Journal of Data Science **9**(2011), 311-330

A Bayesian Multiple Comparison Procedure for Simple Order-Restricted Mixed Models with Missing Values

Junfeng Shang
Bowling Green State University

Abstract: A Bayesian hierarchical model is developed for multiple comparisons in mixed models with missing values where the population means satisfy a simple order restriction. We employ the Gibbs sampling and Metropolis-within-Gibbs sampling techniques to obtain parameter estimates and estimates of the posterior probabilities of the equality of the mean pairs. The latter estimates are used to test whether any two means are significantly different, and to test the global hypothesis of the equality of all means. The performance of the model is investigated in simulations by means of both multiple imputations and ignoring missingness. We also illustrate the utility of the model in a real data set. The results show that the proposed hierarchical model can effectively unify parameter estimation, multiple imputations, and multiple comparisons in one setting.

Key words: Gibbs sampling, hierarchical model, Metropolis-within-Gibbs sampling, mixture prior, posterior probability, simple order restriction.

1. Introduction

The two-way ANOVA balanced mixed model (or repeated measures model) is defined as

$$y_{ij} = \mu_i + b_j + \varepsilon_{ij}, \qquad i = 1, \ldots, k, \qquad j = 1, \ldots, m, \tag{1.1}$$

where y_{ij}, μ_i, b_j, and ε_{ij} are all scalars. Here, y_{ij} denotes the response observed on the jth subject in the ith treatment, μ_i is a fixed treatment effect for the ith treatment, b_j is a random subject effect, and ε_{ij} is an error term. Note that the overall sample size is $m \times k$. We assume that the b_j's are distributed as $\mathcal{N}(0, \sigma_\tau^2)$, that the ε_{ij}'s are distributed as $\mathcal{N}(0, \sigma_\tau^2)$, and that the $m(k+1)$ variables b_j and ε_{ij}, $i = 1, \ldots, k$ and $j = 1, \ldots, m$, are all independent.

The problem of a simple order restriction may exist in quite a few applications. For instance, in a drug efficacy experiment, investigators might study the effect of different dose levels of a compound on a response variable, such as blood pressure

or toxicity. For the dosages considered, the mean response is believed to be a nonincreasing or nondecreasing function of the dose level. Therefore, the simple order assumption (i.e., $\mu_1 \leq \cdots \leq \mu_k$) is rational and realistic. Other orders such as a tree order or a loop order may also be considered, but here we restrict attention to a simple order.

For treatment means that satisfy a simple order restriction, i.e., $\mu_1 \leq \cdots \leq \mu_k$, both frequentist and Bayesian methods have been proposed for multiple comparisons. In the Bayesian paradigm, Gelfand *et al.* (1990) provided Bayesian estimates of order-restricted normal means with arbitrary variances. Gelfand *et al.* (1992) extended these results to other types of inequality constraints. Pauler *et al.* (1999) used Bayes factors to test hypotheses involving inequality constraints. In life-testing models, Kim and Sun (2001) considered the use of intrinsic priors and Bayes factors to choose between the model specified by homogeneity of means and that determined by an order restriction on the means. Molitor and Sun (2002) considered situations in which means and variances simultaneously satisfy order restrictions and provided Bayesian estimates of the relevant parameters.

Mukerjee (1988) and Singh and Wright (1990) discussed frequentist results for a mixed model with a simple ordering on the treatment effects. Mukerjee's results showed that multiple comparison techniques for the one-way ANOVA lead to such techniques for a mixed ANOVA model provided there are no missing data.

However, in practice, the experimental designs often yield the data set with missing values because of subject mortality, non-response, etc. Mukerjee's results only consider the situations without missing data. In a Bayesian context, since repeated measures models with missing data involve the random effects, choosing the priors on the variance components poses a challenge if we need to find the marginal posterior in a close form. To liberate this challenge, this paper proposes a hierarchical model where the parameter estimates and estimates of the posterior probabilities can be obtained by the Gibbs sampling and Metropolis-within-Gibbs sampling algorithms. The model incorporates simple order restriction, parameter estimation, missing value approaches, and multiple comparisons in the same setting. Specifically, a Bayesian hierarchical model is developed for multiple comparisons in mixed models based on a fixed treatment effect and a random subject effect with missing data in the format of (1.1). The successive nonnegative differences of the population means are treated as parameters, for which we choose independent prior distributions that are mixtures of an exponential distribution and a discrete distribution with its entire mass at zero. We employ the Gibbs sampling and Metropolis-within-Gibbs sampling techniques to obtain parameter estimates and estimates of the posterior probabilities of the equality of the mean pairs. The latter estimates are used to test whether any two means are significantly different, and to test the global hypothesis of the equality

of all means.

This paper is arranged as follows. Section 2 introduces the mixed models with missing values and presents the priors and the derived full conditional posterior distributions for the hierarchical model. Section 3 outlines approaches to hypothesis testing for the hierarchical model. In Section 4, based on the missing value mechanisms and patterns, the details of the computations for the hierarchical model are presented. For the mixed models with missing values, we do multiple imputations or simply ignore the missingness within the Gibbs sampling. In the situations where it is not convenient to sample directly from the conditional distributions, we use the Metropolis-within-Gibbs sampling algorithm to update the parameters. In Section 5, we investigate the performance of the model-based inferences with simulated data sets, focusing on parameter estimation and successive-mean comparisons using the posterior probabilities. We then illustrate the utility of the model in an application based on data from a study designed to reduce lead blood concentrations in children with elevated levels. Some concluding remarks and discussion are given in Section 6.

2. A Bayesian Approach to Multiple Comparisons for the Mixed Model with Missing Values

2.1 The Mixed Model with Missing Values

Consider the model (1.1) in conjunction with the simple order restriction for the population (treatment) means $\mu_1 \leq \cdots \leq \mu_k$. We may parameterize each of the means $\mu_2, \ldots, \mu_k$ based on the difference between the preceding mean and itself. Specifically, with $\delta_{i-1} = \mu_i - \mu_{i-1}$ $(2 \leq i \leq k)$, the mean μ_2 can be denoted as $\mu_1 + \delta_1$, the mean μ_3 as $\mu_1 + \delta_1 + \delta_2$, etc. In general, the ith mean μ_i $(2 \leq i \leq k)$ can be denoted as $\mu_1 + \delta_1 + \cdots + \delta_{i-1}$.

Suppose that some of the y_{ij}'s in the model (1.1) are missing. We can conveniently illustrate this missingness by using a two-way table.

Consider, for instance, an incomplete data set with $k = 4$ treatment groups described by the mixed model (1.1). The layout for such a data set is featured in Table 1. Our goal is to eventually compare the treatments means. Let symbol $\otimes$ denote the absence of an observation in a cell. Thus, Table 1 illustrates a setting where y_{21}, y_{32}, etc., are missing.

Intuitively, one may question whether any of the observed responses could serve as a surrogate for a missing response, in the sense that the missing response and its observed surrogate are generated by a common mechanism under (1.1). However, the observations in each column correspond to a different mean. If the simple order restriction holds with $\mu_1 < \mu_k$, then at least some of the means are increasing, and the observations would therefore tend to increase as we move from

left to right across each row. Moreover, the observations in each row correspond to a different subject. The model thereby imposes that observations within a row are correlated, yet observations between rows are uncorrelated. Simply put, since each observation y_{ij} corresponds to a different conditional mean $\mu_i + b_j$, the observations are not exchangeable, and an observed response cannot be viewed as providing a meaningful replacement for a missing value. However, due to the additive structure of the model, a missing value y_{ij} can be predicted once estimates of the fixed effect μ_i and the random effect b_j are obtained. Ideally, such estimates are obtained using a procedure that incorporates the order restriction.

In the setting of the mixed model under missingness, let k_j denote an index set which contains the integers denoting the treatment indices for the observed responses corresponding to the jth subject. Thus, for the data illustrated in Table 1, we would have $k_1 = \{1, 3, 4\}$, $k_2 = \{1, 2, 4\}$, and $k_3 = \{1, 2, 3\}$, etc. Furthermore, we let n_j denote the number of integers in k_j, i.e., the number of observations corresponding to the jth subject.

Table 1: Data layout for the mixed model with missing values ($k = 4$)

subject	treatments ($k = 4$)			
	$i = 1$	2	3	4
$j = 1$	y_{11}	$\otimes y_{21}$	y_{31}	y_{41}
2	y_{12}	y_{22}	$\otimes y_{32}$	y_{42}
3	y_{13}	y_{23}	y_{33}	$\otimes y_{43}$
4	$\otimes y_{14}$	y_{24}	y_{34}	y_{44}
5	$\otimes y_{15}$	y_{25}	$\otimes y_{35}$	y_{45}
6	y_{16}	y_{26}	y_{36}	$\otimes y_{46}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Let y_j denote a $k \times 1$ vector of responses observed on the jth subject, $j = 1, \cdots, m$, and let $Y = (y_1', \ldots, y_m')'$, and define $\delta_0 = 0$. Given that the subject-specific vectors y_j are independent, we have the joint density function for the data vector Y

$$[Y \mid \mu_1, \{\delta_i\}, \sigma_\tau^2, \sigma^2]$$
$$\propto \prod_{j=1}^{m} \left\{ (\sigma^2)^{-\frac{n_j - 1}{2}} (n_j \sigma_\tau^2 + \sigma^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[s_1(j) - \frac{\sigma_\tau^2}{n_j \sigma_\tau^2 + \sigma^2} s_2(j) \right] \right\} \right\}, \quad (2.1)$$

where

$$s_1(j) = \sum_{i \in k_j} \left(y_{ij} - \mu_1 - \sum_{l=0}^{i-1} \delta_l \right)^2$$

and

$$s_2(j) = \left(\sum_{i \in k_j} (y_{ij} - \mu_1 - \sum_{l=0}^{i-1} \delta_l) \right)^2 .$$

Next, we need to specify prior distributions and to obtain the full conditional posterior distributions for the parameters.

2.2 Priors and Full Conditional Posterior Distributions

In what follows, we will choose the priors and hyperpriors for the hierarchical model and present the full conditional posterior distributions for the parameters and hyperparameters.

Because each successive mean difference δ_i is positive ($\mu_{i+1} > \mu_i$) or zero ($\mu_{i+1} = \mu_i$), we choose a prior distribution for δ_i that is a mixture of an exponential distribution and a discrete distribution with its entire mass at $\delta_i = 0$. The discrete component of the mixture allows the difference between two successive means to be zero. If we define $A = [\delta_i > 0]$ and $B = [\delta_i = 0]$, with I_A and I_B denoting the indicator functions corresponding to events A and B, respectively, the density function for δ_i can be represented as

$$[\delta_i \mid \rho_i, \eta_i] = \rho_i I_B + \Delta I_A, \tag{2.2}$$

where $\Delta = (1 - \rho_i)\frac{1}{\eta_i} \exp\left\{-\frac{\delta_i}{\eta_i}\right\}$. Note that ρ_i denotes the probability of event B and η_i denotes the mean of δ_i with the probability density $\frac{1}{\eta_i} \exp\left\{-\frac{\delta_i}{\eta_i}\right\}$ for $\delta_i > 0$.

For $i = 1, \cdots, k-1$, we multiply the density function in (2.1) by the prior of δ_i in (2.2), then we focus on and re-organize the terms involving δ_i to derive its full conditional posterior distribution. First, we can easily have

$$\begin{aligned}
&[\delta_i \mid Y, \mu_1, \sigma^2, \sigma_\tau^2, \{\delta_1, \cdots, \delta_{i-1}, \delta_{i+1}, \cdots, \delta_{k-1}\}, \rho_i, \eta_i] \\
&\propto \exp\left\{ -\frac{\delta_i^2}{2\sigma^2} \left[\sum_{j=1}^{m} s_4(i,j) - \sum_{j=1}^{m} \frac{\sigma_\tau^2}{n_j \sigma_\tau^2 + \sigma^2} s_4^2(i,j) \right] \right\} \\
&\quad \times \exp\left\{ \frac{\delta_i}{\sigma^2} \left[\sum_{j=1}^{m} s_3(i,j) - \sum_{j=1}^{m} \left(\frac{\sigma_\tau^2}{n_j \sigma_\tau^2 + \sigma^2} s_3(i,j) s_4(i,j) \right) \right] \right\} \\
&\quad \times \left[\rho_i I_B + (1 - \rho_i) \frac{1}{\eta_i} \exp\left(-\frac{\delta_i}{\eta_i} \right) I_A \right],
\end{aligned}$$

where

$$s_3(i,j) = \sum_{p \in k_p} \left(y_{pj} - \mu_1 - \sum_{q=1}^{p-1} \delta_q + \delta_i \right)$$

and

$$s_4(i,j) = \sum_{p=i+1}^{k} I_{k_j}(p).$$

Here, we define the set $k_p = \{p \in k_j \mid p \geq i+1\}$. The full conditional distribution of δ_i can therefore be summarized as a mixture of a discrete part and a continuous part, and is expressed as

$$[\delta_i \mid \cdot] = \begin{cases} c\rho_i f(\delta_i \mid \cdot), & \delta_i = 0, \\ c(1-\rho_i)\frac{1}{\eta_i} f(\delta_i \mid \cdot), & \delta_i > 0, \\ 0, & \delta_i < 0, \end{cases}$$

where

$$f(\delta_i \mid \cdot) = \frac{1}{\sqrt{2\pi(\frac{1}{a_i})}} \exp\left\{ -\frac{1}{2(\frac{1}{a_i})} \left(\delta_i - \frac{g_i - \frac{I_A}{\eta_i}}{a_i} \right)^2 \right\} \exp\left\{ \frac{\left(g_i - \frac{I_A}{\eta_i} \right)^2}{2a_i} \right\}, \quad (2.3)$$

with

$$\begin{aligned} a_i &= \frac{1}{\sigma^2} \left[\sum_{j=1}^{m} s_4(i,j) - \sum_{j=1}^{m} \frac{\sigma_\tau^2}{n_j \sigma_\tau^2 + \sigma^2} s_4^2(i,j) \right], \\ g_i &= \frac{1}{\sigma^2} \left[\sum_{j=1}^{m} s_3(i,j) - \sum_{j=1}^{m} \left(\frac{\sigma_\tau^2}{n_j \sigma_\tau^2 + \sigma^2} s_3(i,j) s_4(i,j) \right) \right], \\ c &= \frac{1}{\rho_i f(0 \mid \cdot) + (1-\rho_i)\frac{1}{\eta_i} \int_0^\infty f(\delta_i \mid \cdot) d\delta_i}. \end{aligned}$$

Note that in the preceding notation, the $\mid \cdot$ represents the condition given the data and all the other parameters except δ_i. Therefore, the $f(\delta_i \mid \cdot)$ represents the kernel full conditional posterior distribution of δ_i given the data and all the other parameters except δ_i. Since the full conditional posterior distribution of δ_i relies on its prior and the prior is defined in the format of (2.2) involving the two cases $\delta_i > 0$ and $\delta_i = 0$, the $f(\delta_i \mid \cdot)$ has different expressions with respect to $\delta_i > 0$ and $\delta_i = 0$. The expression of the $f(\delta_i \mid \cdot)$ therefore accommodates the indicator function I_A.

Prior to the continuity of the next derivation, we may comment on this prior. Due to the simple order restriction held between the means, i.e., $\mu_1 \leq \cdots \leq \mu_k$, the utility of a discrete-continuous mixture prior can effectively reflect this relationship and limit the difference between any two successive means to two possible values, zero or a positive number (see Gottardo and Raftery (2008), Shang *et al.* (2008), and Nashimoto and Wright (2008)). Therefore, from this point of view, the proposed mixture prior will play an important role in contributing to handling the problem of multiple comparisons for the simple order restricted means in the mixed modeling setting.

For the mean μ_1, we choose a normal prior. Specifically, we let $\mu_1 \sim \mathcal{N}(\mu_0, \tau_0^2)$. With this prior, we can express the full conditional posterior distribution of μ_1 as

$$\begin{aligned}
&[\mu_1 \mid Y, \sigma^2, \sigma_\tau^2, \{\delta_i\}] \\
&\propto \exp\left\{-\frac{\mu_1^2}{2}\left(\frac{1}{\tau_0^2} + \frac{\sum_{j=1}^m n_j}{\sigma^2} - \sum_{j=1}^m \frac{n_j^2\sigma_\tau^2}{\sigma^2(n_j\sigma_\tau^2+\sigma^2)}\right)\right. \\
&\quad \left. + \mu_1\left(\frac{\mu_0}{\tau_0^2} + \sum_{j=1}^m \frac{1}{n_j\sigma_\tau^2+\sigma^2} s_5(i,j)\right)\right\} \qquad (2.4) \\
&= \exp\left\{-\frac{\mu_1^2}{2}\left(\frac{1}{\tau_0^2} + \sum_{j=1}^m \frac{n_j}{n_j\sigma_\tau^2+\sigma^2}\right) + \mu_1\left(\frac{\mu_0}{\tau_0^2} + \sum_{j=1}^m \frac{1}{n_j\sigma_\tau^2+\sigma^2} s_5(i,j)\right)\right\},
\end{aligned}$$

where

$$s_5(i,j) = \sum_{p\in k_j}(y_{ij} - \sum_{l=0}^{i-1}\delta_l).$$

Setting

$$u = \frac{\mu_0}{\tau_0^2} + \sum_{j=1}^m \frac{1}{n_j\sigma_\tau^2+\sigma^2} s_5(i,j)$$

and

$$v = \frac{1}{\tau_0^2} + \sum_{j=1}^m \frac{n_j}{n_j\sigma_\tau^2+\sigma^2},$$

and completing the square in (2.4) with respect to μ, one can show that the full conditional posterior distribution of μ_1 is

$$[\mu_1 \mid Y, \sigma^2, \sigma_\tau^2, \{\delta_i\}] = \mathrm{N}\left(\frac{u}{v}, \frac{1}{v}\right).$$

We choose an improper noninformative prior $\pi(\sigma_\tau^2, \sigma^2) \propto \frac{1}{\sigma^2(\sigma_\tau^2+\sigma^2)}$, which is used in Chaloner (1987), for the variance components, σ_τ^2 and σ^2. Then we have the joint posterior full conditional distribution of the variance components as

$$\begin{aligned}
&[\sigma_\tau^2, \sigma^2 \mid Y, \mu_1, \{\delta_i\}] \\
&\propto \prod_{j=1}^{m} \left\{ (\sigma^2)^{-\frac{n_j-1}{2}} (n_j\sigma_\tau^2 + \sigma^2)^{-\frac{1}{2}} \right\} \\
&\quad \times \prod_{j=1}^{m} \left\{ \exp\left[-\frac{1}{2\sigma^2} \left(s_1(j) - \frac{\sigma_\tau^2}{n_j\sigma_\tau^2 + \sigma^2} s_2(j) \right) \right] \frac{1}{\sigma^2(\sigma_\tau^2 + \sigma^2)} \right\}. \qquad (2.5)
\end{aligned}$$

Unfortunately, the full conditional posterior distributions of the individual variance components σ_τ^2 and σ^2 are not in a form which is conducive to sampling. Thus, as outlined in Section 4, we will use the Metropolis algorithm to update the estimates of σ_τ^2 and σ^2 within the iterations of the Gibbs sampling.

Employing the hyperprior $\rho_i \sim Beta(\alpha_0, \beta_0)$, the full conditional posterior distribution of ρ_i can be expressed as

$$[\rho_i \mid \delta_i, \eta_i] = \begin{cases} Beta(\alpha_0 + 1, \beta_0), & \delta_i = 0, \\ Beta(\alpha_0, \beta_0 + 1), & \delta_i > 0. \end{cases}$$

With $\alpha_0 = \beta_0 = 1$ (uniform), we have

$$[\rho_i \mid \delta_i, \eta_i] = \begin{cases} Beta(2, 1), & \delta_i = 0, \\ Beta(1, 2), & \delta_i > 0. \end{cases}$$

In Section 3, we will see how we can regulate the Type 1 error rate by adjusting the parameters α_0 and β_0 for the hyperprior on ρ_i.

We choose an inverse-gamma hyperprior $IG(a_0, b_0)$ for η_i, the parameter for the exponential distribution used in the mixture prior for δ_i in (2.2). The full conditional posterior distribution of η_i is then given by

$$[\eta_i \mid \delta_i, \rho_i] = \begin{cases} IG(a_0, b_0), & \delta_i = 0, \\ IG\left(a_0 + 1, \left[\delta_i + \frac{1}{b_0} \right]^{-1} \right), & \delta_i > 0. \end{cases}$$

Because the full conditional posterior distribution for η_i is improper when based on a noninformative prior, we choose a flat informative prior for η_i.

With the preceding independent priors, most of the full conditional posterior distributions are standard distributions, such as normal, inverse-gamma, and beta. The exceptions are the full conditional posterior distributions for the δ_i and the variance components σ^2 and σ_τ^2.

The Gibbs sampling is employed to estimate the parameters and posterior probabilities of the successive-mean differences. However, as mentioned previously, since it is not convenient to sample directly from the full conditional posterior distributions of σ_τ^2 and σ^2, we use the Metropolis algorithm to update the estimates of σ_τ^2 and σ^2 within the Gibbs sampling iterations. In Section 4, we provide a brief overview of this algorithm, and discuss its implementation in our setting.

3. Hypothesis Testing via the Posterior Probability

Consider the hypotheses for successive pairwise comparisons of the means, $H_{0i} : \mu_i = \mu_i + \delta_i$ versus $H_{1i} : \mu_i < \mu_i + \delta_i$, or $H_{0i} : \delta_i = 0$ versus $H_{1i} : \delta_i > 0$ for $i = 1, \ldots, k-1$, and consider the global test, $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ versus $H_1 : \mu_1 \leq \cdots \leq \mu_k$ with $\mu_1 < \mu_k$.

Under the proposed hierarchical model, using the magnitude of the posterior probabilities of $\delta_i > 0$ and of $\delta_i = 0$ $(1 \leq i \leq k-1)$ in conjunction with a decision rule, we can conduct the hypothesis testing for the means.

Following the intuitive rule proposed by Westfall *et al.* (1997), in order to obtain the probability of the null hypothesis for the global test denoted by $\Pr\{H_0\}$ to be 0.5 for the model with a total of $k-1$ mutually independent priors for the δ_i's. Let $\Pr\{H_{0i}\}$ denote the prior probabilities for the null hypotheses H_{0i}, then we calibrate $\Pr\{H_{0i}\}$ via $\Pr\{H_{0i}\} = 0.5^{1/(k-1)}$.

We observe that

$$\begin{aligned}\Pr\{H_{0i}\} = \Pr\{\delta_i = 0\} &= E(E(I_B \mid \rho_i, \eta_i)) \\ &= E(\rho_i) = \frac{\alpha_0}{\alpha_0 + \beta_0}.\end{aligned}$$

Thus, if one wants the initial prior probabilities of $\Pr\{H_{0i}\} = 0.5$, then one could choose $\alpha_0 = \beta_0$ and select 1 as the common value. Note that for any given k, there exist an infinite number of choices for α_0 and β_0 that satisfy $\Pr\{H_{0i}\} = 0.5^{1/(k-1)} = \alpha_0/(\alpha_0 + \beta_0)$.

We adopt the conventional scheme of using 0.5 (or a cutoff value based on a concrete context) as the deciding criterion for the posterior probability. For pairwise comparisons, we declare H_{1i} if

$$\Pr\{\delta_i = 0 \mid Y\} < 0.5, \tag{3.1}$$

where $\Pr\{\delta_i = 0 \mid Y\}$ is the posterior probability of the null hypothesis resulting from the utility of $\Pr\{H_{0i}\}$.

For the global test, we declare H_1 if at least one of the pairwise tests declares H_{1i}, i.e., if

$$\min_{1 \leq i \leq k-1} \{\Pr\{\delta_i = 0 \mid Y\}\} < 0.5. \tag{3.2}$$

We emphasize that the preceding rules for the pairwise and global tests are formulated to be consistent. By employing these rules, the test results are coherent.

Let $\Pr\{\sum_{i=1}^{k-1} \delta_i = 0 \mid Y\}$ denote the joint posterior probability that all δ_i's are simultaneously zero. One could use the global test based on the joint posterior probability of the δ_i's, that is, the test that rejects H_0 if

$$\Pr\left\{\sum_{i=1}^{k-1} \delta_i = 0 \mid Y\right\} < 0.5. \tag{3.3}$$

However, we recommend the test based on (3.2). First, note the test based on (3.3) is not compatible with the tests in (3.1). That is, using (3.3) and the pairwise comparisons in (3.1) may lead to inconsistent decisions. Second, we found that the global test based on (3.3) may be too liberal, i.e., it tends to reject H_0 more often than the test based on (3.2). In particular, our Monte Carlo study showed a strong tendency for the left-hand inequality in the following and the right-hand inequality clearly holds:

$$\Pr\left\{\sum_{i=1}^{k-1} \delta_i = 0 \mid Y\right\} < \prod_{i=1}^{k-1} \Pr\{\delta_i = 0 \mid Y\} \leq \min_{1\leq i\leq k-1} \{\Pr\{\delta_i = 0 \mid Y\}\}$$

(See Shang *et al.* (2008)).

Thus, the test based on (3.2) has smaller Type I error and power than the test based on (3.3). However, the results of our simulation study (not presented here) suggest the procedure based on (3.1) and (3.2) has reasonable power.

4. Details of Computations in the Gibbs Sampling

4.1 The Metropolis-within-Gibbs Algorithm

In our setting, the Metropolis-within-Gibbs algorithm will be utilized to sample from the full conditional posterior distributions for σ_τ^2 and σ^2 since the forms of the distributions are not conducive to sampling. We provide a brief overview of the Metropolis algorithm, and then discuss its implementation in our context.

Suppose we wish to simulate θ from a posterior density $g(\theta \mid Y)$. For simplification, we write the density simply as $g(\theta)$. Beginning with an initial value $\theta^{(0)}$, the Metropolis-Hasting algorithm simulates the t^{th} value in the sequence $\theta^{(t)}$ given the $(t-1)^{th}$ value in the sequence $\theta^{(t-1)}$. A candidate value is simulated from a proposal density, and an acceptance probability P is computed as the probability that the candidate value is accepted. The algorithm is outlined as follows.

- Simulate a candidate value θ^* from a proposal density $p(\theta^* \mid \theta^{(t-1)})$.
- Compute the ratio

$$\frac{g(\theta^*)p(\theta^{(t-1)} \mid \theta^*)}{g(\theta^{(t-1)})p(\theta^* \mid \theta^{(t-1)})}.$$

- Compute the acceptance probability $P = \min\{R, 1\}$.
- Sample a value $\theta^{(t)}$ such that $\theta^{(t)} = \theta^*$ with the probability P. Otherwise $\theta^{(t)} = \theta^{(t-1)}$.

Under suitable regularity conditions on the proposal density $p(\theta^*|\theta^{(t-1)})$, the sequence of simulated values $\theta^{(1)}, \theta^{(2)}, \cdots$ will converge to a random variable that is distributed according to the posterior distribution $g(\theta)$.

In implementing our method, we estimate the variance components, σ_τ^2 and σ^2, using the Metropolis-within-Gibbs algorithm. Suppose that $\theta^{(t-1)}$ represents the current value of θ, and θ refers to a vector consisting of σ_τ^2 and σ^2. Let $g(\theta)$ represent the joint conditional posterior distribution of σ_τ^2 and σ^2, as shown in (2.5). A candidate value for θ is given by $\theta^* = \theta^{(t-1)} + cZ$, where Z is a bivariate normal variate with mean vector 0 and variance-covariance matrix Σ, and c is a fixed positive scale parameter. That is, the proposal density for θ^* is bivariate normal having the form $p(\theta^* \mid \theta^{(t-1)}) = h(\theta^* - \theta^{(t-1)})$, where h is a symmetric density about the origin. Therefore, the ratio R has the simple form $R = \frac{g(\theta^*)}{g(\theta^{(t-1)})}$.

We incorporate the preceding Metropolis-within-Gibbs algorithm after applying a log transformation to each variance, thereby extending the range of the parameters to the entire set of real numbers. Let Z_1 and Z_2 represent a pair of bivariate normal variates with zero means, unit variances, and a covariance of 0.5. The equations $\log(\sigma_\tau^2)^* = \log(\sigma_\tau^2)^{(t-1)} + cZ_1$ and $\log(\sigma^2)^* = \log(\sigma^2)^{(t-1)} + cZ_2$ are used to separately update the variance components σ_τ^2 and σ^2. We choose c so that the acceptance probability is in the $25 - 50\%$ range.

4.2 Multiple Imputations within the Gibbs Sampling

It is well known that Little and Rubin (1987) introduced specific missing data terminology as a standard framework to deal with missing data mechanisms and their effect on data analysis. For the situation where the missing values are completely at random (MCAR) or at random (MAR), imputations can be completed while the Gibbs sampling is executed to obtain the required results for the hierarchical model; that is, in the same setting, we can do both multiple imputations and multiple comparisons.

We first describe the multiple imputation algorithm for parametric Bayesian models in general. Suppose that the complete data, $Y = (Y_{obs}, Y_{mis})$, follows a parametric model $P(Y \mid \theta)$ where θ is the parameter vector and has a prior distribution, and then we have $P(Y_{mis} \mid Y_{obs}) = \int P(Y_{mis} \mid Y_{obs}, \theta)P(\theta \mid Y_{obs})d\theta$. The

starting value for Y_{mis} can be created by first simulating a random draw of the unknown parameters from their observed-data posterior $\theta^* \sim P(\theta \mid Y_{obs})$ followed by a random draw of the missing values from their conditional predictive distribution $Y^*_{mis} \sim P(Y_{mis} \mid Y_{obs}, \theta^*)$. Consider an iterative, two-step process in which we alternately sample missing values from their conditional predictive distribution $Y^{(t)}_{mis} \sim P(Y_{mis} \mid Y_{obs}, \theta^{(t-1)})$ and then sample unknown parameters from a simulated complete-data posterior $\theta^{(t)} \sim P(\theta \mid Y_{obs}, Y^{(t)}_{mis})$. Given the initial values for θ, this defines a Markov chain $\{Y^{(t)}_{mis}, t = 2, 3, \cdots, T\}$. Under quite general conditions, this chain converges to the stationary distribution $P(Y_{mis}, \theta \mid Y_{obs})$. We execute these steps a large number of times ($T = 20,000$, burn-in=5,000). After the burn-in, this process produces a draw from its observed data posterior and a draw of Y_{mis} from $P(Y_{mis} \mid Y_{obs})$, the distribution from which multiple imputations are generated (Schafer, 1999). For each iteration of a Markov chain, a complete data set is constructed and the estimation of the parameters can be executed based on it.

In the framework of the mixed models, the conditional predictive distribution for a missing observation $(y_{ij_{mis}} \mid b_j, \theta)$ is $\mathcal{N}(\mu_i + b_j, \sigma^2)$. Note that $y_{ij_{mis}}$ denotes a missing observation in Y_{mis}.

Given the prior $(b_j \mid \sigma^2_\tau) \sim \mathcal{N}(0, \sigma^2_\tau)$, the full conditional distribution of b_j, $j = 1, \cdots, m$, is obtained as

$$b_j \mid y_j, \mu_1, \{\delta_i\}, \sigma^2_\tau, \sigma^2 \sim \mathcal{N}\left(\sigma^2_\tau z' V^{-1}(y_j - \mu), \left(\frac{k}{\sigma^2} + \frac{1}{\sigma^2_\tau}\right)^{-1}\right),$$

where $V = z\sigma^2_\tau z' + \sigma^2 I$.

At the first iteration of the Gibbs sampling, the initial values for θ are given, b_j's are drawn from the prior distribution $\mathcal{N}(0, \sigma^2_\tau)$, and then missing observations are drawn from $\mathcal{N}(\mu_i + b_j, \sigma^2)$. The complete data is therefore constructed by $Y = (Y_{obs}, Y_{mis})$. According to the complete data set and the full conditional posterior distributions in Section 2.2, we can update the parameters and calculate the posterior probabilities of $\delta_i = 0$ for multiple comparisons.

5. Simulations and Application

5.1 Brief Description of Simulations

As previously mentioned, we compute the posterior probabilities for the mixed models with missing values by multiple imputations in the Metropolis-within-Gibbs algorithm or by discarding the missing values. Since the convergence is not slow, the number of burn-in samples is taken to be 5,000, and the subsequent 10,000 iterations are used to estimate the parameters. In the iterations after burn-in, the frequency of the event $\delta_i = 0$ is recorded, leading to the approximation

of $\Pr(\delta_i = 0 \mid Y)$. Three population means ($k = 3$) will be compared in the simulations.

For the prior of μ_1, the hyperparameters μ_0 and τ_0^2 do not have much effect on the results provided that the normal prior distribution is relatively flat, i.e., τ_0^2 is very large. In our simulations, $\mu_0 = 0$ and $\tau_0^2 = 100$. To make the prior of η_i flat, hyperparameters $a_0 = 2.2$ and $b_0 = 0.05$ are used. To make the equal prior probability of $\Pr\{H_{0i}\}$ with $k = 3$, the probability $\Pr\{H_{0i}\}$ will be $\Pr\{H_{0i}\} = 0.5^{1/(3-1)} = 0.707$. This is closely approximated by the choice $\rho_i \sim Beta(2.4, 1)$, which gives $\Pr\{H_{0i}\} = E(\rho_i) = 0.706$.

For the simulated data, we choose the mean parameters $\mu = (12, 14, 16)'$. Small variances ($\sigma_\tau^2 = 10$ and $\sigma^2 = 2$) are specified. We generate a large-size sample ($m = 100$). The simulations with large variances ($\sigma_\tau^2 = 50$ and $\sigma^2 = 30$) are also completed, yet the conclusions are quite similar to the setting with the small variances, so the results are not presented here.

For the generated sample, two incomplete data sets are randomly produced by applying 10% and 20% missing rates, respectively. The parameter estimates and the posterior probabilities are then computed for each complete or incomplete sample. For reference purposes, we also compute the unrestricted maximum likelihood estimates (MLEs) of the model parameters via the EM algorithm. For the incomplete data sets, multiple imputations and disregard of the missing values are respectively applied in each iteration of the Gibbs sampling chain.

5.2 Simulation Results

For the generated complete samples, Table 2 features the parameter estimates, their variances, and the posterior probabilities of $\delta_i = 0$. Note that the unrestricted MLE of δ_i ($i = 1, 2$) is the difference between the unrestricted MLEs of the mean μ_{i+1} and the preceding mean μ_i.

Table 2: Posterior parameter estimates and MLEs for complete data sets

$\mu = (12, 14, 16)'$,	$\sigma_\tau^2 = 10$,		$\sigma^2 = 2$
parameter	mean	variance	MLE
μ_1	12.32	0.27	12.28
δ_1	1.37	0.27	1.40
δ_2	2.39	0.18	2.37
σ_τ^2	9.87	2.51	10.20
σ^2	2.10	0.10	1.93
posterior probability			
$P(\delta_1 = 0 \mid Y)$	0.04		
$P(\delta_2 = 0 \mid Y)$	0.01		

For the sample with $\sigma_\tau^2 = 10$ and $\sigma^2 = 2$, the mean parameter estimates are $(12.32, 13.69, 16.08)'$, and the variance estimates are 9.87 and 2.10, respectively. These estimates are quite close to the MLEs computed by the EM algorithm, which are $\hat{\mu} = (12.28, 13.68, 16.05)'$, $\hat{\sigma}_\tau^2 = 10.20$, and $\hat{\sigma}^2 = 1.93$. Among the variances of the parameter estimates, the variance of the σ_τ^2 estimate is the largest. Based on the posterior probabilities, the pairwise tests favor H_{11} and H_{12}, indicating that there exist significant differences between μ_1 and μ_2 and between μ_2 and μ_3.

Note that the posterior parameter estimates are evaluated under the simple order restriction, whereas the unrestricted MLEs are not. The MLEs could violate the simple order restriction, yet the posterior estimates obey the restriction.

It is of interest to compare the preceding conclusions to the results of the the frequentist approach (See details in Singh and Wright, 1990). For the sample with $\sigma_\tau^2 = 10$ and $\sigma^2 = 2$, under the frequentist approach, the likelihood ratio test (LRT) statistic for testing $H_0 : \mu_1 = \mu_2 = \mu_3$ is 373.47 ($df = 198$). Based on 198 degree of freedom, with $\alpha = 0.01$, the critical value is 6.940 from Table A.6 of Robertson *et al.* (1988). The LRT therefore rejects H_0. Thus, the Bayesian and frequentist methods both reject H_0.

Table 3 features the results for the incomplete samples with $\sigma_\tau^2 = 10$ and $\sigma^2 = 2$ by multiple imputations and by simply ignoring the missing values. Comparing the corresponding results from the complete to incomplete data sets by multiple imputations, one can observe that the parameter estimates and their variances are very close although differences do exist. Moreover, the posterior probabilities vary between the complete and incomplete data sets, yet the conclusions for multiple comparisons are identical.

Comparing the corresponding results from the complete to incomplete data sets, one can find that the differences of the parameter estimates and their variances are much greater than those between the complete and incomplete data sets by multiple imputations. For instance, for the complete data set in Table 2 with $\sigma_\tau^2 = 10$ and $\sigma^2 = 2$, $\Pr\{\delta_1 = 0 \mid Y\} = 0.04$ and $\Pr\{\delta_2 = 0 \mid Y\} = 0.01$; for the incomplete data sets in Table 3 with 10% missing values, $\Pr\{\delta_1 = 0 \mid Y\} = 0.36$ and $\Pr\{\delta_2 = 0 \mid Y\} = 0.01$; with 20% missing values, $\Pr\{\delta_1 = 0 \mid Y\} = 0.55$ and $\Pr\{\delta_2 = 0 \mid Y\} = 0.01$. For the complete data set and the incomplete data set with 10% missing values, the pairwise tests reject H_{01} implying that there is significant difference between μ_1 and μ_2. However, the incomplete data set with 20% missing values keeps the null hypothesis supporting that μ_1 and μ_2 are equal.

The results of applying the hierarchical model in the simulated data show that the parameter estimates for the imputed data sets are not the same as those for the complete data set. Since the imputed values are simply not the real

Table 3: Posterior parameter estimates for incomplete data sets

$\mu = (12, 14, 16)'$,	$\sigma_\tau^2 = 10$,	$\sigma^2 = 2$		
By Imputations	10% missing		20% missing	
parameter	mean	variance	mean	variance
μ_1	12.32	0.32	12.35	0.33
δ_1	1.26	0.32	1.25	0.40
δ_2	2.41	0.19	2.44	0.22
σ_τ^2	10.76	3.44	10.68	2.29
σ^2	2.01	0.10	2.10	0.13
posterior probability				
$P(\delta_1 = 0 \mid Y)$	0.08		0.09	
$P(\delta_2 = 0 \mid Y)$	0.01		0.01	
With Missingness	10% missing		20% missing	
parameter	mean	variance	mean	variance
μ_1	12.61	0.31	12.65	0.30
δ_1	0.67	0.38	0.46	0.35
δ_2	2.72	0.22	3.05	0.24
σ_τ^2	10.03	1.82	10.20	2.78
σ^2	2.45	0.17	3.06	0.29
posterior probability				
$P(\delta_1 = 0 \mid Y)$	0.36		0.55	
$P(\delta_2 = 0 \mid Y)$	0.01		0.01	

observations, this can be explained by uncertainty rooted in imputed values. However, it appears that multiple imputations are an effective and flexible tool for dealing with missing values in the proposed hierarchical model.

5.3 Application

In this section, we consider a data set from a study described in a 2000 article from *Pediatric Research* written by the TLC (Treatment of Lead-Exposed Children) Trial Group. The data is provided and analyzed in Fitzmaurice *et al.* (2004). This data has been utilized in Shang *et al.* (2008) for demonstrating the performance of a similar hierarchical model in balanced mixed models. Here, we apply the proposed hierarchical model to the data for the illustration of its effectiveness.

The study is designed to reduce lead blood concentrations in children with elevated levels. At the outset of the study, all participants were provided with a month's supply of vitamin and mineral supplements, and their homes were inspected and cleaned based on a TLC regimen designed to suppress exposure to

leaded dust. The participants (children) were randomly assigned to two groups. Participants in the "treatment" group were then provided with succimer capsules, whereas those in the "control" group were provided with a placebo. We consider blood lead levels collected for 50 of the children who did not receive the succimer capsules. Blood lead levels were measured at baseline (week 0), week 1, week 4, and week 6. If we put these four time points in a reverse order, and let μ_1, μ_2, μ_3, and μ_4 denote the lead level means corresponding to week 6, week 4, week 1, and baseline, respectively. Since the homes of these children were cleaned using an established TLC regimen, one might expect that the mean blood lead levels satisfy the simple order restriction, i.e., $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$.

In the context of model (1.1), y_{ij} will denote the blood lead level for child j ($j = 1, \cdots, 50$) at time i ($i = 1, \cdots, 4$).

Our preliminary analyses shows that the compound symmetric structure adequately describes the control subject measurements. Model (1.1) also assumes normality. At each time period, the blood lead levels are slightly skewed right, yet the skewness does not appear to be strong enough to warrant a transformation.

To adjust the prior probability of $\Pr\{H_0\} = 0.5$ with $k = 4$, the prior probability of H_{0i} must be $0.5^{1/(4-1)} = 0.7937$. The choice $\rho_i \sim Beta(4, 1)$ gives $\Pr\{H_{0i}\} = E(\rho_i) = 0.80$.

The results of our data analysis are listed in Table 4. Based on the posterior estimates $\hat{\mu}_1$, $\hat{\delta}_1$, $\hat{\delta}_2$, and $\hat{\delta}_3$, the estimates for the mean blood lead levels $(\mu_1, \mu_2, \mu_3, \mu_4)'$ are (23.9097, 24.1833, 24.5793, 26.0028)$'$. For comparison, the MLEs obtained by the EM algorithm are $\hat{\mu}$ = (23.6211, 24.0451, 24.6352, 26.2472)$'$. Both sets of estimates are reasonably close to the sample means $\bar{y}$ = (23.6460, 24.0700, 24.6600, 26.2720)$'$, and all satisfy the simple order restriction.

The posterior probabilities and the δ_i estimates in Table 4 suggest that $\delta_3 > 0$ ($\mu_4 > \mu_3$), yet imply that $\delta_2 = 0$ ($\mu_3 = \mu_2$) and that $\delta_1 = 0$ ($\mu_2 = \mu_1$). Thus, the baseline mean is significantly higher than the week 1 mean, yet the week 1, week 4, and week 6 means are not significantly different. It seems logical that the most dramatic decrease in the blood lead level should occur during a period of time immediately following the cleaning of the home. Over time, the lead level should stabilize.

Table 4: Posterior parameter estimates for blood lead level data

sample means: $\bar{y} = (23.6460, 24.0700, 24.6600, 26.2720)'$

parameter	μ_1	σ^2_τ	σ^2	δ_1	δ_2	δ_3
mean	23.9097	23.2996	9.3633	0.2736	0.3960	1.4235
variance	1.2530	43.5819	3.642	0.4521	0.5712	1.3298
posterior probability $P(\delta_i = 0 \mid Y)$				0.8210	0.7360	0.3043

For the frequentist test, the value of the test statistic is 36.2550 ($df = 147$), and the $\alpha = 0.01$ critical value is 7.858. Hence, the frequentist method also strongly favors rejecting the null hypothesis.

Table 5 features the results of the incomplete data sets by multiple imputations. The incomplete data are randomly created. One can observe that the parameter estimates and their variances are quite close. Moreover, the posterior probabilities are quite close to those for the complete data set. Therefore, the multiple comparison conclusions are similar.

Table 5: Posterior parameter estimates for incomplete data sets of blood lead level data

By Imputations	10% missing		20% missing	
parameter	mean	variance	mean	variance
μ_1	23.8877	1.2457	24.0572	1.3399
δ_1	0.3299	0.5758	0.3483	0.6470
δ_2	0.3830	0.5364	0.3352	0.4867
δ_3	1.2667	1.3209	1.4101	1.5159
σ^2_τ	23.8824	44.2309	24.1760	25.2496
σ^2	10.6002	5.1022	11.3082	7.3371
posterior probability				
$P(\delta_1 = 0 \mid Y)$	0.7860		0.7844	
$P(\delta_2 = 0 \mid Y)$	0.7314		0.7627	
$P(\delta_2 = 0 \mid Y)$	0.3546		0.3300	
With Missingness	10% missing		20% missing	
parameter	mean	variance	mean	variance
μ_1	23.9965	1.1589	23.9905	1.1740
δ_1	0.2561	0.4294	0.2560	0.4212
δ_2	0.3776	0.5237	0.3973	0.5419
δ_3	1.3381	1.2764	1.3251	1.2994
σ^2_τ	22.3255	17.7588	22.5531	28.1866
σ^2	9.0079	3.4079	8.9905	2.9235
posterior probability				
$P(\delta_1 = 0 \mid Y)$	0.8307		0.8308	
$P(\delta_2 = 0 \mid Y)$	0.7403		0.7282	
$P(\delta_2 = 0 \mid Y)$	0.3294		0.3357	

Table 5 also shows the results of the incomplete data sets for ignoring the missingness and the results are quite similar to those for the complete data set.

It should be mentioned that in this application, the joint probability $\Pr\{\Sigma_{i=1}^3 \delta_i = 0 \mid Y\}$ for the complete data is 23.17%. For the incomplete data with 10% missing values, the probability $\Pr\{\Sigma_{i=1}^3 \delta_i = 0 \mid Y\}$ is 20.42% by multiple imputations

and 24.73% by discarding the missing values, respectively; for the incomplete data with 20% missing values, the probability $\Pr\{\Sigma_{i=1}^{k-1}\delta_i = 0 \mid Y\}$ is 22.75% by multiple imputations and 24.95% by discarding the missing values, respectively. Therefore, for the global test, all approaches result in rejecting the null hypothesis.

6. Conclusions and Discussion

We have focused on a Bayesian hierarchical model for multiple comparisons in unbalanced mixed models with a simple order restriction. The Gibbs sampling and Metropolis-within-Gibbs sampling techniques are utilized to obtain parameter estimates and estimates of the posterior probabilities of the equality of the mean pairs. The simulation and application results demonstrate that the model is an effective tool for multiple comparisons. Furthermore, by employing the model, estimation of parameters, multiple imputations, and multiple comparisons are unified. In the situation where the missing values are MCAR or MAR, multiple imputations and multiple comparisons can be bonded in one setting and it can perform more effectively than just discarding the missing values. However, these are just examples, thus general conclusions may not be assumed.

With respect to Type I errors for the hypothesis testing in the mixed models with missing values, our simulation results (not presented here) demonstrate that using the proposed hierarchical mixed model with the decision rules in (3.1) and in (3.2) can generally keep pairwise Type I errors and experimentwise error rate (EER) quite small (below 0.05) and smaller than pairwise Type I errors and EER via the frequentist method. Meanwhile, this proposed test procedure can retain test powers appropriately large.

Informative prior distributions for the variance components can affect the results of multiple comparisons although the further research about informative priors are not carried out here. The discussion and theoretical results of Hill (1965) indicated that in some settings the likelihood contributes very little to the posterior distribution of the variance components, and the applicability of informative prior distributions for σ_τ^2 and σ^2 was indicated in Chaloner (1987). In situations where a researcher is very confident that the informative priors more closely reflect the true conditions than noninformative ones, the use of informative priors may lead to superior results.

To extend this methodology in our future research, we plan to modify the hierarchical model for the other order-restricted assumptions or to extend the proposed hierarchical model in additional mixed modeling frameworks of practical interest.

Acknowledgments

The author would like to express her appreciation to the referee for providing thoughtful and insightful comments which helped to improve the original version of this manuscript.

References

Chaloner, K. (1987). A Bayesian approach to the estimation of variance components for the unbalanced one-way random model. *Technometrics* **29**, 323-337.

Fitzmaurice, G., Laird, N. and Ware, J. (2004). *Applied Longitudinal Analysis.* New York: John Wiley & Sons.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972–985.

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

Gottardo, R. and Raftery, A. E. (2008). Markov Chain Monte Carlo with mixtures of mutually singular distributions. *Journal of Computational and Graphical Statistics* **17**, 949–975.

Hill, B. M. (1965). Inference about variance components in the one-way model. *Journal of the American Statistical Association* **60**, 806–825.

Kim, S. and Sun, D. (2001). Intrinsic priors for model selection using an encompassing model. *Lifetime Data Analysis* **6**, 251–269.

Little, R. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* John Wiley & Sons, New York.

Molitor, J. and Sun, D. (2002). Bayesian analysis under order functions of parameters. *Environmental and Ecological Statistics* **9**, 179–193.

Mukerjee, H. (1988). Order restricted inference in a repeated measures model. *Biometrika* **75**, 616–617.

Nashimoto, K. and Wright, F. T. (2008). Bayesian multiple comparisons of simply ordered means using priors with a point mass. *Computational Statistics and Data Analysis* **52**, 5143–5153.

Pauler, D. K., Wakefield, J. C. and Kass, R. E. (1999). Bayes factors and approximations for variance component models. *Journal of the American Statistical Association* **94**, 1242–1253.

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

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research* **8**, 3–15.

Shang, J., Cavanaugh, J.E. and Wright, F.T. (2008). A Bayesian multiple comparison procedure for order-restricted mixed models. *International Statistical Review* **76**, 268-284.

Singh, B. and Wright, F. T. (1990). Testing for and against an order restriction in mixed-effects models. *Statistics and Probability Letters* **9**, 195–200.

TLC Trial Group. (2000). Safety and efficacy of succimer in toddlers with blood levels of 20-44 μg/dL. *Pediatric Research* **48**, 593–599.

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

Receied November 18, 2009; accepted May 18, 2010.

Junfeng Shang
Department of Mathematics and Statistics
Bowling Green State University
450 Math Science Building, Bowling Green, OH 43403, USA
jshang@bgsu.edu