A Bayesian Multiple Comparison Procedure for Order-Restricted Mixed Models

Junfeng Shang[1], **Joseph E. Cavanaugh**[2], **Farroll T. Wright**[3],

[1] Department of Mathematics and Statistics, 450 Math Science Building, Bowling Green State University, Bowling Green, OH 43403, USA
E-mail: jshang@bgnet.bgsu.edu

[2] Department of Biostatistics, University of Iowa, Iowa City, IA 52242, USA
E-mail: joe-cavanaugh@uiowa.edu

[3] Department of Statistics, University of Missouri, Columbia, MO 65211, USA
E-mail: WrightF@missouri.edu

Abstract: A Bayesian hierarchical mixed model is developed for multiple comparisons under a simple order restriction. The model facilitates inferences on the successive differences of the population means, 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 Markov Chain Monte Carlo (MCMC) techniques to obtain parameter estimates and estimates of the posterior probabilities that any two of the means are equal. The latter estimates allow one both to determine if any two means are significantly different and to test the homogeneity of all of the means. We investigate the performance of the model-based inferences with simulated data sets, focusing on parameter estimation and successive-mean comparisons using 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. Our results show that the proposed hierarchical model can effectively unify parameter estimation, tests of hypotheses, and multiple comparisons in one setting.

Key words: Hierarchical model, MCMC, simple order restriction

1 Introduction

In certain applications, researchers may believe before the data is collected that the underlying parameters satisfy an order restriction. 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. The investigators may believe that for the dosages considered, the mean response is 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.

In a frequentist context, for the one-way analysis of variance (ANOVA) with a simple order restriction, a number of procedures for hypothesis testing and multiple comparisons have been developed. Hypothesis testing results are discussed in Chapter 2 of Robertson, Wright, and Dykstra (1988). Marcus and Peritz (1976), Williams (1977), and Marcus (1982) obtained lower confidence bounds for monotone contrasts. However, the only pairwise comparison that arises from a monotone contrast is based on $\mu_k - \mu_1$. Considering pairwise comparisons in this setting, Hayter (1990) developed a one-sided Studentized-range test, and also showed the resulting order-restricted multiple comparison procedure is substantially more efficient than its traditional counterpart when the assumed order restriction holds. Liu (2001) developed another one-sided multiple comparison procedure that is very efficient when comparing μ_1 and μ_k, but not when comparing μ_i and μ_{i+1}. Liu, Lee, and Peng (2002) proposed improved two-sided simultaneous confidence intervals ("max-min" multiple comparison procedures) for simply ordered means. Nashimoto and Wright (2005) studied several pairwise multiple comparison procedures for a simple order and recommended a two-stage procedure.

Frequentist results for a mixed model with a simple order on the treatment effects are discussed in Mukerjee (1988) and Singh and Wright (1990). Mukerjee's results show that multiple comparison techniques for the one-way ANOVA lead to such techniques for a mixed

model provided there is no missing data.

In the 1990s, significant developments were made in Bayesian methods for order-restricted inferences. Gelfand, Hills, Racine-Poon, and Smith (1990) provided Bayesian estimates of order-restricted normal means with arbitrary variances. Gelfand, Smith, and Lee (1992) extended these results to other types of inequality constraints. Pauler, Wakefield, and Kass (1999) used Bayes factors to test hypotheses involving inequality constraints. In life-testing models, Kim and Sun (2001) studied 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 means and variances. These Bayesian developments incorporate ordering information, but none treat estimation, testing, and multiple comparisons in a unified setting. However, Dunson and Herring (2003) obtained such results for a life-testing framework.

In this paper, we propose a hierarchical model in the setting of a mixed model based on a fixed treatment effect and a random subject effect. The model assumes that the treatment means satisfy a simple order assumption. On the successive differences of the means, we place independent prior distributions that are mixtures of an exponential distribution and a discrete distribution with its entire mass at zero. We utilize Markov Chain Monte Carlo (MCMC) methods to obtain parameter estimates and estimates of the posterior probabilities that any two of the means are equal. The latter estimates allow one both to determine if any two means are significantly different and to test the homogeneity of all of the means.

This paper is arranged as follows. In Section 2, we present our hierarchical model. Section 3 outlines approaches to hypothesis testing under the hierarchical model and reviews the frequentist method of Mukerjee (1988). In Section 4, we investigate the performance of the model-based inferences with simulated data sets, focusing on parameter estimation and successive-mean comparisons using 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. The results are compared with those obtained using frequentist methods. Some concluding remarks are given in Section 5. The derivations of the full conditional posterior distributions for the hierarchical model are presented in the Appendix, along with the steps of the Gibbs sampling procedure.

2 A Bayesian Hierarchical Mixed Model for Multiple Comparisons

2.1 Model

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

$$y_{ij} = \mu_i + b_j + \varepsilon_{ij}, \qquad i = 1, \ldots, k, \qquad j = 1, \ldots, m, \tag{2.1}$$

where y_{ij}, μ_i, b_j, and ε_{ij} are all scalars. Here, y_{ij} denotes the response observed on the jth subject under 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 are distributed as $\mathrm{N}(0, \sigma_\tau^2)$, that the ε_{ij} are distributed as $\mathrm{N}(0, \sigma^2)$, and that the $m(k+1)$ variables b_j and ε_{ij}, $i = 1, \ldots, k$ and $j = 1, \ldots, m$, are all independent.

We can re-express model (2.1) in a more succinct format by writing

$$y_j = \mu + z b_j + \varepsilon_j, \qquad j = 1, \ldots, m, \tag{2.2}$$

where y_j, μ, z, and ε_j are all vectors, and b_j is a scalar. Specifically, y_j denotes a $k \times 1$ vector of responses observed on the jth subject; μ is the $k \times 1$ vector $(\mu_1, \ldots, \mu_k)'$; z is a $k \times 1$ vector consisting of all 1's; b_j is as defined previously; and ε_j is a $k \times 1$ error vector distributed as $\mathrm{N}(0, \sigma^2 I)$.

With μ_1 denoting the first mean and $\mu_1 \leq \cdots \leq \mu_k$, we parameterize each of the remaining means 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 second mean can be denoted as $\mu_1 + \delta_1$, the third mean as $\mu_1 + \delta_1 + \delta_2$, etc. In general, the ith mean $(2 \leq i \leq k)$ can be denoted as $\mu_1 + \delta_1 + \cdots + \delta_{i-1}$.

Our goal is to compare the population means μ_i, $i = 1, \ldots, k$, in model (2.1) via a Bayesian hierarchical model.

2.2 Priors and Hyperpriors

Because δ_i is positive or zero, 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. With $I_{(A)}$ denoting the indicator function corresponding to the event A, the density function for δ_i can be represented as

$$[\delta_i \mid \rho_i, \theta_i] = \rho_i I_{(\delta_i=0)} + \Delta I_{(\delta_i>0)}, \tag{2.3}$$

where $\Delta = (1-\rho_i)\frac{1}{\theta_i}\exp\left\{-\frac{\delta_i}{\theta_i}\right\}$.

For the hyperparameter ρ_i, which represents the prior probability of $\delta_i = 0$, we will utilize a beta hyperprior distribution, i.e., BETA(α_o, β_o). Note that this hyperprior distribution will be uniform when $\alpha_o = 1$ and $\beta_o = 1$. For θ_i, we consider an inverse-gamma hyperprior, $\theta_i \sim \text{IG}(a_o, b_o)$, which has density function

$$[\theta_i \mid a_o, b_o] = \frac{1}{\Gamma(a_o)(b_o)^{a_o}} \frac{\exp(-\frac{1}{b_o\theta_i})}{{\theta_i}^{a_o+1}}, \quad \theta_i > 0, \quad i = 1, \ldots, k-1. \tag{2.4}$$

For μ_1, we use a conjugate prior, namely

$$\mu_1 \quad \sim \quad \text{N}(\mu_o, \tau_o^2). \tag{2.5}$$

We choose a vague normal prior with an arbitrary mean μ_o and a large variance τ_o^2.

For the variance components, we consider a noninformative prior distribution as a joint prior for σ_τ^2 and σ^2,

$$\pi(\sigma_\tau^2, \sigma^2) \propto \frac{1}{\sigma^2(k\sigma_\tau^2 + \sigma^2)}.$$

This prior was used by Box and Tiao (1992) in the balanced mixed model case. The model considered here also assumes balance, so we can apply this prior to our model. If we let $\tau^2 = k\sigma_\tau^2 + \sigma^2$, the prior distribution will be

$$\pi(\tau^2, \sigma^2) \propto \frac{1}{\sigma^2 \tau^2}. \tag{2.6}$$

With this prior, we need to obtain the full conditional posterior distribution for τ^2 instead of σ_τ^2, and then based on σ^2 and τ^2, we can obtain σ_τ^2.

Kass and Wasserman (1996) provide a good review of prior distributions and relevant historical perspectives, with an emphasis on Jeffreys' rules and a complete discussion of the evolution of his viewpoint. In practice, many Bayesian analyses are performed with noninformative priors to reduce subjectivity and to allow the data to play a dominant role in the analysis. In the setting considered in this paper, it seems quite reasonable to assign a mixture distribution prior to δ_i, $i = 1, \ldots, k-1$, and a flat prior to the first population mean μ_1. Because the full conditional posterior distribution for θ_i is improper when based on a noninformative prior, we choose a flat informative prior for θ_i. For the hyperparameter ρ_i, we can adjust the parameters α_o and β_o in our approach to reduce the Type I error rate.

With the preceding independent priors, all full conditional posterior distributions (except for δ_i) are standard ones, such as normal, inverse-gamma, and beta distributions. The derivations of the full conditional posterior distributions are presented in Appendix A. Gibbs sampling is employed to estimate the parameters and the posterior probabilities of the successive-mean differences. In the Gibbs sampling iterations, τ^2 and σ^2 are generated from the corresponding full conditional posterior distributions, and then σ_τ^2 is updated by $(\tau^2 - \sigma^2)/k$. The details of the Gibbs sampling are presented in Appendix B.

The primary goal for the proposed hierarchical model is to compute the posterior probabilities of $\delta_i > 0$ and of $\delta_i = 0$ $(1 \le i \le k-1)$. Using the magnitude of these two probabilities in conjunction with a decision rule, we can test the hypotheses $H_{oi} : \mu_i = \mu_i + \delta_i$ versus $H_{1i} : \mu_i < \mu_i + \delta_i$ for $i = 1, \ldots, k-1$. Also, with the hypotheses $H_o : \mu_1 = \cdots = \mu_k$ and

$H_1 : \mu_1 \leq \cdots \leq \mu_k$ with $\mu_1 < \mu_k$, we can conduct the global test of H_o versus H_1.

3 Approaches to Hypothesis Testing

3.1 Hypothesis Testing via the Posterior Probability

Under the proposed hierarchical model, we use posterior probabilities to test hypotheses about the means. Consider the hypotheses for successive pairwise comparisons of the means, $H_{oi} : \mu_i = \mu_i + \delta_i$ versus $H_{1i} : \mu_i < \mu_i + \delta_i$, or $H_{oi} : \delta_i = 0$ versus $H_{1i} : \delta_i > 0$ for $i = 1, \ldots, k-1$. If we assign equal prior probabilities to H_{oi} and H_{1i}, we may reject H_{oi} when $\Pr\{\delta_i = 0 \mid Y\}$ is less than some predetermined cutoff value. This cutoff value could be 0.5, as in a traditional scheme, or another value dictated by the context of the application.

For the global test of $H_o : \mu_1 = \cdots = \mu_k$ versus $H_1 : \mu_1 \leq \cdots \leq \mu_k$ with $\mu_1 < \mu_k$, we may naturally accept H_1 if at least one of the pairwise tests accepts H_{1i}, $i = 1, \ldots, k-1$; that is, accept H_1 if $\min_{1 \leq i \leq k-1}\{\Pr\{\delta_i = 0 \mid Y\}\}$ is less than the aforementioned cutoff value for rejecting H_{oi}.

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. Note that for the preceding rules, the initial probabilities of the individual null hypotheses, $H_{oi} : \delta_i = 0$, $1 \leq i \leq k-1$, are traditionally set at $\Pr\{H_{oi}\} = 0.5$. Some might argue that this cutoff value should be even larger because the null hypothesis is often the "established theory" (Berger and Sellke, 1987, page 115).

However, if several tests are conducted in the same experiment, assigning the initial probabilities of the individual null hypotheses to be 0.5, i.e., $\Pr\{H_{oi}\} = 0.5$, may result in too many false rejections of the individual null hypotheses. Therefore, we may need to adjust the prior probability of the null hypotheses so that the Type I error rate can be effectively controlled.

Following the intuitive rule proposed by Westfall, Johnson, and Utts (1997), in order to

obtain $\Pr\{H_o\} = 0.5$ for the model with a total of $k-1$ mutually independent priors for the δ_i's, we calibrate the prior probabilities for H_{oi} via $\Pr\{H_{oi}\}^* = 0.5^{1/(k-1)}$.

We observe that

$$\begin{aligned}\Pr\{H_{oi}\} = \Pr\{\delta_i = 0\} &= E[E(I_{\{\delta_i=0\}} \mid \rho_i, \theta_i)] \\ &= E(\rho_i) = \frac{\alpha_o}{\alpha_o + \beta_o}.\end{aligned}$$

Thus, if one wants the initial prior probabilities of $\Pr\{H_{oi}\} = 0.5$, then one could choose $\alpha_o = \beta_o$ and select 1 as the common value. Note that for any given k, there exist an infinite number of choices for α_o and β_o that satisfy $\Pr\{H_{oi}\}^* = 0.5^{1/(k-1)} = \frac{\alpha_o}{\alpha_o+\beta_o}$.

For pairwise comparisons, we adopt the conventional scheme of using 0.5 as the decision criterion for the posterior probability. For $H_{oi} : \mu_i = \mu_i + \delta_i$ versus $H_{1i} : \mu_i < \mu_i + \delta_i$, with $1 \le i \le k-1$, 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 use of $\Pr\{H_{oi}\}^*$.

For the global test of $H_o : \mu_1 = \mu_2 = \cdots = \mu_k$ versus $H_1 : \mu_1 \le \cdots \le \mu_k$ with $\mu_1 < \mu_k$, we declare H_1 if at least one of the pairwise tests declares H_{1i}, i.e., if

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

One could conduct the global test based on the joint posterior probability of the δ_i's, by rejecting the global null H_o if

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

However, we recommend conducting 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_o 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:

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

Thus, the test based on (3.2) has a smaller Type I error rate and smaller power than the test based on (3.3). However, the results of our simulation study and the results in the application considered in Section 4 suggest the procedure based on (3.1) and (3.2) has reasonable power.

3.2 Hypothesis Testing via the Frequentist Method

Based on Mukerjee's (1988) results, Singh and Wright (1990) proposed a method to conduct the global test for the means in mixed-effects models. Let

$$\begin{aligned}\bar{y}_{i.} &= \frac{1}{m}\sum_{j=1}^{m} y_{ij},\ i = 1,\ldots,k,\\ \bar{y}_{.j} &= \frac{1}{k}\sum_{i=1}^{k} y_{ij},\ j = 1,\ldots,m,\end{aligned}$$

and $\hat{\mu}_T$ be the overall mean for all the observations. Let $\hat{\mu}^* = (\hat{\mu}_1^*,\ldots,\hat{\mu}_k^*)$ denote the maximum likelihood estimator (MLE) of μ subject to the restriction $\mu_1 \leq \cdots \leq \mu_k$, which can be found by the *minimum-lower-sets algorithm* (Brunk, 1955, and Brunk, Ewing, and Utz, 1957) or by the *pool-adjacent-violator algorithm* (Ayer, Brunk, Ewing, Reid, and Silverman, 1955).

If one believes that the fixed effects are nondecreasing, one could test the statistical significance of the fixed effects by testing $H_o : \mu_1 = \cdots = \mu_k$ versus $H_1 : \mu_1 \leq \cdots \leq \mu_k$ with $\mu_1 < \mu_k$. If all the parameters are unknown, with $\nu = (k-1)(m-1)$ and

$$S_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{m}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \hat{\mu}_T)^2,$$

the likelihood-ratio test (LRT) statistic is

$$S_{01} = \frac{\nu \sum_{i=1}^{k} (\hat{\mu}_i^* - \hat{\mu}_T)^2}{\sum_{i=1}^{k} (\hat{\mu}_i - \hat{\mu}_i^*)^2 + S_1^2/m}.$$

The LRT rejects H_o for large values of S_{01}. Critical values for S_{01} can be found in the Appendix of Robertson, Wright, and Dykstra (1988).

4 Examples Based on Simulated Data and an Application

In this section, we evaluate the performance of the model-based inferences using simulated data, where the tests of hypotheses are implemented via the posterior probabilities in (3.1) and (3.2). We then illustrate the utility of the model in a biostatistical application.

4.1 Brief Description of Examples Based on Simulated Data

As previously mentioned, we compute the posterior probabilities for the model by Gibbs sampling. Since the convergence is not slow, the number of burn-in samples is taken to be 3,000, and the subsequent 7,000 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 $\Pr(\delta_i = 0 \mid Y)$. We consider $k = 3$ means.

For the prior on μ_1, the hyperparameters μ_o and τ_o^2 do not have much effect on the results provided that the normal prior distribution is relatively flat, i.e., τ_o^2 is very large. In our simulations, $\mu_o = 0$ and $\tau_o^2 = 100$. To make the prior on θ_i flat, hyperparameters $a_o = 2.2$ and $b_o = 0.05$ are used. With $k = 3$, the prior probability of H_{oi} must be calibrated via $\Pr\{H_{oi}\}^* = 0.5^{1/(3-1)} = 0.707$. This is closely approximated by the choice $\rho_i \sim \text{BETA}(2.4, 1)$, which gives $\Pr\{H_{oi}\}^* = E(\rho_i) = 0.706$.

For the simulated data, we choose the mean parameters $\mu = (10, 11, 12)'$. We consider both small variances with $\sigma_\tau^2 = 2$ and $\sigma^2 = 1$, and large variances with $\sigma_\tau^2 = 100$ and $\sigma^2 = 80$. For each of these variance specifications, we generate a moderate-size sample ($m = 30$) and a large-size sample ($m = 100$). Therefore, four simulated data sets are featured.

For each of the four data sets, we compute the posterior model parameter estimates and their variances, as well as the corresponding posterior probabilities. For reference purposes, we also compute the unrestricted MLEs of the model parameters via the expectation-maximization (EM) algorithm.

4.2 Results of Analyses for Simulated Data

For $m = 30$ and $m = 100$, $\sigma_\tau^2 = 2$ and $\sigma^2 = 1$, Table 1 features the parameter estimates and their variances, and the posterior probabilities of $\delta_i = 0$ for each i. 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.

For $m = 30$, the mean parameter estimates are $(9.9328, 10.8054, 11.5673)'$, and the variance estimates are 2.4211 and 1.3586, respectively. These estimates are quite close to the MLEs computed by the EM algorithm, which are $\hat{\mu} = (9.7233, 10.8451, 11.6919)'$, $\hat{\sigma}_\tau^2 = 2.2451$, and $\hat{\sigma}^2 = 1.0272$. Among the variances of the parameter estimates in Table 1, the variance of the σ_τ^2 estimate is the largest. Note that σ_τ^2 reflects between subject variability; its estimate is very sensitive to changes in the data. Based on the posterior probabilities, the pairwise tests favor H_{11} and H_{12}, which means that there is a significant difference between μ_1 and μ_2 and between μ_2 and μ_3.

It is of interest to compare this conclusion to the result of the frequentist test in subsection 3.2. Assuming the variances are unknown, we obtain $S_{01} = 55.04$ based on 58 degrees of freedom. With $\alpha = 0.01$, the critical value is 7.211 from Table A.6 of Robertson, Wright, and Dykstra (1988). The LRT therefore rejects H_o, the null hypothesis of equal means. Thus, the Bayesian and frequentist methods both reject H_o.

For $m = 100$, we note that the parameter estimates for means and variances are quite close to both the true parameters and the EM algorithm estimates. Because the posterior probabilities are all far less than 0.5, we observe that the Bayesian tests reject $H_{oi}, i = 1, 2$.

For the frequentist approach, the LRT statistic for testing $H_o : \mu_1 = \mu_2 = \mu_3$ is $S_{01} =$

Table 1: Posterior Parameter Estimates and MLEs

$\mu = (10, 11, 12)'$, $\sigma_\tau^2 = 2$, $\sigma^2 = 1$

	$m = 30$			$m = 100$		
parameter	mean	variance	MLE	mean	variance	MLE
μ_1	9.9328	0.2991	9.7233	9.8565	0.0595	9.8489
δ_1	0.8726	0.4952	1.1218	1.0052	0.0664	1.0059
δ_2	0.7619	0.3983	0.8468	1.0567	0.0519	1.0591
σ_τ^2	2.4211	0.6655	2.2451	1.8181	0.1051	1.7672
σ^2	1.3586	0.1215	1.0272	1.2148	0.0182	1.1583
posterior probability						
$\Pr(\delta_1 = 0 \mid Y)$	0.2997			0.0700		
$\Pr(\delta_2 = 0 \mid Y)$	0.3131			0.0011		

Table 2: Posterior Parameter Estimates and MLEs

$\mu = (10, 11, 12)'$, $\sigma_\tau^2 = 100$, $\sigma^2 = 80$

	$m = 30$			$m = 100$		
parameter	mean	variance	MLE	mean	variance	MLE
μ_1	8.3615	7.6138	7.8269	9.7456	2.0450	9.0951
δ_1	0.8989	4.1447	2.0890	0.4971	1.2946	1.0533
δ_2	0.4160	1.6084	-0.3704	0.6670	1.6760	1.5289
σ_τ^2	127.9883	1951.7600	114.3849	85.4398	295.2145	82.5628
σ^2	87.5824	294.5952	82.1771	95.2454	91.0982	92.6610
posterior probability						
$\Pr(\delta_1 = 0 \mid Y)$	0.7643			0.7835		
$\Pr(\delta_2 = 0 \mid Y)$	0.8446			0.7232		

182.27 ($\nu = 198$), and the $\alpha = 0.01$ critical value is 6.940. Therefore, the test results for the Bayesian and frequentist approaches are consistent.

Table 2 features the results when the variance components are $\sigma_\tau^2 = 100$ and $\sigma^2 = 80$. For the two simulated data sets ($m = 30, m = 100$), the posterior probabilities of $\delta_i = 0$ ($i = 1, 2$) are larger than 0.5. Thus, we retain $H_{oi}, i = 1, 2$. When the frequentist method

is applied to these two data sets, the LRT statistic S_{01} is 0.8526 for the smaller sample and 3.6022 for the larger sample. The $\alpha = 0.05$ critical values are 3.931 and 3.852, respectively. Thus, the Bayesian and frequentist global tests lead to similar conclusions, although S_{01} for the larger sample approaches statistical significance.

Comparing the posterior parameter estimates with the unrestricted MLEs in Table 2, relatively large differences are detected. Here, the true variance components are quite large compared to the means, which allows the sample means to violate the simple order restriction. The posterior parameter estimates are evaluated under the order restriction, whereas the unrestricted MLEs are not. These MLEs show that the simple order is violated, but the posterior estimates adhere to the restriction. Of course, the order-restricted MLEs (not featured here) do not violate the simple order restriction.

A simulation study has been performed to study the Type I errors of the proposed decision rules for pairwise and global tests. The simulation results, which are not presented here, are briefly summarized. Our results demonstrate that using the proposed hierarchical mixed model with the decision rules in (3.1) and (3.2) can generally keep the pairwise Type I error rate and experimentwise error rate (EER) quite small (below 0.05) and maintain reasonable test power. For instance, pairwise Type I error rates and EERs can be made smaller than those via the frequentist method, while test powers can be retained at an appropriately high level. As a result, provided that the treatment means in the mixed model satisfy a simple order restriction, we can use the proposed procedure to test hypotheses and avoid logical inconsistencies between the pairwise tests and the global test.

The hierarchical model with informative prior distributions for the variance components could be used for the multiple comparisons, but we do not study informative priors here. The discussion and theoretical results of Hill (1965) indicate that in some settings the likelihood contributes very little to the posterior distribution of the variance components. The applicability of informative prior distributions for σ_τ^2 and σ^2 is considered by Chaloner (1987). In

situations where a researcher is very confident that informative priors are more appropriate than noninformative ones, the use of informative priors may lead to superior results.

An extensive study of informative priors could also be quite valuable. There are many useful types of informative priors that could be applied in the current framework. For instance, inverse-exponential distributions for the variance components could be utilized. One could also study the effects of informative priors (such as inverse-exponential distributions) for the θ_i.

4.3 Application

To illustrate the utility of our model, we present an application based on data from a study designed to reduce lead blood concentrations in children with elevated levels. The study is a placebo-controlled, randomized trial 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, Laird, and Ware (2004).

The study involved children between 12 and 33 months in age who had blood lead levels between 20 and 44 μg/dL. Most of the children were African-American (77%) and lived in deteriorating inner city housing. The participants were randomly assigned to two groups. 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. Participants in the "treatment" group were then provided with succimer capsules, whereas those in the "control" group were provided with a placebo. Blood lead levels were measured at baseline (week 0), week 1, week 4, and week 6.

In our application, we consider blood lead levels collected for 50 of the children who did not receive the succimer capsules. Since the homes of these children were cleaned using an established TLC regimen, one might expect that the mean blood lead levels of these children would decrease over time, or at least remain the same.

In the context of model (2.1), y_{ij} will denote the blood lead level for child j ($j = 1, \ldots, 50$) at time i ($i = 1, \ldots 4$). Thus, μ_i will denote the mean blood lead level at time i. However, since we anticipate that the means are nonincreasing over time, to apply the simple ordering $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$ to our setting, we need to reverse the time order associated with our index i. Accordingly, time 1 ($i = 1$) will correspond to week 6, time 2 ($i = 2$) to week 4, time 3 ($i = 3$) to week 1, and time 4 ($i = 4$) to baseline. The simple ordering $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$ then represents the hypothesis that the mean blood levels are not elevating as time elapses past baseline.

The basic repeated measures model (2.1) implies a compound symmetric structure for the variance-covariance matrix of the subject-specific response vectors represented in (2.2). Although this type of covariance structure is often inappropriate for longitudinal data, our preliminary analyses shows that the compound symmetric structure adequately describes the control subject measurements. (For instance, the Bayesian information criterion favors the compound symmetric covariance structure over the general covariance structure.) Model (2.1) 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 set the prior probability $\Pr\{H_o\} = 0.5$ with $k = 4$, the prior probability of each H_{oi} must be $0.5^{1/(4-1)} = 0.7937$. The choice $\rho_i \sim \text{BETA}(4, 1)$ yields $\Pr\{H_{oi}\}^* = E(\rho_i) = 0.80$.

The results of our data analysis are listed in Table 3. 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 (23.7911, 24.1309, 24.5864, 26.1180)$'$. 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 variance estimates for σ_τ^2 and σ^2 resulting from our analyses are 25.4409 and 6.2763, respectively. For comparison, the MLEs obtained by the EM algorithm are $\hat{\sigma}_\tau^2 = 24.0226$ and

$\hat{\sigma}^2 = 5.3713$. Thus, the between-subject variation is quite high relative to the within-subject variation, reflecting a substantial degree of subject heterogeneity.

Table 3: Posterior Parameter Estimates for Blood Lead Level Application

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	23.7911	25.4409	6.2763	0.3398	0.4555	1.5316
variance	1.2998	32.5558	1.7157	0.5216	0.5493	1.1652
posterior probability $\Pr(\delta_i = 0 \mid Y)$				0.7752	0.6688	0.2346

The posterior probabilities and the δ_i estimates in Table 3 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. From a physiological perspective, this conclusion is sensible. The most dramatic decrease in the blood lead level should occur during a period of time immediately following the cleaning of the home, reflecting the attenuation of environmental exposure to lead particulates. Over time, the level should begin to stabilize.

Based on the results in Table 3, we might conclude that the means are decreasing over time, yet since the rate of decrease is attenuating, the differences between μ_3, μ_2, and μ_1 are not large enough to achieve significance. Of course, our prior assigns positive probability to the outcomes $\mu_3 = \mu_2$ ($\delta_2 = 0$) and $\mu_4 = \mu_3$ ($\delta_3 = 0$). Yet before the collection of data, we would not know whether the cleaning regimen will have a clinically meaningful effect on lowering the blood levels of the children. By placing positive prior probability on the outcomes $\delta_i = 0$, we are requiring that the weight of evidence in favor of $\delta_i > 0$ must be sufficiently strong to reject the null hypothesis of equality. In the case of μ_3, μ_2, and μ_1, the evidence is insufficient to overcome the prior.

For the global test, the values of the posterior probabilities for $\delta_i = 0$ favor the alternative hypothesis H_1. For the frequentist test, since the variance parameters need to be estimated,

we use the test statistic S_{01}. The value of this statistic is $S_{01} = 36.2550$ ($\nu = 147$), and the $\alpha = 0.01$ critical value is 7.858. Hence, the frequentist method also strongly favors H_1.

The global test based on the joint posterior probability $\Pr\{\bigcap_{i=1}^{k-1} \delta_i = 0 \mid Y\}$ favors H_1 as well. This joint probability is 0.1692. The inequality $\Pr\{\bigcap_{i=1}^{k-1} \delta_i = 0 \mid Y\} < 0.5$ therefore holds, providing further reinforcement of our decision to reject the global null hypothesis.

Figure 1 features histograms that illustrate the posterior distributions for δ_1, δ_2 and δ_3. These histograms are constructed based on the 7,000 iterations used to estimate the parameters in the Gibbs sampling (i.e., the iterations after burn-in). The iterates are generated from the full conditional posterior distributions of the δ_i's, and are therefore distributed in accordance with the posterior distributions of the δ_i's.

The frequencies for $\delta_i = 0$ and $\delta_i > 0$ are of particular interest. The heights of the leftmost bars in the histograms suggest the existence of a preponderance of zero iterates for both δ_1 and δ_2, yet not for δ_3. With δ_3, the majority of the iterates correspond to the event $\delta_3 > 0$. Therefore, Figure 1 provides additional evidence to support the previous conclusions: that is, the posteriors imply that $\delta_3 > 0$ ($\mu_4 > \mu_3$), yet also suggest that $\delta_2 = 0$ ($\mu_3 = \mu_2$) and $\delta_1 = 0$ ($\mu_2 = \mu_1$). The posteriors also support the rejection of the global null hypothesis.

5 Discussion

We have focused on multiple comparisons in a two-way ANOVA mixed model, or basic repeated measures model. The simulation and application results demonstrate that the proposed Bayesian hierarchical model is an effective tool for multiple comparisons. Furthermore, by employing the model, estimation of parameters, tests of the global hypotheses, and multiple comparisons are unified. Thus, with a common decision criterion, such as one based on the posterior probability, the use of this hierarchical model can maintain logical consistency between the global test and the pairwise tests.

To extend this methodology in our future research, we plan to modify the hierarchical model for the tree order assumption. For a tree order, we may set the control mean as μ_o,

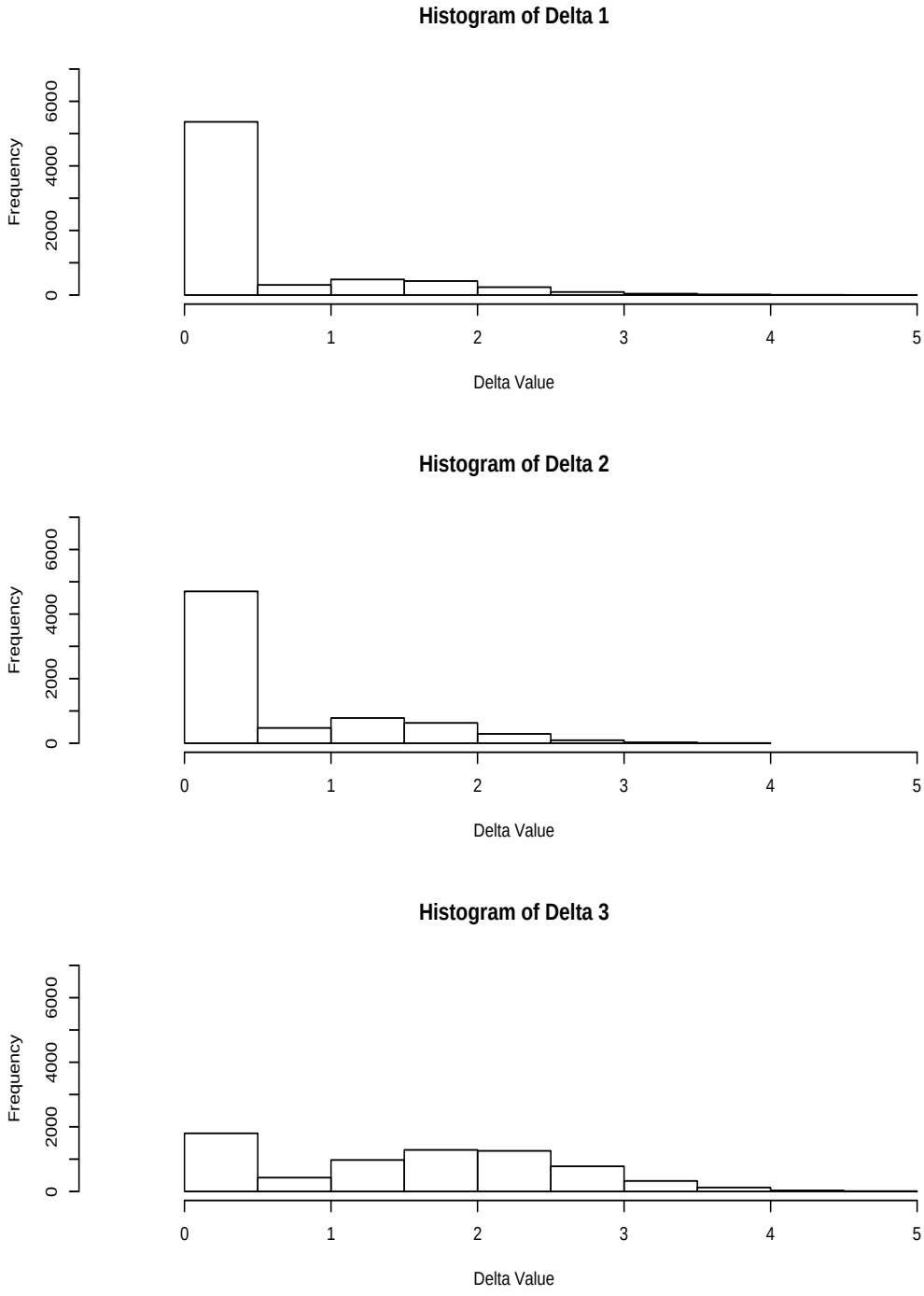

Figure 1: Histograms Representing Posterior Probability Distributions for δ_1, δ_2, and δ_3 in Blood Lead Level Application

and each treatment mean may be defined as $\mu_o + \delta_i$, $1 \leq i \leq k$, with $\delta_i \geq 0$. The same priors can be considered as for the simple order setting.

The hierarchical model we propose can be used to deal with missing values in unbalanced data. For instance, multiple imputation and the computation of inferential results for multiple comparisons can be merged into one MCMC algorithm. This approach is quite feasible provided that the conditional predictive distribution for missing values and the full conditional posterior distributions for parameters are accessible.

Finally, our work in this paper only considers the comparison of multiple treatment means for a specific type of linear mixed model: the basic repeated measures model, or two-way ANOVA mixed model. For additional mixed modeling frameworks of practical interest, it would be desirable to formulate a Bayesian hierarchical model and to derive the appropriate posterior distributions.

Acknowledgements

The authors wish to express their appreciation to an anonymous referee for valuable suggestions that served to improve the original version of this manuscript. This research was supported by a grant from the University of Missouri Research Board, and was also partially supported by a Research Incentive Grant awarded to the first author at Bowling Green State University.

Appendix A: Full Conditional Posterior Distributions

First, we need to obtain the density function for the data. With reference to model (2.2), we have

$$y_j \mid \mu_1, \{\delta_i\}, \sigma_\tau^2, \sigma^2 \;\sim\; \mathrm{N}(\mu, z\sigma_\tau^2 z' + \sigma^2 I), \;\; j = 1, \ldots, m.$$

Let $Y = (y_1', \ldots, y_m')'$ and $\delta_o = 0$. Given that the cases are independent, the joint density function of the data vector Y is given by

$$[Y \mid \mu_1, \{\delta_i\}, \sigma_\tau^2, \sigma^2] \propto (\sigma^2)^{-\frac{m(k-1)}{2}} (k\sigma_\tau^2 + \sigma^2)^{-\frac{m}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[s_1 - \frac{\sigma_\tau^2}{k\sigma_\tau^2 + \sigma^2} s_2 \right] \right\}, \quad \text{(A.1)}$$

where

$$\begin{aligned} s_1 &= \sum_{j=1}^{m} \sum_{i=1}^{k} \left(y_{ij} - \mu_1 - \sum_{l=0}^{i-1} \delta_l \right)^2, \text{ and} \\ s_2 &= \sum_{j=1}^{m} \left(\sum_{i=1}^{k} (y_{ij} - \mu_1 - \sum_{l=0}^{i-1} \delta_l) \right)^2. \end{aligned}$$

Now, we need to obtain the full conditional posterior distributions for the parameters.

1. Full Conditional Posterior Distribution of μ_1

For μ_1, we have

$$\begin{aligned} & \sum_{j=1}^{m} \sum_{i=1}^{k} \left(y_{ij} - \mu_1 - \sum_{l=0}^{i-1} \delta_l \right)^2 \\ \propto \;& \sum_{j=1}^{m} \sum_{i=1}^{k} \left(\mu_1^2 - 2\mu_1 (y_{ij} - \sum_{l=0}^{i-1} \delta_l) \right) = km\mu_1^2 - 2\mu_1 \sum_{j=1}^{m} \sum_{i=1}^{k} \left(y_{ij} - \sum_{l=0}^{i-1} \delta_l \right), \text{ and} \\ & \sum_{j=1}^{m} \left(\sum_{i=1}^{k} (y_{ij} - \mu_1 - \sum_{l=0}^{i-1} \delta_l) \right)^2 = \sum_{j=1}^{m} \left(k\mu_1 - \sum_{i=1}^{k} (y_{ij} - \sum_{l=0}^{i-1} \delta_l) \right)^2 \\ \propto \;& \sum_{j=1}^{m} \left(k^2\mu_1^2 - 2k\mu_1 \left(\sum_{i=1}^{k} (y_{ij} - \sum_{l=0}^{i-1} \delta_l) \right) \right) = k^2 m \mu_1^2 - 2k\mu_1 \sum_{j=1}^{m} \sum_{i=1}^{k} \left(y_{ij} - \sum_{l=0}^{i-1} \delta_l \right). \end{aligned}$$

Let

$$s_3 = \sum_{j=1}^{m} \sum_{i=1}^{k} \left(y_{ij} - \sum_{l=0}^{i-1} \delta_l \right).$$

Then using the density function of Y in (A.1) and the prior for μ_1 in (2.5), we can establish

$$\begin{aligned}
[\mu_1 \quad | \quad & Y, \sigma^2, \sigma_\tau^2, \{\delta_i\}] \\
\propto \quad & \exp\left\{-\frac{\mu_1^2}{2}\left(\frac{1}{\tau_o^2}+\frac{km}{\sigma^2}-\frac{\sigma_\tau^2 k^2 m}{\sigma^2(k\sigma_\tau^2+\sigma^2)}\right)+\mu_1\left[\frac{\mu_o}{\tau_o^2}+\frac{s_3}{\sigma^2}-\frac{\sigma_\tau^2 k s_3}{\sigma^2(k\sigma_\tau^2+\sigma^2)}\right]\right\} \\
= \quad & \exp\left\{-\frac{\mu_1^2}{2}\left(\frac{1}{\tau_o^2}+\frac{km}{k\sigma_\tau^2+\sigma^2}\right)+\mu_1\left(\frac{\mu_o}{\tau_o^2}+\frac{s_3}{k\sigma_\tau^2+\sigma^2}\right)\right\}. \qquad \text{(A.2)}
\end{aligned}$$

Setting

$$\begin{aligned}
u \quad &= \quad \frac{\mu_o}{\tau_o^2}+\frac{s_3}{k\sigma_\tau^2+\sigma^2} \text{ and} \\
v \quad &= \quad \frac{1}{\tau_o^2}+\frac{km}{k\sigma_\tau^2+\sigma^2},
\end{aligned}$$

and completing the square in (A.2), one can show that the full conditional posterior distribution of μ_1 is

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

As mentioned previously, we set $\tau^2 = k\sigma_\tau^2 + \sigma^2$. Making use of the density function of Y in (A.1) and the joint prior for σ^2 and τ^2 in (2.6), one can obtain

$$\begin{aligned}
[\sigma^2, \tau^2 \mid Y, \mu_1, \{\delta_i\}] \quad \propto \quad & (\sigma^2)^{-\frac{k(m-1)}{2}-1}\exp\left\{-\frac{1}{2\sigma^2}\left(s_1-\frac{1}{k}s_2\right)\right\} \\
& (\tau^2)^{-\frac{m}{2}-1}\exp\left\{-\frac{s_2}{2k\tau^2}\right\}. \qquad \text{(A.3)}
\end{aligned}$$

2. Full Conditional Posterior Distribution of σ^2

Based on (A.3), it can be shown that

$$[\sigma^2 \mid Y, \mu_1, \sigma_\tau^2, \{\delta_i\}] = \mathrm{IG}\left(\frac{(k-1)m}{2}, \frac{2}{s_1-\frac{1}{k}s_2}\right).$$

3. Full Conditional Posterior Distribution of τ^2 (for the calculation of σ_τ^2)

As described in Section 2.2, if we let $\tau^2 = k\sigma_\tau^2 + \sigma^2$, it is straightforward to show that

$$[\tau^2 \mid Y, \mu_1, \sigma^2, \{\delta_i\}] = \mathrm{IG}\left(\frac{m}{2}, \frac{2k}{s_2}\right).$$

Note that after knowing τ^2 and σ^2, we can calculate σ_τ^2 by using $\frac{\tau^2-\sigma^2}{k}$.

4. Full Conditional Posterior Distribution of δ_i

For a given δ_i, we have

$$\begin{aligned}
& \sum_{j=1}^{m}\sum_{p=i+1}^{k}\left(y_{pj}-\mu_1-\sum_{q=1}^{p-1}\delta_q\right)^2 \\
= & \sum_{j=1}^{m}\sum_{p=i+1}^{k}\left(\delta_i-(y_{pj}-\mu_1-\sum_{q=1}^{p-1}\delta_q+\delta_i)\right)^2 \\
\propto & (k-i)m\delta_i^2-2\delta_i\sum_{j=1}^{m}\sum_{p=i+1}^{k}\left(y_{pj}-\mu_1-\sum_{q=1}^{p-1}\delta_q+\delta_i\right), \text{ and} \\
& \sum_{j=1}^{m}\left(\sum_{p=i+1}^{k}(y_{pj}-\mu_1-\sum_{q=1}^{p-1}\delta_q+\delta_i-\delta_i)\right)^2 \\
= & \sum_{j=1}^{m}\left((k-i)\delta_i-\sum_{p=i+1}^{k}(y_{pj}-\mu_1-\sum_{q=1}^{p-1}\delta_q+\delta_i)\right)^2 \\
\propto & (k-i)^2m\delta_i^2-2(k-i)\delta_i\sum_{j=1}^{m}\sum_{p=i+1}^{k}\left(y_{pj}-\mu_1-\sum_{q=1}^{p-1}\delta_q+\delta_i\right).
\end{aligned}$$

With regard to (A.1), the preceding relations, and the prior for δ_i in (2.3), for $i=1,\cdots,k-1$, we then have

$$\begin{aligned}
& [\delta_i \mid Y,\mu_1,\sigma^2,\sigma_\tau^2,\{\delta_1,\cdots,\delta_{i-1},\delta_{i+1},\cdots,\delta_{k-1}\},\rho_i,\theta_i] \\
\propto & \exp\left\{-\frac{1}{2\sigma^2}\left[\left((k-i)-\frac{\sigma_\tau^2(k-i)^2}{k\sigma_\tau^2+\sigma^2}\right)m\delta_i^2-2\delta_i\left(1-\frac{(k-i)\sigma_\tau^2}{k\sigma_\tau^2+\sigma^2}\right)s_4\right]\right\} \\
& \left(\rho_i I_{\{\delta_i=0\}}+(1-\rho_i)\frac{1}{\theta_i}\exp\left\{-\frac{\delta_i}{\theta_i}\right\}\right),
\end{aligned}$$

where

$$s_4=\sum_{j=1}^{m}\sum_{p=i+1}^{k}\left(y_{pj}-\mu_1-\sum_{q=1}^{p-1}\delta_q+\delta_i\right).$$

The full conditional distribution of δ_i can therefore be summarized as a mixture of a discrete part and a continuous part.

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

$$
\begin{aligned}
h(\delta_i) &= \frac{1}{\sqrt{2\pi(\frac{1}{a_i})}} \exp\left\{ -\frac{1}{2(\frac{1}{a_i})} \left(\delta_i - \frac{g_i - \frac{\Delta_i}{\theta_i}}{a_i} \right)^2 \right\} \exp\left\{ \frac{\left(g_i - \frac{\Delta_i}{\theta_i} \right)^2}{2a_i} \right\}, \text{ with} \\
\Delta_i &= I_{\{\delta_i > 0\}}, \\
a_i &= (k-i)\frac{m}{\sigma^2}\left(1 - \frac{\sigma_\tau^2 (k-i)}{k\sigma_\tau^2 + \sigma^2} \right), \\
g_i &= \frac{1}{\sigma^2}\left(1 - \frac{\sigma_\tau^2 (k-i)}{k\sigma_\tau^2 + \sigma^2} \right) s_4, \text{ and} \\
c &= \frac{1}{\rho_i h(0) + (1-\rho_i)\frac{1}{\theta_i}\int_0^\infty h(\delta_i) d\delta_i}.
\end{aligned} \tag{A.4}
$$

5. Full Conditional Posterior Distribution of ρ_i

By the prior on ρ_i, BETA(α_o, β_o), and the mixture prior on δ_i in (2.3), the full conditional distribution of ρ_i can be expressed as

$$
[\rho_i \mid \delta_i, \theta_i] = \begin{cases} \text{BETA}(\alpha_\text{o} + 1, \beta_\text{o}), & \delta_i = 0 \\ \text{BETA}(\alpha_\text{o}, \beta_\text{o} + 1), & \delta_i > 0 \end{cases}.
$$

With $\alpha_o = \beta_o = 1$ (uniform), we have

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

6. Full Conditional Posterior Distribution of θ_i

For the prior on θ_i in (2.4) and the mixture prior on δ_i in (2.3), the full conditional distribution of θ_i is given by

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

Appendix B: Gibbs Sampling Steps

Step 0

In this step, set the starting values ${\delta_i}^{(0)}$, ${\rho_i}^{(0)}$, and ${\theta_i}^{(0)}$ for $1 \leq i \leq k-1$, ${\mu_1}^{(0)}$, $\sigma_\tau^{2(0)}$, and $\sigma^{2(0)}$.

Step t

(i) First, compute ${a_i}^{(t)}$ and ${g_i}^{(t)}$ by

$$\begin{aligned} a_i^{(t)} &= (k-i)\frac{m}{\sigma^{2(t-1)}}\left(1 - \frac{\sigma_\tau^{2(t-1)}(k-i)}{k\sigma_\tau^{2(t-1)} + \sigma^{2(t-1)}}\right), \text{ and} \\ g_i^{(t)} &= \frac{1}{\sigma^{2(t-1)}}\left(1 - \frac{\sigma_\tau^{2(t-1)}(k-i)}{k\sigma_\tau^{2(t-1)} + \sigma^{2(t-1)}}\right) s_4^{(t-1)}, \end{aligned}$$

where

$$s_4^{(t-1)} = \sum_{j=1}^{m}\sum_{p=i+1}^{k}\left(y_{pj} - {\mu_1}^{(t-1)} - \sum_{q=1}^{p-1}{\delta_q}^{(t-1)} + {\delta_i}^{(t-1)}\right).$$

Second, compute the conditional posterior probability of ${\delta_i}^{(t)} = 0$ (i.e., ${\lambda_i}^{(t)}$) for $1 \leq i \leq k-1$ by

$$\begin{aligned} {\lambda_i}^{(t)} &= \Pr({\delta_i}^{(t)} = 0 \mid \cdot) \\ &= \frac{{\rho_i}^{(t-1)} h(0; {a_i}^{(t)}, {g_i}^{(t)})}{{\rho_i}^{(t-1)} h(0; {a_i}^{(t)}, {g_i}^{(t)}) + (1 - {\rho_i}^{(t-1)})\frac{1}{{\theta_i}^{(t-1)}}\int_0^\infty h({\delta_i}^{(t-1)}; {a_i}^{(t)}, {g_i}^{(t)}, {\theta_i}^{(t-1)}) d{\delta_i}^{(t-1)}}. \end{aligned}$$

Here, $h(\cdot)$ is defined in (A.4).

Next, sample ${\mathrm{B}_i}^{(t)}$ from a Bernoulli distribution with success probability ${\lambda_i}^{(t)}$, i.e., $\Pr({\mathrm{B}_i}^{(t)} = 1) = {\lambda_i}^{(t)}$. If ${\mathrm{B}_i}^{(t)} = 1$, set ${\delta_i}^{(t)} = 0$. If ${\mathrm{B}_i}^{(t)} = 0$, sample $({\delta_i}^{(t)} \mid \cdot)$ from a truncated normal distribution $(0 < {\delta_i}^{(t)} < \infty)$ with mean $({g_i}^{(t)} - \frac{1}{{\theta_i}^{(t-1)}})/{a_i}^{(t)}$ and variance $1/{a_i}^{(t)}$.

(ii) Sample ${\rho_i}^{(t)}$, ${\theta_i}^{(t)}$, ${\mu_1}^{(t)}$, $\sigma^{2(t)}$, and $\tau^{2(t)}$ from their full conditional posterior distributions, respectively, for $1 \leq i \leq k-1$ and $j = 1, \ldots, m$. Subsequently, $\sigma_\tau^{2(t)}$ can be computed using $\tau^{2(t)}$ and $\sigma^{2(t)}$.

Then, repeat Step t, t =1, 2, ..., and continue.

References

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., & Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics* **26**, 641–647.

Berger, J. O. & Sellke, T (1987). Testing a point null hypothesis: The irreconcilability of p values and evidence (with comments). *Journal of the American Statistical Association* **82**, 112–122.

Box, G.E.P. & Tiao, G.C. (1992). *Bayesian Inference in Statistical Analysis.* New York: John Wiley & Sons.

Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics* **26**, 607–616.

Brunk, H. D., Ewing, G. M., & Utz, W. R. (1957). Minimizing integrals in certain classes of monotone functions. *Pacific Journal of Mathematics* **7**, 833–847.

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

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

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

Gelfand, A. E., Hills, S. E., Racine-Poon, A., & 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., & Lee, T. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association* **87**, 523–532.

Hayter, A. J. (1990). A one-sided Studentized range test for testing against a simple ordered alternative. *Journal of the American Statistical Association* **85**, 778–785.

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

Kass, R. E. & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association* **91**, 1343–1370.

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

Liu, L. (2001). Simultaneous statistical inference for monotone dose-response means. Doctoral dissertation, Memorial University of Newfoundland, St. John's, Canada.

Liu, L., Lee, C. C., & Peng, J. (2002). Max-min multiple comparison procedure for isotonic dose-response curves. *Journal of Statistical Planning and Inference* **107**, 133–141.

Marcus, R. (1982). Some results on simultaneous confidence intervals for monotone contrasts in one-way ANOVA model. *Communications in Statistics* **A 11**, 615–622.

Marcus, R. & Peritz, E. (1976). Some simultaneous confidence bounds in normal models with restricted alternatives. *Journal of the Royal Statistical Society* **B 38**, 157–165.

Molitor, J. & 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. & Wright, F. T. (2005). A note on multiple comparison procedures for detecting differences in simply ordered means. *Statistics and Probability Letters* **73**, 393–401.

Pauler, D. K., Wakefield, J. C., & 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., & Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* New York: John Wiley & Sons.

Singh, B. & Wright, F. T. (1990). Testing for and against an order restriction in mixed-effects models. *Statistics & 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., & Utts, J. M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika* **84**, 419–427.

Williams, D. A. (1977). Some inference procedures for monotonically ordered normal means. *Biometrika* **64**, 9–14.