BAYESIAN ORDER RESTRICTED METHODS WITH BIOMEDICAL APPLICATIONS

by

Laura Hilton Gunn

Institute of Statistics & Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. David B. Dunson, Supervisor

_____
Dr. Dalene K. Stangl, Supervisor

_____
Dr. Merlise Clyde

_____
Dr. David Banks

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

2004

ABSTRACT

(Bayesian Order Restricted Inference)

BAYESIAN ORDER RESTRICTED METHODS WITH BIOMEDICAL APPLICATIONS

by

Laura Hilton Gunn

Institute of Statistics & Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. David B. Dunson, Supervisor

_____
Dr. Dalene K. Stangl, Supervisor

_____
Dr. Merlise Clyde

_____
Dr. David Banks

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

2004

Abstract

This dissertation focuses on Bayesian order restricted inference, with interest in applying new methodology to biomedical examples. The first section considers samples of curves restricted to follow a particular shape. For example, progesterone levels in healthy women increase during the menstrual cycle to a random peak with decreases thereafter. Reproductive epidemiologists are interested in studying the distribution of the peak and the trajectory for women in different groups. Motivated by this application, we propose a simple approach for restricting each woman's mean trajectory to follow an umbrella shape. An unconstrained hierarchical Bayesian model is used to characterize the data, and draws from the posterior distribution obtained using a Gibbs sampler are then mapped to the constrained space. Inferences are based on the resulting posterior distribution for the peak and individual woman trajectories. Methods are applied to a study comparing progesterone trajectories for conception and non-conception cycles.

The second section addresses studies that collect event time data in which it is often appropriate to assume non-decreasing hazards across dose groups, though dose effects may vary with time. Motivated by this application, we propose a Bayesian approach for order restricted inference using a non-proportional hazards model with time-varying coefficients. In order to make inferences on equalities versus increases in hazard functions, a prior is chosen for the time-varying coefficients that assigns positive probability to no dose effect while restricting coefficients to be non-negative. By using a high dimensional piecewise constant model and smoothing functions by coupling Markov beta and gamma processes, we obtain a flexible and computationally tractable approach for identifying sets of dose and age values at which hazards increase. This approach can also be used to estimate dose response and survival

curves. The methods are illustrated through application to data from a toxicology study.

To Papa

Acknowledgements

Shortly after beginning graduate school, I formed an opinion that those earning or who have earned a Ph.D. in one area or another were simply not normal individuals. At the closing of my graduate school experience, I stand by my theory even though its tone has changed dramatically. Indeed, those with or on the way toward earning a Ph.D. are not normal, but I have come to realize that many of them live outside the upper $99th$ percentile when it comes to amazing people I am fortunate to know! Of course, the remaining portion of those I know living outside the upper $99th$ percentile without Ph.D.s (i.e., people I originally thought to be normal) are non-normal, as well. In fact, as a T-shirt design that several of us ISDS folks created states: *Nothing [and may I add No One] Is Normal*. With that said, there are many people I would like to acknowledge for their contribution, in one way or another, to making my path to this summit possible, less painful, and, dare I say, at times enjoyable!

First, I want to thank my advisors, David Dunson and Dalene Stangl, for giving me the direction I needed to complete this dissertation. David, I often wondered if I could have made it through this Ph.D. without you! I cannot thank you enough for your time, interest, and patience with me over these past 2 years. I adopted your passion for contributing methodology to biomedical applications, and it was in seeing the useful connections to real problems that kept me from loosing focus on my goal. In addition, your willingness to help me was very genuine. Not even the best of friends returns emails as promptly as you do, and you always found time to make available for any questions I had. I also believe I have grown as a scientific writer as a result of your influence. I have always been amazed at the articles you contribute for publication, and despite all the red and blue pen marks, I thank you for teaching me to write appropriately for the scientific audience. To top all of this off, you were

not only a joy to work with, but I appreciate you as a person and friend.

Dalene, your guidance through both my preliminary and doctoral research was invaluable. From staying late in the office in a hunt for programming errors, to reviewing slides before presentations, to helpful conversations about relevant topics to my research, I am very grateful for your willingness to take me on as an advisee. Aside from my research, I have learned from you many aspects of being an effective statistics professor, and I am eager to continue developing those characteristics as part of my own teaching style. As icing on the cake, I have enjoyed your friendship over these past 5 years. I was lucky to find a preliminary advisor and doctoral co-advisor with whom I get along well.

In addition to my advisors, I thank Merlise Clyde and David Banks for serving on my committee. Merlise, I have enjoyed working with you through classes, as a TA, with my preliminary research, and now with my doctoral work. Your guidance throughout all of these areas means a lot to me! David, I've enjoyed meeting with you this last year of mine. Your pleasant disposition always brings a smile to my face even on a bad programming day!

Outside of my formal Ph.D. committee, I would like to thank Val Johnson for his contributions to my growth as a researcher. I thank Mike West and Michael Lavine, directors of ISDS and ISDS graduate studies, respectively, when I entered the department. Thank you for trusting that I would be a good candidate for the program! To additional professors within ISDS with whom I may not have worked (or only worked with in classes), I have enjoyed being a part of the ISDS family, and I will miss you when I head further south. I thank Eric Van Gyzen, Pat Johnson, Tonya Rambert, Krista Moyle, Nicole Scott, and Marsha Harrison for their assistance, friendship, and pleasant dispositions over the years in ISDS. I thank Kate Whetten for her interest and friendship. I would also like to thank Michael Reed for introducing

me to the connections between the mathematical sciences and biology during the Mathematical Biology workshop in Duke's math department in the summer of 1998. I thank Marilyn Repsher, Marcelle Bessman, Sidney Kung, Bob Hollister, Deanna Haunsperger, Laura Chihara, and Karen Brucks for their impact on my educational and professional development.

A lifetime friend, Nancy Aldrich, gave me a picture for my wall that states: *Friends are flowers in the garden of life.* My friends and family are my support network, and I have been blessed to have several dear people walking alongside me during the path to the Ph.D. I would like to begin by thanking my best friend, 'magnetic' Mekka Drew. I am so thankful for our friendship throughout the last 6 years, including the many laughs we share! Mekka, I appreciate and love you dearly! I thank Chong Tu for his impact on my life over the past 3 years. Thank you for having so much passion in life and showing it in distinctive ways. You are special, and thank you for your guidance along the way! I would also like to thank Germán Molina. Laughter is one of the best gifts in life, and I thank you for giving me so much of it! Germán, thank you for your unique friendship, and for helping me make it to my 'independence day'! I thank 'joyous' Jingshu Wang and Shu-hua Pai for their friendship, laughter, lunches, and work-outs together while at Duke. I thank Özlem Ergisi for her friendship and support over the past 13 years . . . your friendship is very special to me! I want to thank Joey Elaty for his friendship over the past 9 years, for seeing the best there was in me, and for giving me one of my biggest support systems, Teddy! I would like to thank Matt Navidomskis for being such a special friend these past 5 years, for 'patience', for being a firm believer that there really are no accidents, and for serving as my 'Florida-connection' to make my adjustment to Durham less painful. Also during my first year at Duke, and the years thereafter, I want to thank my first roommate ever, Henrike Lohmeier, for her friendship and patience with my

adjustment to Durham. Matt and Henrike, my first year at Duke would not have been the same without you ... thanks for making it unforgettable! I thank Sarah Murray for being such a fun and fantastic roommate and friend while at Duke! There is never a dull moment with you, Sarah, and I always have so much fun with you! I want to thank the Haines family (Caitlyn, Christopher, Michael, Shannon, Margaret, and Tony), Shaun Smith, Joshua Duncan, the Casselberrys (Cherrie, Dan, Michelle, and Shawn), Kristy Nasser, Emily Kistner, David Loveday, Ryan Butterfield, Florence O'Connell, and Aleksandra Ugljesa for their friendship and encouragement over the past 10 years, in most cases. A special thank you goes to Dr. Kalna for being a wonderful mentor and friend over the past 11 years. You always know how to give the best analogies and lessons in life ... I appreciate you and thank you for your positive influence on my life!

I would also like to thank John Kern, Jenise Swall, Hedibert and Anne Lopes, Kate Calder, Chris Holloman, Ana Rappold, Fabio Rigat (and Monica Lucena), Enrique ter-Horst, Beth Brown, Leanna House, Sining Chen, Christina Guyer, Christine Kohnen, Rosy Luo, and Zhenglei Gao for their friendship and guidance during my years at ISDS. I thank additional friends who have encouraged me through these past five years including: Sandra Garnier, Maria Steffero, Tracy Teixeira, Janina Brandt, Cathy Herzog, Marian Kromberg, Peg Trimble, Sharon Tanner, Patty Otterson, Jane White, Ron Smith, Rick McIntee, the Salazars, Dina Ramadan, Barry Hoffman, Manny Valle, Kristi Brown, Cherrie Henry, Mrs. Duncan, Mrs. Liao, Sarah Sandblom, and Chouki El Hamel.

To friends I have grown to know within ISDS, each of you contributed to a wonderful working and social environment. I will especially never forget my last year in Old Chemistry 222 with the lethal combination of Carlos Carvalho, Scotland Leman, Chris Hans, and one of the most unique individuals I have ever known, Eric Vance.

As many times as I could not believe some of your conversation topics, I want you to know just how many smiles you brought to my face with such crazy conversations, especially on frustrating programming days!

I would like to thank those I saw on a regular basis throughout daily life in Durham. They include: Billy, Milton, and Kim for either starting or ending my day at Duke with a smile. I thank Steven for knowing my drink order at Starbucks before ordering it (extra mocha, mocha frappuccino), and Michael for our weekly chats in the produce section of Harris Teeter.

Saving the best for last, I thank my family for getting me through to this point. Teddy, I thank you for the love, laughter, and joy you have given to me throughout these past 4.5 years! To my sister, Karen C. Gunn, I thank you for our growing friendship and many laughs together over the past 3 years living together in addition to the previous 24 years. To my deceased grandmother, Eva R. Olesen, I thank you for the impression you made on my life and for the love you freely gave ... I love and miss you! I also thank my grandfather, Thomas E. Gunn, who died less than two weeks before my defense. Here's a thousand kisses and hugs for you ... I love you, 'K4BYX'! I thank God, above all else, for blessing me with all the above: my advisors, committee, friends, and family (sounds like a long distance phone commercial), as well as providing me with the right spirit to make it through each day, and finally, for my abilities to conquer this mountain with the faith of a mustard seed. To conclude my acknowledgments (as I know all who read this are waiting anxiously for), I thank my mother, Camellia O. Gunn, and father, Thomas H. Gunn. Thank you for your love, patience, discipline (although I did not need much, right), laughs, and friendship throughout the years. I dedicate this dissertation to both of you!

Aristotle once said, "The roots of education are bitter, but the fruit is sweet." I want to thank all of you for helping the roots taste sweet, too!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Order restrictions occur naturally in numerous applications. In biomedical research, there are many instances when model parameters follow a known ordering. For example in bioassay toxicology studies, the average response, such as tumor onset, is often found to be non-decreasing with increasing dose. As a result, it is reasonable to incorporate an order constraint on the parameter of interest. However, placing a restriction on the parameter makes estimation and inference difficult, particularly in complex problems involving correlated and censored data. Throughout this dissertation, I focus on Bayesian methods that simplify order restricted inference, contributing new methodology to biomedical applications.

As in bioassay toxicology studies, it is often appropriate to assume non-decreasing hazards across dose groups. If we denote λ_j as the hazard function for an individual in the jth dose group ($j = 1, \ldots, h$), then we expect $\boldsymbol{\lambda}$ to follow a simple linear ordering with respect to dose level:

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_h. \tag{1.1}$$

Another biomedical example consists of hormone trajectories in the menstrual cycle. In particular, levels of progesterone increase monotonically to an unknown

peak, then decrease thereafter. Unlike the previous example, the ordering in this case is umbrella-shaped. If we denote μ_{ik} to be the mean level of progesterone for woman i ($i = 1, \ldots, N$) on day k ($k = 1, \ldots, D$), then it is natural for $\boldsymbol{\mu}_i$ to follow an umbrella ordering with respect to day of the women's cycle:

$$\mu_{i,1} \leq \cdots \leq \mu_{i,\kappa-1} \leq \mu_{i,\kappa} \geq \mu_{i,\kappa+1} \cdots \geq \mu_{i,D}, \qquad \text{for } \kappa \neq 1, D, \tag{1.2}$$

where κ is an unknown changepoint. When $\kappa = 1$ or D, a simple non-increasing or non-decreasing order, respectively, results.

There are non-biomedical applications where order constraints occur naturally in the data, as well. One is likely to find a monotonic ordering for employee salaries with respect to job rank. Monotonic trends also occur in temperature patterns, such as that found in global warming. Furthermore, one may find umbrella orderings in such applications as heart-rate readings with respect to time spent exercising, or purchasing patterns associated with new products. Accounting for the known order restriction provides more information about underlying biological, environmental, behavioral, social, and other relationships. In this dissertation, our primary focus is on biomedical applications for order restricted inference.

1.1 Classical Order Restricted Approaches

Under the classical framework, inference on model parameters that follow a known ordering is often accomplished through order restricted hypothesis testing. If interest focuses on a monotonic ordering for a vector of means, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$, then a null hypothesis can be constructed such that $H_0 : \mu_1 = \cdots = \mu_k$ is compared against the non-decreasing ordered alternative hypothesis $H_1 : \mu_1 \leq \cdots \leq \mu_k$, with at least one strict inequality. Although the non-decreasing order restriction is considered

here, there are additional orderings that can be explored, such as non-increasing, umbrella and bath-tub orderings, among other constraints. Robertson, Wright, and Dykstra (1988) and Hwang and Peddada (1994) offer several appealing theoretical properties for inference under order constraints. One aspect involves the power of likelihood ratio tests against H_0, which they show is higher when estimating $\boldsymbol{\mu}$ under the non-decreasing order restriction shown in H_1. They also show a decrease in the mean square error (MSE) of the estimates under the order constraint. The MSE is a good measure of the restricted estimator's performance, since having a small MSE indicates a small combined variance and bias.

Frequentists offer restricted maximum likelihood estimates (RMLEs) as a mode of order restricted inference under the classical ANOVA setting. For example, a response for the ith group ($i = 1, \ldots, k$) under the ANOVA framework typically follows a normal distribution with mean μ_i and variance σ_i^2. The RMLE of μ_i is then computed by simply maximizing the likelihood based on the constraint provided in H_1. The RMLE is calculated by replacing consecutive unconstrained MLEs that do not follow the order constraint with the weighted average of these consecutive "violating" MLEs. It is in averaging these "pooled" estimates that the RMLE also became known as the *pooled adjacent violators* estimate.

It is easy to determine the RMLE using the pooled adjacent violators algorithm (PAVA) by taking a weighted average of consecutive estimates that do not obey the order constraint. The following example demonstrates the PAVA in a simple, small-scale setting. Begin by assuming that three groups exist where the groups have equal sample sizes, denoted by $n_i = 5$ for $i = 1, 2, 3$, and respective sample means $\bar{y}_1 = 2.0$, $\bar{y}_2 = 1.0$, and $\bar{y}_3 = 3.0$. Let μ_i^* denote the RMLE of μ_i and $\bar{y}_i$ represent the unconstrained MLE of μ_i. Using the sample means $\bar{y}_1$, $\bar{y}_2$, and $\bar{y}_3$ and the non-decreasing order restriction in H_1, it is apparent that the violation of the ordering

occurs between $\bar{y}_1 = 2.0$ and $\bar{y}_2 = 1.0$. The weighted average (or simple average since all groups have equal sample sizes) between these two means is $\bar{y}_{12} = 1.5$, with a corresponding weight of $n_{12} = 5 + 5 = 10$, implying that $\bar{y}_1 = \bar{y}_2 = 1.5$. Since the order restriction holds with the two remaining estimates, $\bar{y}_{12} = 1.5$ and $\bar{y}_3 = 3.0$, then the RMLE of $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)'$ is $\boldsymbol{\mu}^* = (1.5, 1.5, 3.0)'$.

Another form the RMLE takes on is in that of a min-max function introduced by Robertson, Wright, and Dykstra (1988). Since methods that incorporate the min-max function for order restricted inference are used throughout this dissertation, it is helpful to walk through a simple example implementing the function. The RMLE is found using:

$$\mu_i^* = \min_{t \geq i} \max_{s \leq i} \left(\frac{\sum_{h=s}^{t} n_h \bar{y}_h}{\sum_{h=s}^{t} n_h} \right), \tag{1.3}$$

where μ_i^* is the RMLE of μ_i, and $\bar{y}_h$ represents the unconstrained MLE of μ_h. In the simple $k = 3$ case, the min-max function (1.3) is not needed to solve for the RMLE, as one can see in utilizing the PAVA in the previous example. However, expression (1.3) is used in less trivial scenarios, such as the case with umbrella order restrictions.

Using the same values of the $k = 3$ example in the PAVA case, we can determine the RMLE, and peak location when the ordering is umbrella with unknown peak, using the min-max transformation in (1.3). Assuming an unknown peak κ ($\kappa = 1, \ldots, k$), it is necessary to consider the three possible constraints: 1) $\boldsymbol{\mu}^{*1} : \mu_1 \geq \mu_2 \geq \mu_3$, 2) $\boldsymbol{\mu}^{*2}$: $\mu_1 \leq \mu_2 \geq \mu_3$, and 3) $\boldsymbol{\mu}^{*3}$: $\mu_1 \leq \mu_2 \leq \mu_3$. Our goal is to find $\boldsymbol{\mu}^{*\kappa} = (\mu_1^{*\kappa}, \mu_2^{*\kappa}, \mu_3^{*\kappa})'$ for each choice of κ. Applying the min-max function to the unknown peak case, we obtain:

$$\mu_i^{*\kappa} = \min_{t \in \{\ ?\ \}} \max_{s \in \{\ ?\ \}} \left(\frac{\sum_{h=s}^{t} n_h \bar{y}_h}{\sum_{h=s}^{t} n_h} \right). \tag{1.4}$$

From here, we ask 'for what t is $\mu_t \geq \mu_i$ (this corresponds to the *min* set) and for

what s is $\mu_s \leq \mu_i$ (this corresponds to the *max* set)?' In order to answer this question, we begin with the ordering where $\kappa = 1$, $\mu_1 \geq \mu_2 \geq \mu_3$.

$$\mu_1^{*1} = \min_{t\in\{1\}} \max_{s\in\{1,2,3\}} \big(2.0, 1.5, 2.0\big) = 2.0.$$

The function takes the maximum over s while fixing t. Then,

$$\mu_2^{*1} = \min_{t\in\{1,2\}} \max_{s\in\{2,3\}} \big(1.5, 1.0, 2.0, 2.0\big) = \min\big\{\max(1.5, 2.0), \max(1.0, 2.0)\big\} = 2.0$$

and

$$\mu_3^{*1} = \min_{t\in\{1,2,3\}} \max_{s\in\{3\}} \big(2.0, 2.0, 3.0\big) = 2.0.$$

As a result, the RMLE for $\boldsymbol{\mu}$ when $\kappa = 1$ is $\boldsymbol{\mu}^{*1} = (2.0, 2.0, 2.0)'$. The same approach is used to find the RMLE for $\boldsymbol{\mu}$ when $\kappa = 2$ and 3. For $\kappa = 3$, $\boldsymbol{\mu}^{*3} = (1.5, 1.5, 3.0)'$. However, when $\kappa = 2$, the process in finding the RMLE requires an additional consideration. When the peak is at 2, the relation between μ_1 and μ_3 is unknown. Therefore, two vectors of $\boldsymbol{\mu}^{*2}$ result: one when $\mu_1 \leq \mu_3$ ($\boldsymbol{\mu}^{*2} = (1.5, 2.0, 2.0)'$), and another when $\mu_1 \geq \mu_3$ ($\boldsymbol{\mu}^{*2} = (2.0, 2.0, 2.0)'$).

We later describe a method for determining the vector representing $\boldsymbol{\mu}^{*2}$ using the Mahalanobis distance measure. The final step in computing $\boldsymbol{\mu}^*$, and implicitly solving for $\boldsymbol{\mu}^{*2}$ in addition to finding the peak location, is to calculate this least squares distance measure. Since all the groups have equal weights, minimizing the Mahalanobis distance measure across different choices for the peak location involves calculating $\big(\boldsymbol{\mu}^{*\kappa} - \boldsymbol{\mu}\big)\ \big(\boldsymbol{\mu}^{*\kappa} - \boldsymbol{\mu}\big)'$. By substituting each of the constrained $\boldsymbol{\mu}^{*\kappa}$'s as well as the unconstrained $\boldsymbol{\mu} = (2.0, 1.0, 3.0)'$, we obtain:

$$\boldsymbol{\mu}^* = \min_{\kappa\in\{1,2,3\}} \Big\{2.0, 2.25, 2.0, 0.5\Big\}.$$

The second value in brackets corresponds to the distance measure using $\boldsymbol{\mu}^{*2} = (1.5, 2.0, 2.0)'$, whereas the third value corresponds to the distance measure using

$\boldsymbol{\mu}^{*2} = (2.0, 2.0, 2.0)'$. Since $2.0 \leq 2.25$, then $\boldsymbol{\mu}^{*2} = (2.0, 2.0, 2.0)'$ is the RMLE when $\kappa = 2$. However, with the minimum value in brackets corresponding to $\boldsymbol{\mu}^{*3} = (1.5, 1.5, 3.0)'$, then the peak location is at 3, and the RMLE of $\boldsymbol{\mu}$ is $\boldsymbol{\mu}^* = \boldsymbol{\mu}^{*3} = (1.5, 1.5, 3.0)'$. This illustrative example provides insight into a classical min-max transformation approach, which will later be used in a Bayesian framework for order restricted inference.

Although we have explored a couple avenues by which one can determine RMLEs, we must not lose sight of the hypotheses to which these RMLEs often contribute. Using the same null and alternative hypotheses previously defined in H_0 and H_1, respectively, the likelihood ratio test statistic for the case with unknown variance σ^2, is (Robertson, Wright, and Dykstra, 1988; Casella and Berger, 1990):

$$LR = \frac{\sum_{i=1}^{k} n_i(\mu_i^* - \bar{y})^2/(k-1)}{\frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2}, \tag{1.5}$$

where $\bar{y}$ is the unconstrained MLE under the null hypothesis H_0, μ_i^* is the RMLE for the ith group determined using expression (1.3), n_i is the sample size corresponding to the ith group, and $N - k = \sum_{i=1}^{k}(n_i - 1)$. A rejection of H_0 occurs when LR is large, whereas when LR is small, one fails to reject H_0. The distribution of LR under the null hypothesis is used to compute p-values, and p-values are a guiding factor in frequentist decision making under an hypothesis testing framework.

In the classical setting, order restricted inference is mostly limited to the ANOVA framework, with few exceptions. Although several methods have been developed outside the ANOVA setting (Robertson, Wright, and Dykstra, 1988; Ramsay, 1998; Morton-Jones et al., 2000; Peddada, Prescott, and Conaway, 2001), it is extremely difficult to conduct inferences using these other frameworks. Whether the order constraint is simple or more complex, or the sample size is large or small, it appears to be nearly impossible to obtain closed-form results that follow known distributions within

the frequentist paradigm. As a result, researchers look to Bayesian methodology for alternative approaches to order restricted inference.

1.2 Why Choose Bayesian Alternatives?

Availability of many computational algorithms is among several attractive features of Bayesian approaches. A widely used and efficient computational algorithm often implemented by Bayesians to estimate posterior distributions is called a *Markov Chain Monte Carlo* (MCMC) algorithm. An MCMC is an iterative updating scheme used to draw samples from a parameter's posterior distribution. Under mild regularity conditions, such as ergodicity of the chain, the chain converges to the parameter's posterior distribution. Once convergence is met, determined visually through trace plots and numerically through various diagnostics, samples collected from the chain form an autocorrelated draw from the posterior distribution of the parameter. Using repeated draws, one can make inferences on the posterior of the parameter being modeled, including calculating posterior means, posterior credible intervals, posterior probabilities, and Bayes factors, among other summaries.

There are several types of MCMC algorithms, including the widely used Gibbs (Gelfand and Smith, 1990; Casella and George, 1992) and Metropolis-Hastings (Hastings, 1970; Chib and Greenberg, 1995) sampling algorithms. Gibbs sampling is used to iteratively draw samples from the full conditional distributions of the modeled parameters. Gibbs sampling is conducted when the conditional posterior distributions are in closed form. Unfortunately, not all posteriors can be written in closed form expressions. When Gibbs sampling cannot be used, the Metropolis-Hastings algorithm is implemented to generate samples from the posterior in a different manner.

One begins with a proposal density, and accepts samples based on an acceptance probability to guarantee chain convergence. The choice of a proposal density can greatly effect whether the samples are accepted. Gelman, et al. (1995) offer helpful techniques in choosing an appropriate proposal density. The Gibbs sampler is the primary sampling algorithm implemented within the context of order restricted problems throughout this dissertation.

With the development of such efficient MCMC sampling algorithms, the use of Bayesian methodology has spread to numerous types of modeling. Among these are several areas that we explore throughout this dissertation, including aspects of survival models (Ibrahim, Chen, and Sinha, 2001), changepoint models (Carlin, 1992), curve estimation (Denison, Mallick, and Smith, 1998), and order restricted problems, which is the main focus of this dissertation.

Another attractive feature of Bayesian methodology consists of the integration of prior information into the model specification, a useful characteristic not only when knowing *a priori* that the data follow certain order restrictions, but especially when historical data is available. Order constraints are often incorporated into the prior specification under the Bayesian framework. There are several ways of specifying prior distributions subject to order constraints, including priors with implicit restricted support, point mass priors combining mass on the boundaries with support on the restricted space to allow for such inferences as Bayesian hypothesis testing and variable selection, and transformation approaches similar to the min-max algorithms discussed earlier.

1.3 Bayesian Order Restricted Inference

Prior distributions having constrained support were the first conventional type of prior for incorporating order restrictions. Gelfand and Kuo (1991) and Ramgopal, Laud, and Smith (1993) proposed methods for nonparametric estimation of potency curves under monotonic shape constraints, where restricted priors are used to accommodate shape restrictions. If the constrained posterior distribution follows the same form as the unrestricted posterior, then specifying a prior distribution with support on the restricted space is somewhat trivial.

In their seminal paper, Gelfand, Smith, and Lee (1992) proposed a general Gibbs sampling strategy for incorporating parameter restrictions on regression parameters. They constrain the priors in a computationally tractable scheme. For example, begin with three parameters μ_1, μ_2, and μ_3, with support on the real line such that the order constraint $\mu_1 < \mu_2 > \mu_3$ is placed on the parameters. Normal prior distributions are typically chosen for $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)'$ such that the joint prior distribution is $\pi(\mu_1, \mu_2, \mu_3) \propto \mathcal{N}(\mu_1; \theta_1, \sigma_1^2)\mathcal{N}(\mu_2; \theta_2, \sigma_2^2)\mathcal{N}(\mu_3; \theta_3, \sigma_3^2)1_{\{\mu_1 < \mu_2 > \mu_3\}}$, where $\theta_1, \theta_2, \theta_3$ and $\sigma_1^2, \sigma_2^2, \sigma_3^2$ are mean and variance hyperparameters specified by the investigator. Gelfand, Smith, and Lee (1992) showed that such a restricted prior taking on a truncated form can be easily incorporated into a model via Gibbs sampling. Posterior computation can proceed by taking draws from the unrestricted parameter's full conditional distribution, which is $\pi(\mu_k | x, \boldsymbol{\mu}_{(-k)})$, where x represents the data and $\boldsymbol{\mu}_{(-k)}$ represents all but the kth element of $\boldsymbol{\mu}$. Samples violating the order restriction $\mu_1 < \mu_2 > \mu_3$ are then discarded. Additional researchers adopted the methodology proposed by Gelfand, Smith, and Lee (1992) to account for monotonic orderings through prior distributions.

Several alternative approaches were also used to accommodate order restrictions. Under a nonparametric Bayesian framework, Lavine and Mockus (1995) used pri-

ors subject to order constraints in an isotonic regression problem. Gelfand and Mallick (1995) proposed an approach where the prior accounts for order constraints in modeling the hazard function. Arjas and Gasbarra (1996) and Gelfand and Kottas (2001) also introduced nonparametric Bayesian approaches for incorporating order constraints through prior distributions. The former paper includes the order restrictions in a joint prior for pairs of survival functions, whereas the latter paper includes the ordering in priors not only in the two sample case, but also in matched pairs problems, ordered regression models, and ordered ANOVA models. Finally, Dunson and Columbo (2003) accounted for an umbrella ordering on the model parameters by choosing constrained priors to accommodate such orderings.

Although the approach of incorporating order constraints through priors with restricted support is appealing for numerous applications, it has historically been used for accommodating restrictions on population parameters. Rather than focusing on population parameters, our interest revolves around subject-specific means, which are functions of population parameters and random effects. When order restrictions occur on higher level parameters in a hierarchical model, a transformation approach can be used to satisfy the order constraint and improve model efficiency. Several frequentist approaches use isotonic regression transformations to constrain parameter estimates. Robertson, Wright, and Dykstra (1988) introduced the min-max transformation expressed in (1.3). Furthermore, Hwang and Peddada (1994) used the min-max transformation formula as a minimal distance mapping to the constrained space for a known peak ordering. Much research has shown that such isotonic transformation methods produce smaller mean square error in most cases (Robertson, Wright, and Dykstra, 1988; Mammen, 1991; Hwang and Peddada, 1994; Dunson and Neelon, 2003).

Under a Bayesian framework, Dunson and Neelon (2003) extended the isotonic

transformation approach to accommodate simple increasing order restrictions in generalized linear models. They apply an extended version of the isotonic regression transformation in (1.3) to unconstrained posterior estimates from an MCMC output, resulting in restricted estimates that are functions of the unrestricted values. Their approach handles simple ordering, so the methods we present offer an additional and more complex ordering under the umbrella-ordered case.

An appealing feature of the transformation approach is that it allows for conventional unconstrained posterior computation. To avoid computational difficulties with restricted priors, the transformation method follows the approach of Gelfand, Smith, and Lee (1992) by choosing priors for model parameters without incorporating an order restriction. Standard unconstrained posterior computation proceeds using typical MCMC sampling methods. The mapping of the unconstrained posterior estimates to the restricted space using the min-max transformation yields restricted estimates.

Another drawback of using priors with restricted support is that they do not allow for equalities in the constraints, which is often of interest for testing hypotheses of no difference in parameter values across levels of a predictor. For example, Gelfand, Smith, and Lee's (1992) approach assigns zero probability to a null hypothesis of homogeneity, since only draws that follow the strict inequality are collected. One of our goals is to conduct a Bayesian hypothesis test for no trend, corresponding to a null hypothesis of $H_0 : \mu_1 = \mu_2 = \mu_3$ in the simple $k = 3$ case. Therefore, previous methods are not appropriate for the framework of our models.

An alternative Bayesian order restricted approach consists of specifying prior distributions that allow for mass on the boundaries to accommodate hypothesis testing as well as model and variable selection. In biomedical applications, it is often uncertain whether the mean response is equal or follows some ordering across levels of a predictor. Scientists use hypotheses to test their conjectures of homogeneity,

$H_0 : \mu_1 = \mu_2 = \mu_3$, versus the simple, non-decreasing case of inequalities in the means, $H_1 : \mu_1 \leq \mu_2 \leq \mu_3$. Applying the approach of Gelfand, Smith, and Lee (1992) under this hypothesis framework allows us to assign a conventional prior, which in many cases can be a normal density, then discard samples not adhering to the constraint $\mu_1 < \mu_2 < \mu_3$. Under this approach, values are never sampled such that $\mu_1 = \mu_2 = \mu_3$, resulting in zero probability allocated to the null hypothesis of no difference in means and probability one assigned to the alternative hypothesis of inequalities in the means across levels of a predictor.

When interest focuses on considering hypotheses such as these, then it is necessary to allocate positive probability to the case of equality across the means. To address this issue, several researchers proposed a mixture prior consisting of a point mass, corresponding to no trend in mean response, and a truncated probability density, representing inequalities in the mean response across levels of a predictor (Dunson and Herring, 2003; 2004; Neelon and Dunson, 2004). Dunson and Herring (2003) introduced the one-inflated gamma prior:

$$\pi(\cdot) = \mathcal{I}_1 - \mathcal{G}_{[1,\infty)}(\cdot; \pi, a, b), \tag{1.6}$$

which consists of a mixture of a point mass at one (with probability π) and a $\mathcal{G}(\cdot; a, b)$ density truncated to the left at one,

$$\mathcal{I}_1 - \mathcal{G}_{[1,\infty)}(z; \pi, a, b) = 1_{\{z=1\}}\pi + 1_{\{z>1\}}(1-\pi)\frac{\mathcal{G}(z; a, b)}{\int_1^\infty \mathcal{G}(u; a, b)du}. \tag{1.7}$$

One goal in developing this prior was to accommodate a comparison involving a null hypothesis of no difference with a monotonically increasing alternative hypothesis across levels of a predictor. Their methods are applied to the Cox proportional hazards model. Holmes and Heard (2003) used mixture priors for assessing evidence of ordering by testing hypotheses of a monotonically increasing dose response function

compared to other dose response curves. Neelon and Dunson (2004) assessed the evidence of an association within a regression function containing non-decreasing order restrictions via a zero-inflated normal distribution, comprised of a mixture of point masses at zero and truncated normal densities.

Mixture priors are also used in variable and model selection (Geweke, 1996; George and McCulloch, 1997; Chipman, George, and McCulloch, 2001). Dunson and Herring (2004) accounted for model uncertainty using a one-inflated gamma prior that allocates positive probability to hazard models with only additive effects, only proportional effects, both additive and multiplicative effects, as well as no association. An advantage of this prior structure lies in its property that full conditional distributions of modeled parameters take a convenient, conjugate form. Having a simple conjugate structure ultimately leads to efficient posterior computation. Furthermore, this prior approach is appealing over frequentist methods because it does not rely upon large sample theory. Many classical testing methods under additive hazard models depend on large sample approximations that may or may not hold true for small to moderate sample sizes.

1.4 Goals and Outline of Dissertation

Motivated by gaps in current methods for incorporating order constraints and conducting inferences under classical and Bayesian frameworks, the goal of this dissertation is to offer improved techniques for order restricted inference in biomedical applications. This work introduces three improvements to curve estimation and trend testing in hierarchical and hazard models, by extending methods of transformation and mixture prior approaches for incorporating order restrictions.

Chapter 2 is motivated by studies of the progesterone hormone in the menstrual cycle. The natural known ordering of progesterone follows an umbrella-shape in which the mean progesterone value increases monotonically to an unknown peak with decreases thereafter. A hierarchical model with women-specific random effects and autocorrelated errors is used to account for the dependency in hormone measurements. Order constraints fall on higher-level parameters in the hierarchy, so a transformation approach is used to satisfy the order restriction.

Generalizing the approaches of Hwang and Peddada (1994) and Dunson and Neelon (2003), we account for umbrella ordering with an unknown peak. Standard unconstrained posterior samples are collected and mapped to the restricted space using a min-max transformation formula extended from equation (1.3). The peak is then found by locating the value that minimizes the distance between the unconstrained and umbrella-ordered parameters. We offer a simulation study providing evidence of smaller bias and greater efficiency, and demonstrate its applicability in modeling progesterone levels during the menstrual cycle.

Chapter 3 is motivated by toxicology studies that assume non-decreasing hazards with increasing dose. Focus is typically on estimation under strict constraints, however we are interested in estimating survival curves as well as assessing evidence of ordering, hence evidence of a dose effect. We generalize the approach of Dunson and Herring (2003) to allow for order-restricted non-proportional hazards models with time-dependent coefficients (Hastie and Tibshirani, 1993). Since one of our goals is to determine whether the time until death decreases or remains the same while increasing the dose level, we propose a prior distribution for the time-varying coefficients which allocates positive probability to no dose effect while constraining the coefficients to be non-negative. The prior is structured to couple Markov gamma and beta process priors (Nieto-Barajas and Walker, 2002) for smoothing the functions to

result in a mixture prior allowing for inferences on equalities versus increases in the hazards across levels of dose and time.

We approximate the hazard function with a high dimensional piecewise constant model. The use of piecewise constant functions, also known as step functions, provides flexibility in constructing hazard functions. In particular, an advantage of using piecewise constant functions for intensity models is that the addition, or multiplication in the case of proportional hazards, of these functions also yields a piecewise constant function. Assume a basic model with three time periods having interval endpoints denoted by τ_1 and τ_2 such that the first interval is from 0 to τ_1, the second interval goes from τ_1 to τ_2, while the third interval begins at τ_2 and continues onward. If an event for subject i occurs in the first interval, then their duration is denoted by t_i. If a subject has an event between τ_1 and τ_2, then that subject contributes a duration of τ_1 to the first interval and a duration of $t_i - \tau_1$ to the second interval. Finally, if a subject has an event occurring in the third time interval, then that subject contributes durations of τ_1 and τ_2 to the first two intervals, respectively, and a duration of $t_i - \tau_1 - \tau_2$ to the final time interval. Notice in the second two cases that $\tau_1 + (t_i - \tau_1) = t_i$ and $\tau_1 + \tau_2 + (t_i - \tau_1 - \tau_2) = t_i$ are simply the total times that subject i is at risk. By constructing the model under the piecewise constant framework, we divide the period at risk for subject i into pieces, where the risk is constant in each interval but is allowed to change from interval to interval.

Since we explore a high dimensional piecewise constant model, efficient posterior computation is difficult to achieve. In Chapter 3, the coupling of Markov beta and gamma process priors is used not only for its appealing theoretical properties but also to simplify posterior computation. In addition to simplifying posterior computation, the prior structure we have chosen also smoothes the hazard curves. Hence, the autocorrelation component of the Markov prior processes allows us to borrow information

across neighboring intervals within baseline hazard, point mass probability, and difference in consecutive hazards parameters. Structuring the prior in this framework while allowing for flat regions corresponding to no effect with increases in dose, one can estimate the dose response more efficiently. One way of accomplishing this is by evaluating the magnitude of the differences in mortality rates corresponding to each dose group. As a result, we can provide policy makers with more appropriate information in setting health guidelines for exposure to certain chemicals.

When the focus is on inference on a null hypothesis of equalities in parameters versus one or more order constrained alternative hypotheses, point mass mixture priors have some appealing properties. However, many researchers think that point null hypotheses representing exact equalities in the parameters are unreasonable, and it is preferable to consider null hypotheses that the parameters are within $\pm\epsilon$, or some small value, which can be chosen as the minimal value such that the difference is scientifically relevant. This new mixture prior is introduced in Chapter 4.

Chapter 4 summarizes research findings and provides directions for future research. In particular, a third improvement to curve estimation and trend testing in hierarchical and hazard models is introduced which proposes an alternative approach for Bayesian inferences on partial orderings, avoiding the point null representation. This approach also limits the problems we have observed in biased estimation under restricted priors. In particular, when there is a moderate to high dimensional parameter space subject to simple ordering and one uses a uniform improper prior as suggested by Gelfand, Smith, and Lee (1992), we observe striking degrees of bias in posterior means. In addition, 95% credible intervals often do not include the true parameter value, in many cases for most of the parameters. By allowing probability to congregate on values close to equalities, we limit this problem, as we discuss in Chapter 4.

Chapter 2

Transformation Method for Umbrella Shape Restrictions

2.1 Motivation and Overview

Reproductive epidemiologists and endocrinologists are interested in studying hormone patterns in the menstrual cycle. Differences in hormone levels among women during the various phases of the cycle may be indicative of reproductive function (Baird et al., 1997; 1999). For certain hormones, such as progesterone, studies show that the trajectories increase monotonically to an unknown peak with decreases thereafter (Yen and Jaffe, 1986). Characteristics of these curves, including peak location, may provide insight into the probabilities of conception and early pregnancy loss (Baird et al., 1997; 1999).

In such biomedical studies, smoothing methods are commonly used to estimate mean trajectories, while limiting parametric assumptions. Zhang et al. (1998) and Zhang, Lin, and Sowers (2000) used semiparametric models for correlated data, and estimated smoothing parameters using restricted maximum likelihood estimation. Brumback and Rice (1998) used smoothing splines with mixed effects models to

analyze progesterone data in the menstrual cycle. Although such approaches have many advantages, they do not incorporate shape constraints on the mean curves. Such constraints provide a means for incorporating biological information into the analysis without introducing restrictive parametric assumptions about the unknown trajectories. Restricting the shapes that are possible to a class consistent with prior information can improve statistical efficiency while producing a more realistic estimate.

Several approaches have been proposed for estimating curves subject to shape constraints. Gelfand and Kuo (1991) and Ramgopal, Laud, and Smith (1993) used restricted nonparametric priors to accommodate shape restrictions in potency curves. Gelfand, Smith, and Lee (1992) proposed a general Gibbs sampling strategy for incorporating parameter constraints by discarding draws inconsistent with the constraint. Lavine and Mockus (1995) proposed a general method for nonparametric Bayesian isotonic regression. Dunson and Columbo (2003) proposed a hierarchical model and smoothing prior that allowed for shape restrictions on mean curves underlying categorical observations. Adaptation of these approaches to the problem of incorporating constraints on higher level parameters in a hierarchical model (e.g., individual-specific means) is not straightforward.

In addition to these Bayesian approaches, frequentist methods are available for monotone curve estimation. Mammen (1991) used smoothing and isotonization steps to efficiently estimate a monotone regression function. Mammen and Thomas-Agnan (1999) proposed a method for constraining smoothing splines by first calculating the unrestricted smoothing spline, and then projecting onto the constrained space using a Sobolev-type norm. Such isotonization tends to yield a smaller mean square error in many cases (Robertson, Wright, and Dykstra, 1988; Mammen, 1991; Hwang and Peddada, 1994).

Applying the isotonization idea to the Bayesian paradigm, Dunson and Neelon (2003) proposed an approach for order restricted inference in generalized linear models. Instead of mapping a single unrestricted estimate to the constrained space, they proposed applying an isotonic regression transformation to draws from the unconstrained posterior density obtained using an MCMC algorithm. Their focus was on simple ordering in regression parameters, and not on ordering with unknown changepoints or constraints on higher level parameters in a hierarchical model.

Motivated by applications to hormone studies, we generalize the method of Dunson and Neelon (2003) to allow for umbrella ordering and constraining of higher level parameters in a hierarchical model. To generalize isotonic regression transformations to the unknown peak case, we choose the peak that minimizes the weighted least squares distance between the unconstrained and umbrella-ordered parameters. We then apply this transformation to individual-specific means generated from a hierarchical model.

Section 2.2 considers umbrella ordered cases when data are independent and normal. Section 2.3 generalizes the approach to constrain individual-specific trajectories characterized by a hierarchical model. Section 2.4 applies the method to progesterone data considered previously by Brumback and Rice (1998) and provides a simulation example, while Section 2.5 discusses the results.

2.2 Umbrella Order Restrictions

2.2.1 Ordered Normal Means

We focus initially on inference for a vector of normal means known *a priori* to follow an umbrella order restriction. Let $\boldsymbol{y}_i \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ is the mean vector and $\boldsymbol{\Sigma} = \sigma^2 \mathrm{I}_{p\times p}$ is the covariance matrix, for $i = 1, \ldots, n$. Ignoring the ordering

initially, a conjugate prior for $\boldsymbol{\mu}$ can be specified as $\pi(\boldsymbol{\mu}|\boldsymbol{\Sigma}_0) = \prod_{j=1}^{p} \mathcal{N}\big(\mu_j; \mu_{0j}, \sigma_{0j}^2\big)$, where $\boldsymbol{\Sigma}_0 = \text{diag}(\sigma_{01}^2, \ldots, \sigma_{0p}^2)$. The posterior of μ_j conditional on σ^2 and $\boldsymbol{y}$ is then $\pi(\mu_j|\sigma^2, \boldsymbol{y}) = \mathcal{N}(\mu_j; \widehat{\mu}_j, \widehat{\sigma}_j^2)$, where $\widehat{\mu}_j = \widehat{\sigma}_j^2(\sigma_{0j}^{-2}\mu_{0j} + \sigma^{-2}n\bar{y}_j)$ and $\widehat{\sigma}_j^2 = (\sigma_{0j}^{-2} + \sigma^{-2}n)^{-1}$. We focus inference on an order-constrained parameter $\boldsymbol{\mu}^*$, which arises from a mapping of the unconstrained mean vector $\boldsymbol{\mu}$ from $\Re^p \to \boldsymbol{\Omega}$, where $\boldsymbol{\Omega} \subset \Re^p$ is a subset defined by the following set of inequalities on the elements of $\boldsymbol{\mu}$ such that $\boldsymbol{\Omega} = \bigcup_{\kappa=1}^{p} \boldsymbol{\Omega}_\kappa$. We define $\boldsymbol{\Omega}_\kappa$ as the set of umbrella ordered means with peak κ:

$$\boldsymbol{\Omega}_\kappa = \big\{\boldsymbol{\mu} : \mu_1 \leq \cdots \leq \mu_\kappa \geq \cdots \geq \mu_p\big\}, \qquad \text{for } \kappa \neq 1, p. \tag{2.1}$$

When the peak occurs at 1 or p, a simple non-increasing or non-decreasing order, respectively, results. The support of $\boldsymbol{\mu}^*$ is therefore the space of $p \times 1$ vectors following an umbrella ordering with unknown peak $\kappa \in \{1, \ldots, p\}$.

Following Dunson and Neelon (2003) who considered simple ordering, we choose a mapping from $\Re^p \to \boldsymbol{\Omega}$ that minimizes the distance between $\boldsymbol{\mu}^*$ and $\boldsymbol{\mu}$. In particular, based on Robertson, Wright, and Dykstra (1988) and Hwang and Peddada (1994), we use an isotonic regression transformation that sets the elements of $\boldsymbol{\mu}^*$ equal to weighted averages of the elements of $\boldsymbol{\mu}$. Assuming a peak at κ, the transformation follows the form shown in Dunson and Neelon (2003):

$$\mu_j^{*\kappa} = g_j(\boldsymbol{\mu}) = \min_{t \in U_j^\kappa} \max_{s \in L_j^\kappa} \left(\frac{\mathbf{1}'_{|t-s|+1} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\mu}|\sigma,\boldsymbol{y}[s:t]} \boldsymbol{\mu}_{[s:t]}}{\mathbf{1}'_{|t-s|+1} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\mu}|\sigma,\boldsymbol{y}[s:t]} \mathbf{1}_{|t-s|+1}} \right), \quad \text{for } j = 1, \ldots, p, \tag{2.2}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\sigma,\boldsymbol{y}}$ is the unconstrained posterior covariance and L_j^κ and U_j^κ denote subsets of $\{1, \ldots, p\}$ such that the ordering $\mu_{j'} \leq \mu_j$ is known for all $j' \in L_j^\kappa$ and the ordering $\mu_{j'} \geq \mu_j$ is known for all $j' \in U_j^\kappa$. The [s:t] subscript represents the submatrices and subvectors corresponding to elements $s, s+1, \ldots, t$. Note that in this simple case, $\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\sigma,\boldsymbol{y}} = \text{diag}(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_p^2)$.

To allow for an unknown peak, κ, we choose $\boldsymbol{\mu}^*$ by minimizing the Mahalanobis distance measure across different choices of peak:

$$\boldsymbol{\mu}^* \quad = \quad \min_{\kappa \in \{1,\ldots,p\}} \left\{ \left(\boldsymbol{\mu}^{*\kappa} - \boldsymbol{\mu}\right) \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\mu}|\sigma,\boldsymbol{y}} \left(\boldsymbol{\mu}^{*\kappa} - \boldsymbol{\mu}\right)' \right\}. \tag{2.3}$$

Using this expression in combination with expression (2.2) produces the value of $\boldsymbol{\mu}^* \in \boldsymbol{\Omega}$, which is as close as possible to the unrestricted value $\boldsymbol{\mu} \in \Re^p$. The weighted least squares distance measure minimizes the distance between $\boldsymbol{\mu}^*$ and $\boldsymbol{\mu}$, requiring the distance to be relatively small for elements of $\boldsymbol{\mu}$ about which much is known (i.e., the posterior variance is small). In this manner, the mapping tends to result in values of $\boldsymbol{\mu}^*$ with high density in the posterior for $\boldsymbol{\mu}$. We follow Dunson and Neelon (2003) in taking the view that $\boldsymbol{\mu}^*$ is an order-restricted functional of the parameter $\boldsymbol{\mu}$, and hence we can consider the draws of $\boldsymbol{\mu}^*$ as draws from a Bayesian posterior.

An alternative approach would be to directly place a prior distribution and define a likelihood based on $\boldsymbol{\mu}^*$. Although this more conventional Bayesian strategy is conceptually appealing, our transformation approach has important practical advantages. In particular, it is typically the case that investigators are interested in both the order-restricted $\boldsymbol{\mu}^*$ and the unrestricted $\boldsymbol{\mu}$, since assessing changes in the estimates due to the incorporation of the constraint is a critical component of the analysis (particularly, in applications in which one may be somewhat skeptical about the constraint). In addition to computational advantages in more complex problems (as described in Sections 2.3 and 2.4) and a tendency to produce estimates that are close to unrestricted ridge-type estimators, the transformation strategy has the advantage that it is straightforward to assess the impact of incorporating the constraint. One can simply examine how inferences change using $\boldsymbol{\mu}^*$ instead of $\boldsymbol{\mu}$.

The approach is also appealing from a frequentist perspective due to the close relationship with traditional isotonic regression transformations. In addition, one can show that the mode of the posterior for $\boldsymbol{\mu}^*$ converges to the restricted maximum

likelihood estimate (RMLE) in the limit as $n \to \infty$. Using this equivalence, it is also straightforward to show that the mode of the posterior for $\boldsymbol{\mu}^*$ is a consistent estimator as long as the true means belong to $\boldsymbol{\Omega}$. This proof follows along the lines described by Turner and Wollan (1997) for their classical estimator, and is one motivation for our use of the Mahalanobis distance measure in choosing the peak. A final comment is that the posterior tends to be centered close to the RMLE when the information in the prior is small relative to the sample size even when the sample size is small to moderate. In complex settings in which it is very difficult to estimate the RMLE, the Bayesian transformation strategy can be viewed as a reasonable alternative.

For additional theoretical properties, we focus on the simple $p = 3$ case, since theorems for general p are very difficult to establish (as in classical order restricted inference problems).

Theorem 1.

For $p \leq 3$, $\boldsymbol{\mu}^*$ has its peak at the same location as $\boldsymbol{\mu}$, for all $\boldsymbol{\mu} \in \Re^p$ (see Appendix A for a proof).

This theorem implies that the transformation from $\Re^p \to \boldsymbol{\Omega}$ maintains the peak location. Since we wish to limit systematic biases in mapping from $\Re^p \to \boldsymbol{\Omega}$, this is an appealing property.

2.2.2 Comparison with Standard Approach

We compare the proposed transformation approach from Section 2.2.1 with the standard Bayesian approach of placing a truncated conjugate prior on the restricted space. A typical prior used to constrain $\boldsymbol{\mu} \in \boldsymbol{\Omega}_\kappa$ follows the truncated normal form:

$$\pi(\boldsymbol{\mu}|\kappa) \propto 1(\boldsymbol{\mu} \in \boldsymbol{\Omega}_\kappa)\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \tag{2.4}$$

where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are defined in Section 2.2.1. The posterior distribution for μ_κ, conditional on $\boldsymbol{\mu}_{-\kappa} = (\mu_1, \ldots, \mu_{\kappa-1}, \mu_{\kappa+1}, \ldots, \mu_p)'$ and the peak location occurring at κ, arising from prior (2.4) follows the truncated normal distribution:

$$\pi(\mu_\kappa|\boldsymbol{\mu}_{-\kappa}, \boldsymbol{\Sigma}, \kappa, \boldsymbol{y}) = \frac{1\big(\mu_\kappa > \max\{\boldsymbol{\mu}_{-\kappa}\}\big) f\big(\max\{\boldsymbol{\mu}_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa\big)}{F\big(-\max\{\boldsymbol{\mu}_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa\big)}, \tag{2.5}$$

where $f(\cdot)$ and $F(\cdot)$ are the normal density and distribution functions, respectively, and $\widehat{\mu}_\kappa$ and $\widehat{\sigma}^2_\kappa$ are defined in Section 2.2.1, with $\widehat{\mu}_\kappa \approx \overline{y}_\kappa$ when the prior variance is large. Sampling from (2.5) is equivalent to drawing from the unrestricted posterior $\mathcal{N}(\mu_\kappa; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa)$ and discarding values that are less than $\max\{\boldsymbol{\mu}_{-\kappa}\}$. This conditional posterior gives zero prior probability to $\mu_\kappa \leq \max\{\boldsymbol{\mu}_{-\kappa}\}$ and has expectation

$$\mathrm{E}(\mu_\kappa|\boldsymbol{\mu}_{-\kappa}, \boldsymbol{\Sigma}, \kappa, \boldsymbol{y}) = \widehat{\mu}_\kappa + \frac{\sqrt{\widehat{\sigma}^2_\kappa} f\big(\max\{\boldsymbol{\mu}_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa\big)}{F\big(-\max\{\boldsymbol{\mu}_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa\big)}. \tag{2.6}$$

As $\max\{\boldsymbol{\mu}_{-\kappa}\}$ increases and the conditional posterior variance decreases, the expectation in (2.6) can be substantially larger than $\widehat{\mu}_\kappa$ and hence, when the prior variance is large, substantially larger than $\overline{y}_\kappa$, the mean of the data.

It is of interest to compare expression (2.5) with the conditional distribution for μ^*_κ obtained under the proposed transformation approach:

$$\begin{aligned} \pi(\mu^*_\kappa|\boldsymbol{\mu}^*_{-\kappa}, \boldsymbol{\Sigma}, \kappa, \boldsymbol{y}) &\stackrel{d}{=} 1\big(\mu^*_\kappa = \max\{\boldsymbol{\mu}^*_{-\kappa}\}\big) F\big(\max\{\boldsymbol{\mu}^*_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa\big) \\ &+ 1\big(\mu^*_\kappa > \max\{\boldsymbol{\mu}^*_{-\kappa}\}\big) \mathcal{N}(\mu^*_\kappa; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa), \end{aligned} \tag{2.7}$$

which is equivalent to sampling from the unrestricted $\mathcal{N}(\mu_\kappa; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa)$ posterior and setting values less than $\max\{\boldsymbol{\mu}^*_{-\kappa}\}$ equal to the boundary value $\max\{\boldsymbol{\mu}^*_{-\kappa}\}$. It follows that the conditional estimator using the transformation approach is:

$$\begin{aligned} \mathrm{E}(\mu^*_\kappa|\boldsymbol{\mu}^*_{-\kappa}, \boldsymbol{\Sigma}, \kappa, \boldsymbol{y}) &= \max\{\boldsymbol{\mu}^*_{-\kappa}\} F\big(\max\{\boldsymbol{\mu}^*_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa\big) + F\big(-\max\{\boldsymbol{\mu}^*_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa\big) \\ &\quad \Big\{\widehat{\mu}_\kappa + \frac{\sqrt{\widehat{\sigma}^2_\kappa}\, f(\max\{\boldsymbol{\mu}^*_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa)}{F(-\max\{\boldsymbol{\mu}^*_{-\kappa}\}; \widehat{\mu}_\kappa, \widehat{\sigma}^2_\kappa)}\Big\}, \end{aligned} \tag{2.8}$$

which is closer to the unrestricted posterior mean, $\widehat{\mu}_\kappa$, than the conditional posterior mean shown in expression (2.6) for the truncated normal prior. This result suggests that when the prior precision, $\sigma_{0\kappa}^{-2}$, is small relative to the information in the data, the proposed transformation approach yields estimates closer to the data and less subject to systematic biases compared with the more conventional approach. This difference in the two estimates can be substantial, particularly when the true value is close to the boundary of $\boldsymbol{\Omega}$, when the sample size is small to moderate, and when strong constraints are assigned. Note that we have focused on the case where the peak is known. When the peak is unknown, the standard strategy would be to place a prior distribution (e.g., uniform) on the peak location and estimate this changepoint simultaneously with the mean values. It is well known that the introduction of an unknown changepoint can greatly contribute to the computational burden. This is not a major hurdle in the simple ordered normal means case but is in cases where the changepoint can vary for each subject, as in the menstrual cycle application.

2.2.3 Illustrative Example

It is useful to consider a simple illustrative example. Suppose $p = 3$ and that the group-specific empirical means are $\overline{y}_1 = 0$, $\overline{y}_2 = c$, and $\overline{y}_3 = 0$, with all three groups having equal sample sizes. We fix the empirical variance of $\boldsymbol{y}$ at 100. It is known *a priori* that there exists an umbrella ordering in $\boldsymbol{\mu}$ with the peak unknown. Figure 2.1 plots the estimated posterior means for μ_1, μ_2, and μ_3 for a range of values of c and n under the approach which assigns a uniform improper prior for $\boldsymbol{\mu}$ with support on $\boldsymbol{\Omega}$ and plots the estimators under our proposed transformation approach.

It is apparent from Figure 2.1 that the estimators under the transformation approach are closer to the empirical means in each of the cases, with the largest differ-

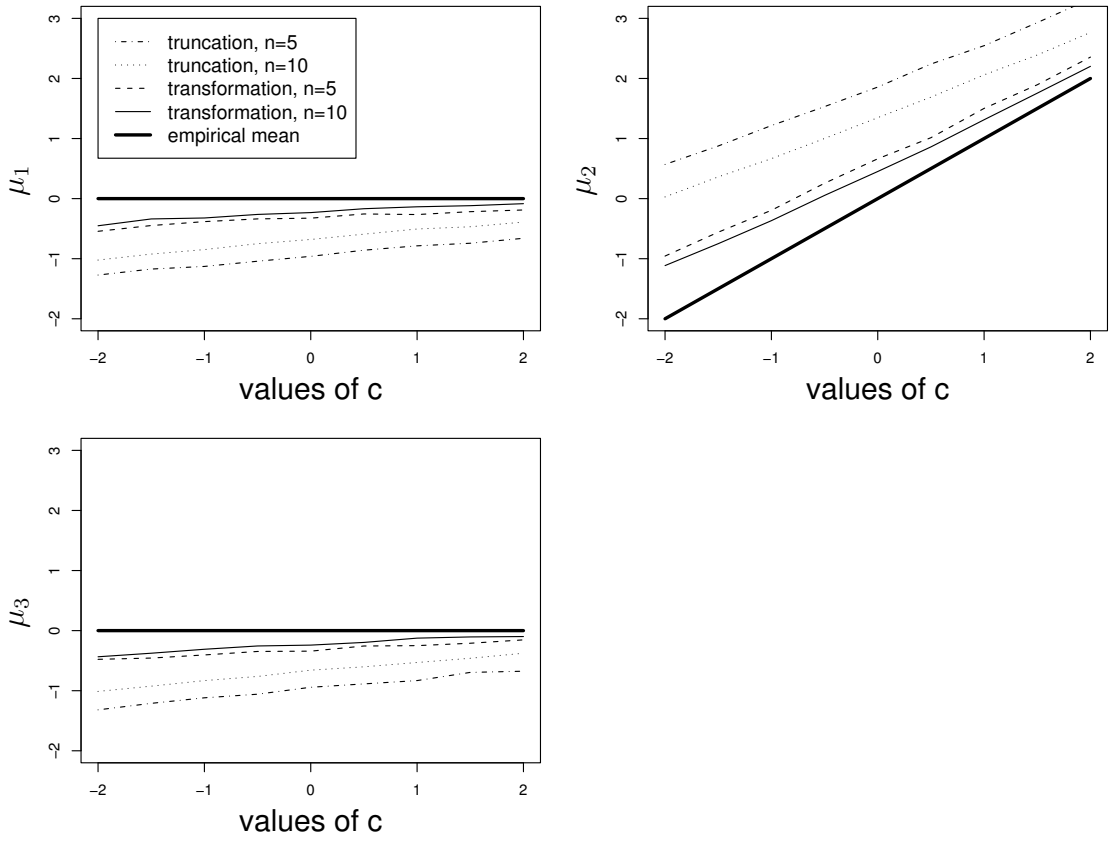

Figure 2.1: Estimated restricted means for μ_1 (top left), μ_2 (top right), and μ_3 (bottom panel) of the illustrative example

ences occurring for small sample sizes and negative values of c. Since having negative values for c produces a bathtub-shape ordering, the umbrella-ordered constraint is clearly violated. This figure suggests that, even in the case where the prior is chosen to be non-informative subject to the order restriction and the data do not violate the order constraint, the posterior densities produced by the standard approach can be centered on values far from the empirical means of the data. The transformation approach is far less sensitive to this problem. Although we have focused on small samples in this illustrative example, this problem is also important in large sample settings in which we are interested in estimating individual-specific means but do not have a large amount of information on each individual.

2.3 Order Restrictions in Hierarchical Models

Until this point, we have focused on the simple case where the data consist of independent observations from a normal distribution. However, it is straightforward to apply the methods to any hierarchical model in which umbrella restrictions are needed. We focus on the case where there are repeated multivariate observations on each of n subjects, so that outcome data on subject i consist of a vector of observations $\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijp})'$ at times $j = 1, \ldots, n_i$. For example, in the menstrual cycle application, i indexes woman and j indexes menstrual cycle within a woman. Let $\boldsymbol{\mu}_{ij} = (\mu_{ij1}, \ldots, \mu_{ijp})'$ denote the linear predictor in an arbitrary hierarchical generalized linear mixed model. This linear predictor is characterized as a function of fixed effects regression parameters, $\boldsymbol{\alpha}$, individual and potentially time-specific random effects, $\boldsymbol{\beta}_i$, variance component parameters, $\boldsymbol{\phi}$, and residual error variances, $\boldsymbol{\epsilon}$. A prior distribution for $\boldsymbol{\mu}_{ij}$ is induced by choosing a random effects density and prior distributions for the population parameters. The usual Bayesian approach of placing constraints on the population parameters by explicitly choosing priors with restricted support is straightforward, with some risk of bias (as discussed in Section 2.2). However, restricting the population parameters is not the same as restricting the linear predictor $\boldsymbol{\mu}_{ij} \in \boldsymbol{\Omega}$, which is our focus.

It is difficult to induce restrictions on $\boldsymbol{\mu}_{ij}$ through restricting the population parameters and random effects, since constraints on the higher-level parameters imply complex restrictions on the lower-level parameters. One can potentially use an MCMC algorithm in which, for each parameter one at a time (including the random effects), the appropriate constraint is calculated conditionally on current values of the other parameters, and the parameter is then updated subject to this constraint.

This approach has several important problems. First, the parameter constraint is often difficult to calculate being dependent on a function of the other parameters and the predictors. This calculation is made more challenging if one wants to restrict the parameters, so that $\boldsymbol{\mu}_{ij} \in \boldsymbol{\Omega}$ for values of the predictors not observed in the study sample. Second, since the constraint on a particular parameter is highly dependent on the current values of the other parameters, there will tend to be high levels of autocorrelation in the Markov chain and hence poor computational efficiency. In addition, we do not want to restrict the peak in $\boldsymbol{\mu}_{ij}$ to occur at the same location for all i, j. The standard Bayesian approach requires the explicit specification of a hierarchical prior for the peak location. In conducting posterior computation, it is then necessary to update the parameters characterizing the distribution of the peak along with the individual-specific changepoint parameters. Since efficient computation is difficult even for models with a single changepoint, computation for umbrella ordered curves with varying peaks is extremely challenging using the standard methods.

The transformation approach proposed in Section 2.2 effectively solves these problems. We begin by collecting draws from the unconstrained posterior distribution of $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\epsilon})$ using a standard Gibbs sampling algorithm without incorporating the constraint. This is straightforward, and the parameters can be updated in blocks to improve computational efficiency. The linear predictor $\boldsymbol{\mu}_{ij}$ is then computed as a function of $\boldsymbol{\Theta}$ for each draw collected from the unconstrained posterior. Once $\boldsymbol{\mu}_{ij}$ is generated, the transformation approach is used to map $\boldsymbol{\mu}_{ij} \rightarrow \boldsymbol{\mu}^*_{ij} \in \boldsymbol{\Omega}$.

2.4 Application to Menstrual Cycle Data

2.4.1 Data Structure

We illustrate the methods with data from a study of progesterone levels in the menstrual cycle, analyzed previously by Brumback and Rice (1998). The data consist of progesterone metabolite measurements in daily urine samples of women with healthy reproductive function. These hormone measurements are recorded on the log scale as a variance stabilizing mechanism. Days are nested within cycles which are nested within women who are in one of two groups - non-conceptive or conceptive. Of the 51 women in the study, 29 of them belong to the non-conceptive group (69 cycles) and 22 belong to the conceptive group (22 cycles). Data are aligned relative to the day of ovulation estimated using urinary luteinizing hormone surge. We focus on the 24 day window starting eight days prior to ovulation and ending 15 days after ovulation, since interest focuses on comparing conceptive and non-conceptive cycles with respect to pre- and post-ovulatory progesterone patterns. Ovulation is referred to as day 0.

Figure 2.2 shows the hormone measurements for two non-conceptive women and for two conceptive women. As expected, progesterone levels tend to increase to a peak and then decrease in non-conceptive cycles, while in conceptive cycles increases continue throughout the window. Note that both of these patterns are consistent with an umbrella ordering, with the peak occurring at the end of the window for conceptive cycles and earlier for non-conceptive cycles.

2.4.2 Hierarchical Model

Let y_{ijk} represent the hormone measurement for day $k = \{1, ..., 24\}$ of cycle $j = \{1, ..., n_i\}$ from woman $i = \{1, ..., 51\}$. We first choose a simple autoregressive hi-

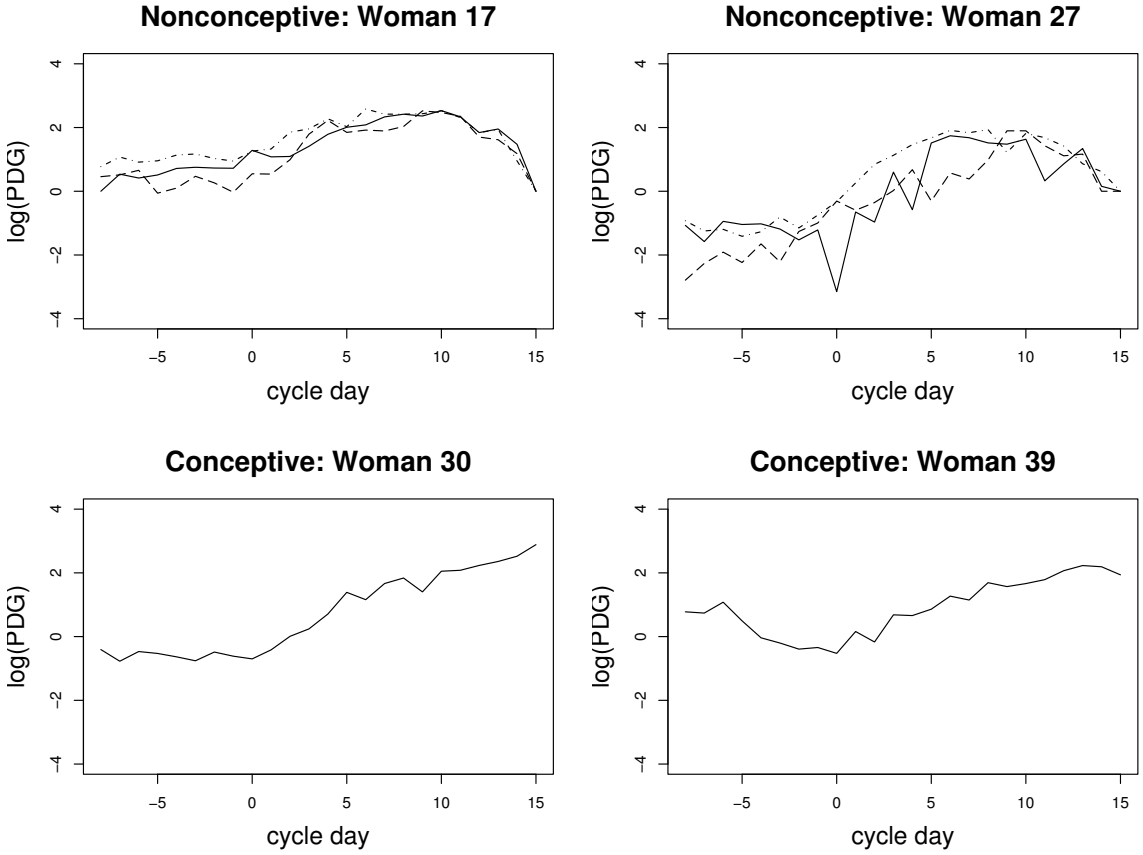

Figure 2.2: Progesterone data for 2 non-conceptive (top panel) and 2 conceptive (bottom panel) women. Each non-conceptive woman has multiple cycles.

erarchical model to characterize the data without considering the constraint. In particular, we let:

$$y_{ijk} \;=\; c_{ij}\alpha_k^{(C)} + (1-c_{ij})\alpha_k^{(N)} + \beta_{ik} + \phi_1\epsilon_{ij,k-1} + \epsilon_{ijk} \tag{2.9}$$

where c_{ij} is an indicator of conception group for the jth cycle of the ith woman, $\alpha_k \sim \mathcal{N}(\alpha_{0k}, 10)$ are day-specific population means (in either the conception (C) or nonconception (N) group), $\beta_{ik} \sim \mathcal{N}(\beta_{i,k-1}, \phi_3^2)$ are women-specific deviations from the population mean, $\phi_1 \sim \mathcal{N}(0.1, 10)$ measures residual autocorrelation, and $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$ are the residual errors. To complete the prior specification, $\sigma^2 \sim$

$\mathcal{IG}(a_{\sigma^2}, b_{\sigma^2})$, $\beta_{i0} \sim \mathcal{N}(0, \phi_2^2)$ is the woman-specific intercept, $\phi_2^2 \sim \mathcal{IG}(a_{\phi_2^2}, b_{\phi_2^2})$ measures heterogeneity in the intercept, and $\phi_3^2 \sim \mathcal{IG}(a_{\phi_3^2}, b_{\phi_3^2})$ measures autocorrelation in the β_{ik}'s. The inverse gamma prior parameters for σ^2, ϕ_2^2, and ϕ_3^2 are chosen to be moderately diffuse with a mean of 5 and variance of 100, and a sensitivity analysis for different prior parameter choices was conducted. The prior mean of $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_0$, is set equal to the empirical mean from a previous study conducted by Baird et al. (1999), with the elements of $\boldsymbol{\alpha}_0$ shown in Table 2.1. The prior variance for $\boldsymbol{\alpha}$ is set at 10 to allow for differences in the studies.

Table 2.1: Prior values for the day-specific means

Cycle Day	Non-Conceptive log(PDG)	Conceptive log(PDG)
-8	-0.65	-0.84
-7	-0.69	-0.84
-6	-0.73	-0.92
-5	-0.80	-0.84
-4	-0.80	-0.87
-3	-0.80	-0.87
-2	-0.73	-0.80
-1	-0.84	-0.92
0	-0.36	-0.48
1	-0.11	-0.16
2	0.41	0.34
3	0.79	0.74
4	1.06	1.06
5	1.19	1.19
6	1.36	1.31
7	1.35	1.39
8	1.39	1.34
9	1.34	1.34
10	1.19	1.36
11	1.03	1.39
12	0.83	1.50
13	0.53	1.55
14	0.41	1.55
15	0.34	1.61

Also notice the choice of prior density for β_{ik}. While it is common to place a

normal prior centered at 0 with a huge variance on the β_{ik}'s, this approach gives high probability to values that are known to be implausible, resulting in slow mixing of the Gibbs sampler. For this reason, we choose the prior specification for β_{ik} as mentioned above. Furthermore, β_{ik} is chosen over β_i or β_{ij} because β_{ik} allows each woman to have a distinct pattern in her day-specific means.

Our focus lies in the order-restricted functional, $\boldsymbol{\mu}_i^* = g_i(\boldsymbol{\mu}_i)$, of the unconstrained women-specific means, where: $\boldsymbol{\Omega} = \left\{\boldsymbol{\mu}_i : \mu_{i,-8} \leq \cdots \leq \mu_{i,\kappa} \geq \cdots \geq \mu_{i,15}\right\}$. The ordering in μ_{ik}^* parallels that of published data and results (Baird et al., 1997; 1999). The restriction on μ_{ik}^* guarantees that the progesterone hormone measurement in a woman's menstrual cycle increases monotonically to an unknown changepoint, that varies from woman to woman, with decreases thereafter.

2.4.3 Gibbs Sampling Algorithm

Since the full conditional distributions of each of the unconstrained parameters follow a simple conjugate form, posterior computation can proceed via a simple Gibbs sampling algorithm which proceeds as follows:

Step 1: Missing observations are sampled from their full conditional posterior distribution under the missing at random assumption.

Step 2: Sample $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, ϕ_1, ϕ_2^2, ϕ_3^2, and σ^2 from their respective full conditional densities.

Step 3: Calculate women-specific means, μ_{ik}, as a deterministic function of the day-specific population means and the women-specific random effects.

Step 4: The unconstrained μ_{ik}'s are mapped to the umbrella-ordered space $\boldsymbol{\Omega}$ using the transformation described in Section 2.2.

The algorithm was run for 110,000 iterations. The first 10,000 iterations were discarded as a burn-in, and every 100th iteration of the following 100,000 samples

were saved to thin the chain. As the number of iterations increases, Figure 2.3 portrays a sample of traceplots for two women- by day- specific means to show how most of the sampled values stabilize to look like autocorrelated samples from a unique stationary distribution. Thus, traceplots indicate convergence of the algorithm and good mixing.

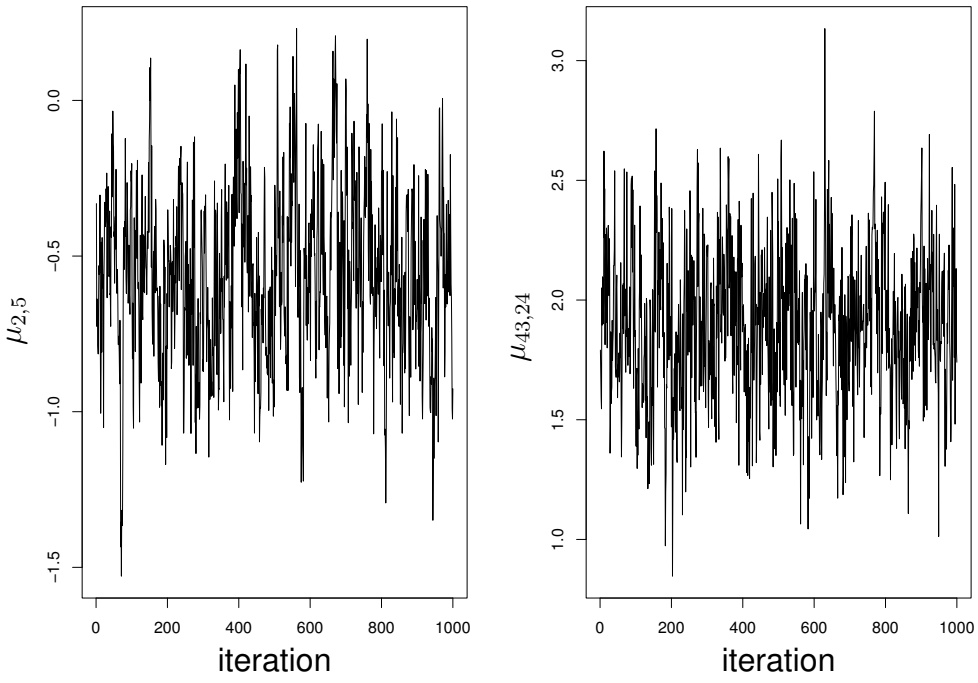

Figure 2.3: Traceplots for $\mu_{2,5}$ (left) and $\mu_{43,24}$ (right)

2.4.4 Results

We show unconstrained and constrained mean estimates and 95% credible intervals for two conceptive and two non-conceptive women in Figure 2.4. Notice in Figure 2.4 and Table 2.2 that the unconstrained estimates have bumps which likely reflect noise in the assay and not true changes in the woman-specific means. Table 2.2 shows day-specific means for women in each group. The constrained estimates do

Table 2.2: Empirical, unconstrained, and constrained estimates of the mean log progesterone levels for non-conception and conception cycles

	Non-conception			*Conception*		
Day	Empirical	Unconstrained	Constrained	Empirical	Unconstrained	Constrained
-8	$-0.83_{(0.88)^\dagger}$	$-0.88_{(0.99)}$	$-1.09_{(0.91)}$	$-0.53_{(0.81)}$	$-0.52_{(0.94)}$	$-0.74_{(0.88)}$
-7	$-0.79_{(0.86)}$	$-0.85_{(0.94)}$	$-1.01_{(0.87)}$	$-0.51_{(0.90)}$	$-0.50_{(0.94)}$	$-0.67_{(0.88)}$
-6	$-0.83_{(0.91)}$	$-0.88_{(0.90)}$	$-0.96_{(0.86)}$	$-0.43_{(0.80)}$	$-0.42_{(0.90)}$	$-0.62_{(0.87)}$
-5	$-0.88_{(0.84)}$	$-0.93_{(0.86)}$	$-0.93_{(0.85)}$	$-0.53_{(0.79)}$	$-0.53_{(0.89)}$	$-0.60_{(0.86)}$
-4	$-0.81_{(0.86)}$	$-0.90_{(0.85)}$	$-0.89_{(0.84)}$	$-0.58_{(0.91)}$	$-0.58_{(0.91)}$	$-0.57_{(0.87)}$
-3	$-0.89_{(0.96)}$	$-1.01_{(0.87)}$	$-0.86_{(0.83)}$	$-0.74_{(1.16)}$	$-0.74_{(0.93)}$	$-0.55_{(0.87)}$
-2	$-0.84_{(0.94)}$	$-0.91_{(0.86)}$	$-0.80_{(0.83)}$	$-0.50_{(0.94)}$	$-0.56_{(0.92)}$	$-0.49_{(0.87)}$
-1	$-0.72_{(0.95)}$	$-0.78_{(0.84)}$	$-0.70_{(0.83)}$	$-0.62_{(1.05)}$	$-0.62_{(0.93)}$	$-0.44_{(0.87)}$
0	$-0.50_{(1.10)}$	$-0.60_{(0.91)}$	$-0.53_{(0.86)}$	$-0.32_{(0.90)}$	$-0.32_{(0.90)}$	$-0.29_{(0.87)}$
1	$-0.06_{(0.78)}$	$-0.18_{(0.85)}$	$-0.17_{(0.83)}$	$-0.41_{(1.14)}$	$-0.32_{(0.96)}$	$-0.18_{(0.89)}$
2	$0.25_{(0.93)}$	$0.16_{(0.86)}$	$0.17_{(0.85)}$	$0.15_{(1.02)}$	$0.15_{(0.93)}$	$0.18_{(0.90)}$
3	$0.86_{(0.83)}$	$0.79_{(0.84)}$	$0.78_{(0.84)}$	$0.73_{(1.01)}$	$0.73_{(0.91)}$	$0.72_{(0.88)}$
4	$1.16_{(1.00)}$	$1.12_{(0.89)}$	$1.12_{(0.88)}$	$0.94_{(1.01)}$	$0.94_{(0.92)}$	$0.95_{(0.89)}$
5	$1.44_{(0.87)}$	$1.42_{(0.83)}$	$1.41_{(0.83)}$	$1.41_{(0.73)}$	$1.40_{(0.87)}$	$1.28_{(0.87)}$
6	$1.55_{(0.97)}$	$1.58_{(0.86)}$	$1.57_{(0.85)}$	$1.31_{(0.86)}$	$1.37_{(0.88)}$	$1.30_{(0.87)}$
7	$1.53_{(1.00)}$	$1.54_{(0.89)}$	$1.57_{(0.89)}$	$1.37_{(1.00)}$	$1.49_{(0.92)}$	$1.44_{(0.89)}$
8	$1.53_{(1.09)}$	$1.55_{(0.91)}$	$1.56_{(0.91)}$	$1.36_{(1.00)}$	$1.48_{(0.92)}$	$1.47_{(0.90)}$
9	$1.37_{(1.19)}$	$1.39_{(0.96)}$	$1.44_{(0.92)}$	$1.27_{(1.50)}$	$1.34_{(1.01)}$	$1.50_{(0.90)}$
10	$1.37_{(1.10)}$	$1.35_{(0.96)}$	$1.37_{(0.92)}$	$1.54_{(1.07)}$	$1.55_{(0.95)}$	$1.61_{(0.91)}$
11	$1.25_{(1.03)}$	$1.29_{(0.92)}$	$1.26_{(0.93)}$	$1.61_{(0.88)}$	$1.74_{(0.90)}$	$1.78_{(0.89)}$
12	$1.00_{(1.05)}$	$1.00_{(0.93)}$	$1.00_{(0.93)}$	$1.89_{(0.94)}$	$2.00_{(0.91)}$	$1.99_{(0.90)}$
13	$0.77_{(0.92)}$	$0.80_{(0.92)}$	$0.79_{(0.92)}$	$1.98_{(0.90)}$	$2.08_{(0.90)}$	$2.09_{(0.90)}$
14	$0.39_{(0.79)}$	$0.49_{(0.96)}$	$0.47_{(0.96)}$	$2.07_{(0.85)}$	$2.18_{(0.90)}$	$2.17_{(0.91)}$
15	$0.09_{(0.58)}$	$-0.06_{(1.04)}$	$-0.07_{(1.04)}$	$1.95_{(1.02)}$	$2.17_{(0.93)}$	$2.14_{(0.94)}$

† values are mean estimates$_{(sd)}$

not vary systematically from the unconstrained estimates, implying minimal bias in the estimates. As expected, the interval estimates under the transformation approach are generally narrower than the unconstrained credible intervals, particularly prior to ovulation, reflecting a possible improvement in efficiency for the constrained approach. These differences in the unconstrained and constrained point and interval estimates are also seen in the simulation study in Section 2.4.5. Furthermore, a likely improvement in efficiency for the constrained approach is also illustrated in Figure 2.5, which shows the ratios of the unconstrained to constrained variances for the women-specific estimates as a function of cycle day. The variation in the women-

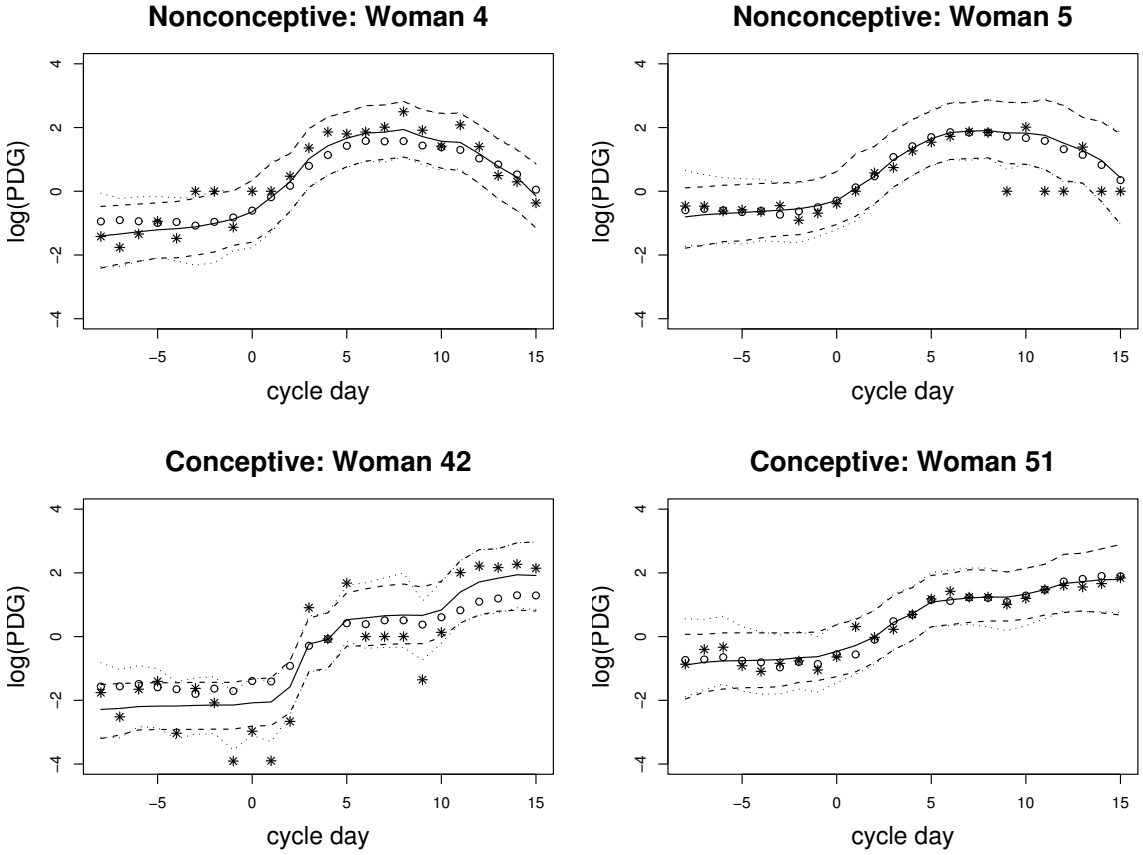

Figure 2.4: Progesterone data for 2 non-conceptive (top panel) and 2 conceptive (bottom panel) women. Data (asterisk), unconstrained women-specific mean estimates (open circle), and constrained women-specific mean estimates using the transformation approach (solid line), with 95% intervals for unconstrained estimates (dotted line) and constrained estimates (broken line).

specific estimates is considerably smaller in the restricted case compared to that of the unrestricted means. As with the narrower interval widths displayed in Figure 2.4 corresponding to the constrained means, Figure 2.5 shows a likely improvement in efficiency, particularly prior to ovulation.

Epidemiologists are also interested in the peak location and in systematic differences in this peak between non-conceptive and conceptive cycles. The day at which progesterone levels peak in non-conceptive cycles is thought to be close to the day of

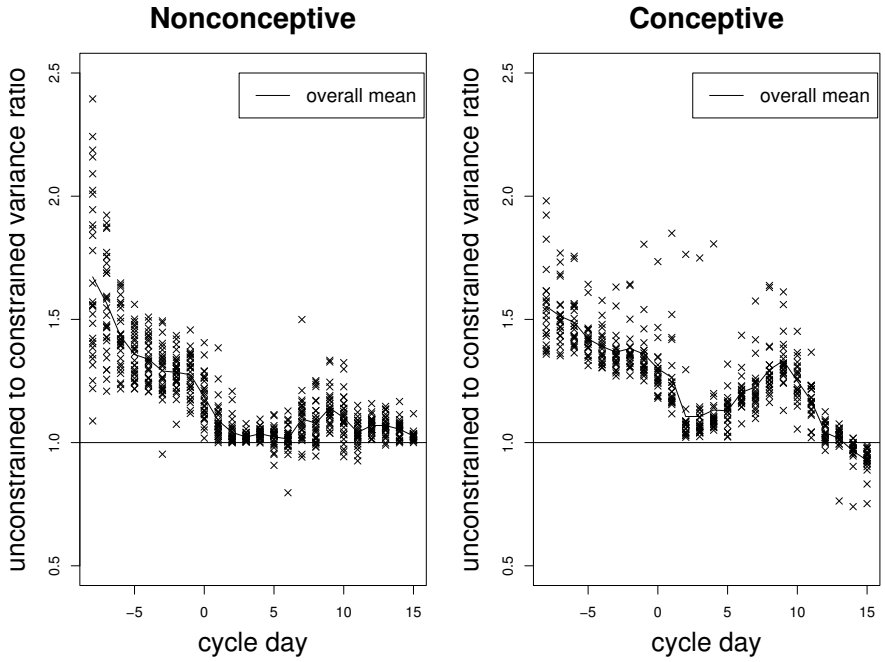

Figure 2.5: Ratios of variances in the women-specific means, unconstrained to constrained, denoted by x

implantation in conceptive cycles. The timing of the peak may be informative about the probability of conception, and women who conceive after progesterone levels fall are at a greater risk of early pregnancy loss and later adverse outcomes (Baird et al., 1999). For this reason, trajectory characteristics, including peak location, are very interesting. Figure 2.6 plots the average peak locations for non-conceptive and conceptive cycles, with the most likely peak day occurring eight days after ovulation for the non-conceptive women and thirteen days after ovulation for the conceptive women. Furthermore, there is an approximate posterior probability of 1 that the peak occurs later on average for conceptive cycles than for non-conceptive cycles.

Another interest of epidemiologists is comparing conceptive trajectories with non-conceptive trajectories during certain days of the cycle. Breaking the cycle days into three intervals: (1) $[-8, 1]$, (2) $[2, 8]$, and (3) $[9, 15]$ (with day 0 corresponding to ovulation day), we find that the estimated probabilities of higher mean progesterone

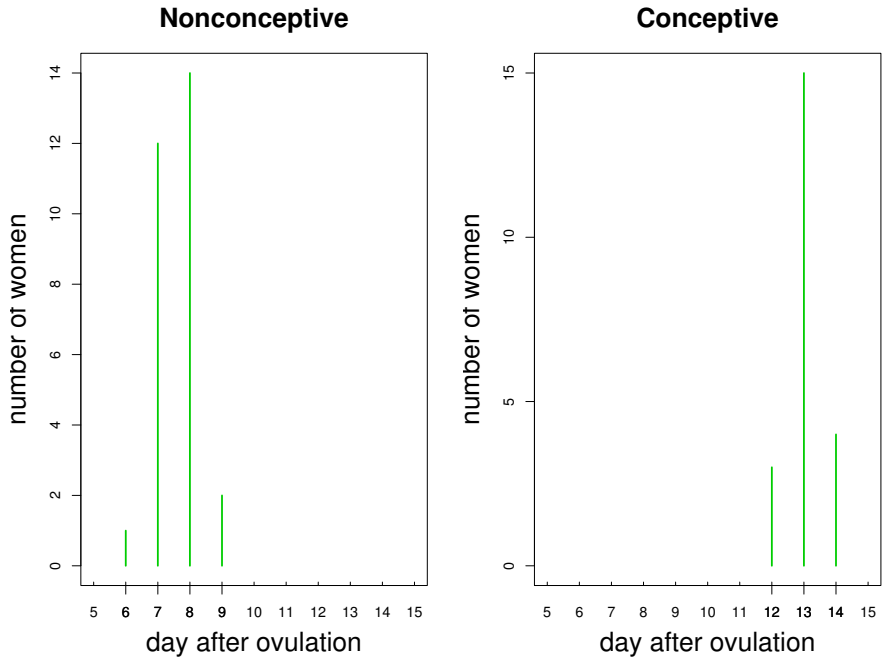

Figure 2.6: Average peak locations for non-conceptive and conceptive women

levels for conception cycles are 0.46, 0.01, and 0.69 for intervals 1-3, respectively. Therefore, conceptive and non-conceptive cycles are similar except during the week following ovulation during which conceptive cycles have significantly lower progesterone levels. There is no evidence of differences in pre-ovulatory progesterone levels between conceptive and non-conceptive cycles, suggesting that progesterone levels during the fertile interval have little effect on fecundability. Note that we have ended the first interval the day after the estimated day of ovulation, which was based on urinary LH surge in this study, to allow for possible measurement error in the LH estimate.

Aside from characteristics of mean trajectories, estimates for σ^2, ϕ_1, ϕ_2^2, and ϕ_3^2 are also computed. The posterior mean for the residual variance is $\widehat{\sigma}^2 = 0.24$, with 95% credible interval (0.23,0.26). The autocorrelation parameter, ϕ_1, has a posterior mean of 0.65, with 95% credible interval (0.60,0.70). The heterogeneity in the intercept,

ϕ_2^2, and in the women-specific random effects, ϕ_3^2, have posterior means of 0.95 and 0.10, with 95% credible intervals (0.56,1.50) and (0.08,0.12), respectively.

Finally, we conducted a sensitivity analysis to assess the robustness of these results to the prior specification. We varied the prior means for σ^2, ϕ_2^2, and ϕ_3^2 from 1 to 7, allowing arbitrarily smaller and larger values than the mean of 5 in the main analysis. Likewise, the prior variances ranged from 10 to 10,000, again allowing smaller and larger variances than that used in the primary analysis. Overall, posterior estimates for these variance component parameters were robust to the prior specification, with the estimate for ϕ_2^2 being the most sensitive. The parameters of primary interest, the restricted woman-specific means, varied approximately 1% on average across the different prior specifications.

2.4.5 Simulation Example

In order to check our computational algorithm and methods as well as explore frequentist operating characteristics, we ran the analysis using 25 simulated datasets having the same structure as the progesterone application data but with simulated progesterone values. The constrained means from the progesterone analysis were used as the true values for the simulated data. In implementing the analysis, we used a burn-in period of 10,000 iterations, with the chain run an additional 50,000 iterations, collecting samples at every 100th iteration.

Based on the results for the 25 simulated datasets, we calculated pointwise bias, defined as the average difference between the restricted estimates and the true values of the parameters, in the constrained day-specific mean estimates as well as the estimate for peak location, and the results are shown in Tables 2.3 and 2.4. There is no evidence of systematic bias in the estimates, and the differences between the averaged restricted means and the true values may reflect Monte Carlo error. Also

Table 2.3: Pointwise bias, average pointwise 95% credible intervals, and pointwise coverage of 95% credible intervals for restricted day-specific estimates in non-conception and conception cycles

	Non-conception			*Conception*		
Day	Bias	PW 95% CI	Coverage	Bias	PW 95% CI	Coverage
-8	-0.11	(-1.35,-1.05)	20/25	-0.16	(-1.11,-0.71)	18/25
-7	-0.08	(-1.20,-0.96)	20/25	-0.12	(-0.97,-0.63)	17/25
-6	-0.04	(-1.11,-0.90)	22/25	-0.09	(-0.87,-0.56)	22/25
-5	-0.01	(-1.04,-0.84)	25/25	-0.04	(-0.79,-0.49)	24/25
-4	0.02	(-0.97,-0.78)	24/25	0.01	(-0.71,-0.42)	23/25
-3	0.05	(-0.91,-0.71)	22/25	0.05	(-0.64,-0.36)	21/25
-2	0.06	(-0.84,-0.64)	21/25	0.05	(-0.58,-0.29)	23/25
-1	0.06	(-0.75,-0.53)	20/25	0.07	(-0.52,-0.22)	20/25
0	0.06	(-0.59,-0.35)	22/25	0.04	(-0.41,-0.08)	23/25
1	0.03	(-0.27,-0.01)	22/25	0.08	(-0.28,0.08)	21/25
2	0.03	(0.07,0.34)	25/25	0.04	(0.02,0.44)	22/25
3	0.00	(0.63,0.91)	24/25	-0.01	(0.50,0.91)	24/25
4	0.00	(0.99,1.26)	25/25	0.01	(0.77,1.16)	24/25
5	-0.01	(1.27,1.54)	24/25	-0.05	(1.04,1.41)	24/25
6	0.00	(1.44,1.71)	23/25	-0.01	(1.16,1.51)	24/25
7	0.02	(1.46,1.72)	24/25	-0.01	(1.26,1.61)	25/25
8	0.01	(1.44,1.70)	25/25	0.05	(1.36,1.70)	22/25
9	0.03	(1.34,1.59)	23/25	0.07	(1.41,1.74)	22/25
10	-0.01	(1.23,1.48)	24/25	0.03	(1.48,1.81)	24/25
11	-0.03	(1.10,1.36)	21/25	0.00	(1.61,1.97)	25/25
12	0.02	(0.90,1.15)	23/25	0.00	(1.79,2.20)	22/25
13	0.00	(0.66,0.92)	25/25	0.03	(1.91,2.34)	22/25
14	-0.03	(0.30,0.58)	22/25	0.01	(1.96,2.40)	23/25
15	-0.01	(-0.22,0.07)	21/25	-0.05	(1.86,2.33)	23/25

notice the minimal bias in peak location across women and analyses in the two groups, with all biases less than 0.75 days. Table 2.3 also shows the average pointwise 95% credible intervals for the restricted day-specific means across the 25 analyses, and coverage. Notice differences in slightly lower and higher coverage prior to and post ovulation, respectively. It appears that coverage of 95% credible intervals is consistent with the nominal level, though a larger simulation study is needed to fully assess this.

Table 2.4: Bias and pointwise coverage of 95% credible intervals for the peak location for non-conception and conception women overall

	Non-conception			*Conception*	
Woman	Bias (days)	Coverage	Woman	Bias (days)	Coverage
1	0.05	25/25	30	-0.29	25/25
2	-0.35	25/25	31	-0.01	25/25
3	0.28	25/25	32	-0.47	24/25
4	-0.16	25/25	33	-0.59	25/25
5	-0.34	25/25	34	-0.08	24/25
6	-0.08	25/25	35	0.26	25/25
7	-0.09	24/25	36	-0.19	25/25
8	0.39	25/25	37	0.74	24/25
9	-0.08	25/25	38	-0.13	24/25
10	-0.31	25/25	39	0.11	25/25
11	-0.11	24/25	40	0.44	24/25
12	-0.51	23/25	41	-0.49	25/25
13	-0.31	25/25	42	-0.09	24/25
14	-0.15	25/25	43	-0.68	24/25
15	-0.06	25/25	44	-0.17	25/25
16	0.20	25/25	45	-0.08	24/25
17	-0.71	21/25	46	-0.07	24/25
18	0.11	25/25	47	-0.01	24/25
19	-0.12	25/25	48	-0.05	24/25
20	0.05	25/25	49	0.26	24/25
21	0.21	25/25	50	0.16	25/25
22	-0.17	25/25	51	-0.14	25/25
23	0.18	25/25			
24	0.03	25/25			
25	0.38	25/25			
26	-0.34	25/25			
27	-0.34	25/25			
28	-0.02	25/25			
29	0.08	25/25			

2.5 Discussion

This chapter introduces a useful and easy-to-implement method for constraining parameters to follow an umbrella ordering. In the progesterone application, each cycle's hormone curve can be viewed as a single random function, with repeated random functions then collected for each woman. This is an example of hierarchical functional data, which refers to correlated samples of functions, as opposed to standard hierarchical data, which refers to data having some natural dependency structure,

such as cycles within women. Although we have applied this approach to hierarchical functional data, there are many other settings in which the methodology is potentially useful.

The transformation approach can be used for inferences on dose response functions with possible downturns at higher dose levels, or hazard functions with bathtub or hill shapes. Applying the approach in other settings is trivial since one can save existing output from an unconstrained analysis (using WinBUGS, for example), and transform the output using a simple function (i.e., coded in S-Plus or SAS). In contrast, following more conventional Bayesian procedures of placing an explicit prior with constrained support requires new coding for each new case. Since computational hurdles are one of the major factors preventing widespread use of order and shape restrictions in analyses of biomedical data, our method is all the more appealing for its promise of computational efficiency.

In addition to computational advantages, our results suggest that we can limit bias induced by the order restriction by using a mapping which minimizes the distance between the restricted and unrestricted estimates. Placing order constraints on parameters is a useful way in which outside information can be incorporated into the analysis in a nonparametric manner without using a highly informative prior for the specific values of the parameters. Such constraints often result in smoothing of the estimates and improvements in efficiency.

Chapter 3

Bayesian Methods for Assessing Ordering in Hazard Functions

3.1 Motivation and Overview

By identifying regions of age and dose where an effect on the hazard exists in studies involving rodents, researchers obtain valuable information regarding the potential adverse effects of a chemical in humans. For example, Figure 3.1 presents Kaplan-Meier survival curves from a National Toxicology Program (NTP) study of ethylbenzene (NTP Technical Report, 1999). It is of interest to assess whether survival decreases with dose, while also identifying thresholds, such as the first dose and age at which an increase in hazard occurs. Because we are interested in decreased survival in response to a potentially adverse exposure, it is natural to consider a one-sided analysis that constrains the hazard functions to be non-decreasing in dose. In addition, it would be appealing to obtain smooth estimates of dose group-specific survival curves.

In a frequentist setting, stochastic ordering methods for survival functions have been introduced (Dykstra, 1982; Feltz and Dykstra, 1985; Dykstra and Feltz, 1989;

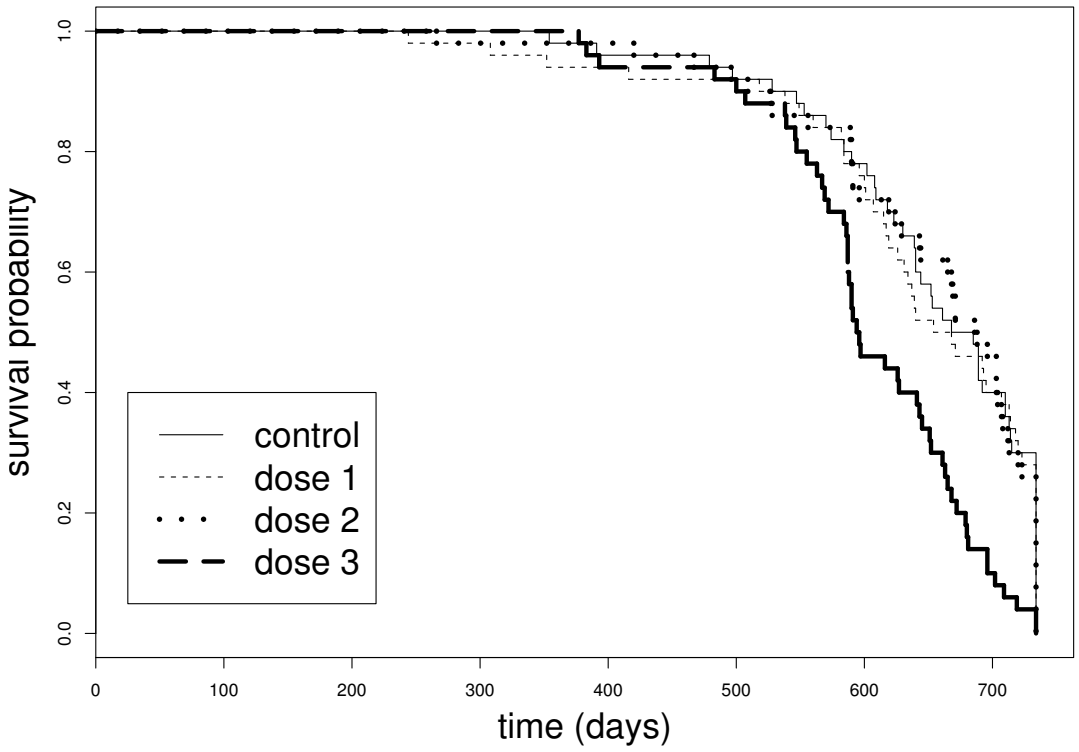

Figure 3.1: Kaplan-Meier estimates of the survival curves for male rats exposed to ethylbenzene

Dykstra, Kochar and Robertson, 1991; 1995). In addition, Arjas and Gasbarra (1996) and Gelfand and Kottas (2001) proposed nonparametric Bayesian methods for accommodating stochastic ordering in two survival functions. Furthermore, Hoff (2003) proposed an innovative Bayesian approach for nonparametric estimation of distributions subject to stochastic ordering. These methods address the issue of estimation subject to strict order constraints, however they are limited in their usefulness for assessing evidence of increasing versus constant hazards. Although several frequentist order-restricted tests are available for proportional hazards models (Sen, 1984; Silvapulle, 1994; Silvapulle and Silvapulle, 1995; Singh and Wright, 1996; 1998), there has been limited consideration of non-proportional hazards cases.

Our primary interest is in identifying dose levels and ages at which there is an increase in hazard. Equalities in hazard functions for different dose groups effectively correspond to multiple age-specific null hypotheses. Following previous authors (cf., Dunson and Herring, 2003; Gönen, Westfall, and Johnson, 2003 for recent references), we treat null hypotheses as if they could be exactly true (refer to Berger and Delampady, 1987, for a discussion and justification of this viewpoint). Because the null hypotheses correspond to different times, it is natural to assume that hypotheses located "close" to each other are highly correlated *a priori*, resulting in smoothing of the model probabilities. Related ideas have been considered by Gönen, Westfall, and Johnson (2003) and Neelon and Dunson (2004) in different settings.

In addition to smoothing of the hypothesis probabilities, it is appealing to smooth the baseline hazard function and time-varying coefficients. There is a rich Bayesian literature on defining priors for hazard functions (for a recent review refer to Ibrahim, Chen, and Sinha, 2001). Most Bayesian analyses have relied on a gamma process prior for the cumulative baseline hazard (Kalbfleisch, 1978) or on independent gamma priors for interval-specific hazards (Walker and Mallick, 1997), though several authors have proposed Markov-type priors (Gamerman, 1991; Arjas and Gasbarra, 1994; Gray, 1994; Sinha and Dey, 1998). The Markov gamma and beta processes proposed by Nieto-Barajas and Walker (2002) have the practical advantage that conjugacy properties are maintained through the use of data augmentation. Such conjugacy properties are particularly important when interest focuses on model or hypothesis selection, and it is necessary to calculate integrating constants in deriving posterior model probabilities.

In order to smooth time-specific hypothesis probabilities and coefficients subject to the non-decreasing hazards constraint, we propose a coupled Markov beta-gamma process prior, which generalizes the structure of Nieto-Barajas and Walker

(2002). The beta process component is used to characterize changes with time in the probability that the hazard curves are constant (i.e., the null hypothesis holds), and the gamma process characterizes changes in a time-varying coefficient occurring while the alternative hypothesis holds. The prior has convenient conjugacy properties which facilitate posterior computation and inferences using a data augmentation Markov chain Monte Carlo (MCMC) algorithm. By allowing time-varying coefficients, this approach generalizes the method of Dunson and Herring (2003) to allow non-proportional hazards.

Section 3.2 considers ordered hazards and introduces the coupled Markov beta and gamma prior. Section 3.3 describes the likelihood structure and approach to posterior computation. Section 3.4 applies the methods to simulation and toxicology examples, while Section 3.5 discusses results.

3.2 Modeling Ordered Hazard Curves

3.2.1 Model and Hypothesis Structure

Let $x_i \in \mathcal{X} = \{1, \ldots, d\}$ be an ordered categorical predictor or group index for subject i ($i = 1, \ldots, n$), and let $\lambda_h(t)$ denote the hazard at time $t \in \mathcal{T}$ for subjects with $x_i = h$. We assume that the hazard curves belong to a restricted functional space, Ω, in which $\lambda_1(t) \leq \lambda_2(t) \leq \cdots \leq \lambda_d(t)$ for all $t \in \mathcal{T}$. The hazard for subject i at time t conditional on x_i can be expressed as:

$$\lambda(t; x_i = h) = \lambda_0(t) + \sum_{k=1}^{h-1} \gamma_k(t), \quad \text{for } h = 1, \ldots, d, \tag{3.1}$$

where $\lambda_0(t) \geq 0$ is the baseline hazard function, and $\gamma_k(t) = \lambda_{k+1}(t) - \lambda_k(t)$ ($\gamma_k(t) \geq 0$) is the increase in the hazard function attributable to changing x_i from k to $k+1$.

Equalities in the hazard curves between groups k and $k+1$ occur at times when $\gamma_k(t) = 0$.

The global null hypothesis of no difference in the hazard curves corresponds to $H_0 : \gamma_1(t) = \cdots = \gamma_{d-1}(t) = 0$ for all $t \in \mathcal{T}$, while the global alternative is $H_1 : \gamma_k(t) \geq 0$ for $k = 1, \ldots, d-1$ and all $t \in \mathcal{T}$, with $\gamma_k(t) > 0$ for at least one k, t. In addition to assessing overall evidence of any increase in the hazard (H_1), we want to identify specific subregions of $\mathcal{X}$ and $\mathcal{T}$ across which increases occur. Such subregions correspond to locations where $\gamma_k(t) > 0$.

Although we focus on the simple case in which the ordering of each group relative to all other groups is known, partial orderings can be accommodated using trivial modifications to the methodology. For example, an interesting case is that of simple tree ordering in which Ω denotes the space of hazard functions with $\lambda_1(t) \leq \lambda_k(t)$ for $k = 2, \ldots, d$ and $t \in \mathcal{T}$. In this case, we can simply modify expression (3.1) so that $\lambda(t; x_i = h) = \lambda_0(t) + \gamma_{k-1}(t)$, where $\gamma_{k-1}(t) \geq 0$ denotes the difference in hazards between $x_i = 1$ and $x_i = k$. Note that the non-negative constraint on the $\gamma(t)$'s now only implies that groups $2, \ldots, k$ (e.g., the treated groups) have hazards at least as big as group 1 (e.g., control group).

3.2.2 Markov Prior Specification

Under the specification of Section 3.2.1, a key step is choosing priors for the baseline hazard, $\lambda_0(t)$, as well as the increments on the hazard function between groups, $\gamma_1(t), \ldots, \gamma_{d-1}(t)$, which have the necessary properties. In particular, each of these functions needs to be non-negative, and the $\gamma(t)$ functions must have a prior which allocates positive probability to regions equivalent to 0. In addition, it is appealing to have a prior which favors relatively smooth functions and penalizes large changes

in time. To facilitate this specification, we approximate the hazard function using a piecewise constant model, with $\lambda_0(t) = \lambda_{0l}$ and $\gamma_k(t) = \gamma_{kl}$ for $t \in (\tau_{l-1}, \tau_l]$ and $t \leq \tau_L$, where $0 = \tau_0 < \tau_1 < \cdots < \tau_L < \tau_{L+1} = \infty$ are pre-specified interval endpoints, which can be chosen to be arbitrarily close to one another. Piecewise constant models are often chosen as a convenient but flexible specification in Bayesian survival models.

For the baseline hazard function, we use the Markov gamma specification of Nieto-Barajas and Walker (2002). In particular, letting $\boldsymbol{\lambda}_0 = (\lambda_{01}, \ldots, \lambda_{0L})'$, we express the prior as follows:

$$\begin{aligned} \lambda_{00} &\sim \mathcal{G}(a_\lambda, b_\lambda) \\ w_l|\lambda_{0l-1} &\sim Poisson(c_\lambda \lambda_{0l-1}) \\ \lambda_{0l}|w_l &\sim \mathcal{G}(a_\lambda + w_l, b_\lambda + c_\lambda), \end{aligned} \tag{3.2}$$

where λ_{00} is a latent baseline term, λ_{0l} has mean $(a_\lambda + w_l)/(b_\lambda + c_\lambda)$ with a_λ, b_λ, and c_λ as fixed hyperparameters specified by the investigator, c_λ is a common smoothing parameter in which more smoothing occurs as c_λ becomes larger, and w_l is a latent process introducing dependence into the prior process. Hence, $\boldsymbol{\lambda}_0$ is calculated using the process $\lambda_{00} \rightarrow w_1 \rightarrow \lambda_{01} \rightarrow w_2 \rightarrow \ldots$. The prior on $\boldsymbol{\lambda}_0$ extends the independent gamma process prior to incorporate dependence with a Markov relation for the baseline hazard. An advantage of using this prior is that it retains its conditionally conjugate form for convenient posterior computation.

The prior for $\gamma_k(t), k = 1, \ldots, d-1$, requires a different approach, since the Markov gamma specification does not allow regions across which $\gamma_k(t) = 0$. For this reason, we propose a novel specification based on coupling Markov gamma and beta processes. First, let $\pi_{kl} = \Pr(\gamma_{kl} = 0)$ denote the probability that there is no difference in the hazard functions for groups k and $k+1$ within the interval $(\tau_{l-1}, \tau_l]$, and let $\gamma_{kl} = (1 - \delta_{kl})\gamma^*_{kl}$ denote the actual difference in hazards within this interval. Following a common technique in the Bayesian variable selection literature

(cf., George and McCulloch, 1997), we express the coefficient, γ_{kl}, as the product of an indicator variable, $\delta_{kl} = 1_{(\gamma_{kl}=0)}$ where $\delta_{kl} = \delta_k(t)$ for $t \in (\tau_{l-1}, \tau_l]$, and the potential value if the coefficient is non-zero, γ^*_{kl}.

We induce a prior on γ_{kl} by specifying priors for the point mass probabilities, π_{kl}, and the potential slopes given non-zero values, γ^*_{kl}. In particular, instead of assuming *a priori* independence in the point mass probabilities $\{\pi_{kl}\}$ (as in Hjort, 1990), we use a Markov beta process (Nieto-Barajas and Walker, 2002) to allow autocorrelation:

$$\begin{aligned} \pi_{k0} &\sim \mathcal{B}e(a_{\pi k}, b_{\pi k}) \\ u_{kl}|\pi_{kl-1} &\sim \mathcal{B}in(c_{\pi k}, \pi_{kl-1}) \\ \pi_{kl}|u_{kl} &\sim \mathcal{B}e(a_{\pi k} + u_{kl}, b_{\pi k} + c_{\pi k} - u_{kl}), \end{aligned} \tag{3.3}$$

where π_{k0} is a latent baseline term included to simplify the analysis, π_{kl} has mean $(a_{\pi k} + u_{kl})/(a_{\pi k} + b_{\pi k} + c_{\pi k})$ with $a_{\pi k}$, $b_{\pi k}$, and $c_{\pi k}$ fixed by the investigator, $c_{\pi k}$ is a smoothing parameter (note $c_{\pi k}$ must be an integer), and u_{kl} is the latent process incorporating dependence into the prior process.

The hyperparameters can be selected so that the probability of the global null hypothesis H_0 is approximately 0.5. This results in a Bayesian correction for multiple comparisons, which is less conservative than the Bonferroni-type correction obtained under *a priori* independence (Westfall, Johnson and Utts, 1997). Of course, setting $\Pr(H_0) = 0.5$ may be overly conservative and may result in too high a degree of shrinkage towards equalities among the groups, particularly when dimensionality is high. Sensitivity of inferences to hyperparameter specification is considered in Section 3.4 in the context of the data example.

It is common in the Bayesian variable selection literature to specify a beta prior for the probability of including a predictor in a regression model. In models involving time-varying coefficients, it is natural to allow autocorrelation in whether a predictor is included. For example, if there is a treatment effect at time t, then it is very likely

that there is a treatment effect at time $t + \epsilon$ for small ϵ. The Markov beta process captures this dependency structure, allowing the correlation in whether a group difference occurs to be decreasing with time. Using the Nieto-Barajas and Walker (2002) structure, which they proposed in a very different context, results in computational simplifications. In particular, closed forms are available for conditional model probabilities, avoiding the need for approximations to the normalizing constants. Details are presented in Section 3.3.

We complete prior specification by defining a Markov gamma process for $\{\gamma^*_{kl}\}$:

$$\begin{aligned} \gamma^*_{k0} &\sim \mathcal{G}(a_{\gamma^* k}, b_{\gamma^* k}) \\ z_{kl}|\gamma^*_{kl-1} &\sim Poisson(c_{\gamma^* k}\gamma^*_{kl-1}) \\ \gamma^*_{kl}|z_{kl} &\sim \mathcal{G}(a_{\gamma^* k} + z_{kl}, b_{\gamma^* k} + c_{\gamma^* k}), \end{aligned} \tag{3.4}$$

where γ^*_{k0} is a latent baseline term, $a_{\gamma^* k}$ and $b_{\gamma^* k}$ are specified by the investigator, γ^*_{kl} has mean $(a_{\gamma^* k} + z_{kl})/(b_{\gamma^* k} + c_{\gamma^* k})$, $c_{\gamma^* k}$ is a smoothing parameter specified by the investigator, and z_{kl} is the latent process introducing dependence into the prior process. By coupling the Markov beta and gamma components, the prior for γ_{kl} has support on the space of non-decreasing hazard functions, Ω, while still allowing for flat regions over which increases in the level of a predictor have no effect on the hazard. The prior for γ_{kl} is directly induced in terms of δ_{kl} and γ^*_{kl}, as defined earlier in this section. As is the case for related priors used in variable selection, specifying γ_{kl} as a function of δ_{kl} and γ^*_{kl} greatly simplifies computation.

Integrating out γ^*_{kl} and δ_{kl}, the induced prior on γ_{kl} can be formally expressed as follows:

$$\gamma_{kl}|\pi_{kl}, z_{kl}, z_{kl+1} = 1_{\{\gamma_{kl}=0\}}\pi_{kl} + 1_{\{\gamma_{kl}>0\}}(1-\pi_{kl})\mathcal{G}(\gamma_{kl}; a_{kl}, b_{kl}), \tag{3.5}$$

where $a_{kl} = a_{\gamma^* k} + z_{kl} + z_{kl+1}$ and $b_{kl} = b_{\gamma^* k} + 2c_{\gamma^* k}$. We refer to density (3.5) with shorthand $\delta_0 - \mathcal{G}(\cdot; \pi, a, b)$, consisting of a mixture of a point mass at 0 (with prob-

ability π) and a $\mathcal{G}(\cdot; a, b)$ density. This prior has the important practical advantage of being conditionally-conjugate, so that closed forms are available for the conditional model probabilities. This property greatly facilitates posterior computation, and avoids the need for approximations to the normalizing constants.

3.2.3 Extension to Multiple Predictors

We can generalize the approach to allow for multiple covariates with different constraints on the regression coefficients. For example, we could extend model (3.1) to incorporate a vector of ordered categorical predictors, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$. Expression (3.1) would become:

$$\lambda(t; \boldsymbol{x}_i) = \lambda_0(t) + \sum_{h=1}^{p} \sum_{k=1}^{x_{ih}-1} \gamma_{hk}(t), \tag{3.6}$$

where $\gamma_{hk}(t)$ is the change in hazard at time t attributable to increasing x_{ih} from k to $k+1$, $\boldsymbol{\gamma}_h = (\gamma_{h1}, \ldots, \gamma_{h,d_h-1})'$ for $h = 1, \ldots, p$, d_h is the total number of categories of x_{ih}, and $\lambda(t; \boldsymbol{x}_i) \geq 0$. We complete a Bayesian model specification by assigning priors to $\boldsymbol{\lambda}_0$ and $\boldsymbol{\gamma}_k$ using the approach described in Section 3.2.2.

We can extend the approach further to account for continuous covariates. Let $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{iq})'$ be a vector of continuous predictors. As an extension to existing additive hazard models with continuous predictors (Lin and Ying, 1994; Lin, Oakes, and Ying, 1998), we introduce:

$$\lambda(t; \boldsymbol{x}_i, \boldsymbol{w}_i) = \lambda_0(t) + \boldsymbol{w}_i'\boldsymbol{\beta} + \sum_{h=1}^{p} \sum_{k=1}^{x_{ih}-1} \gamma_{hk}(t), \tag{3.7}$$

where $\boldsymbol{\beta}$ are regression coefficients, and the remaining expressions are previously defined. For $\lambda(t; \boldsymbol{x}_i, \boldsymbol{w}_i)$ to be non-negative, we apply the constraint $\boldsymbol{w}_i'\boldsymbol{\beta} \geq 0$ into the

prior specification. Following Lin and Ying (1994) and Lin, Oakes, and Ying (1998), we could incorporate a time-dependency on the covariates such that we obtain $\boldsymbol{w}_i(t)$ as a substitute for $\boldsymbol{w}_i$ in model (3.7). We could also potentially include time-varying coefficients, $\boldsymbol{\beta}(t)$, to replace $\boldsymbol{\beta}$.

3.3 Bayesian Inference

3.3.1 Counting Process Likelihood and Data Augmentation

Following the counting process notation introduced by Andersen and Gill (1982), let $N_i(t)$, for subject i ($i = 1, \ldots, n$), represent the process counting the failures occurring up to time t, while $dN_i(t)$ is a small increment of $N_i(t)$ over the interval $[t, t + dt)$. $N_i(t)$ and $dN_i(t)$ equal 1 if the event occurs in $[0, t]$ or $[t, t + dt)$, respectively, and 0 otherwise. Let $Y_i(t) = 1$ if subject i is at risk at time t, and $Y_i(t) = 0$ otherwise. For subject i in group k at time t, the intensity process for $N_i(t)$ under model (3.1) is:

$$\lambda_i(t) = Y_i(t) \left(\lambda_0(t) + \sum_{k=1}^{x_i - 1} \gamma_k(t) \right). \tag{3.8}$$

Under noninformative censoring, the observed counting process likelihood is proportional to:

$$\prod_{i=1}^{n} \left(\prod_{t \geq 0} [Y_i(t)\{\lambda_0(t) + \sum_{k=1}^{x_i - 1} \gamma_k(t)\}]^{dN_i(t)} \right) \exp\left(- \int_{t \geq 0} Y_i(t)\{\lambda_0(t) + \sum_{k=1}^{x_i - 1} \gamma_k(t)\} dt \right) \tag{3.9}$$

following a Poisson form. The infinitesimal counting process increments, $dN_i(t)$, contribute to the likelihood just as with independent Poisson random variables with means $\lambda_i(t)dt$ over the interval $[t, t + dt)$, though $dN_i(t)$ is at most one for all i, t. Defining the model in this framework allows the intensity to be regarded as constant in that interval (Clayton, 1991).

For efficient posterior computation, we implement a data augmentation approach based on the consideration that the $dN_i(t)$ are independent Poisson random variables under the likelihood expression (3.9):

$$dN_i(t) \sim Poisson\left(Y_i(t)\{\lambda_0(t) + \sum_{k:\delta_k(t)=0}^{x_i-1} \gamma_k(t)\}\right) \quad \text{for all } i : Y_i(t) = 1, \quad (3.10)$$

where $\delta_k(t) = \delta_{kj}$ for $t \in (t_{j-1}, t_j]$. Using the additive form of the Poisson sum prevents us from obtaining an efficient computational approach, since full conditional posterior distributions are non-standard. To remedy this problem, we can augment the $dN_i(t)$ in terms of independent Poisson latent variables corresponding to each term of the Poisson mean in (3.10). We introduce latent variables $\boldsymbol{dN}_i(t) = \{dN_{i0}(t), dN_{ik}(t), \text{ for all } t : \delta_k(t) = 0\}$ such that, by integrating out these latent variables, expression (3.10) is equivalent to:

$$dN_i(t) = dN_{i0}(t) + \sum_{k:\delta_k(t)=0}^{x_i-1} dN_{ik}(t), \quad (3.11)$$

with $dN_{i0}(t) \sim Poisson(\lambda_0(t))$ and $dN_{ik}(t) \sim Poisson(\gamma_k(t))$.

Following previous methods (Clayton, 1994; Dunson and Herring, 2004), we define $\boldsymbol{t} = (t_1, ..., t_J)'$ as the union of unique, ordered failure times in the data and the interval endpoints $\boldsymbol{\tau} = (\tau_1, ..., \tau_L)'$. The observed counting process likelihood (3.9) can be expressed as a product across the time intervals, $(t_{j-1}, t_j], j = 1, \ldots, J$:

$$\prod_{i=1}^{n}\prod_{j=1}^{J}\left[\left(\prod_{t\in(t_{j-1},t_j]}[Y_i(t)\{\lambda_0(t) + \sum_{k:\delta_{kj}=0}^{x_i-1}\gamma_k(t)\}]^{dN_i(t)}\right)\right.$$
$$\left.\exp\left(-\int_{t\in(t_{j-1},t_j]} Y_i(t)\{\lambda_0(t) + \sum_{k:\delta_{kj}=0}^{x_i-1}\gamma_k(t)\}dt\right)\right]. \quad (3.12)$$

Let $dN_{ij} = 1$ if subject i fails at time t_j, and 0 otherwise. With the assumption that there is a small amount of risk accumulated in $(t_{j-1}, t_j]$,

$$\int_{t_{j-1}}^{t_j} Y_i(t)\{\lambda_0(t) + \sum_{k:\delta_{kj}=0}^{x_i-1} \gamma_k(t)\}dt \approx 0 \quad \text{for all } i, j, \tag{3.13}$$

the likelihood contribution for a subject at risk in the jth interval is proportional to:

$$\left(d\Lambda_{0j} + \sum_{k:\delta_{kj}=0}^{x_i-1} d\Gamma_{jk}\right)^{dN_{ij}} \exp\left(-(d\Lambda_{0j} + \sum_{k:\delta_{kj}=0}^{x_i-1} d\Gamma_{jk})\right),$$

where $d\Lambda_{0j} = \int_{t_{j-1}}^{t_j} \lambda_0(s)ds$ and $d\Gamma_{jk} = \int_{t_{j-1}}^{t_j} \gamma_k(s)ds$. The Poisson approximation to the likelihood is highly accurate under the assumption that a Poisson random variable with mean (3.13) has a small chance of being greater than one. Assumption (3.13) is justified whether the interval $(t_{j-1}, t_j]$ is narrow or wide. Since at most one event will occur in this interval, by definition, then the hazard will be extremely small, even if the interval is large.

Let $Y_{ij} = 1$ if subject i is at risk at time t_j, and 0 otherwise. Using the approach in (3.11), dN_{ij} is expressed as:

$$dN_{ij} = dN_{ij0} + \sum_{k:\delta_{kj}=0}^{x_i-1} dN_{ijk}, \qquad \text{for all } i, j \text{ and } Y_{ij} = 1, \tag{3.14}$$

with the latent variables having independent Poisson distributions:

$$\Pr(dN_{ij0}) = Poisson(dN_{ij0}; d\Lambda_{0j}) \quad \text{and} \quad \Pr(dN_{ijk}) = Poisson(dN_{ijk}; d\Gamma_{jk}). \tag{3.15}$$

The increment dN_{ij} contributes to the likelihood in a similar manner as $dN_i(t)$, so

we replace expression (3.9) with the augmented data likelihood:

$$\propto \prod_{i=1}^{n}\prod_{j:Y_{ij}=1}^{J}\left[1_{\left(dN_{ij}=dN_{ij0}+\sum_{k:\delta_{kj}=0}^{x_i-1}dN_{ijk}\right)}\frac{d\Lambda_{0j}^{dN_{ij0}}\exp(-d\Lambda_{0j})}{dN_{ij0}!}\right.$$

$$\left.\times\left\{\prod_{k:\delta_{kj}=0}^{x_i-1}\frac{d\Gamma_{jk}^{dN_{ijk}}\exp(-d\Gamma_{jk})}{dN_{ijk}!}\right\}\right]. \tag{3.16}$$

We can now obtain conditionally conjugate posteriors for the MCMC sampling algorithm.

3.3.2 Conditional Posterior Distributions

The necessary conditional distributions are derived in this Section and in Appendix B. In particular, we derive conditional posterior distributions for $\boldsymbol{\lambda}_0$ and $\boldsymbol{\gamma}_k$ using the augmented data likelihood (3.16) and priors (3.2) and (3.5), respectively. The conditional posterior for $\boldsymbol{\lambda}_0$ is:

$$\Pr(\lambda_{0l}|\boldsymbol{\lambda}_{(-l)}, w_l, w_{l+1}, data) = \mathcal{G}(\lambda_{0l}; \tilde{a}_\lambda, \tilde{b}_\lambda), \tag{3.17}$$

where $\boldsymbol{\lambda}_{(-l)}$ is the vector excluding the lth element of $\boldsymbol{\lambda}$, $\tilde{a}_\lambda = a_\lambda + w_l + w_{l+1} + \sum_{i=1}^{n}\sum_{j,Y_{ij}=1:t_j\in(\tau_{l-1},\tau_l]} dN_{ij0}$, and $\tilde{b}_\lambda = b_\lambda + 2c_\lambda + \sum_{i=1}^{n}\sum_{j:t_j\in(\tau_{l-1},\tau_l]} Y_{ij}(t_j - t_{j-1})$. A derivation for expression (3.17) is found in Appendix B. We also show in Appendix B that the conditional posterior for $\boldsymbol{\gamma}_k$ is:

$$\Pr(\gamma_{kl}|\boldsymbol{\gamma}_{(-kl)}, \boldsymbol{\lambda}_0, data) = \delta_0 - \mathcal{G}\left(\gamma_{kl}; \tilde{\pi}_{kl}, \tilde{a}_{kl}, \tilde{b}_{kl}\right), \tag{3.18}$$

where $\boldsymbol{\gamma}_{(-kl)}$ is the vector excluding the klth element of $\boldsymbol{\gamma}$,

$$\tilde{\pi}_{kl} = \frac{\pi_{kl}}{\pi_{kl} + (1-\pi_{kl})\frac{C(\gamma_{kl};a_{kl},b_{kl})}{C(\gamma_{kl};\tilde{a}_{kl},\tilde{b}_{kl})}} \tag{3.19}$$

is the conditional posterior probability of $\gamma_{kl} = 0$, $C(\cdot; a, b)$ is the constant in the gamma density $(b^a/\Gamma(a))$,

$$\begin{aligned} \tilde{a}_{kl} &= a_{kl} + \sum_{i,j:Y_{ij}=1, t_j \in (\tau_{l-1}, \tau_l]} 1_{\{x_i > k\}} dN_{ijk} \\ \tilde{b}_{kl} &= b_{kl} + \sum_{i,j:t_j \in (\tau_{l-1}, \tau_l]} 1_{\{x_i > k\}} Y_{ij}(t_j - t_{j-1}). \end{aligned} \tag{3.20}$$

The conditional posterior density for γ_{kl} follows the same $\delta_0 - \mathcal{G}$ form as prior (3.5), resulting in a conditionally conjugate structure. It is trivial to sample directly from (3.18) by setting $\gamma_{kl} = 0$ with probability $\tilde{\pi}_{kl}$, and sampling γ_{kl} from $\mathcal{G}(\gamma_{kl}; \tilde{a}_{kl}, \tilde{b}_{kl})$ otherwise.

Samples of parameter values are easily obtained using a Gibbs sampling algorithm. Our MCMC algorithm proceeds by alternating among the following steps:

Step 1: Sample the latent Poisson variables $\{dN_{ij0}, dN_{ij1}, \ldots, dN_{ij,d-1}\}$ from their full conditional posterior distribution in (B.1) if $dN_{ij} = 1$. If $dN_{ij} = 0$, then set $dN_{ij0} = dN_{ijk} = 0$ for all $k = 1, \ldots, d-1$.

Step 2: Update the elements of $\boldsymbol{\lambda}_0$ by alternating between sampling the latent process variables $\{w_l\}$ from their full conditional (B.2) and $\boldsymbol{\lambda}_0$ from their gamma full conditional distributions (3.17), for $l = 1, ..., L$.

Step 3: Update the elements of π_{kl} from their beta full conditional distribution (B.4) and $\{u_{kl}\}$ from their full conditionals (B.3), for $l = 1, ..., L$ and $k = 1, \ldots, d-1$.

Step 4: Sample γ^*_{kl} from their gamma full conditional distribution (B.6) and $\{z_{kl}\}$ from their full conditionals (B.5), for $l = 1, ..., L$ and $k = 1, \ldots, d-1$.

Step 5: Calculate $\tilde{\pi}_{kl}$ using (3.19), for $l = 1, ..., L$ and $k = 1, \ldots, d-1$.

Step 6: Sample δ_{kl} from $\mathcal{B}ern(\tilde{\pi}_{kl})$, for $l = 1, ..., L$ and $k = 1, \ldots, d-1$:

If $\delta_{kl} = 1$, then let $\gamma_{kl} = 0$.

If $\delta_{kl} = 0$, then let $\gamma_{kl} = \gamma^*_{kl}$.

Under mild regularity conditions, draws from the above algorithm converge to the joint posterior distribution. We illustrate how samples from the posterior distribution can be used for inferences on differences between groups in the next section.

3.3.3 Inference on Ordering

Since interest focuses on assessing differences in hazards of specific dose groups, we can easily calculate posterior probabilities and Bayes factors for comparisons. For example, we can evaluate the effect of no increase in hazards with increasing dose between those in groups k and $k+1$, compared to the alternative of an increase in hazards for subjects in dose group $k+1$ compared to group k. We use the estimator $\hat{\pi}_k = \sum_{m=1}^{M} 1(\gamma_{k,1}^{(m)} = \gamma_{k,2}^{(m)} = \cdots = \gamma_{k,L}^{(m)} = 0)/M$ to compute the posterior probability of H_{0k}, where $1(\cdot)$ is an indicator function, M is the total amount of iterations saved after detected convergence, and $\gamma_{kl}^{(m)}$ is γ_{kl} at the mth iteration.

We can also compute Bayes factors for making such comparisons. Using the estimate $\hat{\pi}_k$, the Bayes factor measures the weight of evidence in favor of H_{0k} over H_{1k}:

$$\mathrm{BF}_{k,01} = \left(\frac{\hat{\pi}_k}{1-\hat{\pi}_k}\right) / \left(\frac{\pi_k}{1-\pi_k}\right), \tag{3.21}$$

which is simply the posterior divided by prior odds of H_{0k}. There are several Bayesian alternatives to the frequentist p-value, including the posterior probability, posterior odds, and Bayes factors of H_{0k} (Berger and Delampady, 1987; Kass and Raftery, 1995; Lavine and Schervish, 1999). For example, much larger posterior odds than the standard frequentist level of significance, $\alpha = 0.05$, would serve as evidence in favor of increasing hazards with increasing dose from groups k to $k+1$. Similar

methods can be used to determine posterior probabilities, posterior odds, and Bayes factors for overall conclusions about equalities or differences in hazards.

3.4 Application

3.4.1 Data Structure and Background

We apply the approach of Sections 3.2 and 3.3 to data from a toxicology study of ethylbenzene (NTP, 1999), a chemical widely used in manufacturing such products as rubber, plastic, gasoline, insecticide sprays, paints, and polyester fibers. Ethylbenzene belongs to a class of chemicals recognized as possible human carcinogens, and determining safe doses for worker exposure is an important concern. Our interest focuses on assessing evidence of increasing mortality with dose, focusing on male rats for illustrative purposes.

Two hundred male rats were exposed to ethylbenzene particles through inhalation at concentrations of 0 (control), 75, 250, and 750 parts per million (ppm) for six hours per day, 5 days per week. Rats were assigned at random to treatment and control groups, resulting in 50 rats per group. For ethical reasons, very ill rats were sacrificed to ease suffering. We group these moribund sacrifices with natural deaths in the analysis, so that the event of interest is defined as the time to critical illness or death. Animals surviving through the end of the study at two years were killed in a terminal sacrifice, which provides an ignorable censoring mechanism. Table 3.1 provides summaries of the numbers of rats dying from each cause.

3.4.2 Model and Prior Specification

Letting $x_i \in \{1, 2, 3, 4\}$ be a dose group indicator, ranging from $x_i = 1$ for control to $x_i = 4$ for the 750 ppm group, the hazard of death at age t can be expressed as in

Table 3.1: Survival rate of rats exposed to ethylbenzene

	Control	Dose 1	Dose 2	Dose 3
Natural death	7/50	16/50	11/50	22/50
Moribund sacrifice	28/50	20/50	26/50	26/50
Survived to study's end	15/50	14/50	13/50	2/50

equation (3.1):

$$\lambda(t; x_i = h) = \lambda_0(t) + \sum_{k=1}^{h-1} \gamma_k(t), \quad \text{for } h = 1, \ldots, 4,$$

where $\lambda_0(t)$ is the hazard at age t for control rats, and $\gamma_k(t)$ is the increase in hazard attributable to increasing the dose from the level in group k to the level in group $k+1$. In order to place more knots at those ages when changes in the hazard of death are more likely and to avoid having many knots in regions across which few events occur, we use unique event time data from an historical study's control group to define the τ's. To remain consistent with the type, gender, and diet of rodent used in the ethylbenzene data, we chose an inhalation historical study on 2-Butoxyethanol (Ethylene Glycol Monobutyl Ether) for male rats under the same diet.

We used this historical control group in specifying hyperparameters for λ_{00} within the Markov gamma process. Hazards were calculated from survival probabilities of each corresponding time interval in the historical control. Hazards were then averaged over time intervals and used as the mean for the initial element of the baseline hazard, λ_{00}. The variance of the historical control hazards was inflated by a factor of six to allow for study differences in the ethylbenzene example. As a result, λ_{00} is assigned a gamma density with mean 0.09 and variance 0.50. Furthermore, we fix the smoothing parameter, c_λ, at 50 to allow for correlation in the baseline hazard across time intervals, l. Although we fix c_λ, we later discuss its robustness to different choices of values.

In specifying the coupled Markov beta and gamma density for each γ_{kl}, we assign a prior to initial elements of the beta and gamma processes for π_{kl} and γ^*_{kl}, respectively. The hyperparameters on $\boldsymbol{\pi}_k$ give the prior guess at no dose effect when increasing dose from group k to $k+1$. The prior on the initial point mass is specified by choosing hyperparameters such that the prior probability of homogeneity in hazards across levels of dose and time is approximately 0.50. In other words, we want the prior probability of a trend to reflect the fact that we do not have any strong prior belief of an adverse survival effect. Similarly, the initial γ^*_{k0} is assigned a gamma density with mean 0.50 and variance 1.0. Another element to the Markov beta-gamma process prior involves the smoothing parameters for each component process. By including autocorrelation we are smoothing the functions, borrowing information across neighboring intervals. To allow moderate to high autocorrelation in $\boldsymbol{\pi}_k$ and $\boldsymbol{\gamma}_k$, we fix $c_{\pi k}$ and $c_{\gamma^* k}$, respectively, at 50. We later show robustness to choice of $c_{\pi k}$ and $c_{\gamma^* k}$ through a sensitivity analysis.

3.4.3 Simulation Study

We first ran the analysis on simulated data similar in structure to historical control data. Using the mean (660 days) and variance (6621.41) of the historical control survival times, we simulated survival times for 200 subjects, with 50 subjects in each of four groups, from a gamma distribution. Values greater than the 2-year mark were given a value of 731 days and were considered censored, as with terminal sacrifices in the data example. One of our goals in the simulation study is to create an example under the null hypothesis.

We implemented the data augmentation sampling algorithm outlined in Section 3.3.2. A burn-in period of 10,000 iterations was used, with an additional 50,000 iterations obtained to collect every 25th sample for posterior summaries. The burn-

in was sufficient for apparent convergence based on diagnostic plots. Similar to the analysis in Chapter 2, traceplots were thoroughly examined for apparent convergence. The sampled values appeared to stabilize throughout most of the diagnostic plots. Furthermore, the level of autocorrelation in the samples tended to be low to moderate. As a result, a collection interval of 50,000 was judged sufficient to limit Monte Carlo error in estimating posterior summaries.

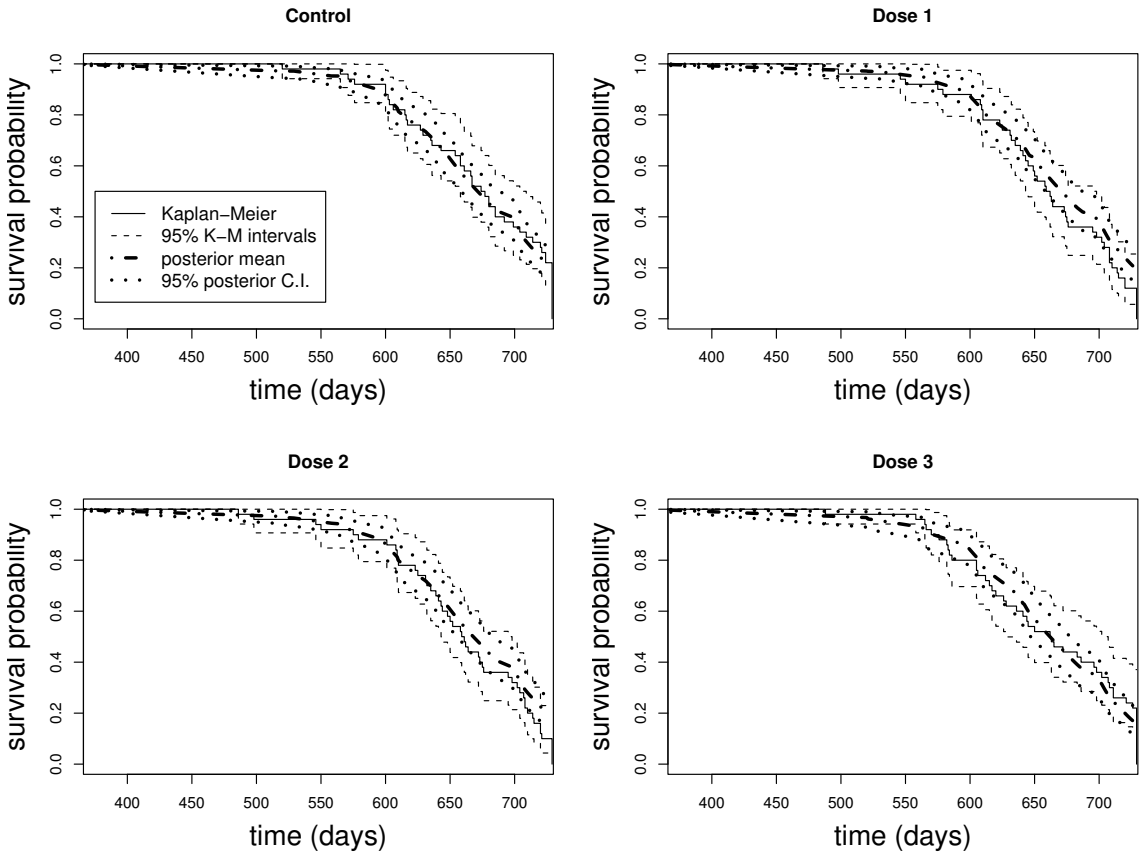

Figure 3.2: Simulation results: posterior means and 95% pointwise credible intervals of the survival probability, after the first year, overlaid on the Kaplan-Meier curves with their 95% intervals

Posterior means and pointwise 95% credible intervals for the survival curves are plotted in Figure 3.2 along with the Kaplan-Meier estimates. The close correspon-

dence between the posterior means and the Kaplan-Meier estimates suggests that our estimates provide a good fit to the data, with minimal bias introduced by the prior structure. However, the posterior mean is smoother than the Kaplan-Meier estimate, due to the Markov prior structure, and provides a more realistic estimate. In addition, the 95% credible intervals are substantially narrower than 95% frequentist confidence intervals, with an average reduction of 0.2 in the interval widths. This result possibly reflects improved efficiency attributable to the order restriction and smoothing prior. Note that the estimated survival curves are essentially identical for the different groups even though we are estimating the curves under a non-decreasing constraint. This is certainly due to the prior structure, which allows equalities with positive prior probability.

Prior and posterior probabilities of an overall dose effect are 0.21, 0.36, and 0.49 for doses 1, 2, and 3, respectively, and 0.21, 0.41, and 0.63, respectively. These values were calculated similarly to the posterior probabilities described in Section 3.3.3. Furthermore, Bayes factors of 1.00, 1.24, and 1.77 for doses 1, 2, and 3, respectively, confirm little evidence of any dose effect or trend. Figure 3.3 portrays posterior probabilities of a biologically-important increase in mortality with increasing dose and time relative to the control. Using a reasonable cutoff of 1% to indicate biological significance (as opposed to statistical significance), we estimate the age- and dose-specific posterior probabilities of increases of at least this magnitude. Although such cutoffs are arbitrary, they are often used in risk assessment and toxicology (U.S. EPA, 1995). For adverse binary outcomes, such as death in our example, one avoids setting a conservative threshold. By choosing a 1% threshold, we combine those samples for which the difference in mortality relative to the control is greater than 0% but less than 1% in the same group as samples with a 0% difference. As expected, since we simulated under the null hypothesis, there is no stronger evidence of an

increase than no increase in mortality in any of the dose groups relative to control. These results parallel those from a frequentist log-rank test (p=0.6). However, our

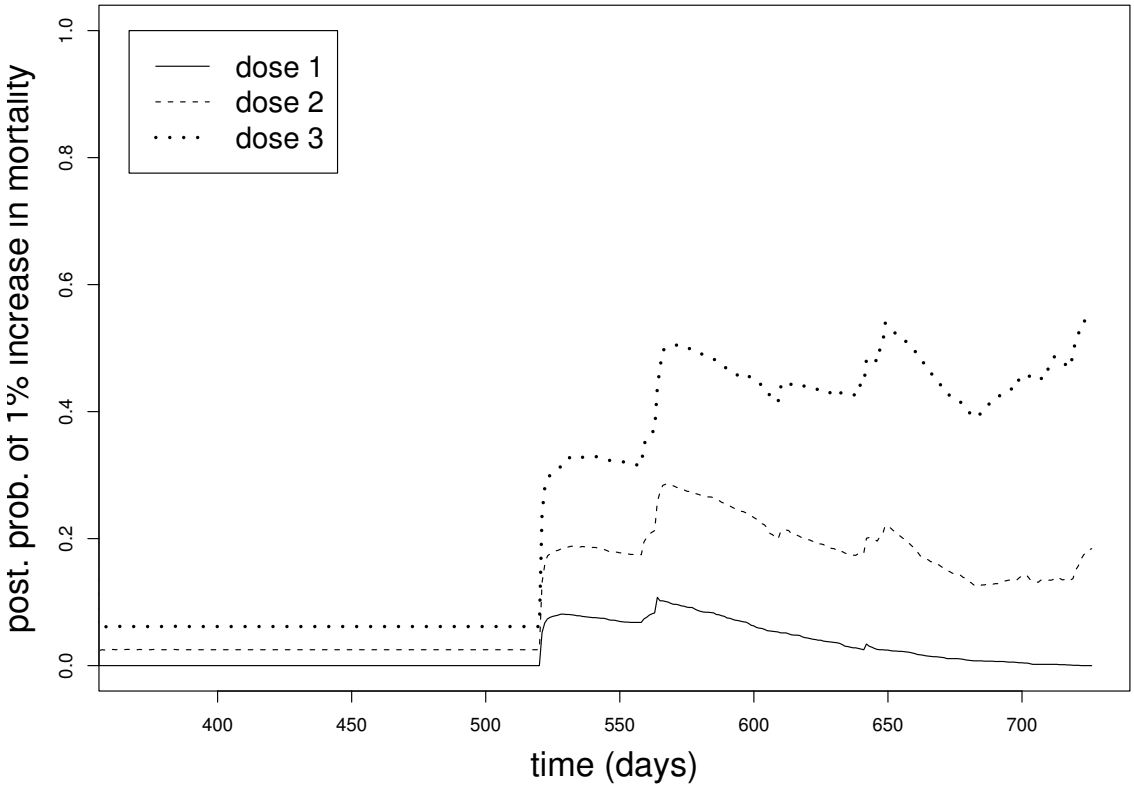

Figure 3.3: Posterior probabilities of a biologically-significant (i.e., 1%) increase in mortality for each dose x age combination

approach provides more realistic estimates through the smoothing and borrowing of information across levels of dose and time. The results suggest that our MCMC algorithm is specified appropriately, and that the prior does not lead to incorrect conclusions of differences when they do not occur.

3.4.4 Toxicology Study Results

We analyzed the real data using the same approach used for the simulated data, and again observed good rates of convergence and mixing based on sampled values that settled down throughout diagnostic plots. Posterior estimates of survival probabilities across levels of time and dose are presented in Figure 3.4. Although pointwise 95% credible intervals are not shown, the average reduction in the interval widths from the frequentist 95% confidence intervals is 0.18, similar to the 0.2 reduction under the simulation study. Figure 3.4 suggests that survival is similar in the 0, 75, and

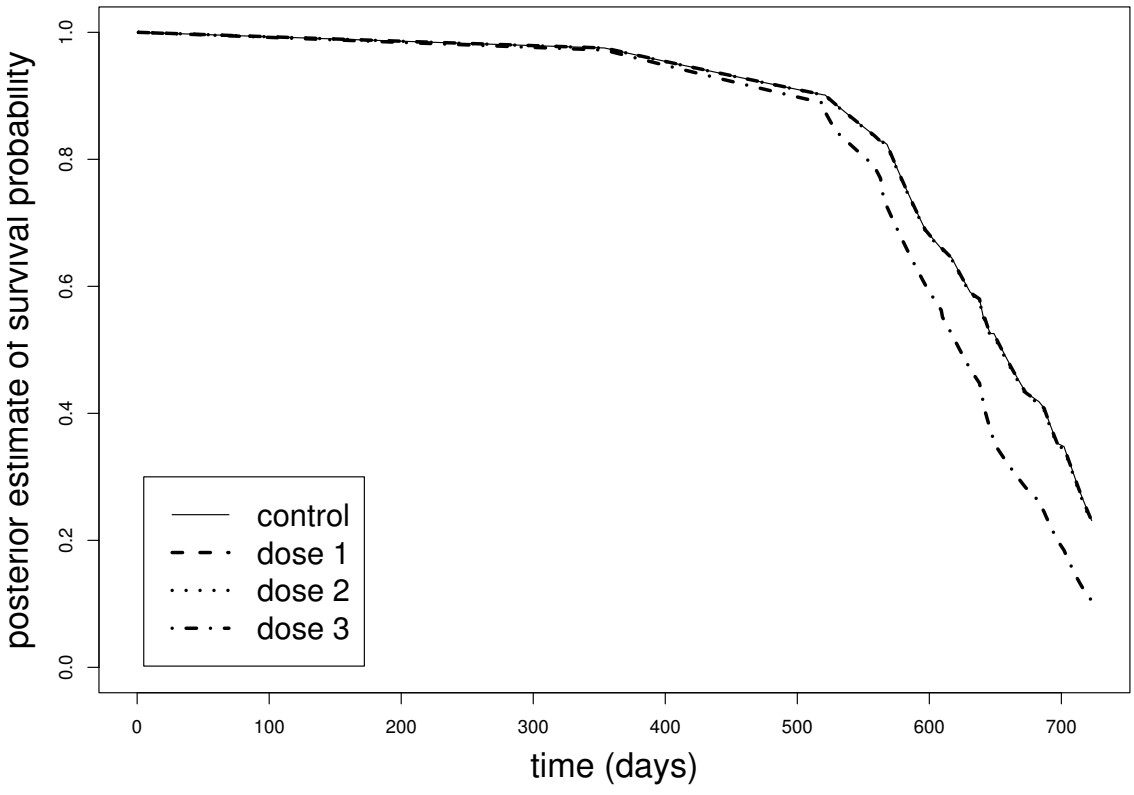

Figure 3.4: Toxicology results: posterior means of the survival probability for male rats exposed to ethylbenzene

250 ppm dose groups, but there may be a decrease in survival in the 750 ppm dose group. Our analysis shows clear evidence of an increase in mortality for rats in the 750 ppm dose group, which parallels results showing a significant change in survival with dose based on a frequentist log-rank test (p=0.001). Also notice how the curve for dose 3 diverges from the other groups shortly after day 500, corresponding to the age at which the Kaplan-Meier estimates begin to diverge (refer to Figure 3.1). Our estimates reveal an apparent lower effect for dose 3, with the survival probabilities for those in the control and doses 1 and 2 pooled close to one another such that their curves are essentially identical.

Figure 3.5 plots posterior probabilities of a 1% increase in mortality for each dose group relative to control, including results of the simulation example for comparison. The posterior probabilities are low for each dose group prior to one year of age when mortality is rare, but then increase dramatically for the high dose group at 521 days of age. This age corresponds to the time at which the Kaplan-Meier curve for dose group 3 starts to diverge from the other curves. By the end of the study at 2 years, there is strong evidence of higher mortality in the high dose group, as evidenced by high posterior probabilities of an overall increase in mortality (shown in Table 3.2). Table 3.2 also presents prior probabilities and Bayes factors. Since our prior is chosen to assign equal probability to the global null and alternative hypotheses, posterior samples of the hazard curves tend to congregate together except at locations of time and dose at which there is evidence in the data of a difference.

In addition to our primary analysis, we assessed whether our results are driven by our prior specification by conducting the analysis for lower and higher values of $c_{\pi k}$. Using values of 10 and 100, results did not change significantly. The estimated survival curves produced no change for different values of $c_{\pi k}$. The bottom two panels of Figure 3.5 show little variation between the posterior probabilities of a 1% increase

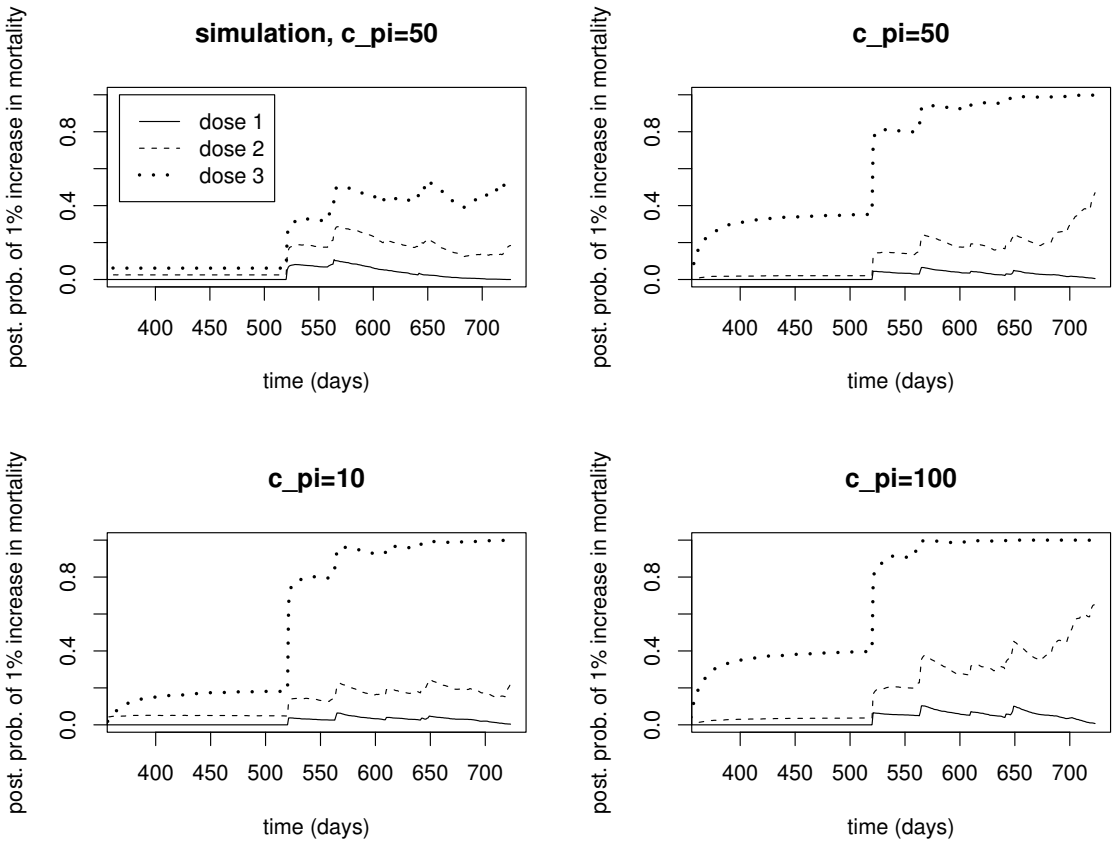

Figure 3.5: Posterior probabilities of a biologically-significant (i.e., 1%) increase in mortality for each dose x age combination (c_λ and $c_{\gamma^* k}$ are set at 50)

Table 3.2: Prior probabilities, posterior probabilities, and Bayes factors for overall effects across dose and time by values of $c_{\pi k}$

$c_{\pi k}$	*Prior*			*Posterior*			*Bayes factors*		
	Dose 1	Dose 2	Dose 3	Dose 1	Dose 2	Dose 3	Dose 1	Dose 2	Dose 3
10	0.19	0.35	0.48	0.20	0.38	0.99	1.07	1.14	> 100
50	0.21	0.36	0.49	0.25	0.48	0.99	1.25	1.64	> 100
100	0.21	0.36	0.49	0.29	0.65	0.99	1.54	3.30	> 100

in mortality across dose and time, compared to that of the primary analysis where $c_{\pi k} = 50$. Notice the difference in results in Table 3.2 and Figure 3.5 compared to

that of the simulation study under the null hypothesis. There clearly appears to be an overall dose effect in the toxicology example, with posterior probabilities close to 1 and Bayes factors greater than 100, compared to no evidence of an overall dose effect for the simulation study based on a posterior probability and Bayes factor of 0.63 and 1.77, respectively. Furthermore, we altered values of c_λ and $c_{\gamma^* k}$ to 10 and 100 and saw no significant changes in posterior estimates of survival curves, posterior probabilities of a change in hazards, or posterior probabilities of a biologically-important increase in mortality. These results indicate that our model and inferences are generally robust to the prior specification.

3.5 Discussion

This chapter proposes a useful and easy-to-implement Bayesian approach for inferences on ordering in hazard functions. Although we assume non-decreasing hazards with increasing dose, we account for flat regions of homogeneity in the hazards with increasing dose levels. In particular, our approach considers an alternative subhypothesis of non-decreasing hazards with increasing dose for subjects in consecutive dose groups compared to a local null of equality when subjects in dose group k are compared to group $k+1$. The Markov beta-gamma structure in our model allows for positive probability to be allocated to such null hypotheses of no effect within specific regions. It is straightforward to make inferences on global hypotheses for hazards across all dose levels. We also include autocorrelation in the model to not only borrow information across neighboring time intervals but to smooth the curve. Furthermore, the conditionally conjugate structure yields efficient posterior computation. It is trivial to find posterior probabilities and Bayes factors of local null and alternative hypotheses.

We consider this approach under the framework of an ordered constraint on the categorical covariate representing the difference in hazards between subjects in dose group k compared to group $k+1$, however there are numerous settings for which this methodology can be applied. We offer a generalization of our approach to the case with multiple ordinal predictors, as well as continuous covariates. In addition, we can extend the methods for inferences on dose response curves with downturns at higher doses where the peak location is unknown, or hazard functions with other shapes, including bathtub orderings.

Chapter 4

Conclusions and Future Work

4.1 Summary and Extensions

The methods developed throughout this dissertation are intended to build on and improve existing approaches for order restricted inference in Bayesian modeling. Since order constraints are appropriate to incorporate in many applications, we offer new techniques of embedding order restrictions in several modeling frameworks, with interest primarily devoted to curve estimation and trend testing. We especially focus on estimation and inference in complex problems involving correlated and censored data. Although we are interested in applications to biomedical examples, our purpose in constructing the methods within this dissertation is to provide a general framework for incorporating order restrictions in hierarchical and hazard models of various forms and throughout numerous applications.

The second chapter focuses on an approach to approximate Bayesian inferences by introducing a transformation strategy for restricting parameters to follow an umbrella ordering. The transformation approach maps draws from the unconstrained posterior density for the subject-specific parameters onto the restricted space, using

an isotonic transformation and minimizing the distance between the unconstrained and constrained estimates. Note that this approach effectively places an implicit prior on the parameters which has support on the restricted space. Since a constraint on the higher level parameters implies complex and dependent constraints on the lower level parameters, it is not straightforward, and hence can be difficult, to implement computation when a prior is explicitly chosen to have constrained support for higher level parameters in the hierarchy. Results from a simulation study and reproductive health data example show that bias introduced through the order constraint can be reduced by incorporating the transformation strategy proposed in Chapter 2, indicating that smoothing occurs in regions with limited data without smoothing out rapid changes in the trajectories. Results of the transformation approach also suggest an improvement in efficiency with over 90% of the restricted estimates' variance being smaller than that of the unrestricted estimates.

There are numerous settings, outside of hierarchical functional data applications, to which the methods in Chapter 2 can be applied. For example, the methods can be used to conduct inferences on dose response functions, even with possible downturns at higher doses, in an epidemiological or toxicological setting. Hazard functions with bathtub or hill shapes are another possible framework for which this approach can be applied. A major advantage to our method is the ease with which the approach can be applied to these other settings. In contrast to requiring new programming code for each new case under the explicit prior with constrained support scenario, it is trivial to take output of the existing MCMC chain for the unrestricted case (using WinBUGS, for example) and transform these estimates using a short S-Plus or SAS function. This is an extremely important consideration, since computational hurdles are one of the major factors preventing the widespread use of order and shape restrictions in analyses of biomedical data.

There are additional extensions to methods described in Chapter 2. Covariates could be included in the model, which offer key information regarding the underlying biological relationship or differences between groups of subjects. For example, timing of sexual activity as a predictor could provide insight into differences between conception and non-conception groups. Furthermore, reproductive epidemiologists may be interested in extending the methodology to the case with multiple peaks, such as that found in estrogen. Additional shapes motivated by biology could also serve as possible inclusions into the model.

The third chapter proposes a Bayesian approach for conducting inferences on orderings in hazard functions. Interest focuses on comparing hypotheses of non-decreasing hazards to regions of equality in the hazards as the dose level increases. In particular, we are interested in identifying dose and age values at which an increase in hazards exists. By considering homogeneity in the hazards across dose groups, age-dependent null hypotheses arise. It is reasonable to assume that hypotheses located near one another are strongly correlated *a priori*, producing smooth model probabilities. Our model also incorporates smoothing of the baseline hazard and time-varying coefficients, which are all modeled using some form of Markov gamma or beta process prior.

We introduce a novel approach by coupling Markov beta and gamma processes in order to smooth the time-dependent hypothesis probabilities and coefficients subject to the order constraint. The beta component characterizes changes with time in the probability that the hazards are equal, while the gamma component describes changes in a time-dependent coefficient occurring when the alternative hypothesis holds. This point mass mixture prior allows positive probability to be allocated to flat regions in which no change in hazard occurs with increasing dose. Furthermore, one of the major advantages to this approach is that conjugacy properties are maintained

through a data augmentation scheme, which facilitates posterior computation and inference using a Gibbs sampling algorithm.

Although we focus on a single ordinal covariate in Chapter 3, we offer several extensions to the methodology. In Section 3.2.3, we provide an outline for incorporating multiple ordinal predictors and continuous covariates for assessing the effect of these predictors on the survival time, with possible time dependencies on the covariates as well potentially including time-varying coefficients. The approach in Chapter 3 can also be used in modeling dose response curves with downturns occurring at high doses in which the peak location is random, similar to that found in Chapter 2.

Since results of many bioassay experiments show that dose response curves are often isotonic, with possible downturns at high doses, it is of interest to epidemiologists and toxicologists to modify the approach in Chapter 3 to incorporate umbrella orderings. Under umbrella ordering, the null hypothesis would be defined subject to homogeneity in the hazards:

$$H_{0,umbrella} : \lambda_1 = \cdots = \lambda_\kappa = \cdots = \lambda_d, \tag{4.1}$$

and the alternative hypothesis would be expressed as the union over κ of the following subsets:

$$H_{1,umbrella} : \lambda_1 \leq \cdots \leq \lambda_\kappa \geq \cdots \geq \lambda_d, \qquad \text{for } \kappa \neq 1, d, \tag{4.2}$$

with at least one strict inequality in the hazards and where κ is the unknown peak parameter. If $\kappa = 1$ or d, then a non-increasing or non-decreasing order, respectively, results. From here, we could place a multinomial prior on the peak location, κ. If the vector of probabilities, corresponding to that in the multinomial distribution, consists of all zeroes, with the last element equal to one, then the result is a simple increasing order, just as with the focus in Chapter 3. In contrast, if the probability vector consists of all zeroes, however the first element is equal to one, then a simple

decreasing order results. Under this scenario, the alternative hypothesis incorporates both orderings through:

$$H_{1,both} : \lambda_1 \leq \cdots \leq \lambda_d \quad \text{or} \quad \lambda_1 \geq \cdots \geq \lambda_d, \tag{4.3}$$

with at least one strict inequality.

4.2 Mixture Priors for Order Restricted Inference

4.2.1 Motivation and Overview

Parameter constraints in biomedical applications can be incorporated into Bayesian models through prior distributions with restricted support. For example, let θ_j be a parameter which is specific to age group j $(j = 1, \ldots, p)$, under the non-decreasing order constraint:

$$\theta_1 \leq \theta_2 \leq ... \leq \theta_p, \tag{4.4}$$

where p may be moderately large. The standard approach would be to choose normal prior distributions for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ subject to restriction (4.4). However, a major problem with Bayesian estimation using traditional truncated conjugate priors and uniform improper priors for placing increasing constraints on parameters is that probability is not allowed to accrue close to the boundary of the constrained region, which can produce substantial bias in the estimates (Gelman, 1996; van Onna, 2002). In particular, when there is a moderate to high dimensional parameter subject to the ordering in (4.4) and one uses a uniform improper prior as suggested by Gelfand, Smith, and Lee (1992), we observe striking degrees of bias in posterior means. In addition, 95% credible intervals will not even include the true parameter values, in most cases. Motivated by this problem, we consider methods for order restricted inference in cases involving shrinkage-type prior distributions.

In contrast to Bayesian methods, classical estimation subject to order restrictions can be accomplished by calculating RMLEs, as we have discussed throughout this dissertation. When RMLEs are difficult to compute for more complex models, restricted estimates can be found using Bayesian machinery with a type of simulated annealing, and where the RMLEs are obtained in the limit of a sequence of Bayes estimators (Robert and Hwang, 1996). In addition to the vast literature on standard Bayesian procedures for conducting order restricted inference that was discussed throughout this dissertation (Gelfand and Kuo, 1991; Ramgopal, Laud, and Smith, 1993; Lavine and Mockus, 1995; among others), Albert (1994) and McDonald and Prevost (1997) also contributed to such procedures. However, the straightforward and easy-to-implement routine of Gelfand, Smith, and Lee (1992) incorporates parameter constraints using Markov chain Monte Carlo (MCMC) analyses of Bayesian models. Unfortunately, such standard, 'noninformative' uniform priors often result in strong prior belief that the true function under consideration is higher than it actually is, producing substantial bias in the estimates (Gelman, 1996; van Onna, 2002; Dunson and Neelon, 2003).

This issue can be partially addressed by choosing a prior incorporating a probability mass on the boundary (Geweke, 1996; Dunson and Herring, 2003; 2004; Neelon and Dunson, 2004), resulting in a type of shrinkage estimator. Shrinkage estimators are used to improve estimation under quadratic loss. Sengupta and Sen (1991) and Sen and Sengupta (1991) showed that their shrinkage estimators subject to an order-restricted space for the normal means case dominated (e.g., produced smaller risk than) MLEs and RMLEs. In fact, Iliopoulos (2000) showed an improvement in risk exceeding 50% near the boundary of the restricted space by using a shrinkage estimator. Ouassou and Strawderman (2002) and Fourdrinier, Ouassou and Strawderman (2003) proposed a new class of shrinkage estimators, which dominate standard MLEs

under quadratic loss, by applying Stein-type shrinkage factors directly to MLEs. In general, when interest focuses on inference on a null hypothesis of equalities in parameters versus one or more order restricted alternative hypotheses, the point mass mixture priors have some appealing properties. However, this does not entirely solve the problem, since some view point nulls as artificial because exact equalities are unlikely.

For illustrative purposes, consider the model with a regression parameter equal to zero, which is practically identical to the model with the parameter equal to some extremely, small positive value. The point mass mixture approach treats these two cases differently. However, to account for the similarity in these two types of models, it is preferable to consider null hypotheses that the parameters are within $\pm\ \epsilon$, or some small value, of zero which can be chosen as the minimal value such that the difference is scientifically relevant.

We propose such an alternative approach for Bayesian inferences on uncertain orderings, which avoids the point null representation. Following a similar prior structure as George and McCulloch (1993) use for variable selection in linear models, this mixture prior consists of a smooth component with support within a small value of zero (with zero denoting equality) and a second component with support to the right of this region. In particular, this approach limits the problems we have observed in biased estimation under restricted priors by allowing probability to congregate on values close to equalities. Based on similar criteria used in Sengupta and Sen (1991) and Fourdrinier, Ouassou, and Strawderman (2003), we can assess the performance of the Bayes estimator under our shrinkage prior.

4.2.2 Priors for Increasing Normal Means

For simplicity, we focus on the case in which data consist of n_j independent normal samples for groups $j = 1, \ldots, p$, so that we have $y_{ij} \sim \mathcal{N}(\theta_j, \sigma^2)$ for $i = 1, \ldots, n_j$. The mean vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ is subject to a non-decreasing constraint: $\boldsymbol{\theta} \in \boldsymbol{\Theta}_+$, where $\boldsymbol{\Theta}_+ = \{\theta : \theta_1, \ldots, \theta_p\}$. By reparameterizing the model to let $\theta_j = \alpha + \sum_{h=1}^{j-1} \beta_h$, for $j = 1, \ldots, p$, then the restriction $\beta_j \geq 0$ for $j = 1, \ldots, p-1$ implies that $\boldsymbol{\theta} \in \boldsymbol{\Theta}_+$. We can assign a uniform improper prior for the baseline, α, for simplicity. However, our main focus is on choosing a prior for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{p-1})'$.

Several prior distributions have been considered for parameters subject to order constraints. Focusing on independence priors for $\boldsymbol{\beta}$ for simplicity, interest has primarily been on incorporating one of the three following priors:

1. Uniform improper on the restricted space:

$$\pi(\boldsymbol{\beta}) \propto 1(\beta_j > 0, j = 1, \ldots, p-1),$$

2. Truncated normal:

$$\pi(\boldsymbol{\beta}) = \prod_{j=1}^{p-1} \mathrm{N}_+(\beta_j; 0, s^2),$$

3. Zero-inflated truncated normal:

$$\pi(\boldsymbol{\beta}) = \prod_{j=1}^{p-1} \Big\{ \pi\delta_0(\cdot) + (1-\pi)\mathrm{N}_+(\beta_j; 0, s^2) \Big\},$$

where s^2 is the variance of the normal distributions prior to truncation, π is the probability of being in the point mass, while $\delta_0(\cdot)$ is an indicator of being in the point mass, which is zero in this scenario.

Gelman (1996) and van Onna (2002) showed that the uniform improper and truncated priors represented by priors 1 and 2 above produce substantially biased

estimates, resulting in unreasonable choices for prior specification. For prior choice 3, consider the case of equality in the true parameters. We can choose a prior which is the mixture of a point mass, to allow equalities, and truncated conjugate priors. Berger and Delampady (1987) discuss how testing point nulls is a reasonable approximation to an analysis considering the point null as 'almost true.' For example, one may test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ and define $\Omega = \{\mu : |\mu - \mu_0| \leq \epsilon\}$, for some small ϵ. Although Berger and Delampady (1987) do not apply asymptotics to their proof, they show that it is reasonable to consider the point null as an approximation to H_0 when ϵ is at most half a sample standard deviation in width. When point nulls are used, however, posterior samples will live part of the time in the point mass and part of the time on really small positive values as the sample size increases. For finite samples and highly constrained models, the bias can be substantial. Hence, even asymptotically, the posterior mean will be a positively biased estimator, leading many researchers to think that point null hypotheses representing exact equalities in the parameters are unreasonable.

An alternative prior specification that allows probability to accrue close to 0 is preferable. We propose a new mixture prior:

$$\pi(\boldsymbol{\beta}) = \prod_{j=1}^{p-1} \left\{ \pi \mathrm{N}_{[-\delta,\delta]}(\beta_j; 0, \delta^2) + (1-\pi)\mathrm{N}_{(\delta,\infty)}(\beta_j; 0, \delta^2/\phi) \right\},$$

where $\mathrm{N}_A(\cdot, \mu, \sigma^2)$ denotes the $\mathrm{N}(\mu, \sigma^2)$ density truncated to the region $A \subset \Re$, δ is a small positive constant controlling the spread of values that β_j can take under the null hypothesis, and ϕ is a scale factor to allow a different degree of shrinkage under the alternative hypothesis, typically chosen to be less than one. Although this specification does not strictly enforce the constraint in that negative values close to zero are admitted, it does allow probability to accrue close to 0. This prior specification addresses the concern that point null hypotheses are unreasonable.

The mixture prior is closely related to variable selection priors of George and McCulloch (1993). Although they do not consider order constraints, they use a mixture of a low variance normal distribution centered at 0 and a high variance normal distribution to conduct variable selection in linear regression. Their framework is less complex than other models with which our mixture prior can accommodate, such as hazard models.

This approach has some appealing theoretical properties. In particular, it ensures that as the sample size increases, the estimated posterior probability of the null hypothesis will converge appropriately to one when the null hypothesis is true. In contrast, the point mass mixture approach has the unappealing property that the posterior probability of the null hypothesis will tend to converge to 0.5 under the null, with half of the probability assigned to values that are very close but not exact equalities. Although these results are based on simulation studies using normal priors, it would be interesting to determine if these statements hold for different choices of prior distributions.

We can consider inference on the posterior distribution of the difference between θ_p and θ_1. If the θ's are independent and normally distributed, then posterior computation of this difference is straightforward. However, with the dependency on a sum of truncated normal densities, computation of the difference is not clear. As a result, we turn to MCMC sampling methods of obtaining posterior draws from the distribution of the difference between θ_p and θ_1. We can perform simulation studies to discover what happens to the posterior as p and s^2 increase. It would also be interesting to consider what happens when the true difference in θ_p and θ_1 is zero, a small positive value, or a large positive value.

Using a quadratic loss function, it would be nice to be able to show lower Bayes risk for the proposed mixture prior compared with the previous three priors, or the

unrestricted MLE. This may be possible analytically for a simple case, however we may need to rely on simulation studies for the multivariate case. We can also run simulation studies comparing the performance under a variety of scenarios. Computing the Bayes risk is the alternative to calculating properties of such frequentist shrinkage estimators for order restricted inference discussed in Section 4.2.1 (Sen and Sengupta, 1991; Sengupta and Sen, 1991; Iliopoulos, 2000; Kuriki and Takemura, 2000; Ouassou and Strawderman, 2002; Fourdrinier, Ouassou and Strawderman, 2003). While the literature focuses on properties of a point estimator, we are interested in the general case of estimation and inferences for data having arbitrary distributions and for arbitrary components in a model.

Appendix A

Outline of Proof for Theorem 1

For p=3, there are six possible inequalities in the means:

$$\begin{array}{ll} 1)\ \mu_1 \leq \mu_2 \leq \mu_3 & 2)\ \mu_1 \geq \mu_2 \geq \mu_3 \\ 3)\ \mu_1 \leq \mu_3 \leq \mu_2 & 4)\ \mu_1 \geq \mu_3 \geq \mu_2 \\ 5)\ \mu_3 \leq \mu_1 \leq \mu_2 & 6)\ \mu_3 \geq \mu_1 \geq \mu_2. \end{array}$$

In general, the values of the constrained parameters given a peak at $\kappa = 1$ are defined as follows:

$$\mu_1^{*1} = \max\left\{\mu_1, \frac{\mu_1\sigma_{11} + \mu_2\sigma_{22}}{\sigma_{11} + \sigma_{22}}, \frac{\mu_1\sigma_{11} + \mu_2\sigma_{22} + \mu_3\sigma_{33}}{\sigma_{11} + \sigma_{22} + \sigma_{33}}\right\}$$

$$\mu_2^{*1} = \min\left\{\max\Big(\frac{\mu_1\sigma_{11} + \mu_2\sigma_{22}}{\sigma_{11} + \sigma_{22}}, \frac{\mu_1\sigma_{11} + \mu_2\sigma_{22} + \mu_3\sigma_{33}}{\sigma_{11} + \sigma_{22} + \sigma_{33}}\Big), \max\Big(\mu_2, \frac{\mu_2\sigma_{22} + \mu_3\sigma_{33}}{\sigma_{22} + \sigma_{33}}\Big)\right\}$$

$$\mu_3^{*1} = \min\left\{\frac{\mu_1\sigma_{11} + \mu_2\sigma_{22} + \mu_3\sigma_{33}}{\sigma_{11} + \sigma_{22} + \sigma_{33}}, \frac{\mu_2\sigma_{22} + \mu_3\sigma_{33}}{\sigma_{22} + \sigma_{33}}, \mu_3\right\},$$

where σ_{hl} are the elements of the unconstrained posterior covariance, $\boldsymbol{\Sigma}_{\boldsymbol{\mu}|\sigma,\boldsymbol{y}}$, for $h, l = 1, \ldots, p$. Next, $\boldsymbol{\mu}^{*1}$ is computed using the above results for each of the six inequalities, respectively:

1) $\boldsymbol{\mu}^{*1} = \left\{\frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}}, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}}, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}}\right\}$

2) $\boldsymbol{\mu}^{*1} = \left\{ \mu_1, \mu_2, \mu_3 \right\}$

3) $\boldsymbol{\mu}^{*1} = \left\{ \mu_1, \min\left[\max\left(\frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}}{\sigma_{11}+\sigma_{22}}, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}} \right), \frac{\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{22}+\sigma_{33}} \right], \frac{\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{22}+\sigma_{33}} \right\}$

4) $\boldsymbol{\mu}^{*1} = \left\{ \max\left(\frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}}{\sigma_{11}+\sigma_{22}}, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}} \right), \max\left(\frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}}{\sigma_{11}+\sigma_{22}}, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}} \right), \right.$

$\left. \min\left(\mu_3, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}} \right) \right\}$

5) $\boldsymbol{\mu}^{*1} = \left\{ \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}}{\sigma_{11}+\sigma_{22}}, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}}{\sigma_{11}+\sigma_{22}}, \mu_3 \right\}$

6) $\boldsymbol{\mu}^{*1} = \left\{ \max\left(\mu_1, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}} \right), \max\left(\frac{\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{22}+\sigma_{33}}, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}} \right), \right.$

$\left. \min\left(\frac{\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{22}+\sigma_{33}}, \frac{\mu_1\sigma_{11}+\mu_2\sigma_{22}+\mu_3\sigma_{33}}{\sigma_{11}+\sigma_{22}+\sigma_{33}} \right) \right\}$

This procedure allows us to find the inequality for which the Mahalanobis distance (2.3) is minimized, hence providing an ordering in the restricted estimates, $\boldsymbol{\mu}^*$. For the ordering in which the peak occurs at 1, this process shows the loss corresponding to the second inequality, $\mu_1 \geq \mu_2 \geq \mu_3$, is zero. No other loss pertaining to the remaining inequalities under the ordering $\mu_1 \geq \mu_2 \geq \mu_3$ in the unconstrained estimates has a value this small. Comparing the distances to find the one that satisfies expression (2.3) is not always as simple as with this ordering in the unconstrained estimates specified here. When the ordering in the unrestricted estimates is different from the one specified here, not all of the inequalities yielding the smallest value of (2.3) attains a loss of zero. Nonetheless, computing (2.3) for each of the orderings in the unconstrained estimates will show that the ordering in $\boldsymbol{\mu}^*$ is equivalent to the ordering in $\boldsymbol{\mu}$ for $p \leq 3$.

Appendix B

Posterior Distributions & Derivations for Chapter 3

Step 1: The full conditional distribution of $(dN_{ij0}, dN_{ij1}, \ldots, dN_{ij,d-1})'$, for $i, j : Y_{ij} = 1$, is proportional to:

$$\mathcal{M}ultinom\ (dN_{ij}; P_{ij0}, P_{ijk}, k = 1, \ldots, d-1), \tag{B.1}$$

where $P_{ij0} = \frac{d\Lambda_{0j}}{d\Lambda_{0j} + \sum_{k:\delta_{kj}=0}^{d-1} d\Gamma_{jk}}$ and $P_{ijk} = \frac{d\Gamma_{jk}}{d\Lambda_{0j} + \sum_{k:\delta_{kj}=0}^{d-1} d\Gamma_{jk}}$.

Step 2: The full conditional distribution of $\{w_l\}$, for $l = 1, \ldots, L$ and $w_l = 0, 1, 2, \ldots,$ is proportional to:

$$\begin{aligned} & \frac{(c_\lambda \lambda_{0l-1})^{w_l}}{w_l!} \frac{(b_\lambda + c_\lambda)^{w_l}}{\Gamma(a_\lambda + w_l)} \lambda_{0l}^{w_l} \\ \propto\ & \frac{\{c_\lambda (b_\lambda + c_\lambda) \lambda_{0l-1} \lambda_{0l}\}^{w_l}}{\Gamma(w_l + 1)\Gamma(a_\lambda + w_l)}. \end{aligned} \tag{B.2}$$

Furthermore, the gamma full conditional of λ_{0l}, for $l = 1, \ldots, L$, is proportional to:

$$\begin{aligned}
&\left\{ \prod_{i,j:Y_{ij}=1} d\Lambda_{0j}^{dN_{ij0}} \exp(-d\Lambda_{0j}) \right\} \lambda_{0l}^{a_\lambda + w_l + w_{l+1} - 1} \exp\{-\lambda_{0l}(b_\lambda + 2c_\lambda)\} \\
\propto \quad &\left[\prod_{i,j:Y_{ij}=1} \left\{ \sum_{l=1}^{L} \lambda_{0l}(t_j - t_{j-1}) 1_{\{t_j \in (\tau_{l-1}, \tau_l]\}} \right\}^{dN_{ij0}} \right. \\
&\left. \times \exp\left\{ - \sum_{l=1}^{L} \lambda_{0l}(t_j - t_{j-1}) 1_{\{t_j \in (\tau_{l-1}, \tau_l]\}} \right\} \right] \lambda_{0l}^{a_\lambda + w_l + w_{l+1} - 1} \exp\{-\lambda_{0l}(b_\lambda + 2c_\lambda)\} \\
\propto \quad &\lambda_{0l}^{\sum_{i,j:Y_{ij}=1; t_j \in (\tau_{l-1}, \tau_l]} dN_{ij0}} \exp\left\{ -\lambda_{0l} \sum_{i,j:t_j \in (\tau_{l-1}, \tau_l]} Y_{ij}(t_j - t_{j-1}) \right\} \lambda_{0l}^{a_\lambda + w_l + w_{l+1} - 1} \\
&\times \exp\{-\lambda_{0l}(b_\lambda + 2c_\lambda)\}
\end{aligned}$$

The above expression can be expressed as posterior density (3.17).

Step 3: The full conditional distribution of $\{u_{kl}\}$, $l = 1, \ldots, L$, $k = 1, \ldots, d-1$, and $u_l = 0, \ldots, c_{\pi k}$, is proportional to:

$$\begin{aligned}
&\frac{\pi_{kl-1}^{u_{kl}} (1 - \pi_{kl-1})^{-u_{kl}}}{(c_{\pi k} - u_{kl})! u_{kl}!} \frac{\pi_{kl}^{u_{kl}} (1 - \pi_{kl})^{-u_{kl}}}{\Gamma(a_\pi + u_{kl}) \Gamma(b_\pi + c_{\pi k} - u_{kl})} \\
\propto \quad &\frac{(\pi_{kl-1} \pi_{kl})^{u_{kl}} \{(1 - \pi_{kl-1})(1 - \pi_{kl})\}^{-u_{kl}}}{\Gamma(c_{\pi k} - u_{kl} + 1) \Gamma(u_{kl} + 1) \Gamma(a_\pi + u_{kl}) \Gamma(b_\pi + c_{\pi k} - u_{kl})}. \qquad \text{(B.3)}
\end{aligned}$$

In addition, the π_{kl}'s have a beta full conditional posterior distribution, for $l = 1, \ldots, L$ and $k = 1, \ldots, d-1$:

$$\begin{aligned}
&\pi_{kl}^{u_{kl+1}} (1 - \pi_{kl})^{c_{\pi k} - u_{kl+1}} \pi_{kl}^{a_\pi + u_{kl} - 1} (1 - \pi_{kl})^{b_\pi + c_{\pi k} - u_{kl} - 1} \\
&\times \left\{ 1_{\{\gamma_{kl} = 0\}} \pi_{kl} + 1_{\{\gamma_{kl} > 0\}} (1 - \pi_{kl}) \right\} \\
\propto \quad &\mathcal{B}e\left(\pi_{kl}; a_\pi + u_{kl} + u_{kl+1} + 1_{\{\gamma_{kl} = 0\}}, b_\pi + 2c_{\pi k} - u_{kl} - u_{kl+1} + 1_{\{\gamma_{kl} > 0\}} \right) \qquad \text{(B.4)}
\end{aligned}$$

Steps 4 & 6: The full conditional of $\{z_{kl}\}$, for $l = 1, \ldots, L$ and $k = 1, \ldots, d-1$, is

proportional to:

$$\begin{aligned} & \frac{(c_{\gamma^* k}\gamma^*_{kl-1})^{z_{kl}}}{z_{kl}!}\frac{(b_{\gamma^* k}+c_{\gamma^* k})^{z_{kl}}}{\Gamma(a_{\gamma^* k}+z_{kl})}(\gamma^*_{kl})^{z_{kl}} \\ \propto \quad & \frac{\left\{c_{\gamma^* k}(b_{\gamma^* k}+c_{\gamma^* k})\gamma^*_{kl-1}\gamma^*_{kl}\right\}^{z_{kl}}}{\Gamma(z_{kl}+1)\Gamma(a_{\gamma^* k}+z_{kl})}, \end{aligned} \tag{B.5}$$

with the gamma full conditional distribution for γ^*_{kl}, for $k = 1, \ldots, d-1$ and $l = 1, \ldots, L$, as:

$$\begin{aligned} & (\gamma^*_{kl})^{a_{\gamma^*}+z_{kl}-1}\exp\{-\gamma^*_{kl}(b_{\gamma^*}+c_{\gamma^*})\}(\gamma^*_{kl})^{z_{kl+1}}\exp(-c_{\gamma^*}\gamma^*_{kl}) \\ \propto \quad & \mathcal{G}\left(\gamma^*_{kl}; a_{\gamma^*}+z_{kl}+z_{kl+1}, b_{\gamma^*}+2c_{\gamma^*}\right). \end{aligned} \tag{B.6}$$

Although the parameters of the above gamma distribution are the same as the parameters of the gamma component in the $\delta_0 - \mathcal{G}$ distribution when $\gamma_{kl} > 0$, we show a formal way of deriving (3.18) as the full conditional distribution of γ_{kl}:

$$\begin{aligned} & \left[1_{\{\gamma_{kl}=0\}}\pi_{kl} + 1_{\{\gamma_{kl}>0\}}(1-\pi_{kl})\mathcal{G}(\gamma_{kl}; a_{kl}, b_{kl})\right] \\ & \times \prod_{i,j:Y_{ij}=1}\left[\left\{\sum_{l=1}^{L}\gamma_{kl}(t_j - t_{j-1})1_{\{t_j\in(\tau_{l-1},\tau_l]\}}\right\}^{1_{\{x_i>k\}}dN_{ijk}}\right. \\ & \left.\times\exp\left\{-\sum_{l=1}^{L}\gamma_{kl}(t_j - t_{j-1})1_{\{t_j\in(\tau_{l-1},\tau_l]\}}\right\}\right]. \end{aligned}$$

This expression is simplified to the form:

$$\begin{aligned} & \left[1_{\{\gamma_{kl}=0\}}\pi_{kl} + 1_{\{\gamma_{kl}>0\}}(1-\pi_{kl})C(\gamma_{kl}; a_{kl}, b_{kl})\right] \\ & \times\gamma_{kl}^{a_{kl}+\sum_{i,j=1,t_j\in(\tau_{l-1},\tau_l]}1_{\{x_i>k\}}Y_{ij}dN_{ijk}-1} \\ & \times\exp\left[-\gamma_{kl}\{b_{kl} + \sum_{i,j:t_j\in(\tau_{l-1},\tau_l]}1_{\{x_i>k\}}Y_{ij}(t_j - t_{j-1})\}\right], \end{aligned}$$

where $C(\cdot; a, b)$ is the constant term of the $\mathcal{G}(\cdot; a, b)$ density. Furthermore, the second line in the above expression can also be written as $\mathcal{G}(\gamma_{kl}; \tilde{a}_{kl}, \tilde{b}_{kl})/C(\tilde{a}_{kl}, \tilde{b}_{kl})$, where $\tilde{a}_{kl}$ and $\tilde{b}_{kl}$ are defined in (3.20). After dividing by the normalizing constant, the conditional posterior density of γ_{kl} is:

$$1_{\{\gamma_{kl}=0\}}\tilde{\pi}_{kl} + 1_{\{\gamma_{kl}>0\}}(1 - \tilde{\pi_{kl}})\mathcal{G}(\gamma_{kl}; \tilde{a}_{kl}, \tilde{b}_{kl}), \tag{B.7}$$

where $\tilde{\pi}_{kl}$ is defined in expression (3.19).

References

Albert, J.H. (1994) A bayesian approach to estimation of gpas of university of iowa freshmen under order restrictions. *Journal of Educational Statistics*, **19,** 1–21.

Andersen, P.K. and Gill, R.D. (1982) Cox regression-model for counting-processes - a large sample study. *Annals of Statistics*, **10,** 1100–1120.

Arjas, E. and Gasbarra, D. (1994) Nonparametric bayesian inference from right censored survival data using the gibbs sampler. *Statistica Sinica*, **4,** 505–524.

Arjas, E. and Gasbarra, D. (1996) Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association*, **91,** 1101–1109.

Baird, D. D., Weinberg, C. R., Zhou, H., Kamel, F., McConnaughey, D. R., Kesner, J. S. and Wilcox, A. J. (1999) Preimplantation urinary hormone profiles and the probability of conception in healthy women. *Fertility and Sterility*, **71,** 40–49.

Baird, D. D., Wilcox, A. J., Weinberg, C. R., Kamel, F., McConnaughey, D. R., Musey, P. I. and Collins, D. C. (1997) Preimplantation hormonal differences between the conception and non-conception menstrual cycles of 32 normal women. *Human Reproduction*, **12,** 2607–2613.

Berger, J.O. and Delampady, M. (1987) Testing precise hypotheses. *Statistical Science*, **2,** 317–352.

Brumback, B. A. and Rice, J. A. (1998) Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, **93,** 961–976.

Carlin, B.P., Gelfand, A.E. and Smith, A.F.M. (1992) Hierarchical bayesian analysis of changepoint problems. *Applied Statistics*, **41,** 389–405.

Casella, G. and Berger, R.L. (1990) *Statistical Inference*. Belmont, CA: Duxbury.

Casella, G. and George, E.I. (1992) Explaining the gibbs sampler. *American Statistician*, **46,** 167–174.

Chib, S. and Greenberg, E. (1995) Understanding the metropolis-hastings algorithm. *American Statistician*, **49,** 327–335.

Chipman, H.A., George, E.I. and McCulloch, R.E. (2001) The practical implementation of bayesian model selection. In *Model Selection*, IMS Lecture Notes - Monograph Series, vol. 38, pp. 67–116.

Clayton, D.G. (1991) A monte-carlo method for bayesian-inference in frailty models. *Biometrics*, **47,** 467–485.

Clayton, D.G. (1994) Bayesian analysis of frailty models. Technical Report. Medical Research Council Biostatistics Unit, Cambridge, U.K.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998) Automatic bayesian curve fitting. *Journal of the Royal Statistical Society* B, **60,** 333–350.

Dunson, D. B. and Colombo, B. (2003) Bayesian modeling of markers of day-specific fertility. *Journal of the American Statistical Association*, **98,** 28–37.

Dunson, D. B. and Neelon, B. (2003) Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics*, **59,** 286–295.

Dunson, D.B. and Herring, A.H. (2003) Bayesian inferences in the cox model for order restricted hypotheses. *Biometrics*, **59,** 916–923.

Dunson, D.B. and Herring, A.H. (2004) Bayesian model selection and averaging in additive and proportional hazards models. Institute of Statistics & Decision Sciences discussion paper, Duke University (available for download at $http://ftp.isds.duke.edu/WorkingPapers/04-16.pdf$).

Dykstra, R.L. (1982) Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association*, **77,** 621–628.

Dykstra, R.L. and Feltz, C.J. (1989) Nonparametric maximum likelihood estimation of survival functions with a general stochastic ordering and its dual. *Biometrika*, **76,** 331–341.

Dykstra, R.L., Kochar, S. and Robertson, T. (1991) Statistical inference for uniform stochastic ordering in several populations. *Annals of Statistics*, **19,** 870–888.

Dykstra, R.L., Kochar, S. and Robertson, T. (1995) Inference for likelihood ratio ordering in the two-sample problem. *Journal of the American Statistical Association*, **90,** 1034–1040.

Feltz, C.J. and Dykstra, R.L. (1985) Maximum likelihood estimation of the survival functions of n stochastically ordered random variables. *Journal of the American Statistical Association*, **80,** 1012–1019.

Fourdrinier, D., Ouassou, I. and Strawderman, W.E. (2003) Estimation of a parameter vector when some components are restricted. *Journal of Multivariate Analysis*, **86,** 14–27.

Gamerman, D. (1991) Dynamic bayesian models for survival data. *Applied Statistics*, **40,** 63–79.

Gelfand, A. E. and Kuo, L. (1991) Nonparametric bayesian bioassay including ordered polytomous response. *Biometrika*, **78,** 657–666.

Gelfand, A. E., Smith, A. F. M. and Lee, T. M. (1992) Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, **87,** 523–532.

Gelfand, A.E. and Kottas, A. (2001) Nonparametric bayesian modeling for stochastic order. *Annals of the Institute of Statistical Mathematics*, **53,** 865–876.

Gelfand, A.E. and Mallick, B.K. (1995) Bayesian analysis of proportional hazards models built from monotone functions. *Biometrics*, **51,** 843–852.

Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85,** 398–409.

Gelman, A. (1996) Bayesian model-building by pure thought: some principles and examples. *Statistica Sinica*, **6,** 215–232.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995) *Bayesian Data Analysis*. London: Chapman & Hall.

George, E.I. and McCulloch, R.E. (1993) Variable selection via gibbs sampling. *Journal of the American Statistical Association*, **88,** 881–889.

George, E.I. and McCulloch, R.E. (1997) Approaches for bayesian variable selection. *Statistica Sinica*, **7,** 339–373.

Geweke, J. (1996) Variable selection and model comparison in regression. In *Bayesian Statistics 5* (eds J.M. Bernado, J.O. Berger, A.P. Dawid and A.F.M. Smith), pp. 609–620. Oxford, UK: Claredon Press.

Gönen, M., Westfall, P.H. and Johnson, W.O. (2003) Bayesian multiple testing for two-sample multivariate endpoints. *Biometrics*, **59,** 76–82.

Gray, R.J. (1994) A bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics*, **50,** 244–253.

Hastie, T. and Tibshirani, R. (1993) Varying-coefficient models. *Journal of the Royal Statistical Society* B, **55,** 757–796.

Hastings, W.K. (1970) Monte carlo sampling methods using markov chains and their applications. *Biometrika*, **57,** 97–109.

Hjort, N.L. (1990) Nonparametric bayes estimators based on beta processes in models for life history data. *Annals of Statistics*, **18,** 1259–1294.

Hoff, P.D. (2003) Bayesian methods for partial stochastic orderings. *Biometrika*, **90,** 303–317.

Holmes, C.C. and Heard, N.A. (2003) Generalized monotonic regression using random change points. *Statistics in Medicine*, **22,** 623–638.

Hwang, J. T. G. and Peddada, S. D. (1994) Confidence interval estimation subject to order restrictions. *Annals of Statistics*, **22,** 67–93.

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

Iliopoulos, G. (2000) A note on decision theoretic estimation of ordered parameters. *Statistics & Probability Letters*, **50,** 33–38.

Kalbfleisch, J.D. (1978) Non-parametric bayesian analysis of survival time data. *Journal of the Royal Statistical Society* B, **40,** 214–221.

Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *Journal of the American Statistical Association*, **90,** 773–795.

Kuriki, S. and Takemura, A. (2000) Shrinkage estimation towards a closed convex set with a smooth boundary. *Journal of Multivariate Analysis*, **75,** 79–111.

Lavine, M. and Mockus, A. (1995) A nonparametric bayes method for isotonic regression. *Journal of Statistical Planning and Inference*, **46,** 235–248.

Lavine, M. and Schervish, M. (1999) Bayes factors: what they are and what they are not. *American Statistician*, **53,** 119–122.

Lin, D.Y., Oakes, D. and Ying, Z. (1998) Additive hazards regression with current status data. *Biometrika*, **85,** 289–298.

Lin, D.Y. and Ying, Z. (1994) Semiparametric analysis of the additive risk model. *Biometrika*, **81,** 61–71.

Mammen, E. (1991) Estimating a smooth monotone regression function. *The Annals of Statistics*, **19,** 724–740.

Mammen, E. and Thomas-Agnan, C. (1999) Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics*, **26,** 239–252.

McDonald, J.W. and Prevost, A.T. (1997) The fitting of parameter-constrained demographic models. *Mathematical & Computer Modeling*, **26,** 79–88.

Morton-Jones, T., Diggle, P., Parker, L., Dickinson, H. O. and Binks, K. (2000) Additive isotonic regression models in epidemiology. *Statistics in Medicine*, **19,** 849–859.

National Toxicology, Program (1999) Toxicology and carcinogenesis studies of ethylbenzene (case no. 100-41-4) in f344/n rats and $b6c3f_1$ mice (inhalation studies). Technical Report. U.S. Department of Health and Human Services, Research Triangle Park.

Neelon, B. and Dunson, D.B. (2004) Bayesian isotonic regression and trend analysis. *Biometrics*, **60,**; in press.

Nieto-Barajas, L.E. and Walker, S.G. (2002) Markov beta and gamma processes for modeling hazard rates. *Scandinavian Journal of Statistics*, **29,** 413–424.

van Onna, M.J.H. (2002) Bayesian estimation and model selection in ordered latent class models for polytomous items. *Psychometrika*, **67,** 519–538.

Ouassou, I. and Strawderman, W.E. (2002) Estimation of a parameter vector restricted to a cone. *Statistics & Probability Letters*, **56,** 121–129.

Peddada, S. D., Prescott, K. E. and Conaway, M. (2001) Tests for order restrictions in binary data. *Biometrics*, **57,** 1219–1227.

Ramgopal, P., Laud, P. W. and Smith, A. F. M. (1993) Nonparametric bayesian bioassay with prior constraints on the shape of the potency curve. *Biometrika*, **80,** 489–498.

Ramsay, J. O. (1998) Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series* B, **60,** 365–375.

Robert, C.P. and Hwang, J.T.G. (1996) Maximum likelihood estimation under order restrictions by the prior feedback method. *Journal of the American Statistical Association*, **91,** 167–172.

Robertson, T., Wright, F. and Dykstra, R. (1988) *Order Restricted Statistical Inference*. New York: Wiley.

Sen, P.K. (1984) Subhypotheses testing against restricted alternatives for the cox regression model. *Journal of Statistical Planning and Inference*, **10,** 31–42.

Sen, P.K. and Sengupta, D. (1991) On characterizations of pitman closeness of some shrinkage estimators. *Communications in Statistics: Theory & Methods*, **20,** 3551–3580.

Sengupta, D. and Sen, P.K. (1991) Shrinkage estimation in a restricted parameter space. *Sankhya* A, **53,** 389–411.

Silvapulle, M.J. (1994) On tests against one-sided hypotheses in some generalized linear models. *Biometrics*, **50,** 853–858.

Silvapulle, M.J. and Silvapulle, P. (1995) A score test against one-sided alternatives. *Journal of the American Statistical Association*, **90,** 342–349.

Singh, B. and Wright, F.T. (1996) Testing order restricted hypotheses with proportional hazards. *Lifetime Data Analysis*, **2,** 363–389.

Singh, B. and Wright, F.T. (1998) Comparing survival times for treatments with those for a control under proportional hazards. *Lifetime Data Analysis*, **4,** 265–279.

Sinha, D. and Dey, D.K. (1998) Survival analysis using semiparametric bayesian methods. In *Practical nonparametric and semiparametric Bayesian statistics* (eds D.K. Dey, P. Müller and D. Sinha). New York: Springer.

Turner, T. R. and Wollan, P. C. (1997) Locating a maximum using isotonic regression. *Computational Statistics & Data Analysis*, **25,** 305–320.

U.S. Environmental Protection, Agency (1995) Use of the benchmark dose approach in health risk assessment. Technical Report. Risk Assessment Forum, Washington, DC. EPA/630/R-94/007.

Walker, S.G. and Mallick, B.K. (1997) Hierarchical generalized linear models and frailty models with bayesian nonparametric mixing. *Journal of the Royal Statistical Society: Series* B, **59,** 845–860.

Westfall, P.H., Johnson, W.O. and Utts, J.M. (1997) A bayesian perspective on the bonferroni adjustment. *Biometrika*, **84,** 419–427.

Yen, S.S.C. and Jaffe, R.B. (1986) *Reproductive Endocrinology: Physiology, Pathophysiology, and Clinical Management*, Second edn. Philadelphia: W.B. Saunders Company.

Zhang, D., Lin, X., Raz, J. and Sowers, M. (1998) Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, **93,** 710–719.

Zhang, D., Lin, X. and Sowers, M. (2000) Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics*, **56,** 31–39.

Biography

Laura Hilton Gunn was born on July 16, 1977 in Boone, North Carolina, USA. She earned her B.A. degree in 1999 from Jacksonville University in Jacksonville, Florida, graduating *magna cum laude* and with honors in mathematics, including a French minor. She obtained her M.S. in Statistics & Decision Sciences and became a Ph.D. candidate in 2001 at Duke University in Durham, North Carolina. She received a two-year *Intramural Research Training* fellowship from the National Institute of Environmental Health Sciences for 2003-04. Her first job out of graduate school begins in 2004 at Georgia Southern University's School of Public Health in Statesboro, Georgia as an Assistant Professor of Biostatistics.