second histogram, that for $P(v < -0.25|240\ pts)$, evinces greater variability, indicating that this quantity is somewhat less predictable, but it is still very likely that v would be less than -0.25. The third histogram shows still more variability for $P(v < -0.5|240\ pts)$.

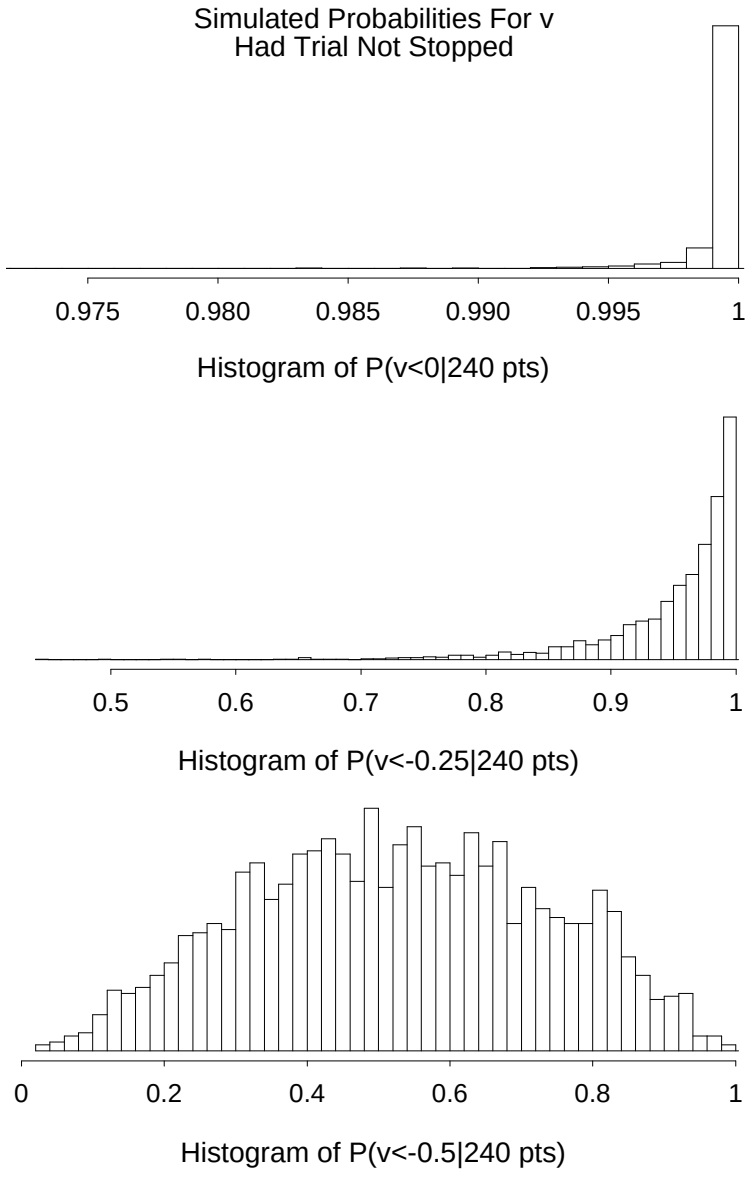

Figure 3.5: Histogram of simulated posterior probabilities of v

Figure 3.6 displays simulated values of the mean survival times for the two therapy groups. From this figure, one can conclude that the mean advantage in survival time of about 11 months would not change much had an additional 85 patients accrued.

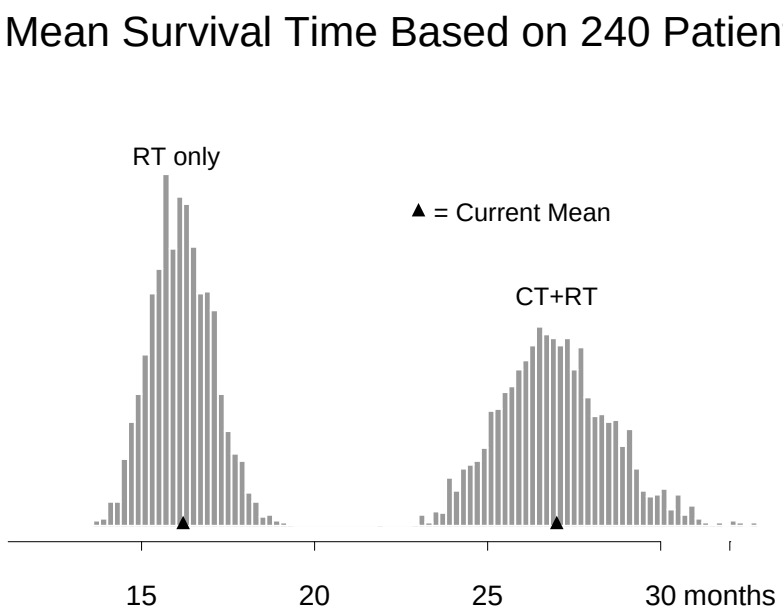

Figure 3.6: Histogram of predictive mean survival time

Chapter 4

A Bayesian Hierarchical Model for Multicenter Trial

4.1 Introduction

Center differences in a multicenter clinical trial are examined in this chapter. A multicenter trial applies therapies to a wider range of centers and patient groups than a trial at a single center does. Different medical centers might have different characteristics, so that their patients' response to one therapy might be different at different centers. I will describe a general Bayesian hierarchical model for analyzing a multicenter clinical trial data, and in section 3, I will briefly introduce the Gibbs sampling technique that will be used for calculating posterior marginal densities.

In this chapter, a multicenter trial is conducted to compare two therapies. Observations are patients' survival times. Exponential distributions will be used for survival times. Our goal is to compare the efficacies of the two therapies in the presence of a possible center effect.

4.2 Statistical Model

In a multicenter trial of two therapies, data are a combination of $I \times 2$ subgroups, where I is the number of medical centers that participated in the trial. For each center, there are two subgroups of data that correspond to the two therapies. In the hierarchical model proposed in this chapter, survival times are modeled in the first level of the hierarchy. In the second level, parameters for individual centers are modeled to be a sample from a center population. In the third level, the parameters in the center population are modeled.

Let t_{ijk} be the survival time of the kth patient on therapy j in center i, $i = 1, \cdots, I$, $j = 1, 2$, $k = 1, \cdots, n_{ij}$. In the first stage, t_{ijk} has distribution:

$$\begin{aligned} f(t_{ijk}|\lambda_{ij}) &= \lambda_{ij} e^{-\lambda_{ij} t_{ijk}}, \quad i = 1, \cdots, I \\ & \qquad\qquad\quad j = 1, 2 \\ & \qquad\qquad\quad k = 1, \cdots, n_{ij} \end{aligned}$$

where $j = 1$ corresponds to the standard therapy, and $j = 2$ corresponds to the experimental therapy. λ_{i1} and λ_{i2} are the two hazard rates in center i. Comparing the two therapies in center i is to compare λ_{i1} and λ_{i2}. Let $v_i = ln(\lambda_{i2}/\lambda_{i1})$ — the log hazard ratio in center i, $i = 1, \cdots, I$. Parameter v_i is positive or negative depending on whether the standard or the experimental therapy is better (in the sense of having smaller hazard rate) in center i, and the value of v_i represents the degree of difference between the two therapies in center i. λ_{i2} is determined by λ_{i1} and v_i through relation $\lambda_{i2} = \lambda_{i1} e^{v_i}$. So ($\lambda_{i1}$, v_i), $i = 1, \cdots, I$, form a full parameter space in the first level of this hierarchical model.

In the second level, λ_{i1}, $i = 1, \cdots, I$ are assumed to be exchangeable with a

gamma distribution; and $v_i,\ i = 1, \cdots, I$ are assumed to be exchangeable with a normal distribution.

$$\begin{aligned} \lambda_{i1}|(a,b) &\overset{i.i.d.}{\sim} Gamma(a,b) \\ f(\lambda_{i1}|a,b) &\propto \lambda_{i1}^{a-1} e^{-b\lambda_{i1}} \qquad i = 1, \cdots, I \\ v_i|(\mu,\sigma^2) &\overset{i.i.d.}{\sim} N(\mu,\sigma^2) \qquad i = 1, \cdots, I \end{aligned}$$

$\lambda_{i1}'s$ and $v_i's$ are independent.

In the third level, parameters are assumed as

$$\begin{aligned} a &\ is\ known \\ b &\sim Gamma(c,d) \\ f(b) &\propto b^{c-1} e^{-db} \\ \mu &\sim N(0,1) \\ \sigma^2 &\sim IG(u,w) \text{ (Inverse Gamma)} \\ f(\sigma^2) &\propto (\sigma^2)^{-(u+1)} e^{-\frac{w}{\sigma^2}} \end{aligned}$$

b, μ and σ^2 are independent. c, d, u and w are known parameters.

Assume that the observations are (t_{ijk}, δ_{ijk}), where δ_{ijk} is the status at the time of the observation.

$$\delta_{ijk} = \begin{cases} 0 & \text{if } t_{ijk} \text{ is right censored} \qquad i = 1, \cdots, I \\ & \qquad\qquad\qquad\qquad\quad j = 1, 2 \\ 1 & \text{otherwise} \qquad\qquad\qquad\quad k = 1, \cdots, n_{ij} \end{cases}$$

The likelihood function in center i is

$$\begin{aligned} L_i(\lambda_{i1}, v_i|data) &= \Pi_{j=1}^{2}\Pi_{k=1}^{n_{ij}} f(t_{ijk}, \delta_{ijk}|\lambda_{ij}) \\ &= \Pi_{j=1}^{2}\Pi_{k=1}^{n_{ij}} \lambda_{ij}^{\delta_{ijk}} e^{-\lambda_{ij} t_{ijk}} \\ &= \Pi_{j=1}^{2} \lambda_{ij}^{\sum_{k=1}^{n_{ij}} \delta_{ijk}} e^{-\lambda_{ij} \sum_{k=1}^{n_{ij}} t_{ijk}} \\ &= \lambda_{i1}^{d_{i1}} e^{-\lambda_{i1} T_{i1}} \lambda_{i2}^{d_{i2}} e^{-\lambda_{i2} T_{i2}} \\ &= \lambda_{i1}^{d_{i1}+d_{i2}} e^{d_{i2} v_i} exp[-\lambda_{i1}(T_{i1} + T_{i2} e^{v_i})] \end{aligned}$$

where

$$\begin{aligned} d_{ij} &= \textstyle\sum_{k=1}^{n_{ij}} \delta_{ijk} \\ T_{ij} &= \textstyle\sum_{k=1}^{n_{ij}} t_{ijk} \ , \quad j = 1, 2, \ i = 1, \cdots, I \end{aligned}$$

are the total number of uncensored observations and total exposure time on therapy j in center i, $j = 1, 2, \ i = 1, \cdots, I$.

The likelihood function is

$$\begin{aligned} L(\underline{\lambda}, \underline{v}, b, \mu, \sigma^2|data) &= \Pi_{i=1}^{I} L_i \\ &= \Pi_{i=1}^{I} \lambda_{i1}^{d_{i1}+d_{i2}} e^{d_{i2} v_i} exp[-\lambda_{i1}(T_{i1} + T_{i2} e^{v_i})] \end{aligned}$$

where $\underline{\lambda} = (\lambda_{11}, \cdots, \lambda_{I1})$, $\underline{v} = (v_1, \cdots, v_I)$.

The sufficient statistics for those parameters are d_{ij}, T_{ij}, $j = 1, 2$, $i = 1, \cdots, I$.

4.3 Gibbs Sampling

In a high dimensional problem, to calculate the marginal posterior distributions of parameters once involved numerical integration. Geman and Geman (1984)

introduced a method of estimating the marginal posterior distributions of random variables using the full conditional distributions, and their method was further developed by Gelfand and Smith (1990) into a popularly used method called Gibbs sampling technique.

Suppose we have m random variables, $X_1, \cdots, X_m$, and know the full conditional distributions $f(X_i|X_j, j \neq i), i = 1, \cdots, m$. We need to find their marginal distributions. To do Gibbs sampling, we begin with m arbitrary starting values $X_1^0, \cdots, X_m^0$ for the m random variables. First, we update X_1^0 by a random observation X_1^1 from the full conditional distribution $f(X_1|X_2^0, \cdots, X_m^0)$. Next, X_2^0 is updated by a random observation X_2^1 from the full conditional distribution $f(X_2|X_1^1, X_3^0, \cdots, X_m^0)$. The process is continued for $X_3, \cdots, X_m$ until X_m^0 is updated by a random observation X_m^1 from the full conditional distribution $f(X_m|X_1^1, \cdots, X_{m-1}^1)$. $X_1^0, \cdots, X_m^0$ is updated by $X_1^1, \cdots, X_m^1$, and we repeat the entire process S times, updating X_i^{s-1} by a random observation X_i^s from the full conditional distribution $f(X_i|X_1^s, \cdots, X_{i-1}^s, X_{i+1}^{s-1}, \cdots, X_m^{s-1})$ at the s-th iteration. Under mild regularity conditions, Geman and Geman showed that as $s \to \infty$, this joint sample tends in probability distribution to a variable with the joint distribution $f(X_1, X_2, \cdots, X_m)$. So $(X_1^s, X_2^s, \cdots, X_m^s)$ can be regarded as a random observation from the joint distribution. We do this sampling R times to get R m-tuples $(X_1^{(r)}, X_2^{(r)}, \cdots, X_m^{(r)}), r = 1, 2, \cdots, R$, and to approximate the marginal density of $f(X_i)$ by the finite mixture density

$$f(X_i) = \frac{1}{R} \sum_{r=1}^{R} f(X_i|X_1^{(r)}, \cdots, X_{i-1}^{(r)}, X_{i+1}^{(r)}, \cdots, X_m^{(r)})$$

Because of the hierarchical structure, the full posterior conditional distributions are relatively easy to be derived in a hierarchical model. The Gibbs

sampling technique greatly simplifies the process of calculating marginal posterior distributions in a hierarchical model.

4.4 Full Posterior Conditional Distributions

To do statistical inference on the parameters $\lambda_{i1},\ v_i,\ i = 1, \cdots, I,\ b,\ \mu$ and σ^2, I applied the Gibbs sampling technique in such a high dimensional problem to get posterior marginal distributions. First, I must derive full posterior conditional distributions. In the following conditional distributions, "*all*" refers to all these parameters, and "*others*" refers to all the other parameters except the one whose conditional distribution is being derived.

The full posterior conditional distributions are

$$\begin{aligned}
f(\lambda_{i1} \mid others, data) &\propto f(\lambda_{i1} \mid a, b) f(data \mid all) \\
&\propto \lambda_{i1}^{a-1} e^{-b\lambda_{i1}} \lambda_{i1}^{d_{i1}+d_{i2}} exp[-\lambda_{i1}(T_{i1} + T_{i2} e^{v_i})] \\
&\propto \lambda_{i1}^{a+d_{i1}+d_{i2}-1} exp[-\lambda_{i1}(b + T_{i1} + T_{i2} e^{v_i})] \\
(\lambda_{i1} \mid others, data) &\sim Gamma(a + d_{i1} + d_{i2}, b + T_{i1} + T_{i2} e^{v_i}) \\
f(v_i \mid others, data) &\propto f(v_i \mid \mu, \sigma^2) f(data \mid all) \\
&\propto e^{-\frac{(v_i - \mu)^2}{2\sigma^2}} e^{d_{i2} v_i} e^{-\lambda_{i1} T_{i2} e^{v_i}} \\
f(b \mid others, data) &\propto f(b) f(\lambda_{11}, ..., \lambda_{I1} | a, b) f(data | all) \\
&\propto f(b) \Pi_{i=1}^{I} f(\lambda_{i1} \mid a, b) \\
&\propto b^{c-1} e^{-db} b^{Ia} e^{-b \sum_{i=1}^{I} \lambda_{i1}} \\
&\propto b^{Ia+c-1} e^{-b(d + \sum_{i=1}^{I} \lambda_{i1})} \\
(b \mid others, data) &\sim Gamma(Ia + c, d + \textstyle\sum_{i=1}^{I} \lambda_{i1})
\end{aligned}$$

$$
\begin{aligned}
f(\mu \mid others, data) &\propto f(\mu) f(v_1, ..., v_k \mid \mu, \sigma^2) f(data \mid all) \\
&\propto f(\mu) \Pi_{i=1}^{I} f(v_i \mid \mu, \sigma^2) \\
&\propto e^{-\frac{\mu^2}{2}} e^{-\frac{\sum_{i=1}^{I} (\mu - v_i)^2}{2\sigma^2}} \\
&\propto exp\{-\frac{(\mu - \frac{\sum_{i=1}^{I} v_i}{\sigma^2 + I})^2}{2(\frac{\sigma^2}{\sigma^2 + I})}\} \\
(\mu \mid others, data) &\sim N(\frac{\sum_{i=1}^{I} v_i}{\sigma^2 + I}, \frac{\sigma^2}{\sigma^2 + I}) \\
f(\sigma^2 \mid others, data) &\propto f(\sigma^2) \Pi_{i=1}^{I} f(v_i \mid \mu, \sigma^2) f(data | all) \\
&\propto (\sigma^2)^{-(u+1)} e^{-\frac{w}{\sigma^2}} (\sigma^2)^{-I/2} e^{-\frac{\sum_{i=1}^{I} (v_i - \mu)^2}{2\sigma^2}} \\
&\propto (\sigma^2)^{-(u + \frac{I}{2} + 1)} e^{-\frac{w + \frac{1}{2} \sum_{i=1}^{I} (v_i - \mu)^2}{\sigma^2}} \\
(\sigma^2 \mid others, data) &\sim IG(u + \tfrac{I}{2},\ w + \tfrac{1}{2} \textstyle\sum_{i=1}^{I} (v_i - \mu)^2)
\end{aligned}
$$

In all these conditional distributions, λ_{i1}'s, b's, μ's and $\sigma^{2\prime}$s are gamma, normal and inverse gamma distributions. It is easy to sample from these distributions. The full posterior conditional distributions of v_i's are not among those well defined distribution families, and the rejection method (Devroye, 1986) needs to be used for getting a random sample of v_i from the posterior conditional distribution of $f(v_i|others, data)$. Applying the Gibbs sampling technique, we can estimate the posterior distributions $f(\lambda_{i1}|data)$, $f(v_i|data)$, $i = 1, \cdots, I$, $f(b|data)$, $f(\mu|data)$ and $f(\sigma^2|data)$.

4.5 Choice of Prior Distributions

When one specifies The parameters in the third level of the hierarchical model, they will reflect the prior belief about the hazard rates of the standard therapy, the relative therapy effects, and their heterogeneity across centers. To choose their values for a real problem, one needs to use the knowledge about the therapies before the trial. Prior knowledge about the standard therapy is usually expressed on λ_{i1}, the marginal prior distribution of λ_{i1} is

$$\begin{aligned} f(\lambda_{i1}|c,d) &= \int_0^\infty f(\lambda_{i1}|a,b)f(b|c,d)db \\ &= \int_0^\infty \frac{b^a}{\Gamma(a)}\lambda_{i1}^{a-1}e^{-b\lambda_{i1}}\frac{d^c}{\Gamma(c)}b^{c-1}e^{-bd}db \\ &= \frac{d^c\lambda_{i1}^{a-1}}{\Gamma(a)\Gamma(c)}\int_0^\infty b^{a+c-1}e^{-b(d+\lambda_{i1})}db \\ &= \frac{d^c\Gamma(a+c)}{\Gamma(a)\Gamma(c)}\frac{\lambda_{i1}^{a-1}}{(d+\lambda_{i1})^{a+c}} \end{aligned}$$

prior mode of λ_{i1} is $\hat{\lambda}_{i1} = \frac{(a-1)d}{c+1}$, $i = 1, \cdots, I$.

The prior means of $\lambda_{i1}, v_i,\ i = 1, \cdots, I$ are

$$\begin{aligned} E(\lambda_{i1}|c,d) &= E(E(\lambda_{i1}|b)|c,d) \\ &= E(\frac{a}{b}\,|\,c,d) \\ &= \frac{ad}{c-1} \\ E(v_i|u,w) &= E(E(v_i|\mu,\sigma^2)|u,w) \\ &= E(\mu|u,w) \\ &= 0 \end{aligned}$$

4.6 Predictive Survival Functions

Information on the patients in the trial is analyzed to compute the predicted survival functions for each therapy in every center participated in the trial. Let X_{i1} and X_{i2} denote the survival times of a new patient on therapy 1 and therapy 2 in center i. The predictive survival functions on the two therapies in center i, $i = 1, \cdots, I$ are

$$\begin{aligned} S_{i1}(t) &= P(X_{i1} > t \,|\, data) \\ &= \int_0^\infty P(X_{i1} > t \,|\, \lambda_{i1}) f(\lambda_{i1} \,|\, data) d\lambda_{i1} \\ &= \int_0^\infty e^{-\lambda_{i1} t} f(\lambda_{i1} \,|\, data) d\lambda_{i1} \end{aligned}$$

After applying Gibbs sampling technique, we have a random sample from any marginal or joint posterior distribution. We can calculate the above predictive survival function by applying Monte Carlo integration method.

Chapter 5

The NSCLC Study with a Hierarchical Model

5.1 Introduction

In Chapter 3, the phase III clinical trial conducted by CALGB for comparing two therapies was presented. It was multicenter trial, but we regarded all the observations in one therapy group as exchangeable, ignoring possible center heterogeneity.

In this chapter, I will consider center heterogeneity. The model proposed in Chapter 4 will be used to analyze the trial data. Analysis will focus on the information available in 1992.

In the CALGB trial, the standard therapy was radiotherapy (RT) alone, and the experimental therapy was chemotherapy followed by radiotherapy (CT+RT). By May 1987, the trial accrued a total of 155 patients, and 22 medical centers participated in the trial. The total number of patients in each of these centers varied from 1 to 22. Many centers had only a few patients each.

Center	n_{i1}	d_{i1}	T_{i1}	n_{i2}	d_{i2}	T_{i2}
1	1	0	83.77	1	1	16.53
2	4	3	87.23	1	1	9.47
3	12	12	189.73	10	7	323.37
4	11	10	203.80	11	10	200.67
5	2	2	16.60	1	1	48.07
6	3	2	44.33	4	3	103.70
7	4	4	26.67	3	2	82.23
8	2	2	22.77	2	2	26.03
9	2	2	24.50	2	1	81.43
10	4	3	83.40	4	4	56.43
11	6	6	59.93	6	3	211.97
12	1	1	2.30	3	3	33.70
13	7	6	147.77	11	11	106.00
14	6	6	40.90	7	7	110.17
15	1	1	3.57	1	1	47.83
16	3	3	27.40	3	3	45.97
17	1	1	8.30	1	0	56.67
18	1	1	1.27	2	2	18.53
19	3	3	33.93	3	2	91.03
20	2	2	10.03	0	0	0
21	0	0	0	1	1	12.03
22	1	1	17.50	1	0	55.77
total	77	71	1135.7	78	65	1737.6

Table 5.1: Sufficient statistics by center

5.2 Data

The sufficient statistics $d_{ij},\ T_{ij},\ j = 1,2,\ i = 1,\cdots,I$, represent the total number of uncensored observations and the total exposure time observed on therapy j in center i. $j = 1$ denotes the patients treated with RT only, and $j = 2$ denotes the CT+RT group. Table 5.1 displays d_{ij}, T_{ij} and n_{ij} (the total number of patients) from all the centers. The time unit is one month. Only five centers have more than 10 patients.

5.3 Choice of Prior Distributions

Before applying the hierarchical model introduced in Chapter 4, we must specify the parameter a in the second level and the parameters in the third level distributions of the hierarchy. These parameters reflect our prior belief about the value of the hazard rate for the RT therapy and the log hazard ratio, and also the heterogeneity between different centers. The first analysis is performed with values of a, c, d, u and w that are chosen according to our prior knowledge. Then analyses will be performed with other values for these parameters to determine the sensitivity of the analysis to the prior distribution.

In this trial, our prior belief is that there would not be a large variation among the v_i's across centers. That means σ is small. So we choose $u = 3$ and $w = 0.32$, which gives an IG(3,0.32) prior distribution to σ^2, which has prior mean and variance of

$$E(\sigma^2 \mid u, w) = \frac{w}{u-1} = 0.16$$

$$V(\sigma^2 \mid u, w) = \frac{w^2}{(u-1)^2(u-2)} = 0.0256$$

The gamma distribution of λ_{i1} in the second level is conjugate to the exponential distribution in the first level. From the posterior distribution of λ_{i1}, we see that a acts as a prior number of uncensored observations and b is like the prior total exposure time for the patients treated with RT only. The information available before this trial suggests that the median survival time on RT is 8 to 10 months, with a two-year survival rate of 10 to 20 percent and a three-year survival rate of 5 to 10 percent. This suggests that λ_{i1} is likely to be approximately 0.075. We set a at 6, and choose $c = 4$ and $d = 0.075$ to

approximate this information; that gives a Gamma$(4, 0.075)$ prior distribution to b, and the prior mode of λ_{i1} is $\hat{\lambda}_{i1} = \frac{(a-1)d}{c+1} = 0.075$. Prior distributions are

$$\begin{aligned} a &= 6 \\ b &\sim Gamma(4, 0.075) \\ \mu &\sim N(0, 1) \\ \sigma^2 &\sim IG(3, 0.32) \end{aligned}$$

Given a and b, expectations of the population mean and variance of λ_{i1} are

$$\begin{aligned} E(\lambda_{i1} \,|\, a, b) &= \frac{a}{b} \\ V(\lambda_{i1} \,|\, a, b) &= \frac{a}{b^2} \end{aligned}$$

Figure 5.1 presents the prior distributions of a/b, a/b^2, μ and σ^2.

5.4 Posterior Distributions

The results are presented in the following tables and figures. Figure 5.2 displays posterior distributions of v_i at those 22 centers and a random new center (Center 23). In the model, $v_i = 0$ indicates $\lambda_{i1} = \lambda_{i2}$, and that means the RT only regimen and the CT+RT regimen perform equally well at center i. In many centers, zero is close to the middle of the posterior distribution of v_i. There is not enough evidence to conclude which therapy is better in many centers because of small sample sizes in these centers. For the same reason, the posterior distributions of v_i in many centers are disperse. In several centers with relatively larger sample size, the posterior densities of v_i are more concentrated than in other centers with smaller sample sizes.

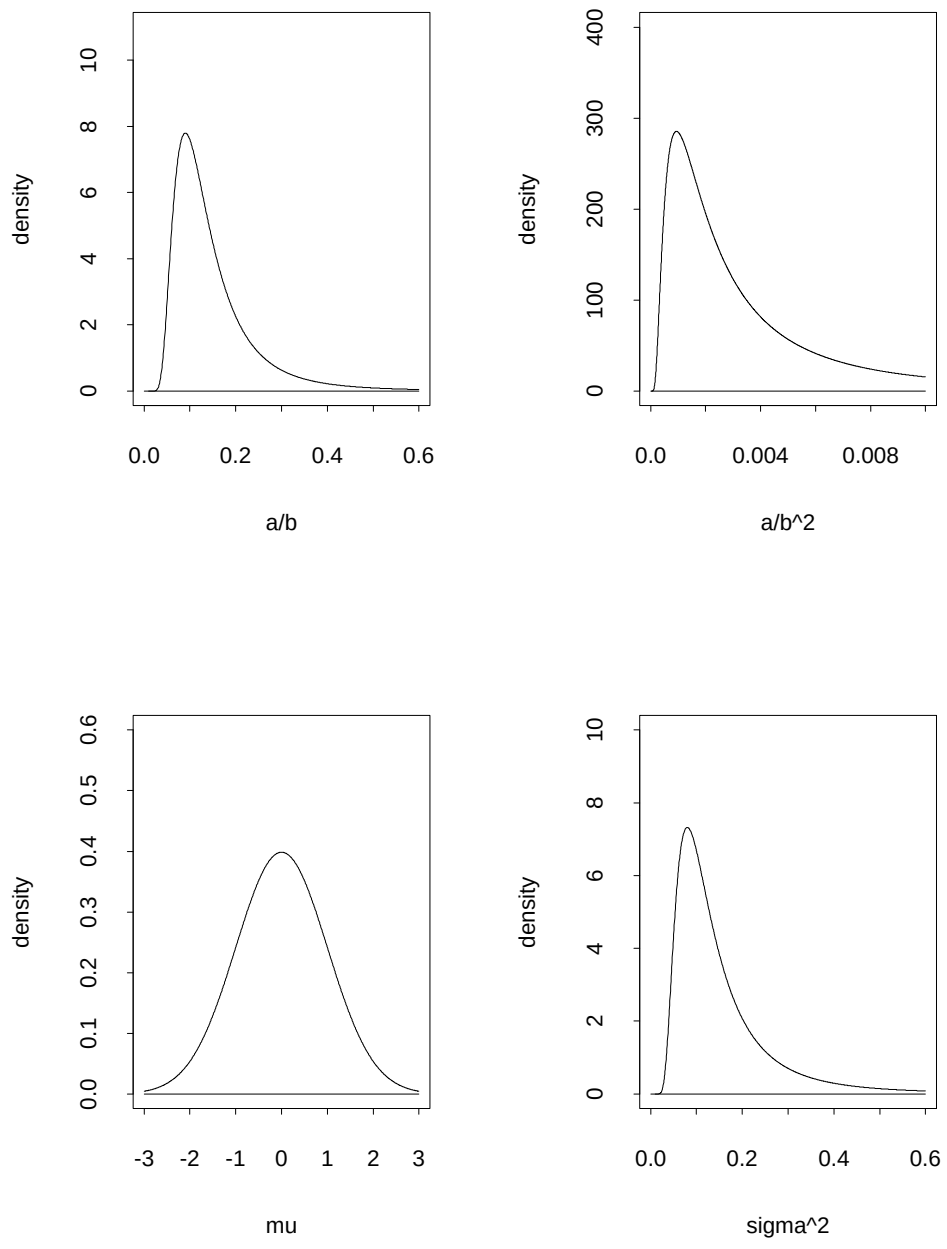

Figure 5.1: Prior distributions of a/b, a/b^2, μ and σ^2

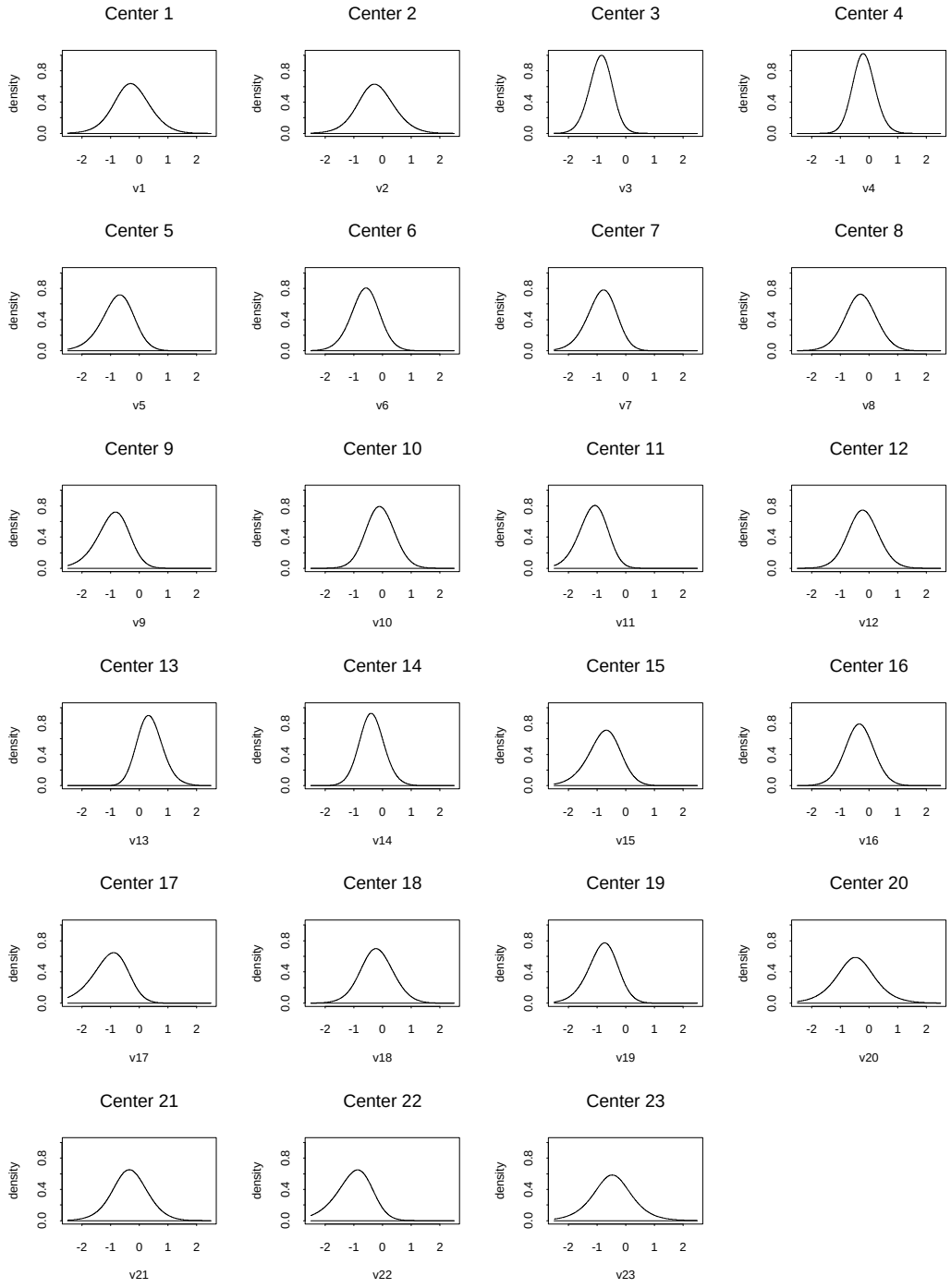

Figure 5.2: Posterior distribution of v_i at each center

Figure 5.3 displays the posterior distributions of $\lambda_{i1},\ i = 1, \cdots, 22$. Similar to the posterior distributions of v_i, posterior distributions of λ_{i1} are disperse in many centers because of small sample size. In several centers with relative larger sample size, The posterior distributions of λ_{i1} are more concentrated than in other centers with smaller sample sizes.

Without hierarchical structure, parameters in different individual centers are independent. The maximum likelihood estimators of v_i and λ_{i1} are

$$\begin{aligned} \lambda_{i1\,MLE} &= \frac{d_{i1}}{T_{i1}} \\ v_{i\,MLE} &= ln(\frac{d_{i2}/T_{i2}}{d_{i1}/T_{i1}}), \qquad i = 1, \cdots, 22 \end{aligned}$$

In a multicenter hierarchical model, statistical inference about each individual center not only depends on the information from that center, but also on the information from all the other centers. Information from the other centers will affect the inference about that center because of the hierarchical structure. That is usually called "borrowing strength." A hierarchical model will pull the parameters in individual centers to each other. Extreme or small centers borrow more strength than other centers do.

For a given b, the mode of λ_{i1} in population is $(a-1)/b$, when we substitute b in its posterior mode $\hat{b}$, our estimator of the mode of λ_{i1} in population is $(a-1)/\hat{b} = 0.057$. For a given μ, the mode of v_i in population is μ, when we substitute μ in its posterior mode $\hat{\mu}$, our estimator of the mode of v_i in population is $\hat{\mu} = -0.472$.

Table 5.2 displays individual center log hazard ratios (their MLE), the posterior mode of v_i ($\hat{v}_i$), the pooled population log hazard ratio (posterior mode $\hat{\mu}$), the posterior mean and variance of v_i, and the posterior probability

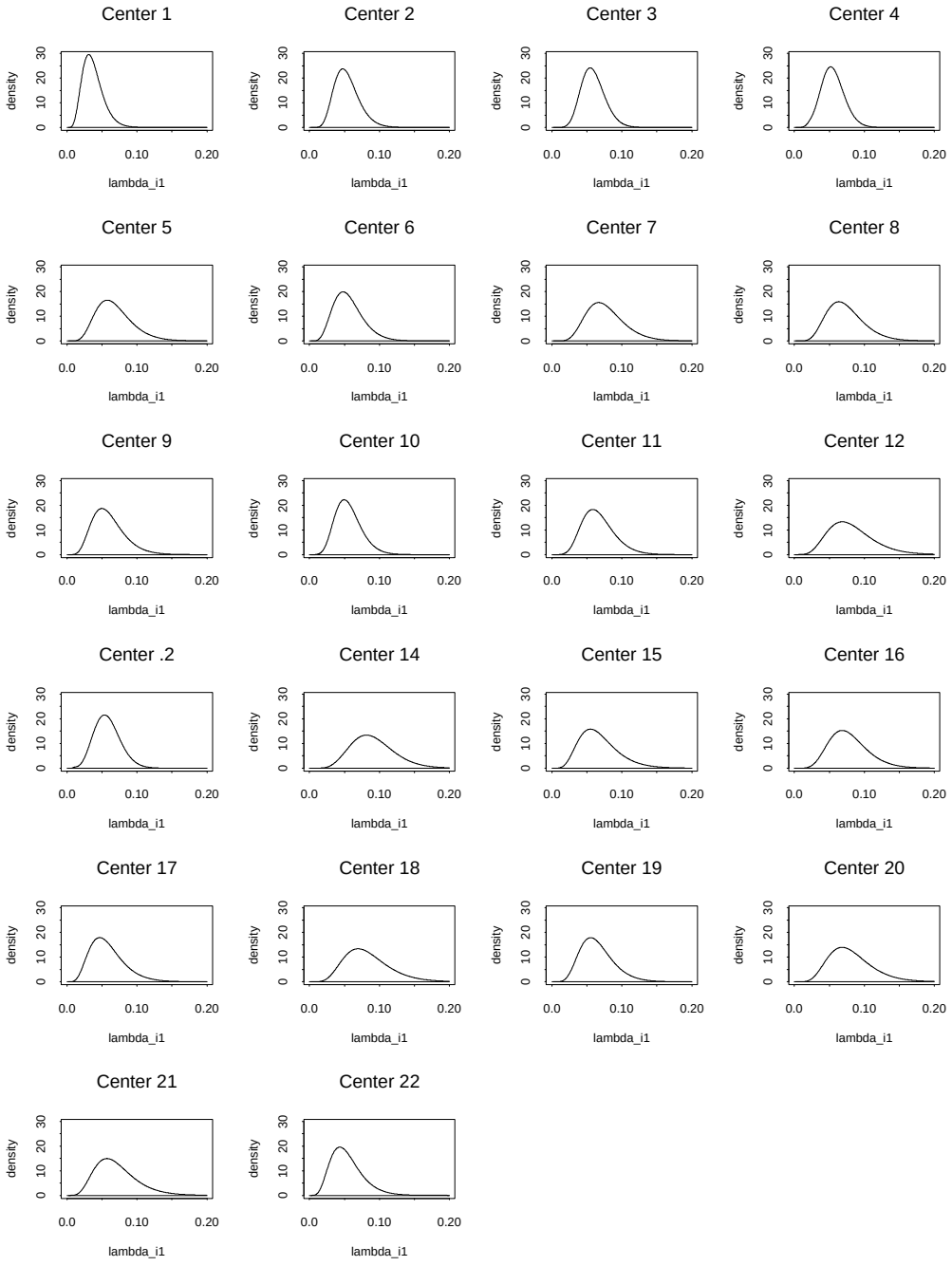

Figure 5.3: Posterior distribution of λ_{i1} at each center

Center	v_{iMLE}	$\hat{v}_i$	$\hat{\mu}$	$\mathrm{E}(v_i \mid D)$	$\mathrm{V}(v_i \mid D)$	$P(v_i < 0 \mid D)$
1	∞	-0.302	-0.472	-0.266	0.442	0.637
2	1.122	-0.281	-0.472	-0.234	0.454	0.636
3	-1.072	-0.844	-0.472	-0.877	0.166	0.931
4	0.015	-0.202	-0.472	-0.173	0.160	0.598
5	-1.756	-0.680	-0.472	-0.777	0.345	0.856
6	-0.444	-0.574	-0.472	-0.607	0.268	0.817
7	-1.819	-0.776	-0.472	-0.854	0.284	0.898
8	-0.134	-0.298	-0.472	-0.295	0.327	0.636
9	-1.894	-0.828	-0.472	-0.943	0.338	0.909
10	0.678	-0.106	-0.472	-0.080	0.264	0.537
11	-1.956	-1.079	-0.472	-1.152	0.255	0.963
12	-1.586	-0.222	-0.472	-0.201	0.306	0.619
13	0.938	0.323	-0.472	0.385	0.210	0.270
14	-0.837	-0.397	-0.472	-0.373	0.199	0.719
15	-2.595	-0.679	-0.472	-0.772	0.352	0.850
16	-0.517	-0.339	-0.472	-0.339	0.273	0.699
17	$-\infty$	-0.895	-0.472	-1.060	0.410	0.931
18	-1.987	-0.227	-0.472	-0.201	0.352	0.605
19	-1.392	-0.738	-0.472	-0.818	0.294	0.874
20	$-$	-0.473	-0.472	-0.474	0.572	0.743
21	$-$	-0.348	-0.472	-0.337	0.422	0.681
22	$-\infty$	-0.868	-0.472	-1.025	0.412	0.921

Table 5.2: The MLE, posterior mode, pooled population mode, posterior mean, variance of v_i and $P(v_i < 0|D)$ from all centers

of $v_i < 0$ for each center.

The parameters for each individual center are pulled toward each other, as is clearly shown in columns 2 to 4 in Tables 5.2. For all the centers that their center MLE of v_i (column 2) is greater than the pooled population log hazard ratio (column 4), the posterior modes of v_i in these centers are pulled down from their center's log hazard ratio. For all centers with MLE of v_i is less than the pooled population value, the posterior mode of v_i in these centers are pulled up from their center's log hazard ratio. The center-to-center variation

between the posterior modes of v_i is much less than that between the MLEs of v_i, because the hierarchical model connects v_i in individual centers together and pulls them closer.

Table 5.3 shows the individual center hazard rates on the RT therapy (their MLE), the posterior mode of λ_{i1} ($\hat{\lambda}_{i1}$), the pooled population hazard rate ($(a-1)/\hat{b}$), and the posterior mean and variance of λ_{i1}. In the same manner as the $v_i's$, the individual center hazard rates on the RT therapy are pulled toward the pooled population hazard rate on the RT therapy. For all the centers that their center MLE of λ_{i1} are greater than the pooled population value, their posterior modes of λ_{i1} are pulled down toward the pooled population value. For all the centers that their center MLE of λ_{i1} are less than the pooled population value, their posterior modes of λ_{i1} are pulled up toward the pooled population vale. The variation between the posterior modes of λ_{i1} in different centers is much less than that between the MLE of λ_{i1} in different centers, because the hierarchical model connects λ_{i1} in different centers together and pulls them closer.

Let us now see how this shrinkage happens at some individual centers. In center 1, there are two patients, one treated with RT only and one receiving CT+RT. The latter patient died after 16.5 months in the study, and the former patient was still alive after 83.8 months in the study. Based only on the information from center 1, it seems that patients can survive much longer on a RT only regimen than on a CT+RT regiman, and that the maximum likelihood estimators of v_1 and λ_{11} are

$$v_{1\,MLE} = 1.6, \quad \lambda_{11\,MLE} = 0$$

However, in the hierarchical model, the evidence from the other centers suggests

Center	$\lambda_{i1\,MLE}$	$\hat{\lambda}_{i1}$	$(a-1)/\hat{b}$	$\mathrm{E}(\lambda_{i1} \mid D)$	$\mathrm{V}(\lambda_{i1} \mid D)$
1	0	0.031	0.057	0.037	0.00021
2	0.034	0.047	0.057	0.054	0.00032
3	0.063	0.055	0.057	0.059	0.00028
4	0.049	0.052	0.057	0.055	0.00027
5	0.120	0.058	0.057	0.068	0.00070
6	0.045	0.048	0.057	0.056	0.00046
7	0.150	0.069	0.057	0.077	0.00076
8	0.088	0.064	0.057	0.074	0.00074
9	0.082	0.050	0.057	0.060	0.00055
10	0.036	0.050	0.057	0.056	0.00036
11	0.100	0.059	0.057	0.067	0.00054
12	0.435	0.069	0.057	0.082	0.00108
13	0.041	0.054	0.057	0.058	0.00035
14	0.147	0.081	0.057	0.091	0.00096
15	0.280	0.055	0.057	0.068	0.00080
16	0.109	0.068	0.057	0.078	0.00078
17	0.012	0.047	0.057	0.058	0.00062
18	0.787	0.069	0.057	0.083	0.00109
19	0.088	0.055	0.057	0.065	0.00059
20	0.200	0.068	0.057	0.081	0.00098
21	–	0.057	0.057	0.071	0.00090
22	0.057	0.043	0.057	0.054	0.00052

Table 5.3: The MLE, posterior mode, pooled population mode, posterior mean and variance of λ_{i1} from all centers

that patients survive longer when treated with CT+RT than when treated with RT only. The posterior mode of v_1 is $\hat{v}_1 = -0.302$, it is shrunk toward the pooled population log hazard ratio from the individual center value $v_{1MLE} = 1.6$, and the posterior distribution of v_1 indicates that v_1 is more likely to be negative than to be positive ($P(v_1 < 0|D) = 0.637$).

Because center 1 is a small center, the posterior distribution of v_1 is greatly affected by the information from the other centers. Similar shrinkage happens to λ_{11}, and its posterior mode of 0.021 shrank toward the pooled population hazard rate for patients receiving RT only. This kind of strong "*strength borrowing*" happens in many other small centers, too.

Take center 3 as an example of the large centers. Center 3 has 12 patients treated with RT only and 10 patients received CT+RT. the sufficient statistics are

$$d_{31} = 12,\ T_{31} = 189.73,\ d_{32} = 7,\ T_{32} = 323.37$$

Based on the information from this center, the maximum likelihood estimators of v_3 and λ_{31} are

$$v_{3MLE} = -1.07, \quad \lambda_{31MLE} = 0.063$$

v_{3MLE} is very small, and $v_3 = -1.07$ means that the mean survival time for patients treated with CT+RT is as much as 2.9 times that of those received RT only. Evidence from the other centers suggests that the population log hazard ratio would not be so small, so the hierarchical model pulled v_3 up toward the pooled population log hazard ratio. The resulting posterior mode of v_3 is -0.844, up from its individual center value. For λ_{31}, its MLE is above the population hazard rate for patients treated with RT only, but information from other centers pulled $\hat{\lambda}_{31}$ toward the population value. The resulting posterior

mode is 0.055, down from the individual center value. Since center 3 has a relatively larger number of patients, the inference about this center is relatively stronger, and the posterior variances of v_i and λ_{i1} in this center are smaller than those in small centers.

Let us now look at another large center. Center 13 has 7 patients on the RT only regimen and 11 patients on the CT+RT regimen. The sufficient statistics are

$$d_{13,1} = 6,\ T_{13,1} = 147.77,\ d_{13,2} = 11,\ T_{13,2} = 106.0$$

Based on the information in this center, the maximum likelihood estimators of v_{13} and $\lambda_{13,1}$ are

$$v_{13MLE} = 0.938, \quad \lambda_{13,1\,MLE} = 0.041$$

v_{13MLE} is far greater than the pooled population log hazard ratio, which indicates that a patient's mean survival time on CT+RT is less than that on RT only. $v_{13} = 0.938$ means that in center 13 a patient's mean survival time on CT+RT is only 39% of that on RT only. After borrowing strength from other centers, v_{13} is pulled down toward the pooled population log hazard ratio. The posterior mode of v_{13} is 0.323, down from its individual center value. $v_{13} = 0.323$ means that in center 13 a patient's mean survival time on CT+RT is about 72% of that on RT only. The individual center value of λ_{131} is below the pooled population value, but information from other centers pulled λ_{131} up toward the population value. The posterior mode of λ_{131} is 0.054, up from the MLE of λ_{131}.

The solid curves in Figure 5.4 show posterior distributions of population parameters a/b, a/b^2, μ and σ^2. The posterior distributions of a/b and μ are much more concentrated than their prior distributions, indicating that the

Parameter	Mode	Mean	Variance
a/b	0.066	0.068	0.00011
a/b^2	0.00114	0.00079	6.62×10^{-8}
μ	-0.472	-0.473	0.075
σ^2	0.304	0.514	0.102

Table 5.4: Posterior mode, mean and variance of a/b, a/b^2, μ and σ^2

study contains much information about these parameters. Table 5.4 displays the posterior mode, mean and variance of a/b, a/b^2, μ and σ^2. The posterior mode of μ is -0.472. $\mu = -0.472$ corresponds to a 60% increase in mean survival time for the population on CT+RT compared to the population treated only with RT. With posterior probability $P(\mu < 0|D) = 0.963$, it is very likely that the population mean of v_i is negative.

5.5 Inference in Large Centers

The five centers that have the most patients are centers 3, 4, 11, 13 and 14. The posterior variances of v_i in these centers are smaller than those in other centers. Figure 5.5 displays the posterior distributions of v_i and λ_{i1} in these five centers. This figure shows the degree of heterogeneity among these large centers. The posterior distributions of these v_i's demonstrate some variation. The posterior modes of v_i in these centers vary from -1.079 to 0.323, and the posterior variances of v_i in these centers vary from 0.160 to 0.255. The v_i's in center 3 and 11 are very likely to be negative ($P(v_3 < 0|D) = 0.931$ and $P(v_{11} < 0|D) = 0.963$), whereas v_{13} is likely to be positive ($P(v_{13} > 0|D) = 0.730$), Thus, the CT+RT regimen is very likely to be more effective than the RT only regimen in centers 3 and 11, whereas the RT only regimen is likely

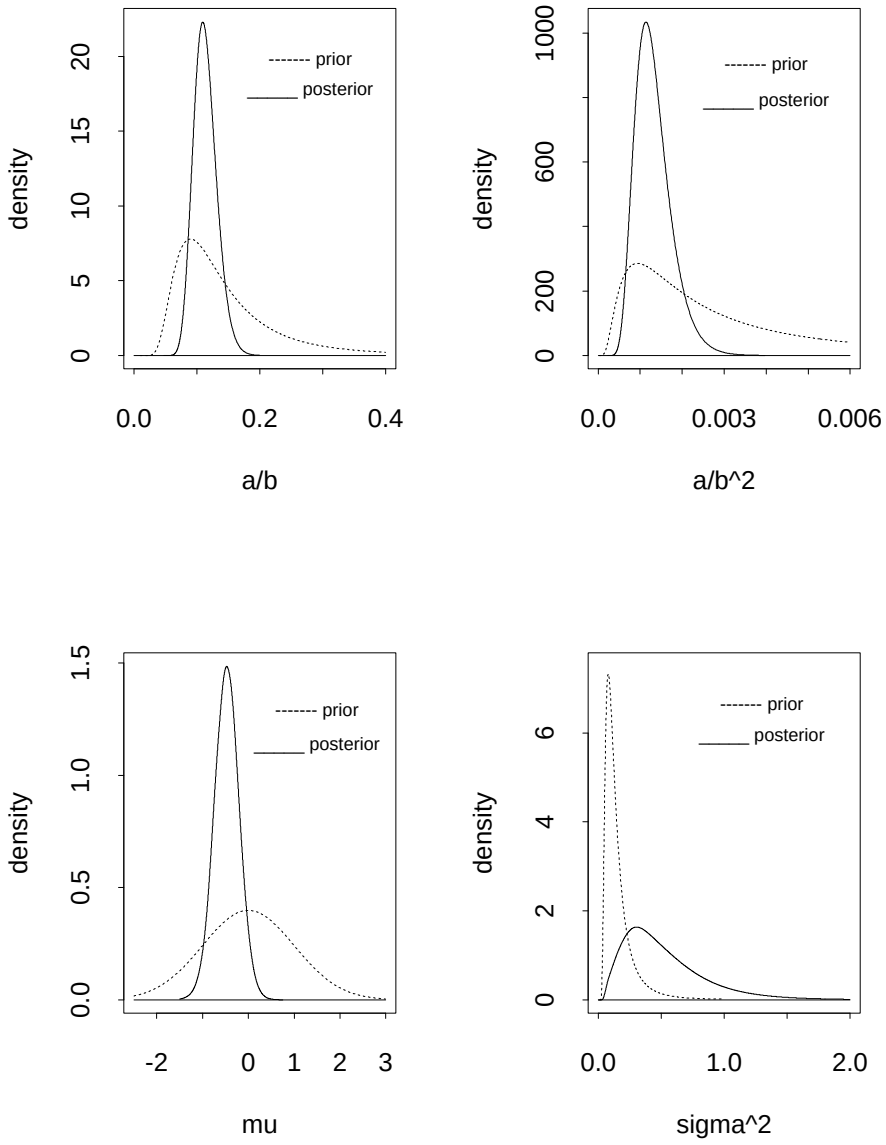

Figure 5.4: Distributions of a/b, a/b^2, μ and σ^2

to be more effective than the CT+RT regimen in center 13. Center 13 is the only center whose v_i is more likely to be positive than negative, which means that center 13 is the only center in which it is likely that the RT only regimen performs better than the CT+RT regimen.

There is less heterogeneity among the $\lambda_{i1}{}'s$ at these centers than there is among the $v_i{}'s$. The posterior modes of λ_{i1} at these centers vary from 0.052 to 0.081, and the posterior variances of λ_{i1} at these centers vary from 0.00027 to 0.00096. The posterior distributions of λ_{i1} at centers 3, 4, 11 and 13 are very close, the posterior modes of λ_{i1} at these four centers change only from 0.052 to 0.059, and the posterior variances of λ_{i1} at these four centers vary from 0.00027 to 0.00054. The posterior distribution of λ_{i1} in center 14 is quite different from those at the other four centers, the posterior mode of λ_{i1} at center 14 is much greater than those at the other four centers, and the posterior distribution of λ_{i1} at center 14 is flatter than those at the other four centers. Therefore, it is quite likely that the hazard rate on the RT only at centers 3, 4, 11, and 13 are very close, and that the λ_{i1} at center 14 is greater than those at the other four centers, which means that the RT only regimen performed equally well at centers 3, 4, 11, and 13, and it performed better at those four centers than in center 14.

5.6 Predictive Survival Functions

The predictive survival functions for patients treated with each therapy are always among the things that a new patient wants to know. Figure 5.6 displays the predictive survival functions on RT only and CT+RT at all 22 centers and a new random center (Center 23).

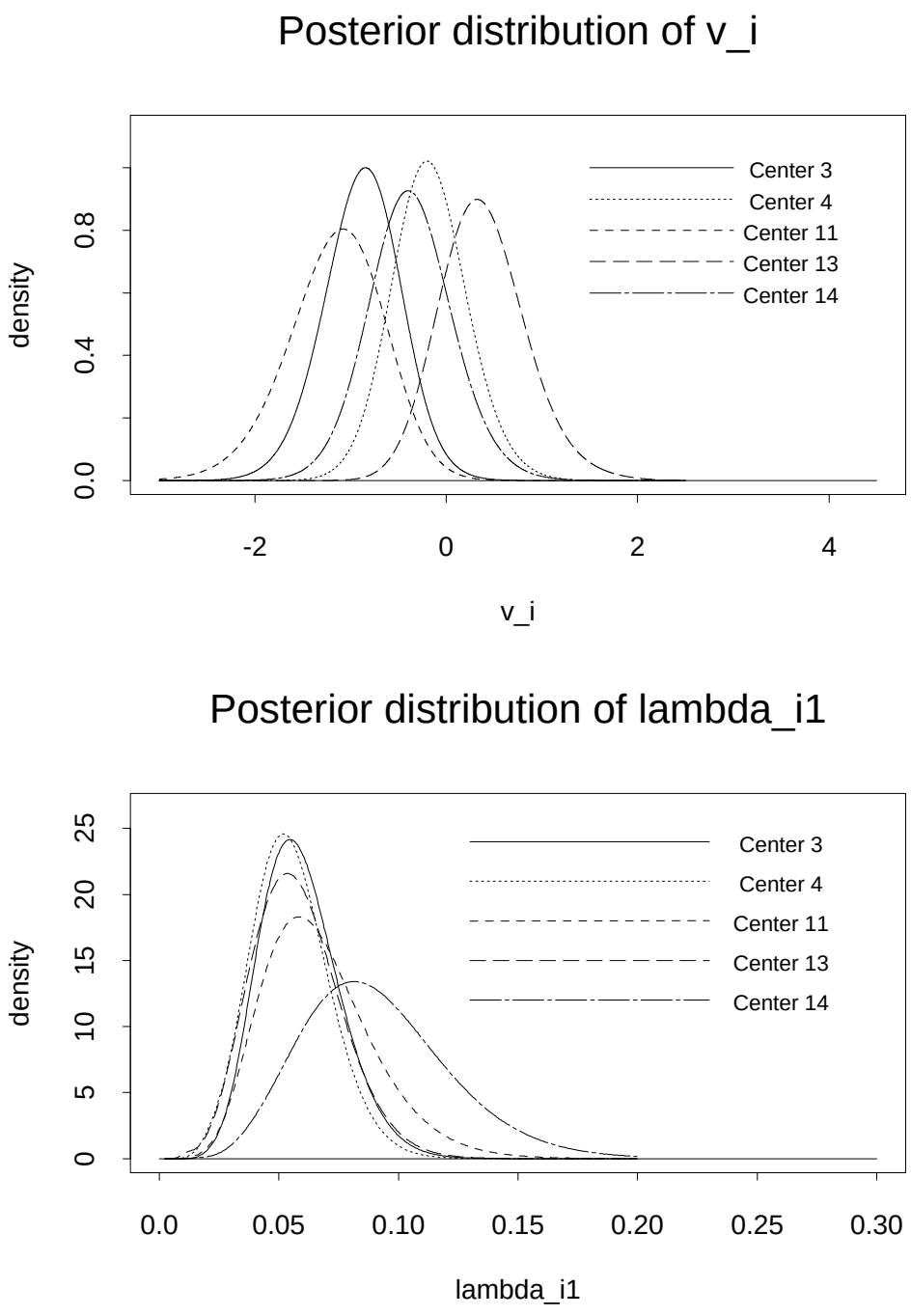

Figure 5.5: Posterior distributions of v_i and λ_{i1} in large centers

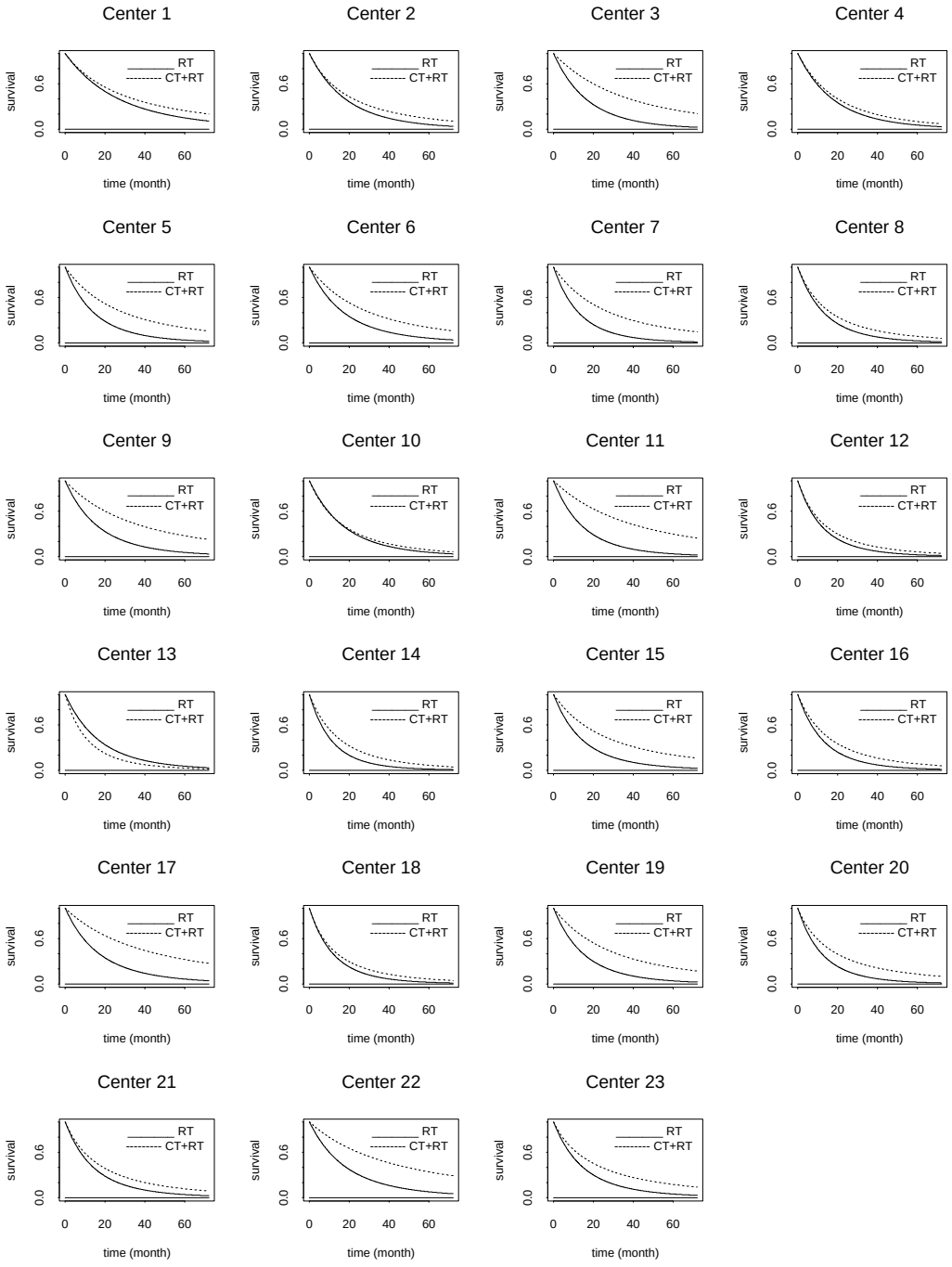

Figure 5.6: Predictive survival functions at each center

There are some variations among the different centers for the difference between the two survival functions at each center. The survival function for patients on CT+RT is always greater than that for those on RT only at all the centers except center 13. There is a large difference between the two survival functions at centers 3, 5, 7, 9, 11, 15, 17, 19 and 22, because the v_i's for these centers are small. For example, in center 3, the probability of surviving up to two years on the RT only regimen is 0.263, whereas it is 0.547 on the CT+RT regimen. At center 11, the two-year survival probability is 0.229 on the RT only regimen and 0.579 on the CT+RT regimen. The difference between the two survival functions on the two therapies is greatest at center 11, because v_{11} is likely to be the smallest among all the $v_i,\ i = 1, \cdots, 22$.

Center 13 is a relatively large center, and it is the only center at which the RT only regimen likely performs better than the CT+RT regimen. The reason could be the randomness in sampling; or it might be that the patients who received CT+RT were in worse condition than those treated with the RT therapy; or perhaps, center 13 performed the therapies slightly differently from the other centers etc. Certainly, a careful review of the patients at center 13 and how this center performs the therapies should be performed.

Table 5.5 displays the predicted survival probabilities of six months, one year, two years and three years on the RT therapy in all centers.

Table 5.6 displays the predicted survival probabilities of six months, one year, two years and three years for patients treated with the CT+RT therapy at each center.

	Predicted survival probability			
Center	six months	one year	two years	three years
1	0.806	0.654	0.439	0.302
2	0.729	0.537	0.300	0.173
3	0.706	0.504	0.263	0.142
4	0.723	0.527	0.288	0.162
5	0.671	0.459	0.228	0.120
6	0.717	0.522	0.289	0.168
7	0.637	0.415	0.187	0.091
8	0.647	0.428	0.199	0.098
9	0.704	0.505	0.271	0.154
10	0.718	0.521	0.284	0.161
11	0.675	0.464	0.229	0.120
12	0.622	0.400	0.180	0.088
13	0.712	0.514	0.276	0.155
14	0.595	0.365	0.149	0.067
15	0.672	0.463	0.234	0.127
16	0.634	0.412	0.185	0.090
17	0.710	0.515	0.285	0.167
18	0.618	0.394	0.174	0.083
19	0.680	0.472	0.239	0.128
20	0.623	0.400	0.178	0.086
21	0.666	0.455	0.227	0.122
22	0.730	0.542	0.312	0.189

Table 5.5: Predicted survival probability on the RT therapy

	Predicted survival probability			
Center	six months	one year	two years	three years
1	0.818	0.686	0.505	0.387
2	0.742	0.577	0.376	0.262
3	0.851	0.730	0.547	0.419
4	0.742	0.564	0.344	0.221
5	0.804	0.660	0.464	0.340
6	0.812	0.671	0.475	0.349
7	0.800	0.652	0.452	0.327
8	0.685	0.498	0.292	0.185
9	0.848	0.722	0.549	0.427
10	0.706	0.519	0.306	0.194
11	0.863	0.751	0.579	0.456
12	0.650	0.452	0.247	0.148
13	0.596	0.380	0.176	0.091
14	0.682	0.485	0.268	0.161
15	0.803	0.658	0.464	0.342
16	0.696	0.508	0.296	0.187
17	0.864	0.755	0.590	0.474
18	0.639	0.445	0.245	0.149
19	0.818	0.680	0.486	0.361
20	0.710	0.538	0.343	0.237
21	0.711	0.536	0.334	0.225
22	0.870	0.765	0.605	0.491

Table 5.6: Predicted survival probability on the CT+RT therapy

5.7 Sensitivity to Prior Distributions

5.7.1 Three Prior Distributions

In previous sections, the analyses are based on the set of prior distributions given in section 5.3, with parameters

$$a = 6,\ c = 4,\ d = 0.075,\ u = 3,\ w = 0.32$$

The set of prior distributions is labeled as Prior I. In this section, analyses will be done based on some different prior distributions to see how prior distributions affect analysis result. We choose two other sets of prior distributions, which will be called Prior II and Prior III. The three sets of prior distributions represent different beliefs in the degree of heterogeneity among the different centers.

In Prior II, the parameters are chosen such that our prior belief about center heterogeneity is vague. We choose a small a and uniform distributions for $1/b$ and σ^2. $1/b$ has a uniform prior is equivalent to that b has a prior of $p(b) \propto 1/b^2$. Prior II is

$$\begin{aligned} a &= 3 \\ p(b) &\propto \frac{1}{b^2}, \quad b > 0 \\ \mu &\sim N(0,1) \\ p(\sigma^2) &\propto constant, \quad \sigma^2 > 0 \end{aligned}$$

The prior distributions of b and σ^2 are improper distributions. With this prior, our belief about the the center heterogeneity is flat, and we are open-mind about it.

In Prior III, the parameters are chosen such that our prior belief is that centers are likely homogeneous. The parameters are

$$a = 15, \quad c = 6, \quad d = 0.0375, \quad u = 5 \quad \text{and} \quad w = 0.25$$

that is

$$\begin{aligned} a &= 15 \\ b &\sim Gamma(6, 0.0375) \\ \mu &\sim N(0, 1) \\ \sigma^2 &\sim IG(5, 0.25) \end{aligned}$$

which sets the prior mean of a/b, a/b^2, and the prior mean and variance of σ^2 to be

$$\begin{aligned} E(\frac{a}{b} \,|\, c, d) &= \frac{ad}{c-1} = 0.1125 \\ E(\frac{a}{b^2} \,|\, c, d) &= \frac{ad^2}{(c-1)(c-2)} = 0.0010 \\ E(\sigma^2 \,|\, u, w) &= \frac{w}{u-1} = 0.0625 \\ V(\sigma^2 \,|\, u, w) &= \frac{w^2}{(u-1)^2(u-2)} = 0.0013 \end{aligned}$$

For comparison, Figure 5.7 displays the prior distributions of a/b, a/b^2, μ and σ^2 in Prior I and Prior III. The distributions of population mean a/b in Prior I and Prior III are very close, and the distributions of population mean μ are identical in Prior I and Prior III. The distributions of center population variances a/b^2 and σ^2 are quite different in Prior I and Prior III.

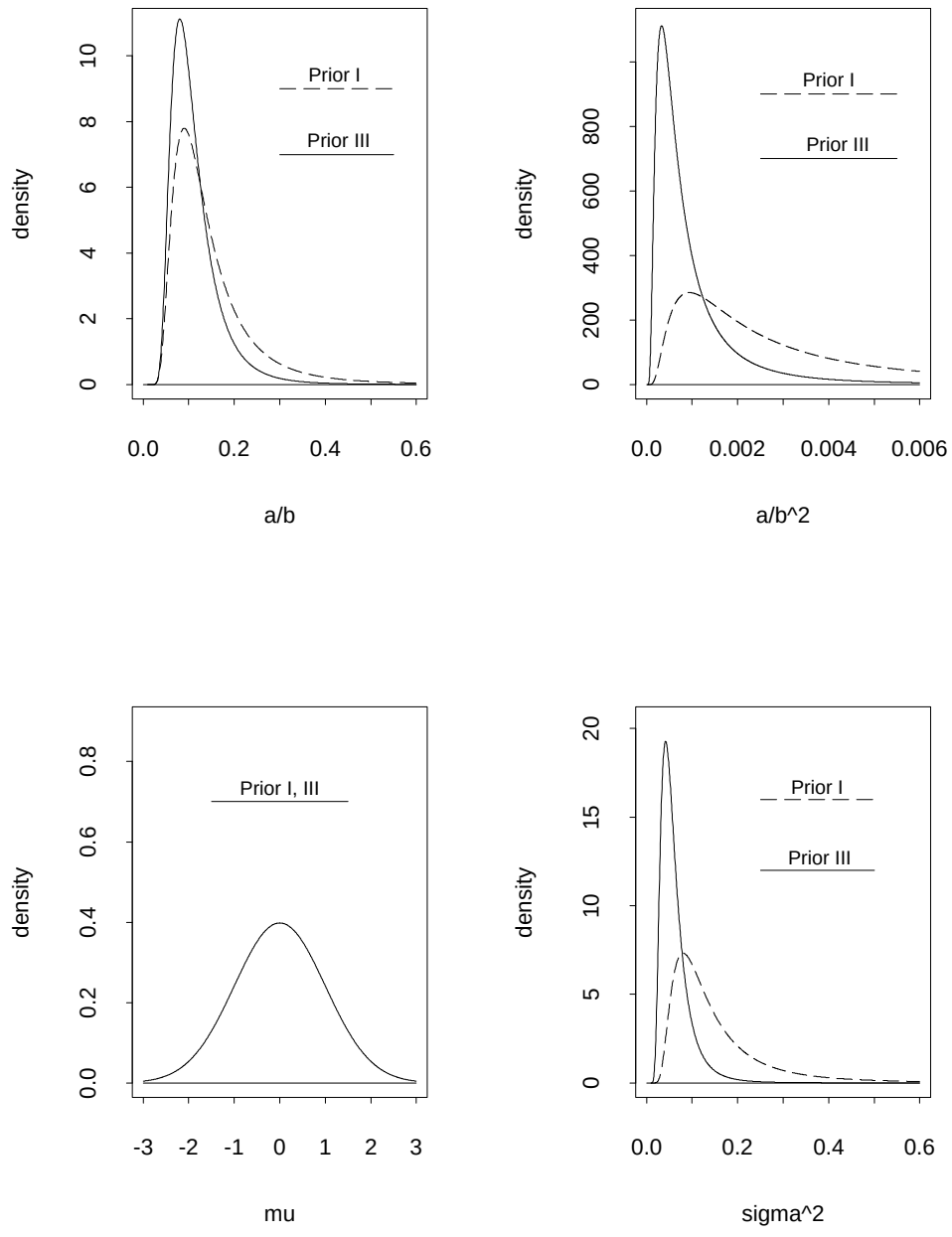

Figure 5.7: Prior distributions of a/b, a/b^2, μ and σ^2

5.7.2 Results

In this section, the posterior distributions based on Prior II and Prior III will be presented. Results based on all the three sets of prior distributions will be compared.

Prior II

Dotted curves in Figures 5.8 and 5.9 display the posterior distributions of v_i based on Prior II for all centers. The prior distributions of $1/b$ and σ^2 are flat in Prior II, so the "*strength borrowing*" is not strong in this case. The posterior distribution of v_i is relatively flat for every center, and there is a relatively large discrepancy among these posterior distributions of v_i for all centers.

Table 5.7 displays the individual center v_i (MLE), posterior mode of v_i ($\hat{v_i}$), pooled population mode ($\hat{\mu}$), posterior mean, posterior variance of v_i and posterior probability of $v_i < 0$ based on Prior II for all the centers. The individual center log hazard ratio shank toward the pooled population value by borrowing strength from other centers, but this shrinkage is not strong, because the prior distribution of σ^2 is flat. The posterior probability $P(v_i < 0|D)$ varies from 0.202 in center 13 to 0.955 in center 11; however, $P(v_i < 0|D)$ is close to 0.5 in many centers, so for many centers, there seems to be no strong evidence to conclude that either therapy is better. Because the number of patients is very small at many centers, and because there is not strong *strength borrowing*, there is no enough information to conclude which therapy performs better at those centers.

The dotted curves in Figures 5.10 and 5.11 display posterior distributions of λ_{i1} based on Prior II for all centers. Table 5.8 displays the individual center

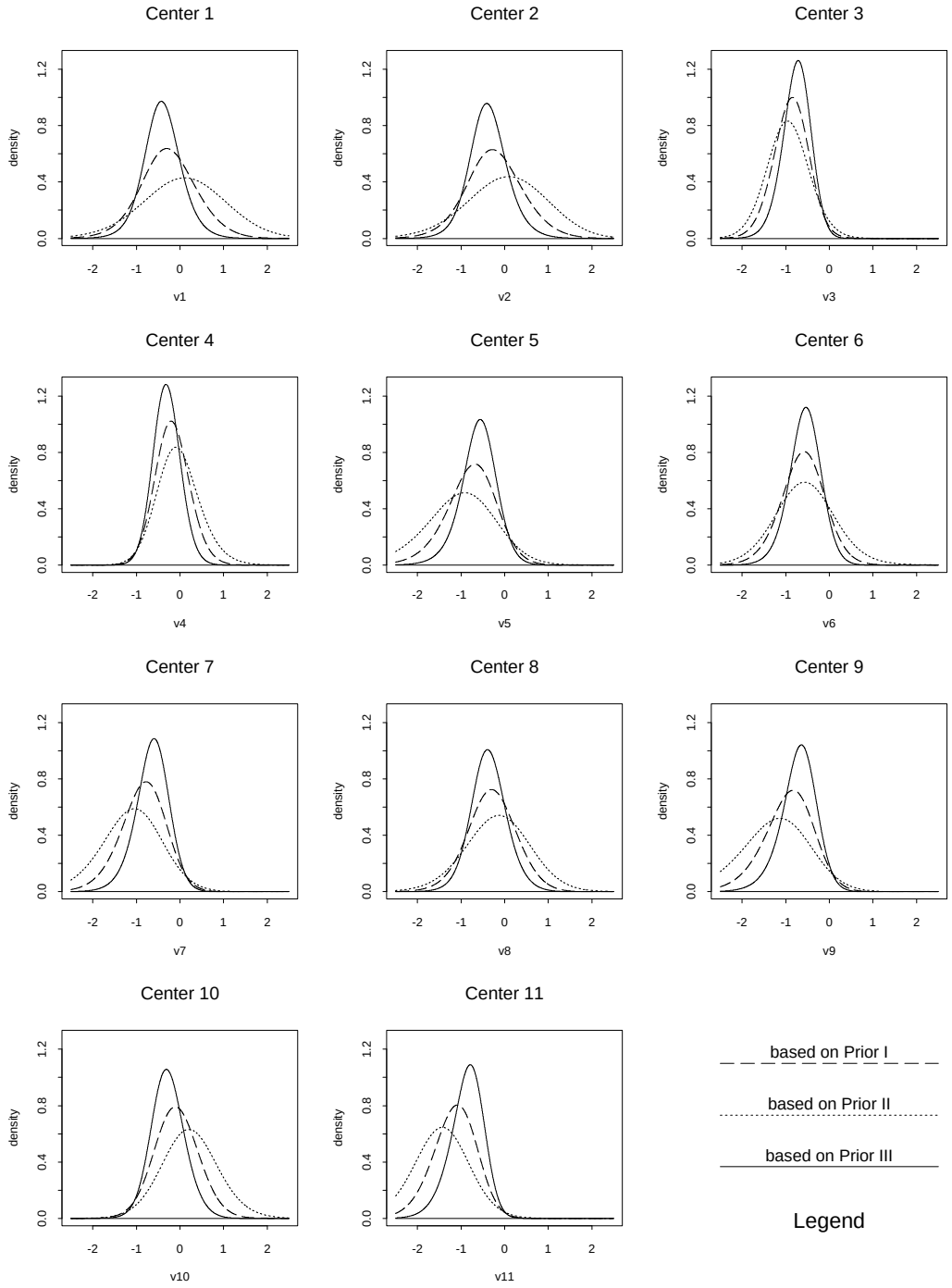

Figure 5.8: Posterior distributions of v_i, $i = 1, \cdots, 11$

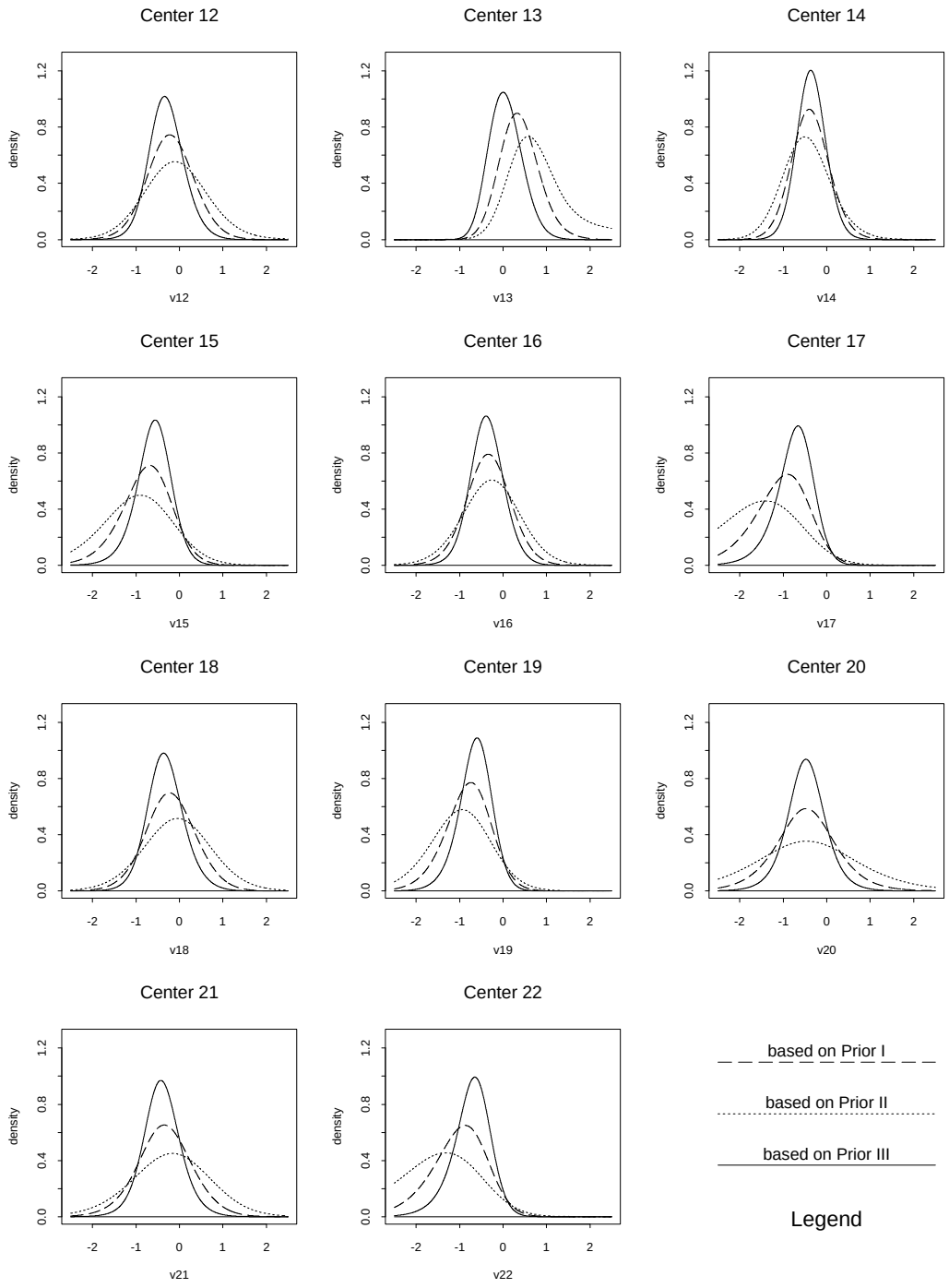

Figure 5.9: Posterior distributions of v_i, $i = 12, \cdots, 22$

Center	v_{iMLE}	$\hat{v}_i$	$\hat{\mu}$	$\mathrm{E}(v_i \mid D)$	$\mathrm{V}(v_i \mid D)$	$P(v_i < 0 \mid D)$
1	∞	0.108	-0.459	0.074	0.903	0.440
2	1.122	0.095	-0.459	0.034	0.855	0.455
3	-1.072	-0.957	-0.459	-0.946	0.241	0.895
4	0.015	-0.093	-0.459	-0.028	0.250	0.531
5	-1.756	-0.914	-0.459	-0.994	0.606	0.816
6	-0.444	-0.567	-0.459	-0.553	0.501	0.706
7	-1.819	-1.043	-0.459	-1.098	0.467	0.867
8	-0.134	-0.128	-0.459	-0.165	0.576	0.539
9	-1.894	-1.132	-0.459	-1.204	0.563	0.881
10	0.678	0.205	-0.459	0.232	0.427	0.389
11	-1.956	-1.427	-0.459	-1.435	0.379	0.955
12	-1.586	-0.104	-0.459	-0.076	0.557	0.522
13	0.938	0.581	-0.459	0.809	0.374	0.202
14	-0.837	-0.501	-0.459	-0.445	0.325	0.708
15	-2.595	-0.896	-0.459	-0.967	0.636	0.813
16	-0.517	-0.258	-0.459	-0.276	0.460	0.574
17	$-\infty$	-1.411	-0.459	-1.442	0.640	0.922
18	-1.987	-0.042	-0.459	-0.054	0.639	0.486
19	-1.392	-0.937	-0.459	-0.986	0.491	0.835
20	$-$	-0.456	-0.459	-0.428	1.293	0.647
21	$-$	-0.157	-0.459	-0.205	0.845	0.533
22	$-\infty$	-1.316	-0.459	-1.376	0.659	0.911

Table 5.7: The MLE, posterior mode, pooled population mode, posterior mean, variance of v_i and $P(v_i < 0|D)$ based on Prior II in all centers

Center	$\lambda_{i1\,MLE}$	$\hat{\lambda}_{i1}$	$(a-1)/\hat{b}$	$\mathrm{E}(\lambda_{i1} \mid D)$	$\mathrm{V}(\lambda_{i1} \mid D)$
1	0	0.018	0.047	0.026	0.00020
2	0.034	0.040	0.047	0.047	0.00037
3	0.063	0.055	0.047	0.060	0.00040
4	0.049	0.047	0.047	0.052	0.00037
5	0.120	0.053	0.047	0.072	0.00136
6	0.045	0.039	0.047	0.052	0.00072
7	0.150	0.068	0.047	0.087	0.00152
8	0.088	0.056	0.047	0.074	0.00130
9	0.082	0.045	0.047	0.061	0.00101
10	0.036	0.040	0.047	0.048	0.00047
11	0.100	0.062	0.047	0.074	0.00096
12	0.435	0.060	0.047	0.087	0.00245
13	0.041	0.042	0.047	0.050	0.00048
14	0.147	0.086	0.047	0.103	0.00196
15	0.280	0.047	0.047	0.072	0.00171
16	0.109	0.064	0.047	0.080	0.00144
17	0.012	0.037	0.047	0.059	0.00122
18	0.787	0.063	0.047	0.070	0.00233
19	0.088	0.052	0.047	0.068	0.00105
20	0.200	0.070	0.047	0.093	0.00200
21	–	0.044	0.047	0.070	0.00176
22	0.057	0.033	0.047	0.050	0.00085

Table 5.8: The MLE, posterior mode, pooled population mode, posterior mean and variance of λ_{i1} based on Prior II in all centers

λ_{i1} (MLE), posterior mode of λ_{i1} ($\hat{\lambda}_{i1}$), pooled population mode ($(a-1)/\hat{b}$), posterior mean and variance of λ_{i1} at each center. The individual center hazard rate was pulled toward the pooled population value by *borrowing strength* from other centers. The posterior mode of λ_{i1} varies from 0.018 in center 1 to 0.086 in center 14. The posterior distribution of λ_{i1} is relatively concentrated in some centers at which there is a relatively large number of patients. There is a relatively large discrepancy among these posterior distributions of λ_{i1} at the different centers.

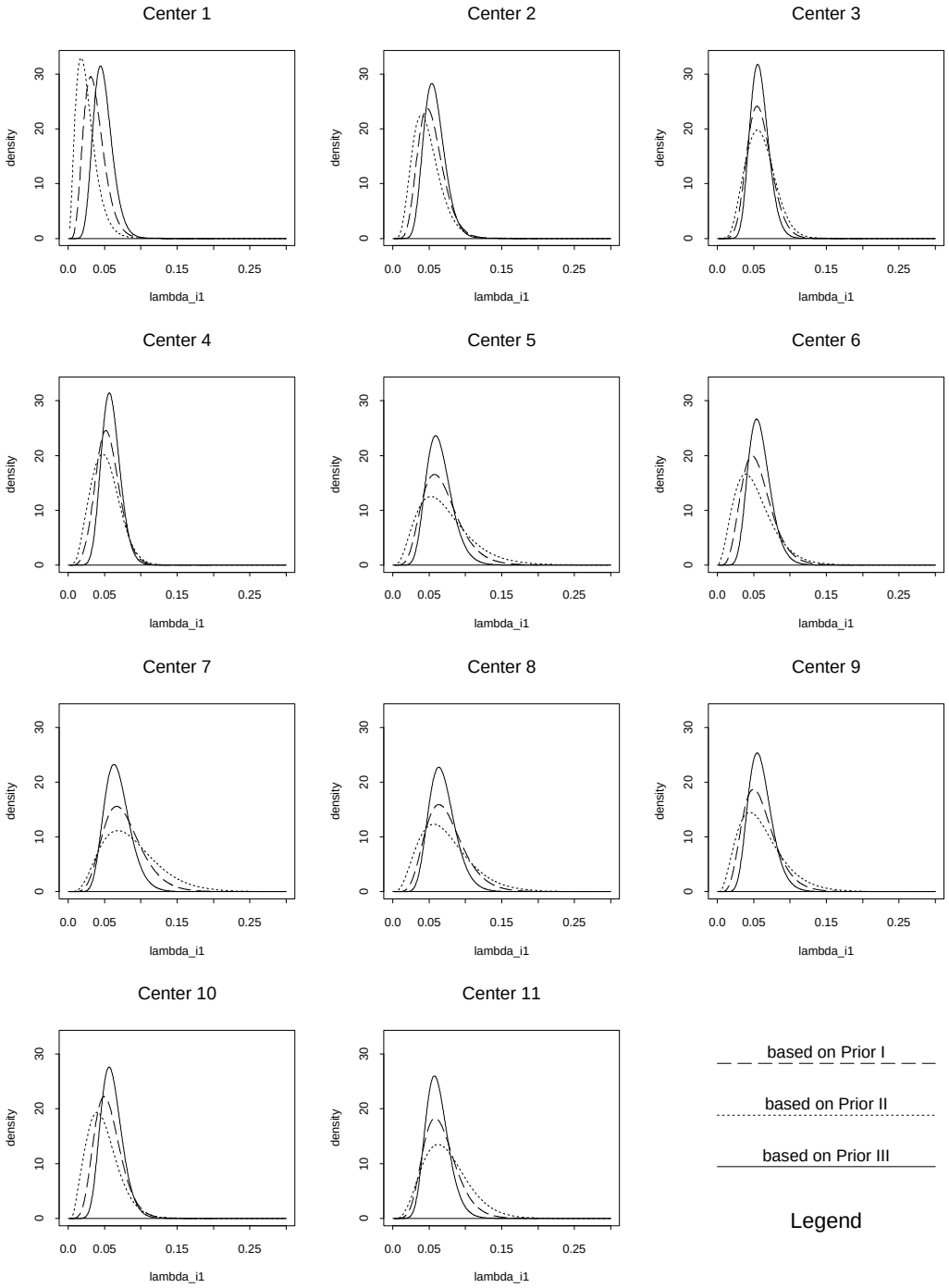

Figure 5.10: Posterior distributions of λ_{i1}, $i = 1, \cdots, 11$

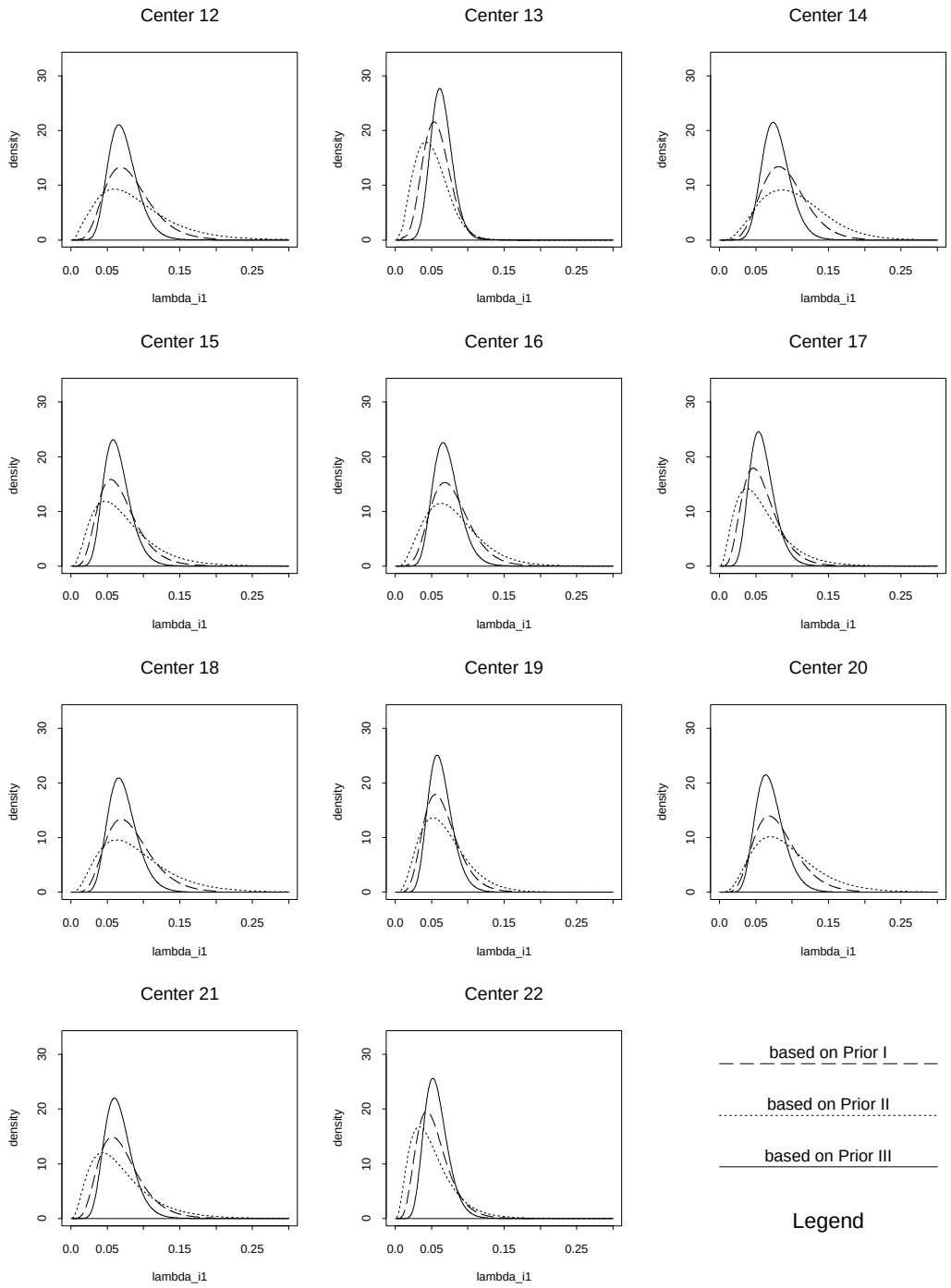

Figure 5.11: Posterior distributions of λ_{i1}, $i = 12, \cdots, 22$

Parameter	Mode	Mean	Variance
a/b	0.065	0.069	0.00022
a/b^2	0.00126	0.00168	5.65×10^{-7}
μ	-0.459	-0.457	0.130
σ^2	1.087	1.578	0.783

Table 5.9: Posterior mode, mean and variance of a/b, a/b^2, μ and σ^2 based on Prior II

The dotted curves in Figure 5.12 display posterior distributions of a/b, a/b^2, μ and σ^2 based on Prior II. Table 5.9 displays the posterior mode, mean and variance of a/b, a/b^2, μ and σ^2. The first two rows represent the population mean and variance of λ_{i1}, and the last two rows represent the population mean and variance of v_i. The posterior mode of a/b is 0.065, which corresponds to a 15.4-month mean survival time on the RT only regiman. The posterior mode of μ is -0.459, which corresponds to a 58% increase in the mean survival time for the patients on therapy 2 over the patients on therapy 1.

Prior III

The solid curves in Figures 5.8 and 5.9 show the posterior distributions of v_i based on Prior III for all centers. The values of a/b^2 and σ^2 are small in Prior III, so the strength borrowing is very strong in this case, and the posterior distribution of v_i is relatively concentrated in every center. The posterior densities of v_i in all centers are very close, and the variations between those posterior densities are very small.

Table 5.10 displays the individual center v_i (MLE), posterior mode of v_i ($\hat{v_i}$), pooled population mode ($\hat{\mu}$), posterior mean, posterior variance of v_i, and posterior probability of $v_i < 0$ based on Prior III for all centers. The

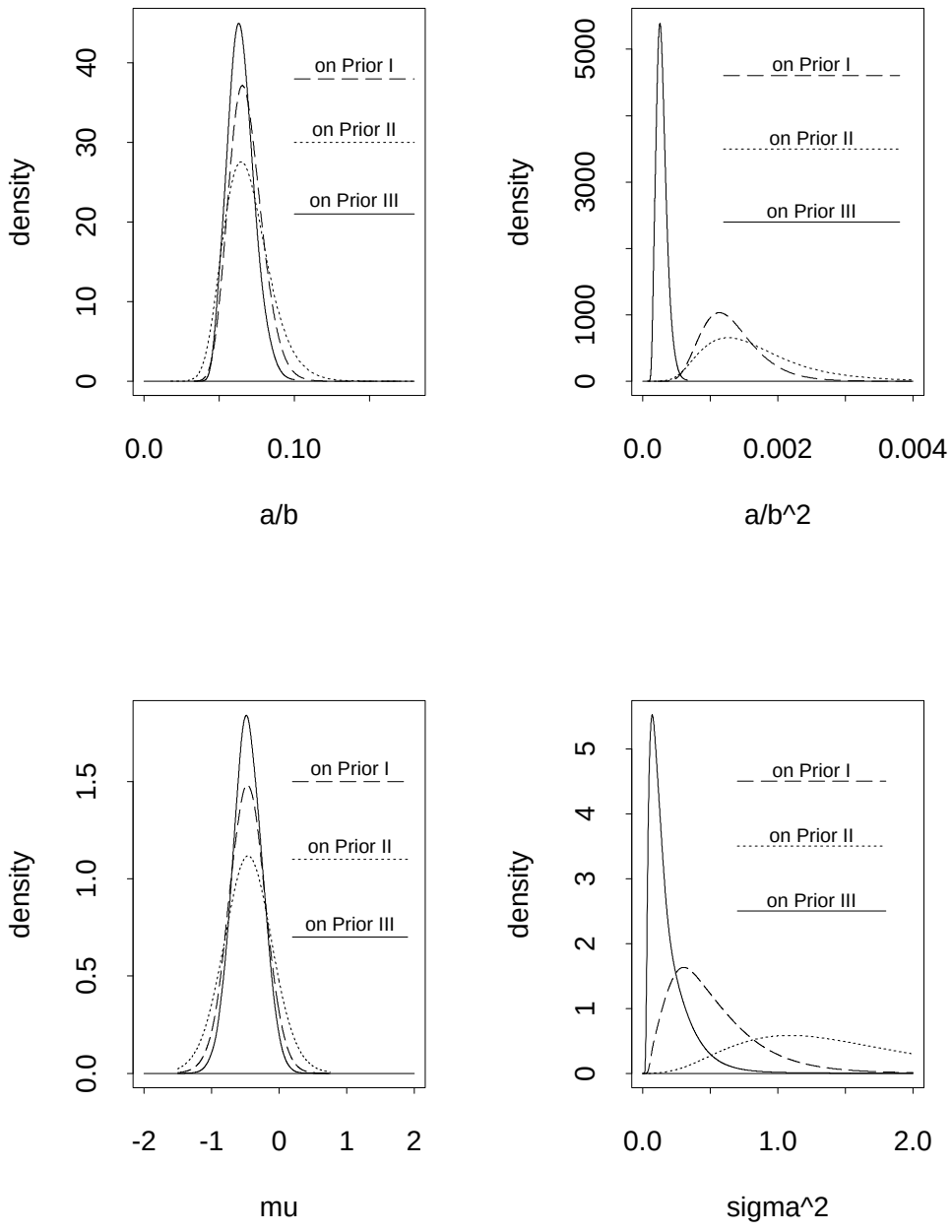

Figure 5.12: Posterior distributions of a/b, a/b^2, μ and σ^2

individual center log hazard ratio shrank toward the pooled population value by borrowing strength from other centers, and that shrinkage was very strong. The individual center MLE of v_i varies from -2.595 to 1.122 (except in the centers without MLE or with an infinity MLE), and the posterior mode of v_i varies only from -0.787 to 0.002. $P(v_i < 0|D)$ is far above 0.5 in all centers except center 13, which indicates that therapy 2 performs better than therapy 1 in all the centers except center 13. The posterior variance of v_i is small for all the centers because of much strength borrowing.

The solid curves in Figures 5.10 and 5.11 display the posterior distributions of λ_{i1} based on Prior III for all centers. Table 5.11 shows each individual center hazard rate (MLE), posterior mode of λ_{i1} ($\hat{\lambda}_{i1}$), pooled population mode ($(a-1)/\hat{b}$), posterior mean and variance of λ_{i1}. The individual center hazard rates were pulled toward each other by borrowing strength, and this shrinkage is strong. The individual center MLE of λ_{i1} varies from 0 to 0.787, and the posterior mode of λ_{i1} varies only from 0.045 to 0.074. The posterior variance of λ_{i1} is small in all centers.

The solid curves in Figure 5.12 display posterior distributions of a/b, a/b^2, μ and σ^2 based on Prior III. Table 5.9 shows the posterior mode, mean and variance of a/b, a/b^2, μ and σ^2. The first two rows represent the population mean and variance of λ_{i1}, the last two rows represent the population mean and variance of v_i. The posterior variances of these parameters are small. The posterior mode of a/b is 0.063, which corresponds to a 15.9-month mean survival time. The posterior mode of μ is -0.485, which corresponds to a 62% increase in mean survival time for patients on therapy 2 over those on therapy 1.

Center	v_{iMLE}	$\hat{v}_i$	$\hat{\mu}$	$\mathrm{E}(v_i \mid D)$	$\mathrm{V}(v_i \mid D)$	$P(v_i < 0 \mid D)$
1	∞	-0.426	-0.485	-0.400	0.202	0.811
2	1.122	-0.409	-0.485	-0.370	0.211	0.798
3	-1.072	-0.712	-0.485	-0.752	0.108	0.963
4	0.015	-0.320	-0.485	-0.302	0.101	0.783
5	-1.756	-0.557	-0.485	-0.602	0.177	0.906
6	-0.444	-0.534	-0.485	-0.559	0.141	0.894
7	-1.819	-0.593	-0.485	-0.642	0.155	0.918
8	-0.134	-0.392	-0.485	-0.362	0.179	0.787
9	-1.894	-0.639	-0.485	-0.710	0.175	0.947
10	0.678	-0.311	-0.485	-0.271	0.154	0.736
11	-1.956	-0.787	-0.485	-0.874	0.153	0.975
12	-1.586	-0.341	-0.485	-0.295	0.170	0.749
13	0.938	0.002	-0.485	0.061	0.147	0.477
14	-0.837	-0.366	-0.485	-0.352	0.117	0.791
15	-2.595	-0.555	-0.485	-0.599	0.176	0.901
16	-0.517	-0.390	-0.485	-0.369	0.155	0.813
17	$-\infty$	-0.657	-0.485	-0.751	0.203	0.958
18	-1.987	-0.363	-0.485	-0.313	0.192	0.761
19	-1.392	-0.597	-0.485	-0.647	0.155	0.927
20	$-$	-0.476	-0.485	-0.469	0.234	0.848
21	$-$	-0.426	-0.485	-0.399	0.204	0.825
22	$-\infty$	-0.648	-0.485	-0.739	0.203	0.942

Table 5.10: The MLE, posterior mode, pooled population mode, posterior mean, variance of v_i and $P(v_i < 0|D)$ based on Prior III in all centers

Center	$\lambda_{i1\,MLE}$	$\hat{\lambda}_{i1}$	$(a-1)/\hat{b}$	$\mathrm{E}(\lambda_{i1} \mid D)$	$\mathrm{V}(\lambda_{i1} \mid D)$
1	0	0.045	0.061	0.049	0.00018
2	0.034	0.054	0.061	0.058	0.00022
3	0.063	0.056	0.061	0.058	0.00017
4	0.049	0.057	0.061	0.059	0.00017
5	0.120	0.059	0.061	0.065	0.00032
6	0.045	0.054	0.061	0.059	0.00025
7	0.150	0.063	0.061	0.069	0.00033
8	0.088	0.064	0.061	0.069	0.00034
9	0.082	0.055	0.061	0.060	0.00028
10	0.036	0.056	0.061	0.061	0.00023
11	0.100	0.058	0.061	0.062	0.00026
12	0.435	0.066	0.061	0.073	0.00040
13	0.041	0.061	0.061	0.064	0.00022
14	0.147	0.074	0.061	0.079	0.00037
15	0.280	0.058	0.061	0.064	0.00034
16	0.109	0.066	0.061	0.071	0.00034
17	0.012	0.054	0.061	0.059	0.00030
18	0.787	0.066	0.061	0.072	0.00041
19	0.088	0.058	0.061	0.063	0.00028
20	0.200	0.064	0.061	0.070	0.00039
21	–	0.060	0.061	0.067	0.00037
22	0.057	0.052	0.061	0.057	0.00027

Table 5.11: The MLE, posterior mode, pooled population mode, posterior mean and variance of λ_{i1} based on Prior III in all centers

Parameter	Mode	Mean	Variance
a/b	0.063	0.064	0.00011
a/b^2	0.00025	0.00028	6.97×10^{-9}
μ	-0.485	-0.469	0.050
σ^2	0.071	0.185	0.023

Table 5.12: Posterior mode, mean and variance of a/b, a/b^2, μ and σ^2 based on Prior III

5.7.3 Comparison of the Three Prior Distributions

When we compare the posterior distributions derived from the three sets of prior distributions that we chose, we can see how the posterior distributions are affected by the prior distributions. By comparing the corresponding plots from Figure 5.8 to Figure 5.12, we can see that the shrinkage of the individual center parameters increased as the priors changed from Prior II to Prior I, and from Prior I to Prior III. Prior belief in center heterogeneity has a strong impact on the posterior distributions of the individual center parameters. It also has a strong impact on the posterior distributions of the population variances of v_i and λ_{i1}, but it does not have a strong impact on the posterior distributions of the population means of v_i and λ_{i1} — these were clearly shown in those posterior distribution figures.

By comparing corresponding numbers from Table 5.2 to Table 5.12, we can see the range of the posterior mode of the individual center parameters changes with the variation of prior belief in center heterogeneity. The range of the posterior mode of v_i across the centers is 2.008 based on Prior II, it decreased to 1.302 based on Prior I, and it further decreased to 0.789 based on Prior III. The range of the posterior mode of λ_{i1} across the centers is 0.068 based on Prior II, it decreased to 0.050 based on Prior I, and it further decreased to 0.029 based on Prior III.

By comparing the corresponding posterior variances of the individual center parameters and the population parameters derived from the three prior distributions, we can clearly see that the posterior variance decreased as prior belief in center heterogeneity decreased. For example, the posterior variance of v_2 is 0.855 based on Prior II, it decreased to 0.454 based on Prior I, and

it further decreased to 0.211 based on Prior III. The posterior variance of λ_{51} is 0.00136 based on Prior II, it decreased to 0.00070 based on Prior I, and it further decreased to 0.00032 based on Prior III. The posterior variance of μ is 0.130 based on Prior II, it decreased to 0.075 based on Prior I, and it further decreased to 0.050 based on Prior III.

5.8 Simulation with Homogeneous Centers

Since many centers have very small number of patients in this trial, someone might wonder that the differences among the posterior distributions of v_i and λ_{i1} at different centers are resulted from the small sample size. For determining this, I performed analysis on some simulated data.

In my simulation, there is not center heterogeneity, that is all the v_i's are equal, and all the λ_{i1}'s are equal, too. I used the posterior modes of v and λ_1 in Chapter 3 as the values of common v_i and λ_{i1} respectively in my simulation. So the two hazard rates in simulation are

$$\lambda_{i1} = 0.062, \quad \lambda_{i2} = \lambda_{i1} e^{v_i} = 0.037, \quad i = 1, \cdots, 22$$

I simulated the same number of patients in every center on each therapy as in the trial. For all the simulated patients whose survival time are longer than 60 months, their survival times are censored at 60 months. I applied the hierarchical model to the simulated data. Analysis results show that if our prior belief on the variances of v_i and λ_{i1} are not concentrated on very small values, then there are some differences among the posterior distributions of v_i and λ_{i1} at different centers. In such small center sample size trial, very strong prior belief on center homogeneity is needed to protect center homogeneity in statistical inference.

Center	n_{i1}	d_{i1}	T_{i1}	n_{i2}	d_{i2}	T_{i2}
1	1	1	10.40	1	1	27.79
2	4	4	75.89	1	1	7.50
3	12	12	144.68	10	10	211.12
4	11	10	294.23	11	11	173.17
5	2	2	14.52	1	1	5.03
6	3	3	33.46	4	4	86.97
7	4	4	91.59	3	3	78.77
8	2	2	15.98	2	1	69.24
9	2	1	65.78	2	2	51.06
10	4	4	32.62	4	3	72.02
11	6	6	49.90	6	6	94.34
12	1	1	13.17	3	3	37.69
13	7	7	84.48	11	9	316.52
14	6	6	135.66	7	5	172.42
15	1	1	19.39	1	1	18.59
16	3	3	13.18	3	3	43.76
17	1	1	16.62	1	0	60.00
18	1	1	1.40	2	2	54.45
19	3	3	38.81	3	3	77.02
20	2	2	55.95	0	0	0.00
21	0	0	0.00	1	1	38.42
22	1	1	1.59	1	1	0.94
total	77	75	1209.3	78	71	1696.8

Table 5.13: Sufficient statistics from simulation

I present one simulation result in the following tables and figures. Table 5.13 displays the sufficient statistics from one simulated data.

Prior I in section 5.3 is used in analyzing the simulated data. Figure 5.13 displays the posterior distributions of v_i at each center. We still can see some variation among those posterior distributions. For example, centers 3 and 4 are two large centers, but the posterior distributions of v_i at centers 3 and 4 are not very close. Posterior distributions of v_i are more concentrated at large centers than at small centers.

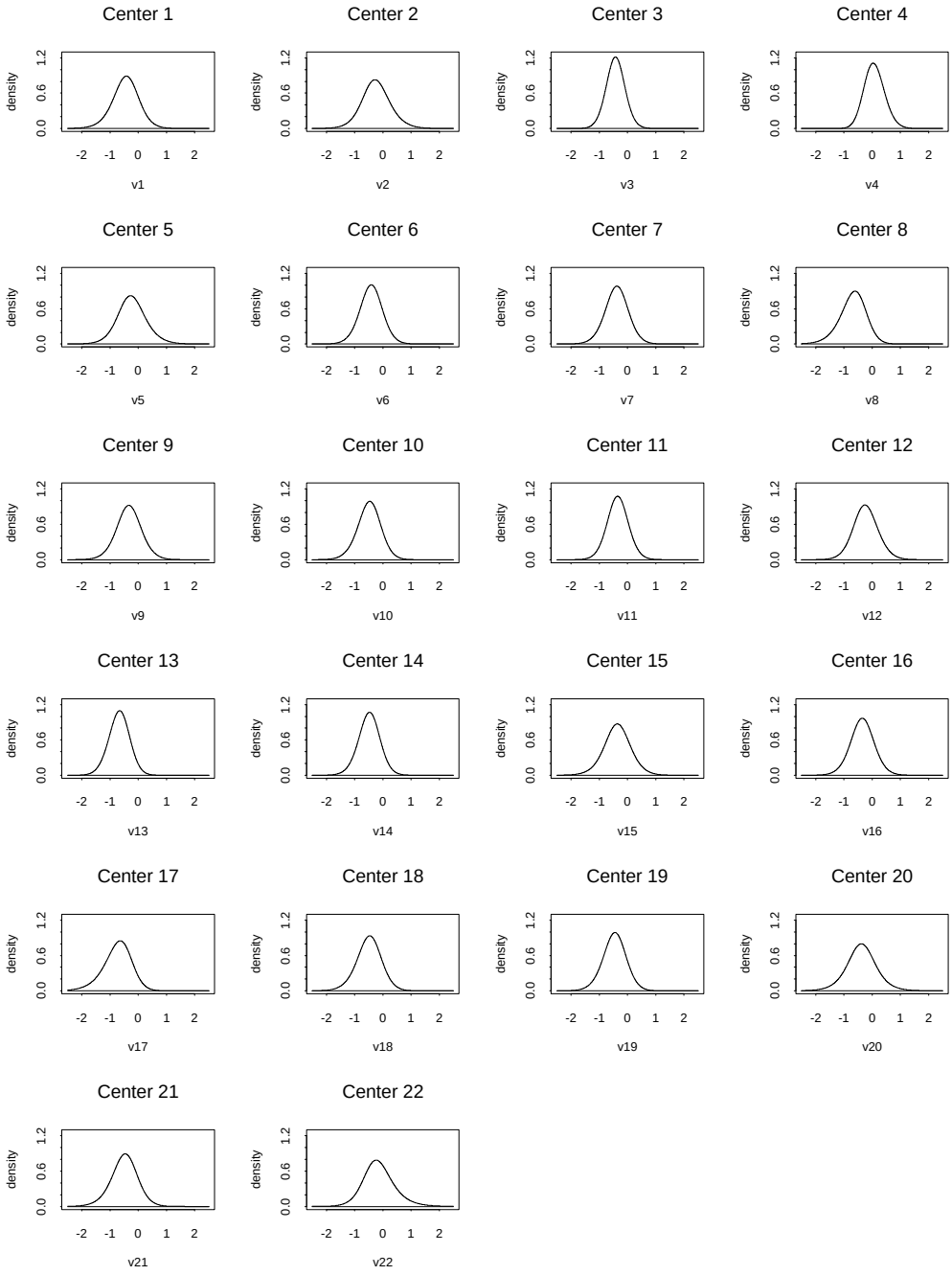

Figure 5.13: Posterior distribution of v_i at each center derived from the simulated data

Center	v_{iMLE}	$\hat{v}_i$	$P(v_i < 0\|D)$
1	-0.983	-0.418	0.815
2	0.928	-0.280	0.673
3	-0.560	-0.425	0.806
4	0.625	0.040	0.452
5	0.367	-0.272	0.652
6	-0.668	-0.409	0.776
7	-0.137	-0.374	0.774
8	-2.159	-0.604	0.901
9	0.946	-0.332	0.737
10	-1.080	-0.462	0.822
11	-0.637	-0.348	0.756
12	0.047	-0.255	0.677
13	-1.070	-0.657	0.910
14	-0.422	-0.470	0.836
15	0.042	-0.358	0.756
16	-1.200	-0.348	0.754
17	$-\infty$	-0.634	0.913
18	-2.968	-0.464	0.819
19	-0.685	-0.442	0.809
20	—	-0.379	0.769
21	—	-0.458	0.826
22	0.526	-0.235	0.634

Table 5.14: The MLE, the posterior mode of v_i, and $P(v_i < 0|D)$ from the simulated data

Table 5.14 shows the MLE, the posterior mode ($\hat{v}_i$) of v_i and the posterior probability $P(v_i < 0|D)$ at each center. The posterior mode of v_i ranges from -0.657 at Center 13 to 0.040 at Center 4. The posterior probability $P(v_i < 0|D)$ ranges from 0.452 at Center 4 to 0.913 at Center 17. Center 4 is a large center, but $P(v_4 < 0|D)$ is below 0.5.

Figure 5.14 displays the posterior distributions of λ_{i1}, $i = 1, \cdots, 22$. We see some variation among those posterior distributions. Posterior distribution of λ_{i1} is concentrated in some large centers, whereas it is relatively flat in many

Center	$\lambda_{i1\,MLE}$	$\hat{\lambda}_{i1}$
1	0.096	0.058
2	0.053	0.058
3	0.083	0.071
4	0.034	0.045
5	0.138	0.074
6	0.090	0.063
7	0.044	0.049
8	0.125	0.054
9	0.015	0.040
10	0.123	0.069
11	0.120	0.079
12	0.076	0.065
13	0.083	0.059
14	0.044	0.046
15	0.052	0.056
16	0.228	0.079
17	0.060	0.042
18	0.714	0.061
19	0.077	0.059
20	0.036	0.048
21	—	0.051
22	0.629	0.076

Table 5.15: The MLE and the posterior mode of λ_{i1} at each center from the simulated data

small centers.

Table 5.15 shows the MLE and posterior mode ($\hat{\lambda}_{i1}$) of λ_{i1} at each center. The range of posterior mode $\hat{\lambda}_{i1}$ is 0.039. $\hat{\lambda}_{i1}$ changes from 0.040 at Center 9 to 0.079 at Center 11. $\lambda_{i1} = 0.040$ corresponds to a 25-month mean survival time at center i, whereas $\lambda_{i1} = 0.079$ corresponds to only a 12.7-month mean survival time at center i in our model.

Figure 5.15 displays the prior and posterior distributions of population parameters. The posterior distributions of a/b and μ are much more concentrated

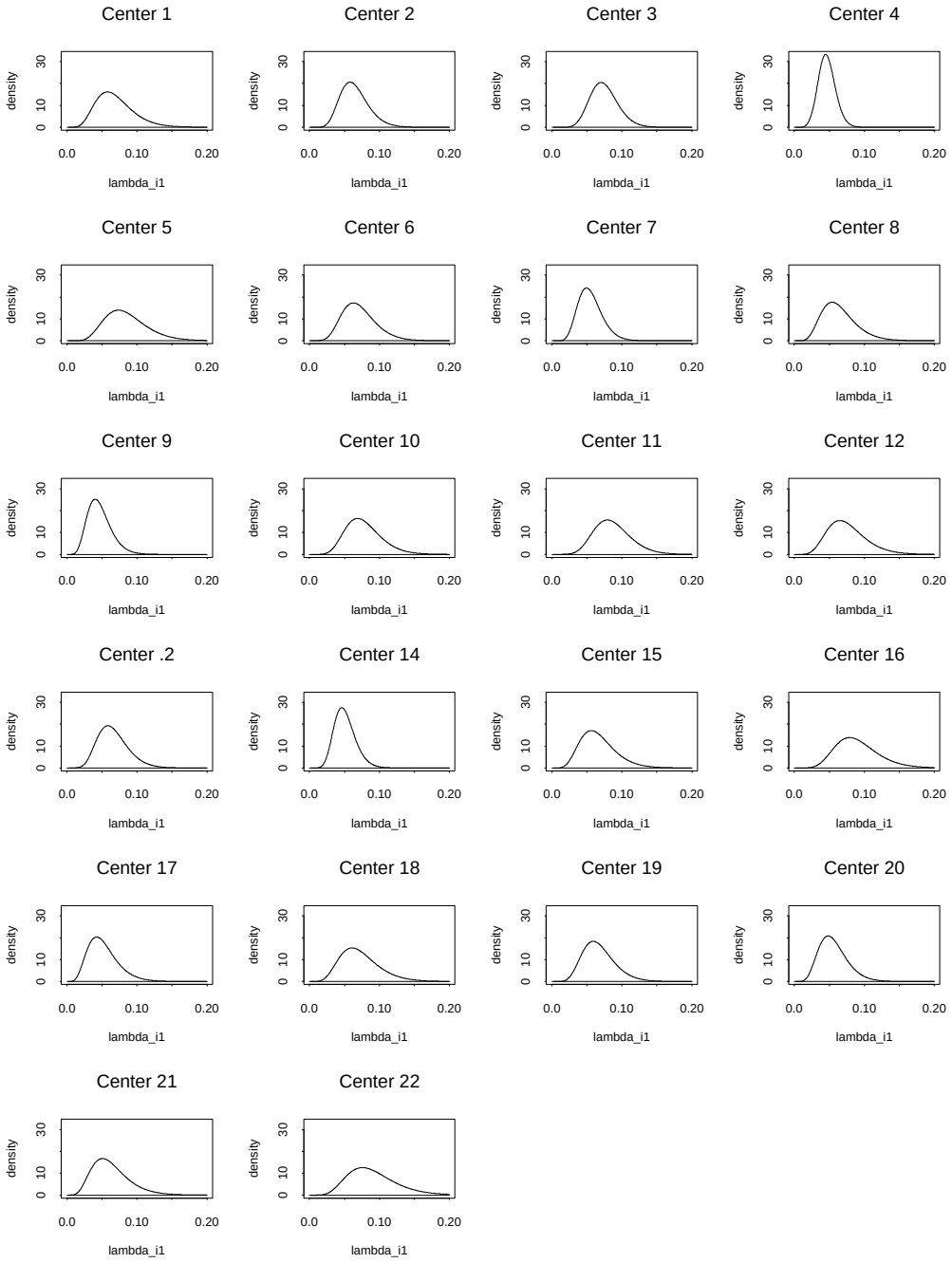

Figure 5.14: Posterior distribution of λ_{i1} at each center derived from the simulated data

than their prior distributions, indicating that we learned a lot from the data for those population means. The posterior distribution of a/b^2 is much more concentrated than the prior distribution of a/b^2, because the prior distribution of a/b^2 is quite disperse. The posterior distribution of σ^2 is more disperse than the prior distribution of σ^2, and σ^2 is likely to be greater on its posterior distribution than on its prior distribution. So our inference indicates that there is certain variation among the v_i's at different centers.

5.9 Summary

In this chapter, we applied the hierarchical model proposed in Chapter 4 to the NSCLC trial data. Since the trial included many centers that had only a few patients, many centers borrowed a lot of strength from other centers, and the posterior distribution is sensitive to the prior distribution of the population parameters. If our prior belief is that the centers are homogeneous, so will be our posterior belief. If we are open-minded about the heterogeneity among the different centers, then each individual center borrows less strength from the other centers. When centers do not borrow strong strength from other centers, the centers that have small number of patients do not have much information on which to make conclusion about center parameters.

In next chapter, I will analyze a data set from a National Institute of Mental Health collaborative study. In that study, the number of patients at each center is relatively large.

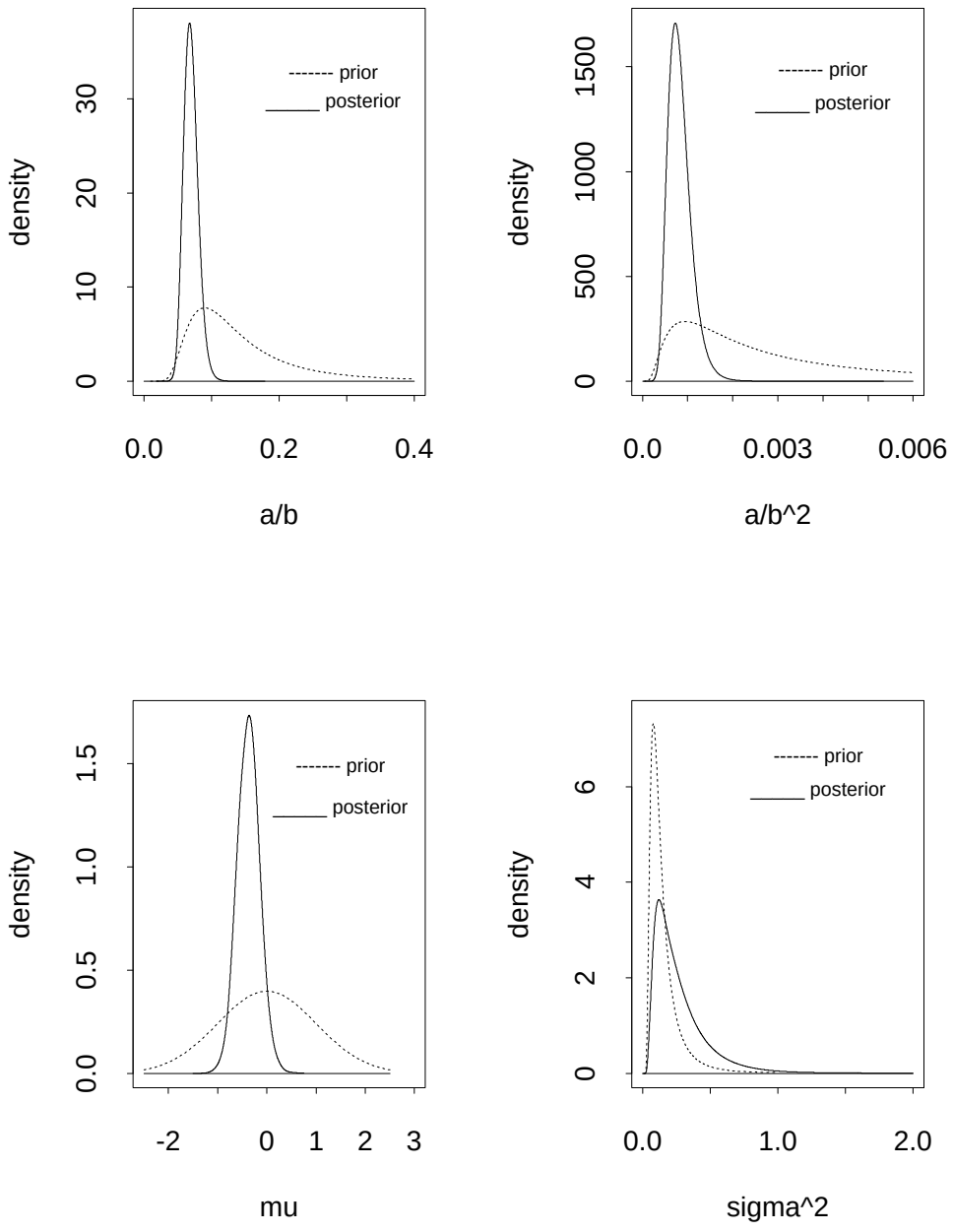

Figure 5.15: Distribution of a/b, a/b^2, μ and σ^2 for the simulated data

Chapter 6

A NIMH Collaborative Study

6.1 Introduction

6.1.1 Purpose of the Study

Depression is a recurrent illness. Psychiatric research has shown that most patients who experience an initial episode of depression will recover, but will then go on to suffer one or more recurrences. Medical researchers have long tried to improve the efficacy of medications designed to prevent the occurrence of new episodes of illness. The multicenter clinical trial analyzed here was conducted to determine the comparative efficiencies of lithium carbonate, imipramine hydrochloride, and a combination of lithium and imipramine in preventing the occurrence of unipolar and bipolar affective disorders. For a detailed description of the Pharmacologic and Somatic Treatments Research Branch of the National Institute of Mental Health (NIMH-PRB) Collaborative Study of Long-Term Maintenance Drug Therapy in Recurrent Affective Illness, see Prien et al. (1984). The following analysis includes only the patients with unipolar depression.

6.1.2 Design of the Study

The NIMH-PRB collaborative study had two phases: a preliminary phase and a maintenance phase. The purpose of the preliminary phase was to control the index episode, stabilize the patient's clinical condition, and establish stable maintenance dose levels of lithium carbonate and imipramine in preparation for the maintenance phase. Upon patients met specified entrance criteria that ensured that they were experiencing an acute episode, they received the treatment of choice of the psychiatrist responsible for their care during this preliminary phase of the illness. Approximately 90% of these patients were treated with imipramine. After their acute symptoms were controlled, these patients received a combination of both lithium and imipramine. Once the patient remained on predetermined medication dosages and met specified "wellness" criteria for two consecutive months, he or she entered the maintenance phase of the study.

The major experimental phase of the study, The maintenance phase involved two years of double-blind comparison testing which treatment regimen prolonged the recurrence of affective disorder. The 150 patients who were eligible for the maintenance phase of the study were randomly assigned to one of two groups. One group remained on imipramine; the other group was withdrawn from imipramine. After randomization, the patients were followed for up to two years (until the end of the study), or until they experienced a recurrence of depression. The response variable of interest is time between randomization and the first recurrence of a depression episode.

6.1.3 Analysis of Recurrence Time

Five medical centers participated in the collaborative study. Figure 6.1 presents the Kaplan-Meier estimates of the survival functions for each therapy with all centers combined. The p-values for testing the homogeneity of these survival curves over therapy are about 0.0001 for both the logrank test and the Wilcoxon test. Figure 6.1 clearly showns that the two survival curves diverge at a early time. At time point about 10 weeks, the difference between the two curves is large; after that time, the difference remains roughly stable. The data show that there are many more early recurrences in the *off imipramine* group than in the *on imipramine* group.

6.2 Bayesian Hierarchical Analysis of the NIMH Collaborative Study

6.2.1 The Model

In this section, I will use the Bayesian hierarchical model proposed in Chapter 4 to analyze the data from the collaborative study. For this study, a patient's "survival time" refers to the time until the patient suffers a recurrence of depression. The standard therapy is the *off imipramine* therapy, and the experimental therapy is the *on imipramine* therapy. The goal is to determine the relative efficacy of the two therapies, and to discover the heterogeneity among the different centers.

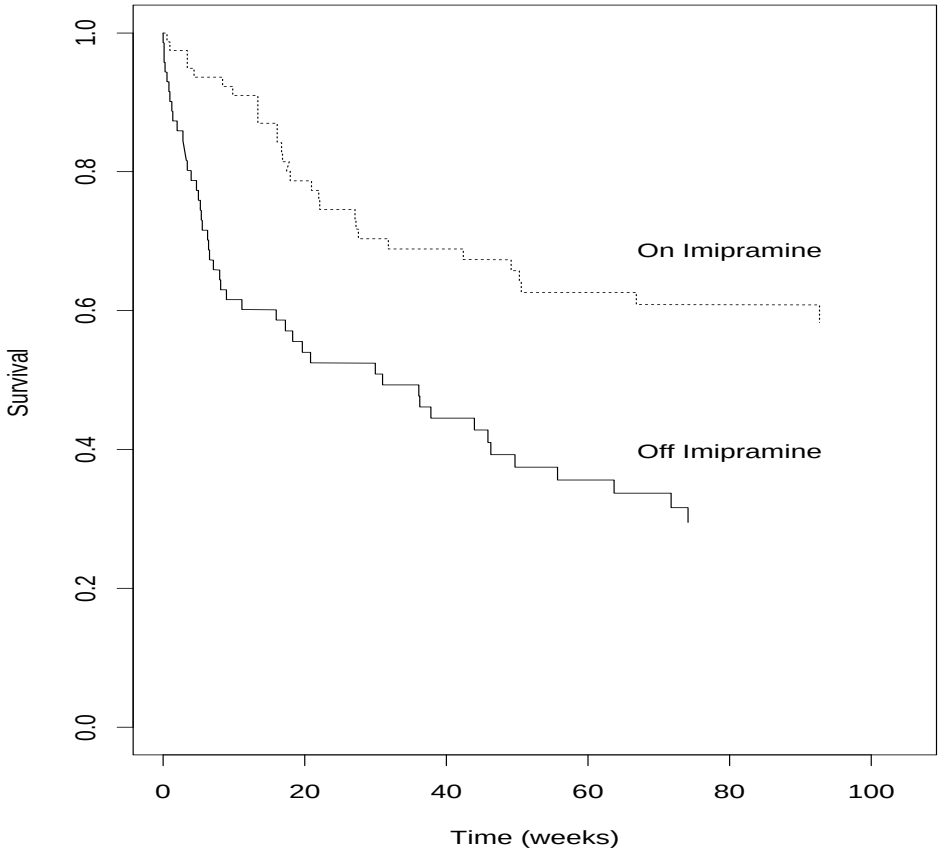

Figure 6.1: Kaplan-Meier plot for the two therapies

Center	n_{i1}	d_{i1}	T_{i1}	n_{i2}	d_{i2}	T_{i2}
A	11	6	498.86	15	3	1205.43
B	10	6	390.72	15	6	769.86
C	8	7	176.86	5	3	332.00
D	25	12	1367.29	26	8	1503.37
E	17	15	240.00	18	9	845.57
total	71	46	2673.73	79	29	4656.23

Table 6.1: Sufficient statistics by center

6.2.2 Data

Appendix B presents the data from the five centers. In the hierarchical model, $d_{ij},\ T_{ij},\ i = 1, \cdots, 5,\ j = 1, 2$ form the sufficient statistics. d_{ij} is the number of patients who experienced a recurrence of depression on therapy j in center i; and T_{ij} is the total time observed for all the patients on therapy j in center i, $i = 1, \cdots, 5,\ j = 1, 2$. Here, $j = 1$ corresponds to the *off imipramine* therapy, and $j = 2$ corresponds to the *on imipramine* therapy. Table 6.1 displays these sufficient statistics and the total number of patients by center-therapy (n_{ij}). The time unit is one week.

6.2.3 Choice of Prior Distributions

Parameters a, c, d, u and w need to be specified for applying the model to the data. The gamma distribution of λ_{i1} in the second level of hierarchy is conjugate to the exponential distribution in the first level of hierarchy. From the posterior distribution of λ_{i1}, one see that a acts as a prior uncensored number of observations and b is the prior total exposure time on the *off imipramine* therapy. The values of a, c and d can be chosen according to some prior information about the *off imipramine* therapy. These parameter values not only

represent our prior belief about the hazard rate on the *off imipramine* therapy, but also represent our prior belief about the center heterogeneity among those λ_{i1}'s in different centers. Note that $V(\lambda_{i1} \mid a, b) = a/b^2$. For fixed a, if our choice of c and d makes b likely to be large, then our prior belief is that center heterogeneity is small.

The second level parameter σ^2 represents the center heterogeneity among those log hazard ratio v_i's at different centers. A smaller σ^2 reflects less center heterogeneity, and a larger σ^2 reflects larger center heterogeneity among those v_i's at different centers. The prior distribution of σ^2 reflects our prior belief of this heterogeneity.

The available prior information about the *off imipramine* therapy is from the history of those patients treated with the therapy, which suggested that the recurrence occurred mainly between 8 and 122 weeks, and the weighted average of the mean recurrence times across centers was 25 weeks.

We choose $a = 3$, $c = 4$. Note that

$$\begin{aligned} E(\lambda_{i1}) &= E(E(\lambda_{i1} \mid a, b))) \\ &= E(\frac{a}{b}) \\ &= \int_0^\infty \frac{a}{b} \frac{d^c}{\Gamma(c)} b^{c-1} e^{-db} db \\ &= \frac{ad}{c-1} \end{aligned}$$

and given λ_{i1}, the mean survival time on the *off imipramine* therapy is $1/\lambda_{i1}$. We choose d such that the value of the prior $E(\lambda_{i1})$ is the reciprocal of the weighted average of mean recurrence time, that is

$$\frac{ad}{c-1} = \frac{1}{25}$$

The following values for a, c and d are chosen:

$$a = 3, \quad c = 4 \quad d = 0.04$$

The mean and variance of λ_{i1} are

$$\begin{aligned} E(\lambda_{i1} \mid a, b) &= \frac{a}{b} \\ V(\lambda_{i1} \mid a, b) &= \frac{a}{b^2} \end{aligned}$$

λ_{i1}'s mean a/b has its prior distribution

$$\begin{aligned} f(x \mid a, c, d) &= \frac{d^c}{\Gamma(c)} \left(\frac{a}{x}\right)^{c-1} e^{-\frac{da}{x}} \cdot \frac{a}{x^2} \\ &= \frac{(da)^c}{\Gamma(c)} \left(\frac{1}{x}\right)^{c+1} e^{-\frac{da}{x}} \end{aligned}$$

and λ_{i1}'s variance a/b^2 has its prior distribution

$$\begin{aligned} f(x \mid a, c, d) &= \frac{d^c}{\Gamma(c)} \left(\frac{\sqrt{a}}{\sqrt{x}}\right)^{c-1} e^{-\frac{d\sqrt{a}}{\sqrt{x}}} \cdot \frac{\sqrt{a}}{2x\sqrt{x}} \\ &= \frac{(d\sqrt{a})^c}{2\Gamma(c)} \frac{1}{x^{1+c/2}} e^{-\frac{d\sqrt{a}}{\sqrt{x}}} \end{aligned}$$

We chose a relatively flat prior distribution of v_i in this study, the values of u and w are chosen as:

$$u = 4, \quad w = 1.5$$

which set the prior mean and variance of σ^2 to be

$$\begin{aligned} E(\sigma^2 \mid u, w) &= \frac{w}{u-1} = 0.5 \\ V(\sigma^2 \mid u, w) &= \frac{w^2}{(u-1)^2(u-2)} = 0.125 \end{aligned}$$

Center	A	B	C	D	E
$P(v_i < 0 \mid D)$	0.986	0.915	0.962	0.889	0.991

Table 6.2: Posterior probability of $v_i < 0$ in all centers

Figure 6.2 displays these prior distributions. The first row shows the distributions of the mean and variance of λ_{i1} — a/b and a/b^2 respectively. The second row displays the distributions of the population mean and variance of v_i — μ and σ^2 respectively.

6.2.4 Posterior Distributions

The posterior distributions of v_i $(i = 1, \cdots, 5)$ are displayed in the first plot of Figure 6.3. From this figure, one can see that the domain of the posterior density of v_i is mainly at the left of $v_i = 0$ at each center. Table 6.2 shows the posterior probabilities of $v_i < 0$; those probabilities are high, which suggests that in all the centers the *on imipramine* therapy prevents the recurrence of depression longer than the *off imipramine* therapy does. At some different centers, the posterior distributions of v_i are quite different, so the relative improvement of *on imipramine* over *off imipramine* are likely to be different across centers.

Table 6.3 displays individual center log hazard ratio (MLE), posterior mode of v_i $(\hat{v}_i)$, pooled population log hazard ratio (posterior mode $\hat{\mu}$), posterior mean and variance of v_i at each center. The table clearly shows that the parameters at each individual centers are shrunk toward the pooled population value through borrowing strength from other centers. The individual center log hazard ratios for centers A, C and E fell below the pooled population log

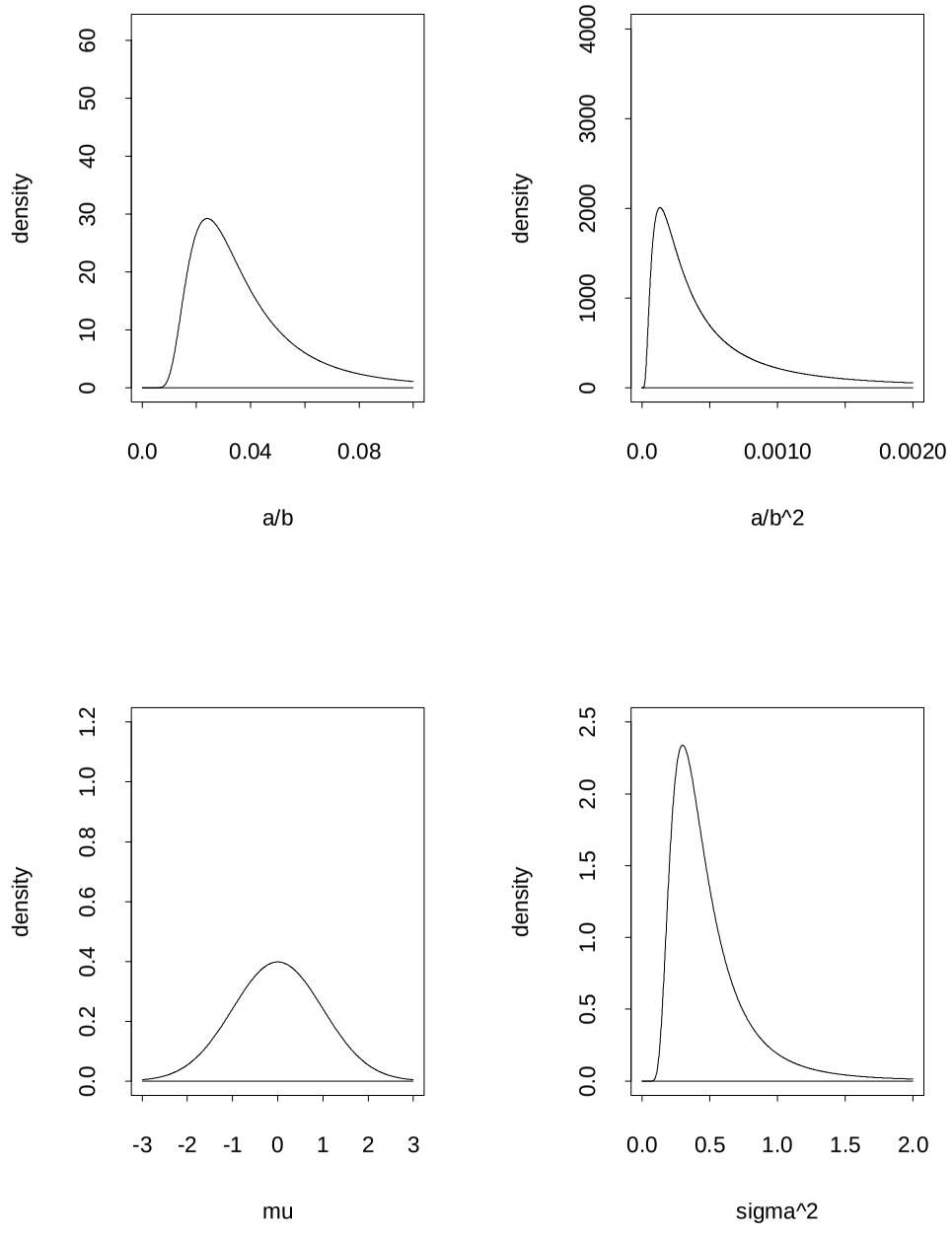

Figure 6.2: Prior distributions of a/b, a/b^2, μ and σ^2

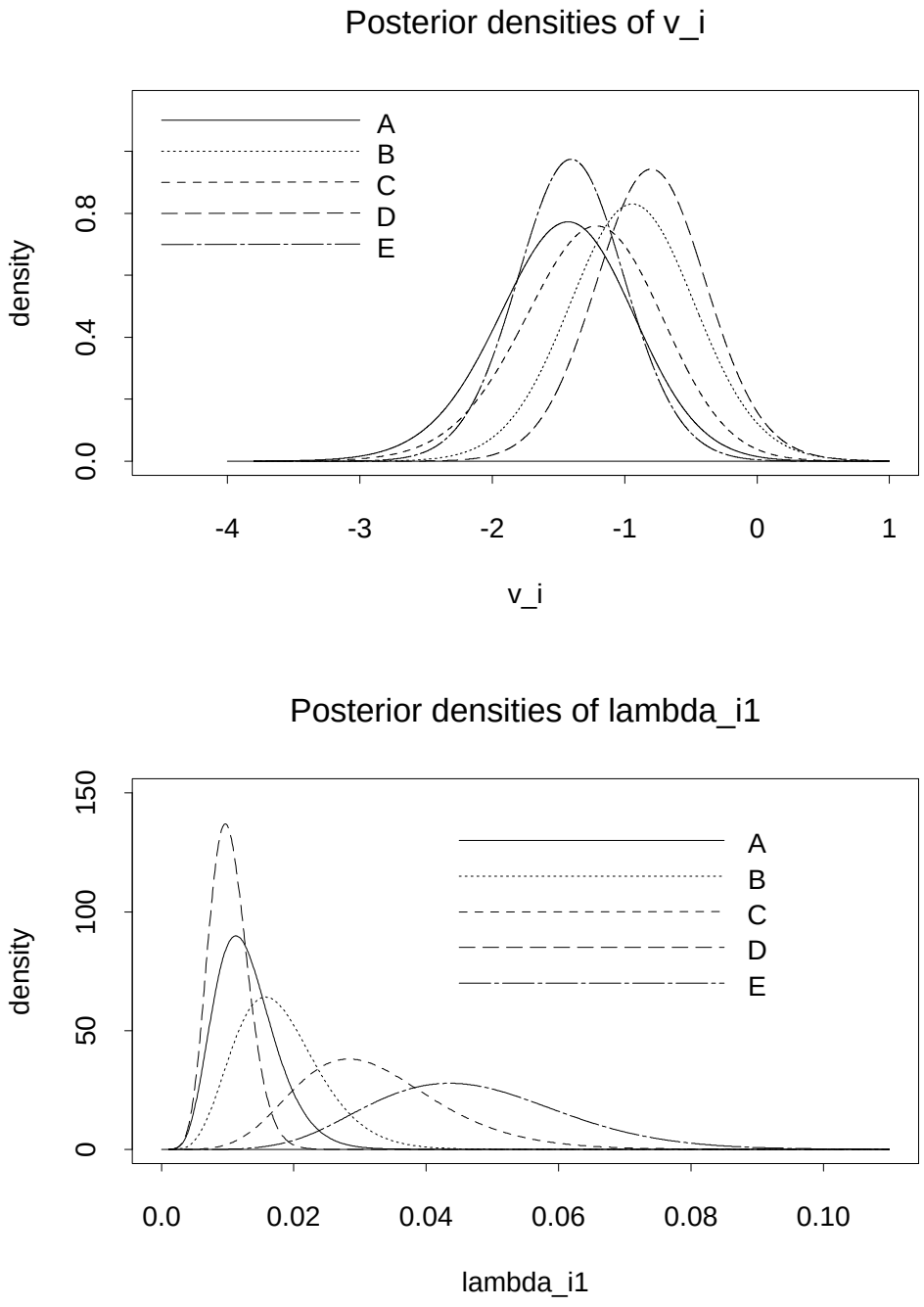

Figure 6.3: Posterior distributions of v_i and λ_{i1} in each center

Center	v_{iMLE}	$\hat{v}_i$	$\hat{\mu}$	$\mathrm{E}(v_i \mid D)$	$\mathrm{V}(v_i \mid D)$
A	-1.575	-1.428	-1.045	-1.462	0.274
B	-0.678	-0.942	-1.045	-0.941	0.239
C	-1.477	-1.215	-1.045	-1.263	0.285
D	-0.500	-0.799	-1.045	-0.797	0.187
E	-1.770	-1.405	-1.045	-1.404	0.174

Table 6.3: The MLE, posterior mode, pooled population mode, posterior mean and variance of v_i

hazard ratio, and the posterior modes of v_i at these centers were shrunk up toward the population value. At centers B and D, the individual center v_i's were above the pooled population value, and their posterior modes were shrunk down toward the pooled population value. The range of the MLE of individual center v_i is 1.27 (from -1.77 at center E to -0.5 at center D), and the range of posterior modes of v_i decreased greatly to 0.629 (from -1.428 at center A to -0.799 at center D). This decrease occurred because the hierarchical model connected individual center parameters together, and pulled them close.

The inference about the posterior distribution of v_i at each center borrowed strength from the other centers. The posterior mode $\hat{v}_i$ reaches its maximum -0.799 at center D. In the model, $v_i = -0.799$ corresponds to $\lambda_{i1}/\lambda_{i2} = e^{0.799} = 2.22$, which represents a 122% increase in median and mean survival time on therapy 2 over therapy 1 at center i. The $\hat{v}_i$ reaches its minimum -1.428 at center A. When $v_i = -1.428$ corresponds to $\lambda_{i1}/\lambda_{i2} = e^{1.428} = 4.17$, it represents a 317% increase in median and mean survival time on therapy 2 over therapy 1 at center i. So the posterior distributions of v_i in individual centers suggest that the *on imipramine* therapy can prolong the time of recurrence of depression better than the *off imipramine* therapy does, but the average

relative increase of recurrence time for patients received the *on imipramine* therapy varies from center to center, ranging from 122% at center D to 317% at center A.

The posterior distributions of λ_{i1} $(i = 1, \cdots, 5)$ displayed in the second plot of Figure 6.3 show a large variation. The domains of the posterior densities of λ_{i1} at centers D and E shared only a tiny area. The evidence strongly suggests that the hazard rate of the *off imipramine* therapy changes from center to center. Center D and E are the two largest centers, but the posterior distribution of λ_{i1} is relatively concentrated at center D and relatively flat at center E, because there is a larger discrepancy on the observations at center E than at center D. The posterior distributions of λ_{i1} at these two centers strongly indicate that the λ_{i1} at center D is much less than that at center E, so the survival time on the *off imipramine* therapy at center D is much longer than that at center E.

Table 6.4 displays the individual center hazard rate (MLE), posterior mode of λ_{i1}, $(\hat{\lambda}_{i1})$, pooled population hazard rate $((a-1)/\hat{b})$, posterior mean and variance of λ_{i1} at each center. Just like the log hazard ratios, the hazard rates at each individual centers are pulled toward each other through borrowing strength from other centers. At centers A, B and D, the MLE of individual center hazard rates were below the pooled population hazard rate, and the posterior modes of λ_{i1}, at these centers were shrunk up toward the pooled population value. At centers C and E, the MLE of individual center hazard rates were above the pooled population hazard rate, the posterior modes of λ_{i1} at these center were shrunk down toward the pooled population value. The range of MLE of λ_{i1} across centers is 0.054 (from 0.009 at center D to 0.063 at center E), and the range of the posterior mode of λ_{i1} across centers decreased

Center	$\lambda_{i1\,MLE}$	$\hat{\lambda}_{i1}$	$(a-1)/\hat{b}$	$\mathrm{E}(\lambda_{i1} \mid D)$	$\mathrm{V}(\lambda_{i1} \mid D)$
A	0.012	0.012	0.016	0.013	0.000022
B	0.015	0.016	0.016	0.018	0.000042
C	0.040	0.028	0.016	0.032	0.000124
D	0.009	0.010	0.016	0.010	0.000009
E	0.063	0.044	0.016	0.047	0.000208

Table 6.4: The MLE, posterior mode, pooled population mode, posterior mean and variance of λ_{i1}

greatly to 0.033 (from 0.010 at center D to 0.044 at center E). This decrease occurred because the hierarchical structure connected the λ_{i1} at the different centers together, and pulled them close. Thus, the inference about a center's λ_{i1} borrowed strength from other centers. The posterior mode $\hat{\lambda}_{i1}$ reaches its minimum 0.010 at center D, and $\lambda_{i1} = 0.010$ corresponds to a 100-week mean survival time on therapy 1 at center i. The $\hat{\lambda}_{i1}$ reaches its maximum 0.044 at center E, and $\lambda_{i1} = 0.044$ corresponds to a 22.7-week mean survival time on therapy 1 at center i. The posterior distributions of λ_{i1} at the individual centers indicate that there is a large difference among the hazard rates of the *off imipramine* therapy at different centers.

Figure 6.4 shows the posterior distributions of the population mean and variance of λ_{i1} and v_i. The first row displays the posterior distributions of the population mean and variance of λ_{i1} — a/b and a/b^2 respectively. The second row shows the posterior distributions of the population mean and variance of v_i — μ and σ^2 respectively.

Table 6.5 presents the posterior mode, mean and variance of b, μ and σ^2. The posterior mode of μ is -1.045, which is the mode of population log hazard ratio. $e^{1.045} = 2.84$, so in the population, the mean survival time of patients

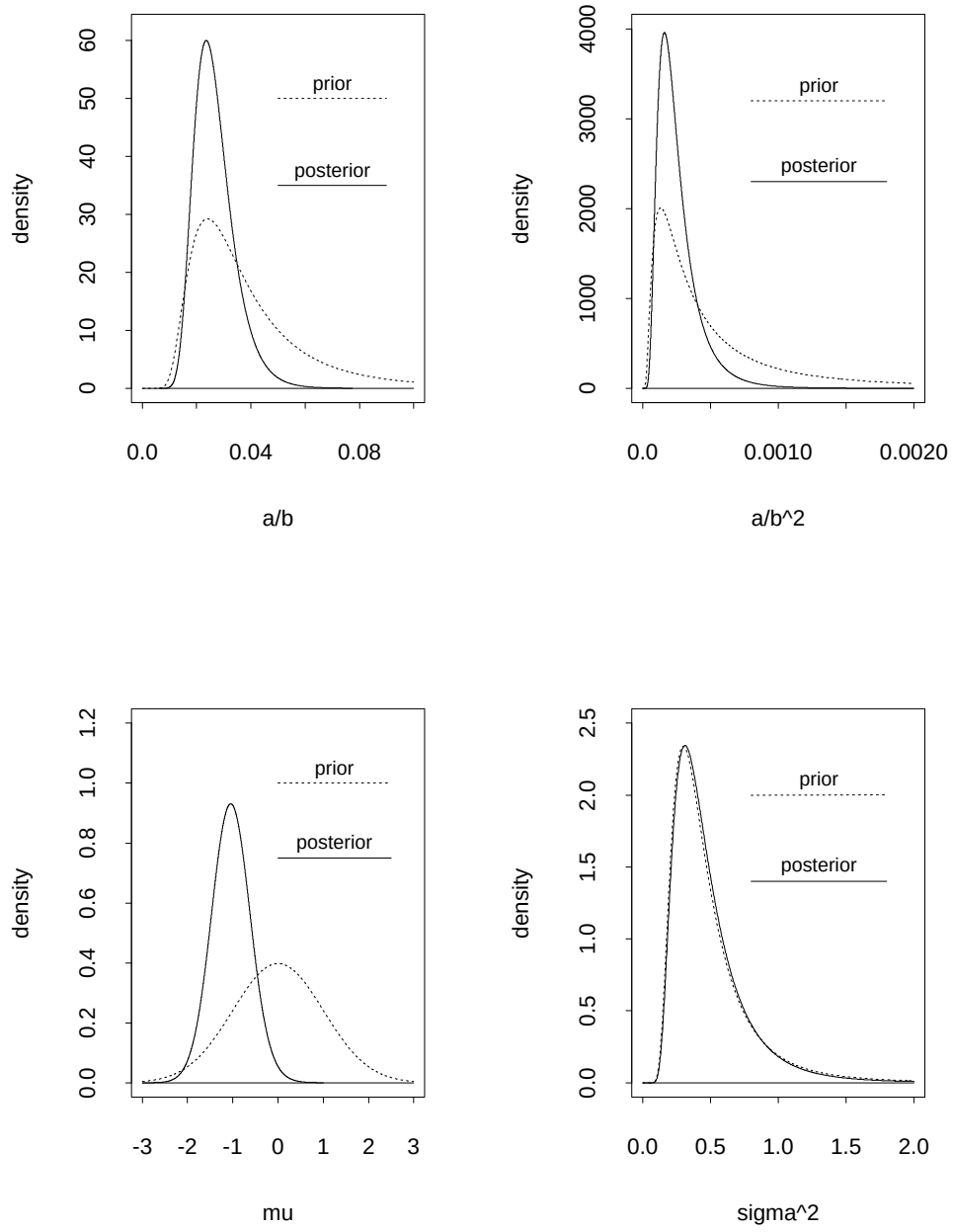

Figure 6.4: Distributions of a/b, a/b^2, μ and σ^2

Parameter	Mode	Mean	Variance
a/b	0.024	0.027	0.0000580
a/b^2	0.0001609	0.0002591	2.6×10^{-8}
μ	-1.045	-1.038	0.188
σ^2	0.313	0.476	0.067

Table 6.5: Posterior mode, mean and variance of a/b, a/b^2, μ and σ^2

receiving the *on imipramine* therapy is about 2.84 times that of those on the *off imipramine* therapy. The posterior probability $P(\mu < 0 \mid D) = 0.989$, which means that it is almost certain that the mean of v_i across centers is negative. The posterior density functions of a/b and μ are much more concentrated than their prior density functions, indicating that the study contains much information about these parameters. The posterior density functions of a/b^2 and σ^2 did not change much from those of their prior distributions, and the data did not change our belief much about the variation among the v_i's and λ_{i1}'s at different centers.

6.2.5 Predictive Survival Functions

New patients who can choose a therapy are primarily interested in how well they will respond to the two therapies. Using the posterior distributions of the parameters at each center, we can calculate the predictive survival function for each therapy at every center. Figure 6.5 shows the predictive survival functions for each center. The survival functions for the two therapies differ from center to center, but patients at all centers always have a higher survival probability on the *on imipramine* therapy than on the *off imipramine* therapy at point in time. The area between the two curves represents the improvement of the

mean survival time between the two therapies, which is large at all the centers. The heterogeneity among the different centers is clearly shown in the figure. For example, the solid curve at center D is very different from that at center E, which indicates that the survival probability on the *off imipramine* therapy at center E is very different from that at center D.

Table 6.6 displays the predicted survival probabilities at time of half-year, one-year, two-years and three-years for the patients on the *off imipramine* therapy. Table 6.7 displays the predicted survival probabilities at time of half-year, one-year, two-years and three-years for the patients on the *on imipramine* therapy. Comparing different rows in these tables, we see the heterogeneity between different centers for the same therapy. For instance, the predicted one-year survival probability for patients on the *off imipramine* therapy is 0.590 at center D, whereas it is only 0.114 at center E. For another example, the predicted two-year survival probability for patients on the *on imipramine* therapy is 0.715 at center A, whereas it is only 0.331 in center E. Comparing corresponding rows in these two tables, we see the increase in survival probabilities for patients the *on imipramine* therapy over those on the *off imipramine* therapy. For example, the predicted two-year survival probability at center E is 0.114 for patients on the *off imipramine* therapy, and it increased to 0.551 for patients on the *on imipramine* therapy.

Table 6.8 shows the predicted mean survival times on the two therapies at each center. At every center, we a very large increase in predicted mean survival time for patients on the *off imipramine* therapy compared to those on the *on imipramine* therapy.

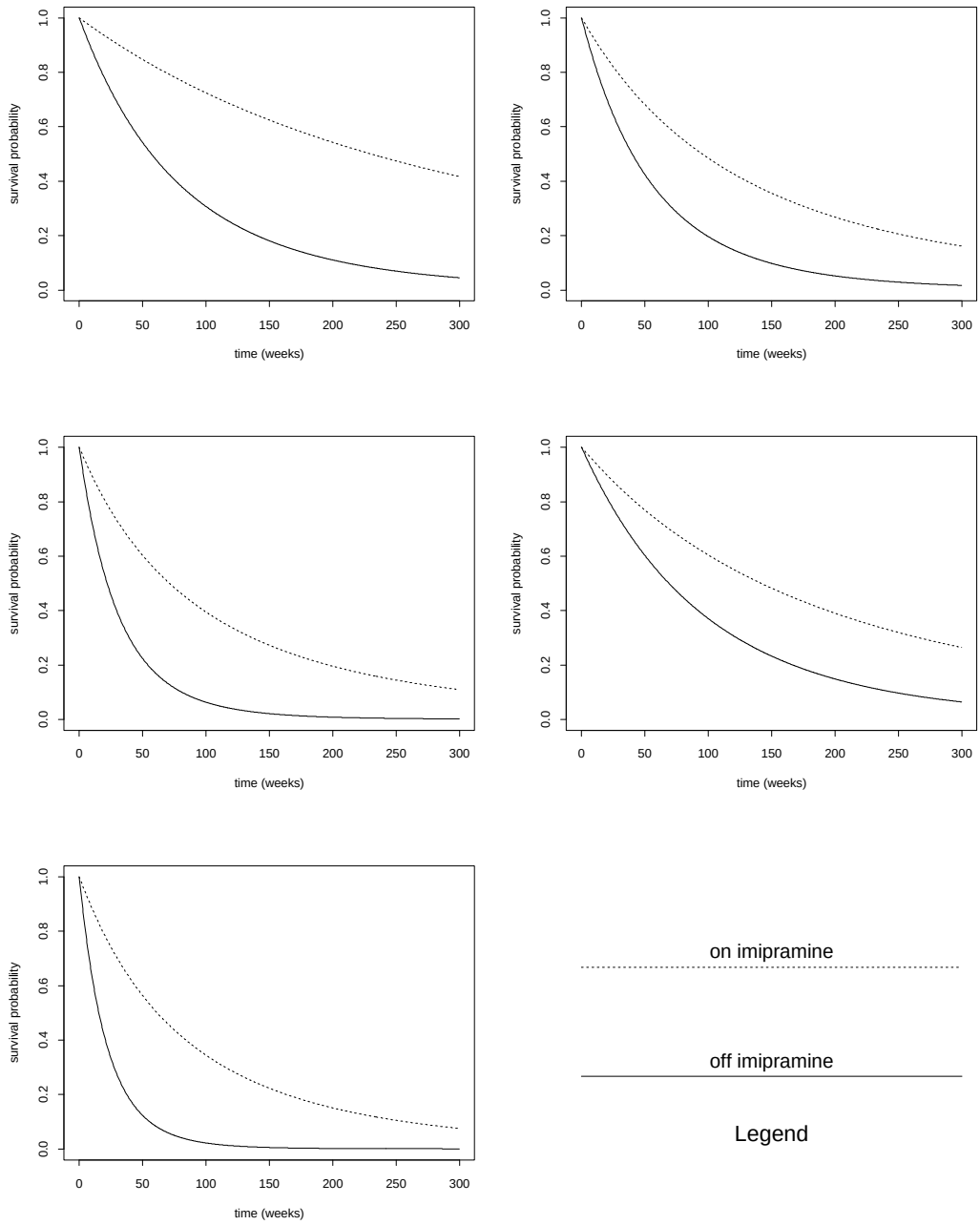

Figure 6.5: Predictive survival functions in each center

	Predicted survival probability			
Center	half-year	one-year	two-year	three-year
A	0.723	0.529	0.392	0.293
B	0.632	0.410	0.272	0.185
C	0.443	0.212	0.107	0.057
D	0.766	0.590	0.457	0.356
E	0.318	0.114	0.045	0.019

Table 6.6: Some predicted survival probabilities on the *off imipramine* therapy

	Predicted survival probability			
Center	half-year	one-year	two-year	three-year
A	0.916	0.841	0.775	0.715
B	0.814	0.671	0.560	0.472
C	0.758	0.592	0.472	0.382
D	0.870	0.762	0.670	0.593
E	0.732	0.551	0.423	0.331

Table 6.7: Some predicted survival probabilities on the *on imipramine* therapy

	Predicted Mean Survival	
Center	*off imipramine*	*on imipramine*
A	89.8	420.0
B	63.3	169.6
C	34.6	132.2
D	105.7	248.6
E	23.8	107.4

Table 6.8: The predicted mean survival time (weeks)

6.3 Sensitivity Analysis

6.3.1 Three Prior Distributions

In this section, I will analyze the data with some different prior distributions to see how prior distribution affects the analysis results. Prior I refers to the set of prior distributions used previously. Prior II and Prior III are two other sets of prior distributions. The three sets of prior distributions represent different beliefs in the degree of heterogeneity among different centers.

Prior II has a small a and uniform distributions for $1/b$ and σ^2, it represents disperse prior distributions for v_i and λ_{i1} in center population. $1/b$ has a uniform prior is equivalent to that b has the prior $p(b) \propto 1/b^2$. So, Prior II is

$$\begin{aligned} a &= 3 \\ p(b) &\propto \frac{1}{b^2}, \quad b > 0 \\ \mu &\sim N(0,1) \\ p(\sigma^2) &\propto constant, \quad \sigma^2 > 0 \end{aligned}$$

The prior distributions of b and σ^2 are improper distributions. In this prior, our belief about the the center heterogeneity is vague. Consequently, data will have strong impact on the inference about center heterogeneity.

In Prior III, our prior belief is that there is only a small variation among centers. The parameter values are

$$a = 10, \quad c = 3, \quad d = 0.008, \quad u = 5 \quad \text{and} \quad w = 0.8$$

that is

$$a = 10$$

$$
\begin{aligned}
b &\sim Gamma(3, 0.008) \\
\mu &\sim N(0, 1) \\
\sigma^2 &\sim IG(5, 0.8)
\end{aligned}
$$

which sets the prior mean of a/b, a/b^2 and prior mean and variance of σ^2 to be

$$
\begin{aligned}
E(\frac{a}{b} \,|\, c, d) &= \frac{ad}{c-1} = 0.04 \\
E(\frac{a}{b^2} \,|\, c, d) &= \frac{ad^2}{(c-1)(c-2)} = 0.00032 \\
E(\sigma^2 \,|\, u, w) &= \frac{w}{u-1} = 0.20 \\
V(\sigma^2 \,|\, u, w) &= \frac{w^2}{(u-1)^2(u-2)} = 0.0133
\end{aligned}
$$

For comparison, Figure 6.6 displays the prior distributions of a/b, a/b^2, μ and σ^2 in Prior I and Prior III.

6.3.2 Results

In this subsection, the posterior distributions and other results based on Prior II and Prior III will be presented. We will compare all the results based on the three sets of prior distributions.

Prior II

The first graph in Figure 6.7 displays posterior distributions of v_i based on Prior II, $i = 1, \cdots, 5$. Because the prior distributions of $1/b$ and σ^2 are flat, the strength borrowing is not strong in the inference. The posterior distribution of v_i is relatively flat at all the centers, and there is a relatively large discrepancy

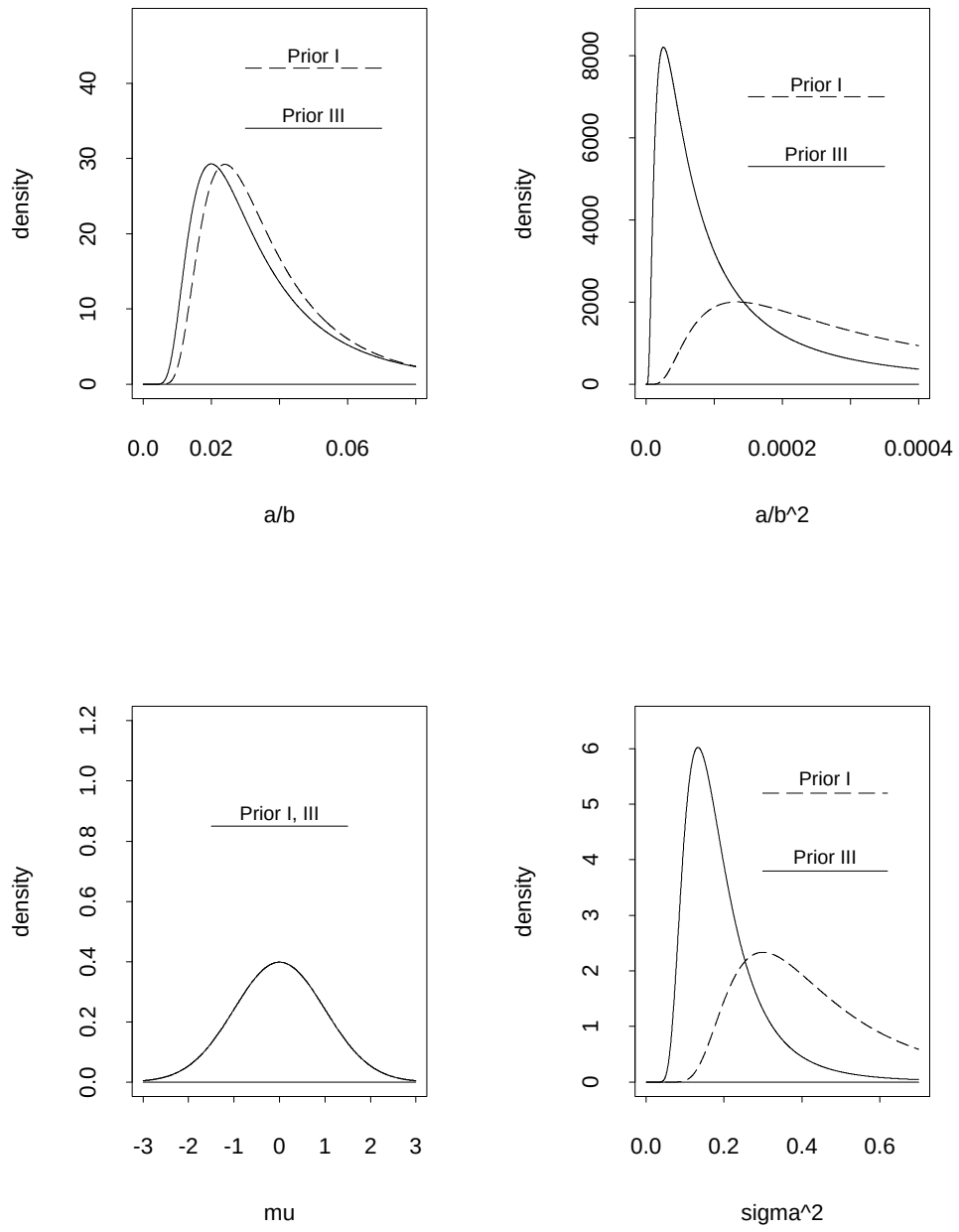

Figure 6.6: Prior distributions of a/b, a/b^2, μ and σ^2 in Prior I, III

Center	A	B	C	D	E
$P(v_i < 0 \mid D)$	0.965	0.806	0.912	0.799	0.976

Table 6.9: Posterior probability of $v_i < 0$ based on Prior II

Center	v_{iMLE}	$\hat{v}_i$	$\hat{\mu}$	$\mathrm{E}(v_i \mid D)$	$\mathrm{V}(v_i \mid D)$
A	-1.575	-1.463	-0.908	-1.545	0.381
B	-0.678	-0.838	-0.908	-0.783	0.332
C	-1.477	-1.182	-0.908	-1.251	0.403
D	-0.500	-0.682	-0.908	-0.645	0.243
E	-1.770	-1.451	-0.908	-1.434	0.224

Table 6.10: The MLE, posterior mode, pooled population mode, posterior mean and variance of v_i based on Prior II

among these posterior distributions. Table 6.9 displays the posterior probability of $v_i < 0$ for all the centers. At some centers, such as centers A and E, there is strong evidence that indicates that v_i is negative. At other centers, such as center B and D, we are less certain that v_i is negative. It is still likely that v_i is negative at all the centers, because the probability $P(v_i < 0 \mid D)$ is far above 0.5 in all centers.

Table 6.10 shows the individual center log hazard ratio (MLE), posterior mode of v_i ($\hat{v}_i$), pooled population log hazard ratio ($\hat{\mu}$), posterior mean and variance of v_i in each center based on Prior II. Column 2 to column 4 still indicate that the individual center log hazard ratios were shrunk toward the pooled population log hazard ratio. The posterior mode of v_i varies from -1.432 at center A to -0.680 at center D, which respectively correspond to a 319% and a 97% increase in mean and median survival time on therapy 2 over therapy 1.

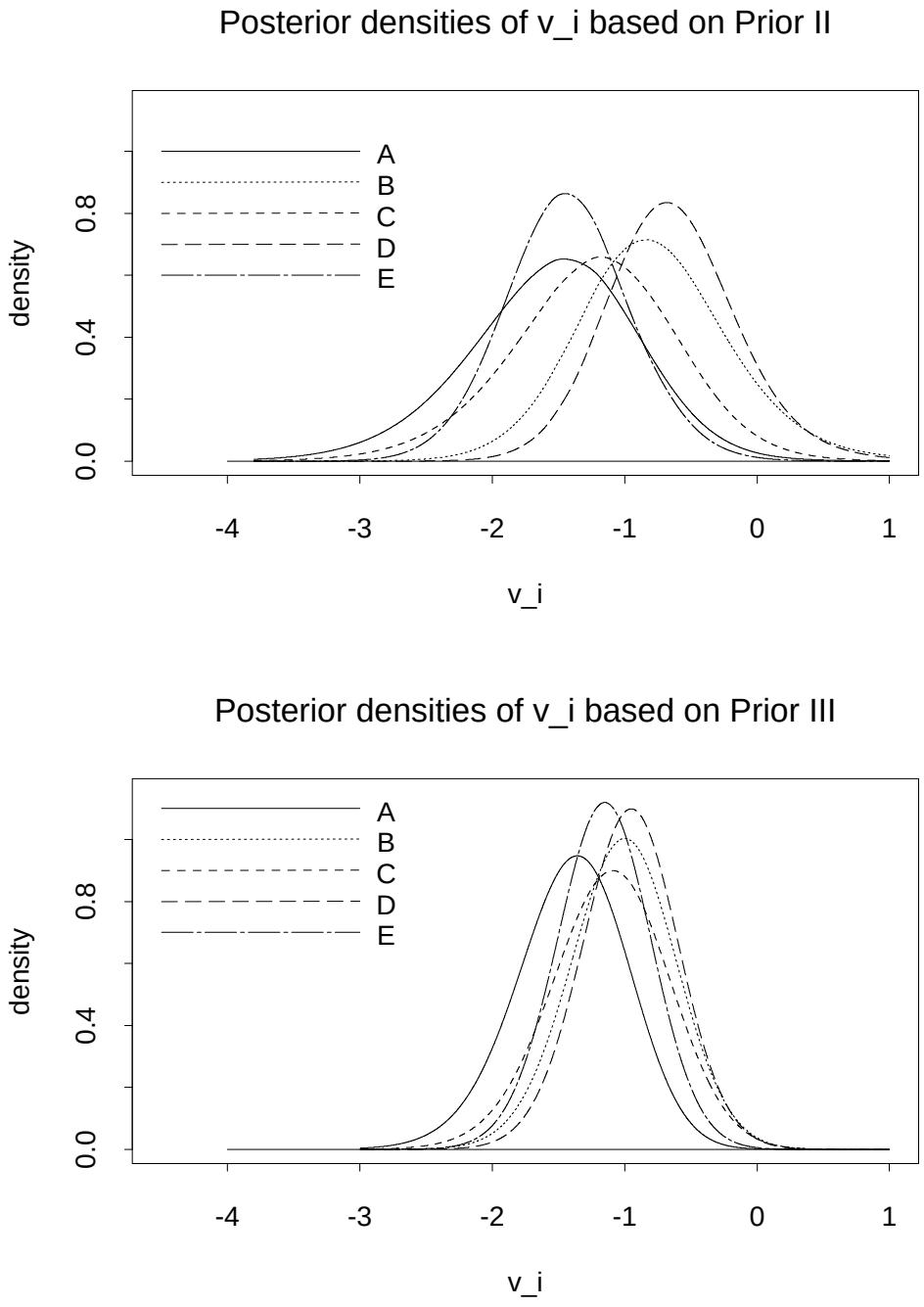

Figure 6.7: Posterior distribution of v_i based on Prior II and Prior III

Center	$\lambda_{i1\,MLE}$	$\hat{\lambda}_{i1}$	$(a-1)/\hat{b}$	$\mathrm{E}(\lambda_{i1} \mid D)$	$\mathrm{V}(\lambda_{i1} \mid D)$
A	0.012	0.012	0.019	0.013	0.000027
B	0.015	0.015	0.019	0.017	0.000054
C	0.040	0.028	0.019	0.032	0.000152
D	0.009	0.009	0.019	0.010	0.000011
E	0.063	0.042	0.019	0.047	0.000263

Table 6.11: The MLE, posterior mode, pooled population mode, posterior mean and variance of λ_{i1} based on Prior II

The first plot of Figure 6.8 displays the posterior distributions of λ_{i1} ($i = 1, \cdots, 5$) based on Prior II. As in the v_i's, we see a large variation among the five posterior distributions of λ_{i1}. These posterior distributions were relatively flat; and the posterior distribution of λ_{i1} at center E was extremely flat, because there was a large discrepancy among the observations at that center. The flat prior distributions on $1/b$ and σ^2 made each center borrow little strength from other centers.

Table 6.11 displays the individual center hazard rate (MLE), posterior mode of λ_{i1} ($\hat{\lambda}_{i1}$), pooled population hazard rate ($(a-1)/\hat{b}$), posterior mean and variance of λ_{i1} in each center based on Prior II. At centers C and E, the MLEs of individual center λ_{i1} are far above the pooled population value, and the posterior modes of λ_{i1} at these centers were shrunk down toward the population value. In this flat prior distribution case, only in very extreme centers, the $\hat{\lambda}_{i1}$ were shrunk to that about the population value. The posterior mode of λ_{i1} varies from 0.009 at center D to 0.042 at center E, which correspond to a 111.1-week and a 23.8-week mean survival time, respectively, on the *off imipramine* therapy.

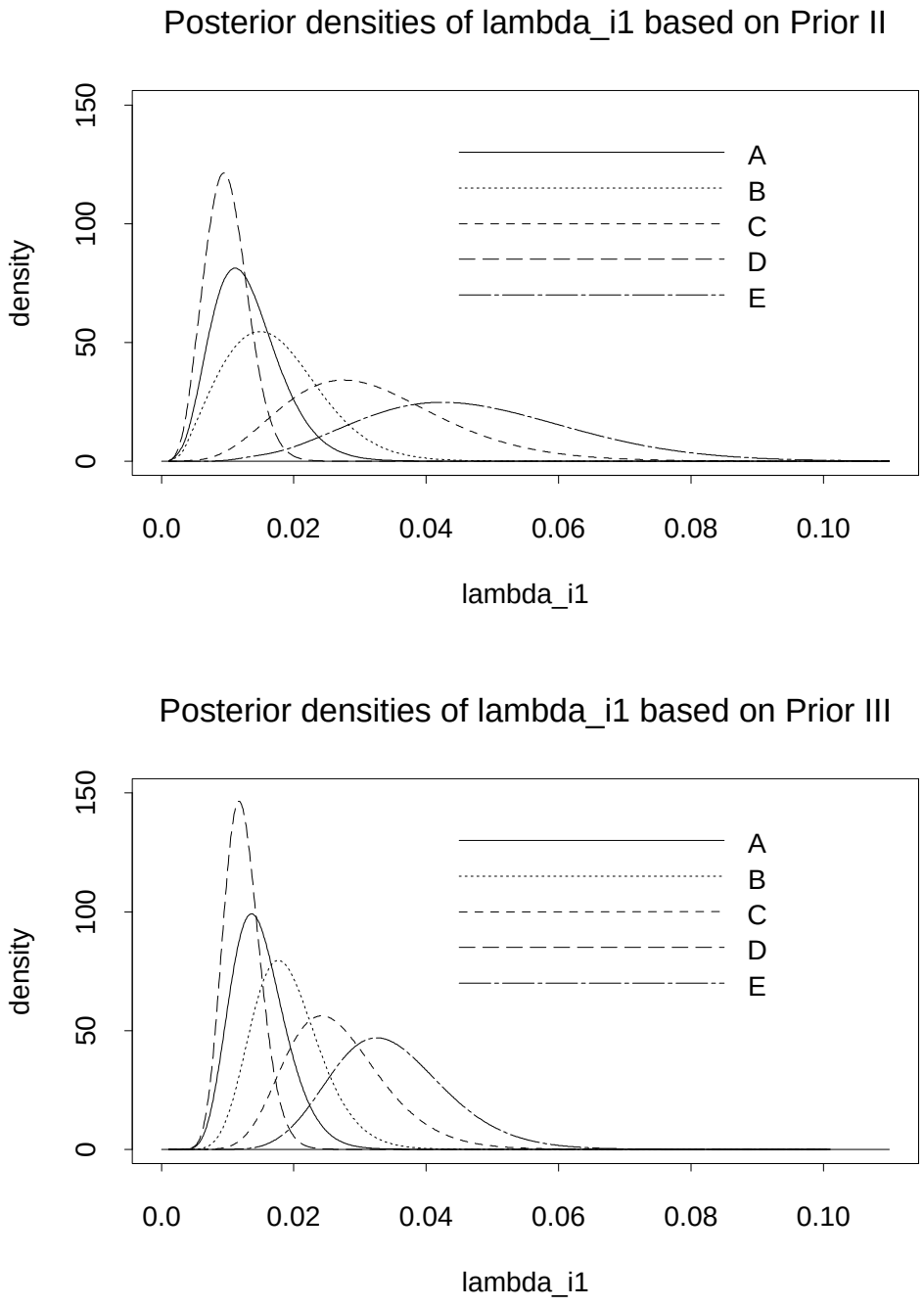

Figure 6.8: Posterior distribution of λ_{i1} based on Prior II and Prior III

Parameter	Mode	Mean	Variance
a/b	0.0288	0.0276	0.001617
a/b^2	0.0002763	0.0002860	2.2×10^{-7}
μ	-0.908	-0.745	0.440
σ^2	0.348	3.234	15.23

Table 6.12: Posterior mode, mean and variance of a/b, a/b^2, μ and σ^2 based on Prior II

The dotted curves in Figure 6.9 show the posterior distributions of population parameters a/b, a/b^2, μ and σ^2 based on Prior II. Our prior distributions of a/b and σ^2 are uniform distributions; and the prior distribution of μ is flat too. The posterior distributions of a/b and μ are much more concentrated than their prior distributions, but the posterior density of σ^2 is still relatively flat. Thus, the data we observed provided much information on population means a/b and μ, but they did not provide much information about σ^2.

Table 6.12 shows the posterior mode, mean and variance of the population mean and variance of λ_{i1} and v_i based on Prior II. The first two rows represent the population mean (a/b) and variance (a/b^2) of λ_{i1}. μ and σ^2 are the population mean and variance of v_i. The posterior mode of μ is -0.908, which corresponds to a 148% increase in mean survival time on therapy 2 over therapy 1 in the population.

Prior III

The second plot in Figure 6.7 displays the posterior distributions of v_i based on Prior III, $i = 1, \cdots, 5$. Since Prior III only assumes very little center heterogeneity, the strength borrowing is very strong, and we only see little variation among these posterior distributions. The posterior density curves of v_i at center B, C, D and E are very close; and the posterior density curve of v_i in center

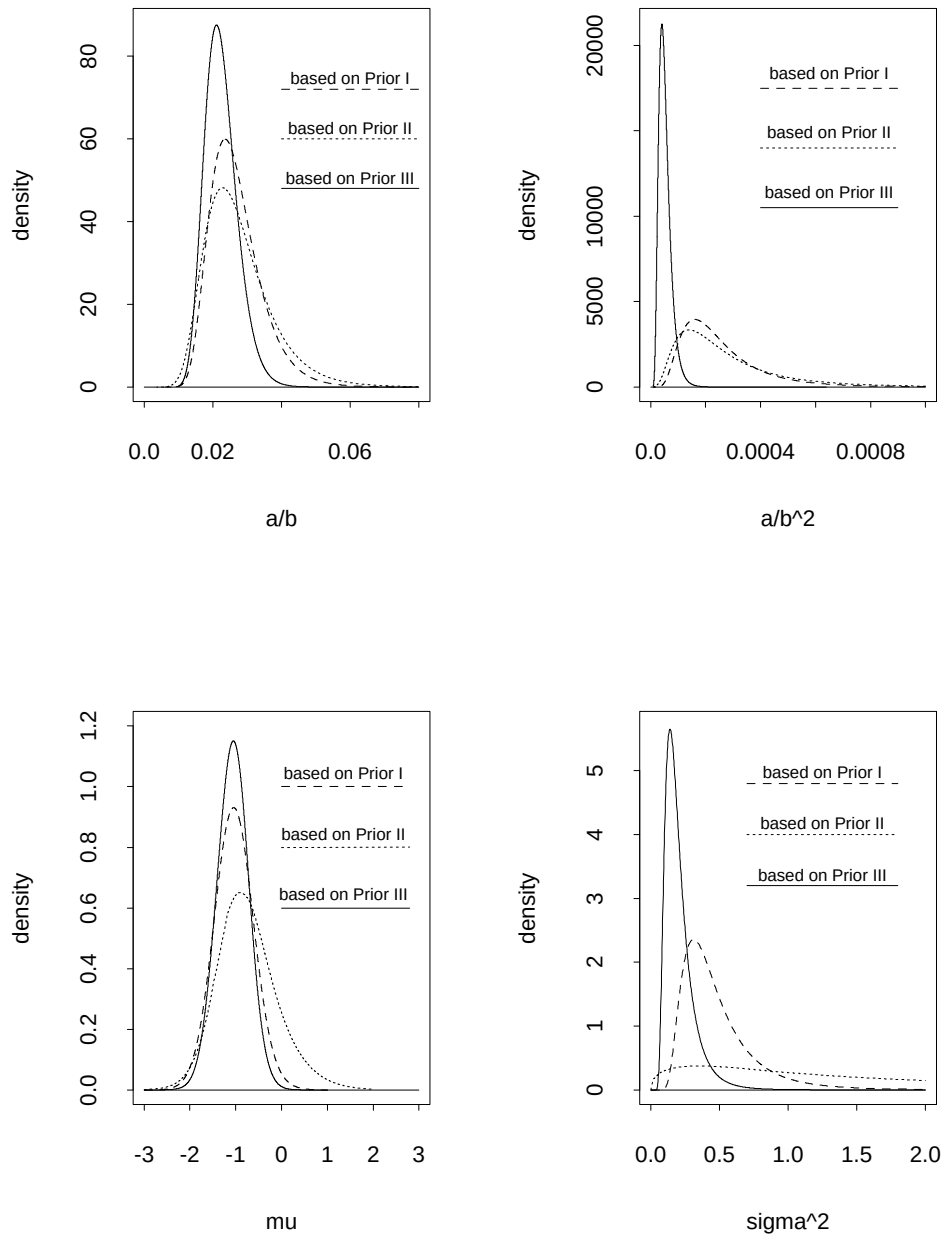

Figure 6.9: Posterior distributions of a/b, a/b^2, μ and σ^2

Center	A	B	C	D	E
$P(v_i < 0 \mid D)$	0.999	0.979	0.984	0.968	0.992

Table 6.13: Posterior probability of $v_i < 0$ based on Prior III

Center	v_{iMLE}	$\hat{v}_i$	$\hat{\mu}$	$\mathrm{E}(v_i \mid D)$	$\mathrm{V}(v_i \mid D)$
A	-1.575	-1.360	-1.052	-1.403	0.181
B	-0.678	-1.002	-1.052	-1.013	0.160
C	-1.477	-1.090	-1.052	-1.114	0.200
D	-0.500	-0.951	-1.052	-0.962	0.135
E	-1.770	-1.151	-1.052	-1.161	0.129

Table 6.14: The MLE, posterior mode, pooled population mode, posterior mean and variance of v_i based on Prior III

A shifted to the left a bit compared to the others. These curves suggest that centers B, C, D and E have almost the same log hazard ratio, and center A has a smaller log hazard ratio than other centers. Table 6.13 displays posterior probabilities of $v_i < 0$ at each center, and those posterior probabilities strongly suggest that v_i is negative for all the centers, which means that patients on therapy 2 survive longer than those on therapy 1 at all the centers.

Table 6.14 shows the individual center log hazard ratio (MLE), posterior mode of v_i ($\hat{v}_i$), pooled population log hazard ratio ($\hat{\mu}$), posterior mean and variance of v_i at each center based on Prior III. From column 2 to column 4, we can see that the individual center log hazard ratio at each center was strongly shrunk toward the population value. The posterior mode of v_i changes from -1.259 at center A to -0.982 at center B, which correspond to a 252% and a 167% increase, respectively, in mean survival time on therapy 2 over therapy 1.

Center	$\lambda_{i1\,MLE}$	$\hat{\lambda}_{i1}$	$(a-1)/\hat{b}$	$\mathrm{E}(\lambda_{i1} \mid D)$	$\mathrm{V}(\lambda_{i1} \mid D)$
A	0.012	0.014	0.019	0.015	0.000018
B	0.015	0.018	0.019	0.019	0.000027
C	0.040	0.024	0.019	0.027	0.000056
D	0.009	0.012	0.019	0.012	0.000008
E	0.063	0.033	0.019	0.034	0.000077

Table 6.15: The MLE, posterior mode, pooled population mode, posterior mean and variance of λ_{i1} based on Prior III

The second graph of Figure 6.8 displays posterior distributions of λ_{i1} $(i = 1, \cdots, 5)$ based on Prior III. Because of the strong strength borrowing, the posterior distribution of λ_{i1} is relatively less flat. We can still clearly see some variation among these posterior density curves. Centers D and E are two centers with extreme λ_{i1}'s, and the posterior density curves of λ_{i1} at the two centers shared only a small area.

Table 6.15 shows the individual center hazard rate (MLE), posterior mode of λ_{i1} $(\hat{\lambda}_{i1})$, pooled population hazard rate $((a-1)/\hat{b})$, posterior mean and variance of λ_{i1} at each center based on Prior III. From column 2 to column 4, we can see that the individual center λ_{i1} was strongly shrunk toward the population value. The posterior mode of λ_{i1} changes from 0.013 at center D to 0.026 at center E, which correspond to a 76.9-week and a 38.5-week mean survival time respectively. The mean survival time doubles from center E to center D.

Table 6.16 displays the posterior mode, mean, and variance of the population mean and variance of λ_{i1} in its second and third rows. The posterior mode, mean and variance of μ and σ^2 are displayed in the last two rows of the table. We see very small values for a/b^2 and σ^2, which indicate strong

Parameter	Mode	Mean	Variance
a/b	0.021	0.023	0.0000228
a/b^2	0.0000406	0.0000528	5.4×10^{-10}
μ	-1.052	-1.068	0.122
σ^2	0.139	0.209	0.014

Table 6.16: Posterior mode, mean and variance of a/b, a/b^2, μ and σ^2 based on Prior III

center homogeneity. The posterior mode of μ is -1.037, which corresponds to a 182% increase in mean survival time on the *on imipramine* therapy over the *off imipramine* therapy.

6.3.3 Comparison of the Three Prior Distributions

When we compare the posterior distributions derived from the three sets of prior distributions that we chose, we can see how the posterior distributions are affected by the prior distributions. By comparing the corresponding plots in the graphs from Figure 6.3 to Figure 6.9, we can see that the shrinkage of the individual center parameters increased as the priors changed from Prior II to Prior I, and from Prior I to Prior III. As we expected, the posterior distributions of individual center parameters derived from a flatter prior are less homogeneous and less concentrated. Prior belief in the center heterogeneity has strong impact on the posterior distributions of individual center parameters, but it has a relatively lighter impact on the posterior distributions of the population mean of the individual center parameters.

When we compare the corresponding numbers in the tables from Table 6.3 to Table 6.16, we can see that the range of the posterior modes of individual center parameters changes with the variation of prior belief in center hetero-

geneity. The range of posterior modes of v_i from center to center is 0.752 based on Prior II, it decreased to 0.629 based on Prior I, and it further decreased to 0.277 based on Prior III. A similar thing happens to λ_{i1}; the range of posterior modes of λ_{i1} from center to center is 0.033 based on Prior II and Prior I, but it decreased to 0.013 based on Prior III.

By comparing the corresponding posterior variances of the individual centers and population parameters derived from the three sets of prior distributions, we can clearly see that the posterior variance decreased as the prior variation between centers decreases. For example, the posterior variance of v_i at center A is 0.381 based on Prior II. It decreased to 0.274 based on Prior I, and it further decreased to 0.181 based on Prior III. For other examples, the posterior variance of λ_{i1} in center C is 0.000152 based on Prior II, it decreased to 0.000124 based on Prior I, and it further decreased to 0.000056 based on Prior III. The posterior variance of population parameter μ is 0.440 based on Prior II, it decreased to 0.188 based on Prior I, and it further decreased to 0.122 based on Prior III.

The data of this study strongly suggest that there is a certain center heterogeneity between different centers, even when it is our prior belief that there is only very small center heterogeneity (Prior III), the posterior distributions still showed a strong heterogeneity on the center hazard rate λ_{i1} for patients on the *off imipramine* therapy.

Appendix A

Data from the NSCLC Study

Variable Codes

Center:	Medical Center
Time:	Months until a death
Status:	A= Alive
	D= Death
Therapy:	1=RT
	2=CT+RT

Center	Time	Status	Therapy
1	83.77	A	1
1	16.53	D	2
2	4.40	D	1
2	7.57	D	1
2	62.50	A	1
2	12.77	D	1
2	9.47	D	2
3	10.67	D	1
3	13.23	D	1
3	6.10	D	1
3	45.40	D	1
3	8.57	D	1
3	10.47	D	1
3	21.43	D	1
3	2.27	D	1
3	2.73	D	1
3	16.23	D	1
3	19.73	D	1
3	32.90	D	1
3	23.00	D	2
3	18.13	D	2
3	62.37	D	2
3	18.53	D	2
3	3.97	D	2
3	59.03	A	2
3	10.97	D	2
3	66.07	A	2
3	6.27	D	2
3	55.03	A	2
4	44.13	D	1
4	8.53	D	1

Table A.1: Data from the NSCLC Study — Part 1

Center	Time	Status	Therapy
4	19.93	D	1
4	4.23	D	1
4	23.30	D	1
4	4.50	D	1
4	18.10	D	1
4	15.83	D	1
4	9.67	D	1
4	42.47	A	1
4	13.10	D	1
4	4.03	D	2
4	62.40	D	2
4	9.80	D	2
4	6.70	D	2
4	16.53	D	2
4	8.33	D	2
4	11.67	D	2
4	12.83	D	2
4	8.73	D	2
4	3.90	D	2
4	55.73	A	2
5	1.27	D	1
5	15.33	D	1
5	48.07	D	2
6	31.20	A	1
6	2.60	D	1
6	10.53	D	1
6	18.93	D	2
6	39.47	A	2
6	40.27	D	2
6	5.03	D	2
7	6.00	D	1

Table A.2: Data from the NSCLC Study — Part 2

Center	Time	Status	Therapy
7	6.77	D	1
7	1.23	D	1
7	12.67	D	1
7	4.50	D	2
7	11.40	D	2
7	66.33	A	2
8	12.63	D	1
8	10.13	D	1
8	7.00	D	2
8	19.03	D	2
9	15.00	D	1
9	9.50	D	1
9	28.83	D	2
9	52.60	A	2
10	0.37	D	1
10	7.50	D	1
10	64.87	A	1
10	10.67	D	1
10	20.67	D	2
10	13.73	D	2
10	6.47	D	2
10	15.57	D	2
11	19.77	D	1
11	8.13	D	1
11	7.67	D	1
11	4.83	D	1
11	5.33	D	1
11	14.20	D	1
11	57.43	A	2
11	69.13	A	2
11	52.67	A	2

Table A.3: Data from the NSCLC Study — Part 3

Center	Time	Status	Therapy
11	23.43	D	2
11	1.83	D	2
11	7.47	D	2
12	2.30	D	1
12	6.57	D	2
12	21.20	D	2
12	5.93	D	2
13	6.90	D	1
13	31.93	D	1
13	15.20	D	1
13	73.43	A	1
13	10.83	D	1
13	6.10	D	1
13	3.37	D	1
13	17.23	D	2
13	6.33	D	2
13	20.47	D	2
13	2.70	D	2
13	16.40	D	2
13	13.30	D	2
13	7.97	D	2
13	5.20	D	2
13	3.13	D	2
13	6.07	D	2
13	7.20	D	2
14	7.63	D	1
14	2.97	D	1
14	6.87	D	1
14	16.87	D	1
14	3.63	D	1
14	2.93	D	1

Table A.4: Data from the NSCLC Study — Part 4

Center	Time	Status	Therapy
14	16.87	D	2
14	46.90	D	2
14	7.53	D	2
14	0.20	D	2
14	10.03	D	2
14	14.57	D	2
14	14.07	D	2
15	3.57	D	1
15	47.83	D	2
16	2.40	D	1
16	8.90	D	1
16	16.10	D	1
16	17.47	D	2
16	9.43	D	2
16	19.07	D	2
17	8.30	D	1
17	56.67	A	2
18	1.27	D	1
18	9.50	D	2
18	9.03	D	2
19	10.27	D	1
19	0.27	D	1
19	23.40	D	1
19	73.93	A	2
19	10.10	D	2
19	7.00	D	2
20	2.87	D	1
20	7.17	D	1
21	12.03	D	2
22	17.50	D	1
22	55.77	A	2

Table A.5: Data from the NSCLC Study — Part 5

Appendix B

Data from the NIMH Collaborative Study

Variable Codes

Center:	**Medical Center**
Time:	**Weeks until a recurrence**
Status:	**0= Censored**
	1= Recurrence
Therapy:	**1=Off Imopramine**
	2=On Imopramine

Center	Recurrence	Status	Therapy
1	36.143	1	1
1	49.714	1	1
1	5.000	1	1
1	2.857	1	1
1	55.714	1	1
1	5.571	1	1
1	14.429	0	1
1	104.857	0	1
1	102.429	0	1
1	105.857	0	1
1	16.286	0	1
1	8.429	1	2
1	13.429	1	2
1	27.286	1	2
1	105.143	0	2
1	74.571	0	2
1	102.143	0	2
1	108.857	0	2
1	106.429	0	2
1	105.143	0	2
1	83.000	0	2
1	104.000	0	2
1	83.000	0	2
1	98.000	0	2
1	98.000	0	2
1	88.000	0	2

Table B.1: Data from the NIMH Collaborative Study — Center A

Center	Recurrence	Status	Therapy
2	1.286	1	1
2	4.000	1	1
2	74.143	1	1
2	0.143	1	1
2	1.429	1	1
2	45.857	1	1
2	2.143	0	1
2	104.857	0	1
2	78.429	0	1
2	78.429	0	1
2	27.143	1	2
2	9.857	1	2
2	42.429	1	2
2	17.429	1	2
2	18.000	1	2
2	66.857	1	2
2	100.000	0	2
2	52.143	0	2
2	78.000	0	2
2	78.857	0	2
2	54.857	0	2
2	78.286	0	2
2	78.143	0	2
2	52.000	0	2
2	15.857	0	2

Table B.2: Data from the NIMH Collaborative Study — Center B

Center	Recurrence	Status	Therapy
3	9.000	1	1
3	3.286	1	1
3	30.000	1	1
3	7.143	1	1
3	31.000	1	1
3	17.286	1	1
3	0.143	1	1
3	79.000	0	1
3	27.571	1	2
3	49.143	1	2
3	16.714	1	2
3	32.571	0	2
3	206.000	0	2

Table B.3: Data from the NIMH Collaborative Study — Center C

Center	Recurrence	Status	Therapy
4	3.286	1	1
4	19.714	1	1
4	8.000	1	1
4	71.714	1	1
4	63.714	1	1
4	36.286	1	1
4	8.143	1	1
4	16.000	1	1
4	37.857	1	1
4	11.143	1	1
4	44.000	1	1
4	0.286	1	1
4	96.286	0	1
4	50.857	0	1

Table B.4: Data from the NIMH Collaborative Study — Center D (I)

Center	Recurrence	Status	Therapy
4	102.571	0	1
4	165.000	0	1
4	124.571	0	1
4	68.000	0	1
4	39.571	0	1
4	131.000	0	1
4	42.000	0	1
4	115.000	0	1
4	77.857	0	1
4	12.429	0	1
4	22.000	0	1
4	16.143	1	2
4	4.429	1	2
4	21.000	1	2
4	16.143	1	2
4	50.571	1	2
4	3.429	1	2
4	13.429	1	2
4	92.714	1	2
4	1.571	0	2
4	126.714	0	2
4	13.000	0	2
4	155.000	0	2
4	39.571	0	2
4	112.571	0	2
4	115.571	0	2
4	28.000	0	2
4	38.000	0	2
4	111.571	0	2
4	26.000	0	2
4	108.000	0	2
4	106.714	0	2
4	55.000	0	2
4	75.000	0	2
4	52.714	0	2
4	86.000	0	2
4	34.517	0	2

Table B.5: Data from the NIMH Collaborative Study — Center D (II)

Center	Recurrence	Status	Therapy
5	5.429	1	1
5	6.286	1	1
5	5.286	1	1
5	3.429	1	1
5	6.571	1	1
5	1.000	1	1
5	0.857	1	1
5	4.714	1	1
5	46.286	1	1
5	0.571	1	1
5	6.429	1	1
5	0.000	1	1
5	20.857	1	1
5	18.286	1	1
5	2.000	1	1
5	67.000	0	1
5	45.000	0	1
5	3.429	1	2
5	1.000	1	2
5	50.286	1	2
5	16.857	1	2
5	0.571	1	2
5	22.143	1	2
5	31.857	1	2
5	22.000	1	2
5	13.429	1	2
5	5.000	0	2
5	15.000	0	2
5	128.143	0	2
5	109.571	0	2
5	106.000	0	2
5	9.143	0	2
5	102.000	0	2
5	104.000	0	2
5	105.143	0	2

Table B.6: Data from the NIMH Collaborative Study — Center E

Appendix C

Computer Programs

This appendix contains the programs used in numerical integration and the Gibbs sampling algorithms. These programs were written in FORTRAN 77 using the following IMSL subroutines and functions for random number generators and calculation:

- `drnun` – subroutine for Uniform $[0, 1]$ generation
- `drnnoa` – subroutine for Normal $(0, 1)$ generation
- `drngam` – subroutine for Gamma$(a, 1)$ generation
- `dgamma` – function for $\Gamma(x)$
- `dlgams` – function for $ln(\Gamma(x))$

C.1 Numerical Integration in Chapter 3

```
C     this is for calculating p(v|data), p(\lambda_1|data)

      integer d1,d2
      real*8 t1,t2,x,s,a,b,y,v(3000),c,l(2000)
```

```
      real*8 y1,s1,s2,c2

      read *, d1,d2,t1,t2    ! the sufficient statistics
      a=2.0
      b=20.0

C     First is for p(v|data)
      s=0.0
      do 10 i=1,3000
      x=-3.+0.0014*i
      y=d2*x-x*x*0.5-(a+d1+d2)*dlog(b+t1+t2*dexp(x))
      v(i)=dexp(y)
      s=s+v(i)
 10   continue
      c=1.0/(s*0.0014)      ! normalizing constant

      do 20 i=1,3000
      x=-3.0+0.0014*i
      v(i)=v(i)*c
      print '(2f15.5)', x, v(i)
 20   continue

C     Forrowing is For Posterior of Lambda_1
      do 30 i=1,2000
       x=0.00008+(i-1)*0.0001
       s=0.0
       do 40 j=1,3000
        y=-3.0+(j-1)*0.0013
        s=s+dexp(d2*y-y*y*0.5-x*t2*dexp(y))
 40    continue
       s1=s*c*0.0013
       y1=(a+d1+d2-1.)*dlog(x)-(b+t1)*x+450.0
       l(i)=s1*dexp(y1)
       s2=s2+l(i)
 30   continue

      c2=1.0/(s2*0.0001)
      do 50 i=1,2000
        x=0.00008+(i-1)*0.0001
        l(i)=l(i)*c2
        print '(2f15.5)', x, l(i)
 50   continue

      end
```

C.2 Numerical Methods in Hierarchical Model

This section contains programs in the hierarchical model used in a multicenter trial.

C.2.1 Gibbs Sampling

The following Fortran program does Gibbs sampling in a multicenter trial.

```
C     This is for doing Gibbs sampling in a multicenter trial

C     m --- the number of centers
C     u(i,j) is the number of uncensored observations in center i
C             on treatment j (-- d(i,j) in chapters 4,5,6 )
C     n(i,j) is the total number of observations in center i on
C             treatment j
C     t(i,j) is the total exposure time observed in center i on
C             treatment j

C     prior of mu is N(0,1)
C     prior of b is gamma(c,d)
C     prior of sigma^2 is IG(aa,bb)

C     a1(i), a2(i) are the parameters of the gamma dis. for lambda_i1
C     a1(m+1), a2(m+1) are the parameters of the gamma dis. for b
C     y(i) is lambda_i1, z(i) is v(i), sig is sigma^2

      parameter(m=22,mm=2000)
      integer u(m,2), ll,l
      real*8  y(m+1), z(m), t(m,2), mu, sig, mv
      real*8  s, f, f1, w, v, vhat, x0, x1, dd, err
      real*8 a1(m+1),a2(m+1),a,c,d, aa, bb, z1, z2
      real*8 drngam, drnnoa, drnun
      external drngam, drnnoa, drnun

C     ut:            sum of u(i,1)

      a=6.0
      c=4.0
      d=0.075  ! prior of b is gamma(c,d)
      a1(m+1)=m*a+c
```

```
      aa=3.0
      bb=.32   ! prior of sig is inverse gamma(aa,bb)

      do 10 i=1,m
        read *, u(i,1),t(i,1),u(i,2),t(i,2)
 10   continue

      do 5 i=1,m
         a1(i)=a+u(i,1)+u(i,2)
 5    continue

      do 100 ll=1,mm       ! ll is the count of Gibbs sampling
        do 20 k=1, m
           a2(k)=20.0                    ! initial value
           call drngam(1,a1(k),y(k))
           y(k)=y(k)/a2(k)
           call drnnoa(1, z(k))
20      continue
           call drngam(1, c, y(m+1))
           y(m+1)=y(m+1)/d
           call drnnoa(1,mu)          ! inisial mu from N(0,1)
           call drngam(1,aa, sig)
           sig=bb/sig               ! prior of sig: Inverse G(aa,bb)

        do 101 l=1,30             ! l is the number of iterations
          a2(m+1)=0.0
          mv=0.0
          do 65 i=1,m
            a2(i)=t(i,2)*exp(z(i))+t(i,1)+y(m+1)
            call drngam(1,a1(i),y(i))
            y(i)=y(i)/a2(i)
            a2(m+1)=a2(m+1)+y(i)

C  following is the process of sampling v(i) with rejection method
C  first is to find the mode of a normal distribution and the
C    envelop function of the conditional distribution of v(i)
C  f is the logrithm of the factor function in the conditional
C    distribution of v(i)
C  f1 is the logrithm of the envelop function

            x0=-0.1
 25           z1=x0+y(i)*t(i,2)*sig*exp(x0)-mu-u(i,2)*sig
              z2=1.0+y(i)*t(i,2)*sig*exp(x0)
              x1=x0-z1/z2
```

```
            err= abs(x1-x0)
              if ( err .le. .000001) goto 35
               x0=x1
               go to 25
35          vhat=x1                    ! vhat is the mode
            dd=u(i,2)-(vhat-mu)/sig
30          call drnun(1,w)
            call drnnoa(1,v)
            v=vhat+v*sqrt(sig)
            f1=dd*(vhat-1.0-0.5*(v-vhat)**2-(v-vhat)**3/6.0)
            f=dd*v-t(i,2)*y(i)*exp(v)
            if (f-f1 .lt. log(w)) go to 30
            z(i)=v
            mv=mv+v
  65      continue

          a2(m+1)=a2(m+1)+d
          call drngam(1,a1(m+1),y(m+1))
          y(m+1)=y(m+1)/a2(m+1)

          mv=mv/(sig+m)
          call drnnoa(1,mu)
          mu=mv+mu*sqrt(sig/(sig+m))

          s=0.0
          do 40 k=1, m
            s=s+(z(k)-mu)**2
40        continue
          call drngam(1, aa+0.5*m, sig)
          sig=(bb+0.5*s)/sig
 101    continue

          print '(7f10.5)', (y(i),i=1,7)
          print '(7f10.5)', (y(i),i=8,14)
          print '(8f9.5)', (y(i),i=15,22)
          print '(7f10.5)', (z(i),i=1,7)
          print '(7f10.5)', (z(i),i=8,14)
          print '(8f9.5)', (z(i),i=15,22)
          print '(3f14.8)', y(m+1), mu, sig
100     continue

      end
```

C.2.2 Posterior density of v_i

Following is the program of calculating the posterior density functions of v_i's.

```
C     this is to calculate posterior density in multicenter trial
C     for the posterior of the v_i's -- log hazard ratio in each center
C     c(m,mm) are the normalizing constant for the conditional density
C     of v_i's
C     m --- the number of centers,
C     mm --- the sample size in Gibbs Sampling
C     y(i,j) --- j-th observation of lambda_i1 in Gibbs sampling
C     z(i,j) --- j-th observation of v_i in Gibbs sampling
C     sig(j) --- j-th observation of sigma^2 in Gibbs sampling
C     fv(i) --- posterior density function of v_i

      parameter(m=22,mm=2000)
      real*8 y(m,mm),z(m,mm),b(mm),mu(mm),sig(mm),c(m,mm)
      real*8  fv(m)
      integer d(m,2)
      real*8 t(m,2),x,a,w,s

      a=15.0
      do 5 i=1,m               ! read in sufficient statistics
        read *, d(i,1),t(i,1),d(i,2),t(i,2)
 5    continue

      do 10 k=1,mm              ! read in the Gibbs sampler
        read *, (y(i,k), i=1,7)
        read *, (y(i,k), i=8,14)
        read *, (y(i,k), i=15,22)
        read *, (z(i,k), i=1,7)
        read *, (z(i,k), i=8,14)
        read *, (z(i,k), i=15,22)
        read *, b(k),mu(k),sig(k)
 10   continue

      do 20 i=1,m     ! calculate normalize constant for v_i's
        do 30 j=1,mm
          s=0.0
          do 40 k=1,5000
             x=-3+(k-1)*.001
             w=-.5*(x-mu(j))**2/sig(j)+d(i,2)*x-y(i,j)*t(i,2)*dexp(x)
             s=s+dexp(w+45.0)
 40       continue
```

```
          c(i,j)=5000.0/(s*5.0)
30      continue
20    continue

      do 50 i=1,2751
        x=-3.0+(i-1)*.002
           do 60 j=1,m
              fv(j)=0.0
              do 70 k=1,mm
                w=-0.5*(x-mu(k))**2/sig(k)+d(j,2)*x
     &              -y(j,k)*t(j,2)*dexp(x)+45.0
                fv(j)=fv(j)+c(j,k)*dexp(w)
70            continue
              fv(j)=fv(j)/mm
60          continue
        print '(8f9.4)', x,(fv(kk),kk=1,7)
        print '(8f9.4)', (fv(kk),kk=8,15)
        print '(7f9.4)', (fv(kk),kk=16,22)
50    continue

      end
```

C.2.3 Posterior density of λ_{i1}

Following is the program of calculating the posterior density functions of λ_{i1}'s.

```
C     this is to calculate posterior density of lambda_i1's in a
C     multicenter trial with a hierarchical model
C     n --- number of points at which density function is calculated
C     m --- the number of centers,
C     mm --- the sample size in Gibbs Sampling
C     y(i,j) --- $j$th observation of lambda_i1 in Gibbs sampling
C     z(i,j) --- $j$th observation of v_i in Gibbs sampling
C     sig(j) --- $j$th observation of sigma^2 in Gibbs sampling
C     flam(i) --- posterior density function of $\lambda_{i1}$

      parameter(m=22,mm=2000,n=800)
      real*8 y(m,mm),z(m,mm),b(mm),mu(mm),sig(mm),flam(m)
      integer d(m,2),u(m)
      real*8 t(m,2),a,w,w1,logal,s,x
      real*8 dlgams, dgamma
```

```
      external dlgams, dgamma

      a=15.0
      do 5 i=1,m               ! read in sufficient statistics
        read *, d(i,1),t(i,1),d(i,2),t(i,2)
        u(i)=d(i,1)+d(i,2)
5     continue

      do 10 k=1,mm              ! read in the Gibbs sampler
        read *, (y(i,k), i=1,7)
        read *, (y(i,k), i=8,14)
        read *, (y(i,k), i=15,22)
        read *, (z(i,k), i=1,7)
        read *, (z(i,k), i=8,14)
        read *, (z(i,k), i=15,22)
        read *, b(k),mu(k),sig(k)
10    continue

      do 20 i=1,n
        x=0.002+i*.298/800.
          do 40 j=1,m
            flam(j)=0.0
            call dlgams(a+u(j),logal,s)
            do 50 k=1,mm
              w=b(k)+t(j,1)+t(j,2)*dexp(z(j,k))
              w1=dlog(w)-x*w+(a+u(j)-1)*dlog(x*w)-logal
              flam(j)=flam(j)+dexp(w1)
50          continue
            flam(j)=flam(j)/mm
40        continue
        print '(8f9.4)', x,(flam(kk),kk=1,7)
        print '(8f9.4)', (flam(kk),kk=8,15)
        print '(7f9.4)', (flam(kk),kk=16,22)
20    continue

      end
```

C.2.4 Posterior density of b

Following is the program of calculating the posterior density function of b.

```
C     this is to calculate the posterior density of b
```

```
C     n --- number of points at which density function is calculated
C     m --- the number of centers,
C     mm --- the sample size in Gibbs Sampling
C     y(i,j) --- $j$th observation of lambda_i1 in Gibbs sampling
C     z(i,j) --- $j$th observation of v_i in Gibbs sampling
C     sig(j) --- $j$th observation of sigma^2 in Gibbs sampling
C     fb --- posterior density function of b

      parameter(m=22,mm=2000,n=2000)
      real*8 y(m,mm),z(m,mm),b(mm),mu(mm),sig(mm)
      real*8  x, fb, sum, gamma,dlgams
      real*8 sd,ss,a,c,a1,b1,d,s,loga1
      external dgamma,dlgams

      a=6.0
      c=4.0
      d=0.075
      a1=m*a+c
      call dlgams(a1,loga1,s)

      do 10 k=1,mm              ! read in the Gibbs sampler
        read *, (y(i,k), i=1,7)
        read *, (y(i,k), i=8,14)
        read *, (y(i,k), i=15,22)
        read *, (z(i,k), i=1,7)
        read *, (z(i,k), i=8,14)
        read *, (z(i,k), i=15,22)
        read *, b(k),mu(k),sig(k)
 10   continue

      do 20 i=1,n
        x=20+(i-1)*.15
        fb=0.0
        do 30 k=1,mm
           sum=0.0
           do 40 j=1,m
              sum=sum+y(j,k)
 40        continue
           b1=d+sum
           ss=a1*log(b1)+(a1-1.0)*dlog(x)-b1*x-loga1
           fb=fb+dexp(ss)
 30     continue
        fb=fb/mm
        print '(2f20.10)',x,fb
```

```
 20   continue

      end
```

C.2.5 Posterior density of μ

Following is the program of calculating the posterior density function of μ.

```
C     this is to calculate posterior density of mu
C     m --- the number of centers,
C     mm --- the sample size in Gibbs Sampling
C     y(i,j) --- $j$th observation of lambda_i1 in Gibbs sampling
C     z(i,j) --- $j$th observation of v_i in Gibbs sampling
C     sig(j) --- $j$th observation of sigma^2 in Gibbs sampling
C     fmu --- posterior density function of mu

      parameter(m=22,mm=2000)
      real*8 y(m,mm),z(m,mm),b(mm),mu(mm),sig(mm)
      real*8  fmu,aveg
      real*8 sd,x,ss


      do 10 k=1,mm              ! read in the Gibbs sampler
        read *, (y(i,k), i=1,7)
        read *, (y(i,k), i=8,14)
        read *, (y(i,k), i=15,22)
        read *, (z(i,k), i=1,7)
        read *, (z(i,k), i=8,14)
        read *, (z(i,k), i=15,22)
        read *, b(k),mu(k),sig(k)
 10   continue

      do 50 i=1,1500
        x=-1.5+i*.0015
           fmu=0.0
           do 60 k=1,mm
             aveg=0.0
             do 70 j=1,m
               aveg=aveg+z(j,k)
 70          continue
             aveg=aveg/(m+sig(k))
             sd = sig(k)/(sig(k)+m)
```

```
            ss = -0.5*(x-aveg)*(x-aveg)/sd
            fmu=fmu+dexp(ss)/sqrt(sd)
60         continue
          fmu=fmu/(sqrt(2*3.141596)*mm)
       print '(2f15.6)',x,fmu
50   continue

     end
```

C.2.6 Posterior density of σ^2

Following is the program of calculating the posterior density function of σ^2.

```
C     this is to calculate posterior density of sigma^2
C     n --- number of points at which density function is calculated
C     m --- the number of centers,
C     mm --- the sample size in Gibbs Sampling
C     y(i,j) --- j-th observation of lambda_i1 in Gibbs sampling
C     z(i,j) --- j-th observation of v_i in Gibbs sampling
C     sig(j) --- j-th observation of sigma^2 in Gibbs sampling
C     fsig --- posterior density function ofsigma^2

      parameter(m=22,mm=2000,n=2000)
      real*8 y(m,mm),z(m,mm),b(mm),mu(mm),sig(mm)
      real*8  fsig,mean,dgamma,s
      real*8 w,x,w1,u,s1,dlgams, logal
      external dgamma,dlgams

      u=3.0    ! parameters in prior of sigma^2
      w=.32

      call dlgams(u+.5*m,logal,s1)  ! calculate log-gamma(u+.5*m)

      do 10 k=1,mm              ! read in the Gibbs sampler
        read *, (y(i,k), i=1,7)
        read *, (y(i,k), i=8,14)
        read *, (y(i,k), i=15,22)
        read *, (z(i,k), i=1,7)
        read *, (z(i,k), i=8,14)
        read *, (z(i,k), i=15,22)
        read *, b(k),mu(k),sig(k)
10    continue
```

```
      do 50 i=1,n
        x=0.001+(i-1)*.001
           fsig=0.0
           do 60 k=1,mm
             mean=0.0
             do 70 j=1,m
               mean=mean+(z(j,k)-mu(k))**2
70           continue
             w1=mean*0.5+w
             s = (u+0.5*m)*dlog(w1)-w1/x -logal-(u+.5*m+1)*dlog(x)
             fsig=fsig+dexp(s)
60          continue
           fsig=fsig/mm
        print '(2f20.8)',x,fsig
50    continue

      end
```

C.2.7 Predictive Survival Functions

Following is the program of calculating the predictive survival functions on each therapy at every center.

```
C     this is the program of calculating the predictive survival functions
C     m --- the number of centers
C     mm --- the sample size in Gibbs Sampling
C     y(i,j) --- j-th observation of lambda_i1 in Gibbs sampling
C     z(i,j) --- j-th observation of v_i in Gibbs sampling
C     sig(j) --- j-th observation of sigma^2 in Gibbs sampling
C     s1(i) --- predictive survival function on therapy 1 at center i
C     s2(i) --- predictive survival function on therapy 2 at center i

      parameter(m=22,mm=2000)
      real*8 y(m,mm),z(m,mm),b(mm),mu(mm),sig(mm)
      real*8 x, s1(m),s2(m)

      do 10 k=1,mm              ! read in the Gibbs sampler
        read *, (y(i,k), i=1,7)
        read *, (y(i,k), i=8,14)
        read *, (y(i,k), i=15,22)
```

```
      read *, (z(i,k), i=1,7)
      read *, (z(i,k), i=8,14)
      read *, (z(i,k), i=15,22)
      read *, b(k),mu(k),sig(k)
10    continue

      do 20 j=0,720
        x=j*0.1
        do 30 k=1,m
          s1(k)=0.0
          s2(k)=0.0
          do 15 i=1,mm
            s1(k)=s1(k)+dexp(-x*y(k,i))
            s2(k)=s2(k)+dexp(-x*y(k,i)*dexp(z(k,i)))
15        continue
          s1(k)=s1(k)/mm
          s2(k)=s2(k)/mm
30      continue
          print '(8f9.4)', x,(s1(k),k=1,7)
          print '(8f9.4)', (s1(k),k=8,15)
          print '(7f9.4)', (s1(k),k=16,m)
          print '(8f9.4)', (s2(k),k=1,8)
          print '(8f9.4)', (s2(k),k=9,16)
          print '(6f9.4)', (s2(k),k=17,m)
20    continue

      end
```

Bibliography

[1] Donald A. Berry, S. M. Berry (1993), Bayesian Metaanalysis for Treatment Comparisons: Dichotomous Responses, in process.

[2] S. R. Chakrovorti, J. E. Grizzle, (1975), Analysis of data from multiclinic experiments, *Biometrics* **31**, 325-338.

[3] D. G. Clayton (1991), A Monte Carlo method for bayesian inference in frailty models, *Biometrics* **47**, 467-485.

[4] L. Devroye (1986), Non-uniform Random Variate Generation, *Springer-Verlag*

[5] R. O. Dillman, S. L. Seagren, K. J. Propert, J. Guerra, W. L. Eaton, M. C. Perry, R. W. Carey, E. F. Frei, III, 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.

[6] William DuMouchel (1990), Bayesian Metaanalysis, In *Statistical Methodology in the Pharmaceutical Sciences*, Donald A. Berry (ed.), Marcel Dekker, New York.

[7] Joseph L. Fleiss (1986), Analysis of Data from Multiclinic Trials, *Controlled Clinical Trials*, **7**, 267-275.

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

[9] S. Geman, D. Geman (1984), Stochastic Relation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

[10] Robert J. Gray (1993), A Bayesian Analysis of Institutional Effects in a Multicenter Cancer Clinical Trial, *Biometrics*, to appear.

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

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

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

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

[15] 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.

[16] A. M. Skene, J. C. Wakefield, (1990), Hierarchical Models for Multicenter Binary Response Studies, *Statistics in Medicine*, **9**, 919-929.

[17] Dalene K. Stangl (1991), Modeling heterogeneity in multi-center clinical trials using bayesian hierarchical survival models, *Ph.D. Thesis*, Carnegie Mellon University.

Biography

Chengchang Li was born in Zhejiang, China on July 28, 1962. He obtained a Bachelor's degree in Mathematics from Hangzhou Teachers College, Hangzhou, China in 1984 and a Master's degree in Statistics from Fudan University, Shanghai, China in 1986. From fall 1986 to summer 1989, he served as an assistant professor of statistics at Fudan University. In September 1989, he came to Duke University to pursue his Ph.D. in statistics. During the five-year study at Duke University, he received one graduate fellowship award, one year teaching assistantship and three year research assistantship.