Biometrics 60, 418–426
June 2004

# Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model

**Scott M. Berry**

Berry Consultants, 5124 Bellerive Bend, College Station, Texas 77845, U.S.A.
*email:* scott@berryconsultants.com

**and**

**Donald A. Berry**

Department of Biostatistics, MD Anderson Cancer Center, Houston, Texas 77030, U.S.A.
*email:* dberry@mdanderson.org

Summary. Multiple comparisons and other multiplicities are among the most difficult of problems that face statisticians, frequentists, and Bayesians alike. An example is the analysis of the many types of adverse events (AEs) that are recorded in drug clinical trials. We propose a three-level hierarchical mixed model. The most basic level is type of AE. The second level is body system, each of which contains a number of types of possibly related AEs. The highest level is the collection of all body systems. Our analysis allows for borrowing across body systems, but there is greater potential—depending on the actual data—for borrowing within each body system. The probability that a drug has caused a type of AE is greater if its rate is elevated for several types of AEs within the same body system than if the AEs with elevated rates were in different body systems. We give examples to illustrate our method and we describe its application to other types of problems.

Key words: Drug safety analysis; Mixture model; Multiple comparisons; Three-level hierarchical model.

## 1. Introduction

Problems of multiple comparisons are common in statistical applications. In each such problem there are several levels of experimental units. A typical example is assessing the safety of an experimental drug in pharmaceutical clinical trials. We develop our methodology in the context of this example, but it applies more generally.

Mehrotra and Heyse (2001) describe three categories of adverse events (AEs). Tier 1 AEs are those thought to be caused by the drug and this hypothesis is being specifically tested in the trial. Tier 1 is empty in many trials. Tier 2 AEs are those routinely collected in clinical trials but about which no specific hypotheses have been formulated in advance. The goal is to identify any unexpected deleterious effects of the drug and to quantify their rates. The observed rates of Tier 2 AEs are reported along with some statistical measure, such as a confidence interval or a $p$-value in comparison with a control rate. There are typically many types of Tier 2 AEs. Tier 3 AEs are rare spontaneous reports of serious events that require specific clinical evaluation.

We consider Tier 2 AEs and clinical trials in which drug and control rates are being compared. The question is which if any AEs should be flagged as probably due to the drug. The multiplicity of AEs makes this problem difficult, and this is so regardless of one's statistical philosophy. Random variability means that the rates of some AEs will be markedly higher in the drug group than in the placebo group, even if the drug is harmless. As is well known (Hochberg and Tamhane, 1987; Tukey, 1991; Benjamini and Hochberg, 1995), ignoring multiplicities gives a much higher than planned overall size of type I error. On the other hand, adjusting for multiplicities using a standard technique such as Bonferroni may fail to flag important differences.

In the standard analysis of efficacy endpoints, controlling for overall type I error rate means using a lower nominal significance level, perhaps as low as 0.001. Such adjustments are conservative with respect to individual type I error comparisons. But they are liberal for the analysis of safety endpoints, probably too liberal. A regulatory agency is unlikely to accept such a liberal adjustment for rates of AEs because important drug effects may slip through such a large filter. In assessing safety, type II errors are at least as important as type I errors.

Suppose the rates of three types of AEs are significantly higher on drug than on control ($p < 0.05$). Should these three types of AEs be flagged? There are at least four considerations in answering such a question: (1) the actual significance levels; (2) the total number of types of AEs being considered; (3) the rates for those AEs not considered for flagging, including their similarity with the three that are being considered; and (4) the biological relationships among the various AEs.

The first two of these are standard considerations in the frequentist approach to multiple comparisons. The second two are not, but they are relevant in the Bayesian approach (Berry, 1988; Berry and Stangl, 1996; Gopalan and Berry, 1998; Berry and Hochberg, 1999).

We take a Bayesian approach and so we consider all four of the above considerations. As regards consideration (4), in the example with three types of AEs with elevated rates on drug, it matters whether the three are in the same body system. Appropriate conclusions may well be different if they are nausea, vomiting, and diarrhea, say, than if they were nausea, rash, and wheezing. As regards consideration (3) above, one's conclusion would be different if the rates for other types of AEs in the same body system were consistent with the three at issue, even though the others may have $p > 0.05$. We demonstrate this by examples.

A principal focus of the frequentist approach in the presence of multiplicities is to suitably adjust the type I error rate. The Bayesian approach is less tied to type I error. Its focus is assessing the probability that the drug causes an AE on the basis of all available information (in the trial and beyond the trial as well, although we consider a single trial in the present article). According to Chi, Hung, and O'Neill (2002) who take a regulatory perspective, "Safety assessment is one area where frequentist strategies have been less applicable. Perhaps Bayesian approaches in this area have more promise." We show that they do.

We present a three-stage hierarchical mixture model for simultaneously addressing many types of AEs that are categorized into body systems. This model provides an explicit method for borrowing information across types of AEs. The hierarchical nature of the model gives rise to a regression effect—which is appealing in the context of multiplicities because it modulates extremes. We use a mixture for the prior distribution of the treatment parameters by assigning a point mass on equality of the treatment and the control rates.

Many body systems have several different types of AEs that are collected in clinical trials. AEs in the same body system may or may not be related but rates of AEs are more likely to be similar within than across body systems. A hierarchical model allows for this possibility, but it does not impose it. As a consequence, borrowing across AEs within body systems may be stronger than across different body systems, depending on the actual data.

In the next section we present an example in which there are 40 types of AEs within eight body systems. Several AE rates are elevated in the treatment group. In Section 4 we find the probability that the difference is due to treatment using the model developed in Section 3. The question is whether—and by how much—an AE rate is elevated in the treatment group. In addition, in Subsection 4.1 we illustrate the model's workings by applying it to modifications of the example. In Section 5 we consider a variety of hypotheses and present simulations showing the operating characteristics of the model.

We deal with a unique assignment of AEs to a body system. The assignment of AEs to body systems is based on biological or regulatory grounds and not on empirical observation. We model AEs in the same body system as being exchangeable. We take the assignment of AEs to body systems as given. If there is uncertainty in this assignment then the model can be applied under various other possible assignments. Such sensitivity analyses allow for presenting a range of conclusions. We conduct a sensitivity analysis in our example, demonstrating that the assignment of AEs to body systems can be important. However, we do not investigate whether the assignment of the various AEs to body systems is "reasonable" or "correct."

## 2. Example

Mehrotra and Heyse (2001) give results from a vaccine trial. The trial involved a quadrivalent vaccine containing measles, mumps, rubella, and varicella (MMRV). Participants were 296 healthy toddlers aged 12–18 months who were randomly assigned to two groups. The "treatment" group received MMRV on day 0 and controls received MMR on day 0 followed by V on day 42. All participants received PedvaxHIB on day 0. Safety follow-up used standard AE reporting and the primary question was to assess local and systemic reactions for the varicella component. The comparison of AEs was between the treatment group during days 0–42 with the control group during days 42–84. The AE counts for all 40 types of AEs are given in Table 1. Designation of "body system" was made in advance of and separate from the data from the trial. The sample size was smaller for controls ($N_C = 132$) than for treatment ($N_T = 148$) because only those participants who received a varicella injection were considered.

The (unadjusted) Fisher's exact test two-sided $p$-value in Table 1 is from Mehrotra and Heyse (2001). Four types of AEs have $p < 0.05$ and these are identified by asterisks. Bonferroni adjustment to the significance level means that $p < 0.0013$ would be required to achieve significance in the context of multiple comparisons. None of the AEs achieve this level of significance.

We do not have access to the raw data for this trial and so we cannot model within-patient correlations of the types of AEs shown in Table 1. A patient-level parameter and possible dependencies among the various AEs could be incorporated into the model. Such per-patient data would also elucidate any relationships among types of AEs.

## 3. Model

In this section we present the three-stage hierarchical mixed model. There are $B$ body systems. Within body system $b$ there are $k_b$ types of AEs labeled $A_{bj}$, where $b = 1, \ldots, B$ and $j = 1, \ldots, k_b$. Of the $N_C$ controls, $X_{bj}$ experience $A_{bj}$ and of the $N_T$ patients in the treatment group, $Y_{bj}$ experience $A_{bj}$. The probabilities of experiencing $A_{bj}$ are $c_{bj}$ and $t_{bj}$, for control and treatment patients, respectively. We use logistic transformations:

$$\gamma_{bj} = \log\left(\frac{c_{bj}}{1 - c_{bj}}\right) \quad \text{and} \quad \theta_{bj} = \log\left(\frac{t_{bj}}{1 - t_{bj}}\right) - \gamma_{bj}.$$

We present the hierarchical prior in its three stages. The following are the stage 1 priors. The $\gamma$'s have a normal prior distribution:

$$\gamma_{bj} \sim N\left(\mu_{\gamma b}, \sigma_\gamma^2\right) \quad \text{for } b = 1, \ldots, B \quad \text{and} \quad j = 1, \ldots, k_b.$$

Parameters $\theta_{bj}$ are the log-odds ratios. If $\theta_{bj} = 0$ then the probability that a patient experiences $A_{bj}$ is the same for

**Table 1**
*Example clinical trial from Mehrotra and Heyse* (2001). *Body system number b. Types of AEs are numbered separately (j) within body system. $Y_{bj}$ and $X_{bj}$ are the numbers of patients experiencing the indicated AE $A_{bj}$ in the treatment and control groups, respectively. Fisher's exact test p-values* $< 0.025$ *have asterisks.*

| | | | Treatment ($N_T = 148$) | | Control ($N_C = 132$) | | Fisher's |
|---|---|---|---|---|---|---|---|
| $b$ | $j$ | Type of AE $A_{bj}$ | $Y_{bj}$ | Rate | $X_{bj}$ | Rate | exact $p$ |
| 1 | 1 | Astenia/fatigue | 57 | 0.385 | 40 | 0.303 | 0.167 |
| 1 | 2 | Fever | 34 | 0.230 | 26 | 0.197 | 0.561 |
| 1 | 3 | Infection, fungal | 2 | 0.014 | 0 | 0.000 | 0.500 |
| 1 | 4 | Infection, viral | 3 | 0.020 | 1 | 0.008 | 0.625 |
| 1 | 5 | Malaise | 27 | 0.182 | 20 | 0.152 | 0.525 |
| 3 | 1 | Anorexia | 7 | 0.047 | 2 | 0.015 | 0.179 |
| 3 | 2 | Cendisiasis, oral | 2 | 0.014 | 0 | 0.000 | 0.500 |
| 3 | 3 | Constipation | 2 | 0.014 | 0 | 0.000 | 0.500 |
| 3 | 4 | Diarrhea | 24 | 0.162 | 10 | 0.076 | 0.029* |
| 3 | 5 | Gastroenteritis | 3 | 0.020 | 1 | 0.008 | 0.625 |
| 3 | 6 | Nausea | 2 | 0.014 | 7 | 0.053 | 0.089 |
| 3 | 7 | Vomiting | 19 | 0.128 | 19 | 0.144 | 0.730 |
| 5 | 1 | Lymphadenopathy | 3 | 0.020 | 2 | 0.015 | 1.000 |
| 6 | 1 | Dehydration | 0 | 0.000 | 2 | 0.015 | 0.221 |
| 8 | 1 | Crying | 2 | 0.014 | 0 | 0.000 | 0.500 |
| 8 | 2 | Insomnia | 2 | 0.014 | 2 | 0.015 | 1.000 |
| 8 | 3 | Irritability | 75 | 0.507 | 43 | 0.326 | 0.003* |
| 9 | 1 | Bronchitis | 4 | 0.027 | 1 | 0.008 | 0.375 |
| 9 | 2 | Congestion, nasal | 4 | 0.027 | 2 | 0.015 | 0.375 |
| 9 | 3 | Congestion, respiratory | 1 | 0.007 | 2 | 0.015 | 0.603 |
| 9 | 4 | Cough | 13 | 0.088 | 8 | 0.061 | 0.497 |
| 9 | 5 | Infection, upper respiratory | 28 | 0.189 | 20 | 0.152 | 0.431 |
| 9 | 6 | Laryngotracheobronchitis | 2 | 0.014 | 1 | 0.008 | 1.000 |
| 9 | 7 | Pharyngitis | 13 | 0.088 | 8 | 0.061 | 0.497 |
| 9 | 8 | Rhinorrhea | 15 | 0.101 | 14 | 0.106 | 1.000 |
| 9 | 9 | Sinusitis | 3 | 0.020 | 1 | 0.008 | 0.625 |
| 9 | 10 | Tonsillitis | 2 | 0.014 | 1 | 0.008 | 1.000 |
| 9 | 11 | Wheezing | 3 | 0.020 | 1 | 0.008 | 0.625 |
| 10 | 1 | Bite/sting | 4 | 0.027 | 0 | 0.000 | 0.125 |
| 10 | 2 | Eczema | 2 | 0.014 | 0 | 0.000 | 0.500 |
| 10 | 3 | Pruritis | 2 | 0.014 | 1 | 0.008 | 1.000 |
| 10 | 4 | Rash | 13 | 0.088 | 3 | 0.023 | 0.021* |
| 10 | 5 | Rash, diaper | 6 | 0.041 | 2 | 0.015 | 0.288 |
| 10 | 6 | Rash, measles/rubella-like | 8 | 0.054 | 1 | 0.008 | 0.039* |
| 10 | 7 | Rash, varicella-like | 4 | 0.027 | 2 | 0.015 | 0.687 |
| 10 | 8 | Urticaria | 0 | 0.000 | 2 | 0.015 | 0.221 |
| 10 | 9 | Viral exanthema | 1 | 0.007 | 2 | 0.015 | 0.603 |
| 11 | 1 | Conjunctivitis | 0 | 0.000 | 2 | 0.015 | 0.221 |
| 11 | 2 | Otitis media | 18 | 0.122 | 14 | 0.106 | 0.711 |
| 11 | 3 | Otorrhea | 2 | 0.014 | 1 | 0.008 | 1.000 |

control and treatment; that is, $c_{bj} = t_{bj}$. We assign positive probability to this possibility using the following mixture prior distribution:

$$\theta_{bj} \sim \pi_b I_{[0]} + (1-\pi_b) N\left(\mu_{\theta b}, \sigma^2_{\theta b}\right) \quad \text{for } b = 1, \ldots, B;\ j = 1, \ldots, k_b.$$

As in a standard Bayesian hierarchical model (Morris and Normand, 1992; Gelman et al., 1995; Smith, Spiegelhalter, and Thomas, 1995; Stangl, 1995; Berry, 2000) we assign a prior distribution to the hyperparameters, which creates the second stage of the prior structure:

$$\mu_{\gamma b} \sim N\left(\mu_{\gamma 0}, \tau^2_{\gamma 0}\right) \quad \text{for } b = 1, \ldots, B \quad \text{and} \quad \sigma^2_\gamma \sim IG\left(\alpha_{\sigma\gamma}, \beta_{\sigma\gamma}\right).$$

The probability $\pi_b$ that $\theta_{bj} = 0$ is the same for all AEs $j$ in body system $b$. When treatment effect $\theta_{bj}$ is not zero then its distribution is normal with mean of $\mu_{\theta b}$ and standard deviation $\sigma_{\theta b}$. This distribution also varies from one body system to the next. Each $\pi_b$ has a beta prior distribution:

$$\pi_b \sim \text{Beta}\left(\alpha_\pi, \beta_\pi\right), \quad b = 1, \ldots, B.$$

For the hyperparameters of the normal portion of the mixture we assume that

$$\mu_{\theta b} \sim N\left(\mu_{\theta 0}, \tau^2_{\theta 0}\right) \quad \text{for } b = 1, \ldots, B \quad \text{and} \quad \sigma^2_{\theta b} \sim \text{IG}(\alpha_\theta, \beta_\theta).$$

In the third level of the hierarchical model the parameters of these distributions (hyper-hyperparameters) themselves have a probability distribution, as follows:

$$\mu_{\gamma 0} \sim N\left(\mu_{\gamma 00}, \tau^2_{\gamma 00}\right) \quad \text{and} \quad \tau^2_{\gamma 0} \sim \text{IG}\left(\alpha_{\tau\gamma}, \beta_{\tau\gamma}\right).$$

Likewise, we assign prior distributions to the hyperparameters of the beta distribution, restricting both parameters to be greater than 1. The beta density is positive at $\pi = 0$ and $\pi = 1$ when $\alpha_\pi < 1$ and $\beta_\pi < 1$. Restricting the parameters to greater than 1 prevents the posterior density of $\pi$ from becoming too heavily concentrated at one of its edges. We assign independent left-truncated exponential prior distributions to $\alpha_\pi$ and $\beta_\pi$:

$$\alpha_\pi \sim \frac{\lambda_\alpha \exp(-\alpha\lambda_\alpha)}{\exp(-\lambda_\alpha)} I_{[\alpha>1]} \quad \text{and} \quad \beta_\pi \sim \frac{\lambda_\beta \exp(-\beta\lambda_\beta)}{\exp(-\lambda_\beta)} I_{[\beta>1]}.$$

Taking $\lambda_\alpha = \lambda_\beta$ induces a symmetry in the prior distributions of $\alpha$ and $\beta$ and means that 0.5 is the a priori probability that $\theta_{bj} = 0$. The hyperparameters for the normal prior distribution of $\mu_{\theta b}$ have the following fixed distributions:

$$\mu_{\theta 0} \sim N\left(\mu_{\theta 00}, \tau^2_{\theta 00}\right) \quad \text{and} \quad \tau^2_{\theta 0} \sim \text{IG}\left(\alpha_{\theta 0}, \beta_{\theta 0}\right).$$

Hyperparameters $\mu_{\gamma 00}$, $\tau^2_{\gamma 00}$, $\alpha_{\tau\gamma}$, $\beta_{\tau\gamma}$, $\alpha_{\sigma\gamma}$, and $\beta_{\sigma\gamma}$ are fixed constants.

The calculations for this model are carried out using Markov chain Monte Carlo (MCMC) methods. The complete conditionals and the details of the MCMC methods are presented in the Appendix. We simulate 10,000 observations from the posterior after a burn-in of 1000 observations.

## 4. Results

In Subsection 4.1 we present the results of the model for the example of Section 2. In Subsection 4.2 we show the results when the data in the example are altered. We assume the following parameter values: $\mu_{\theta 00} = 0$, $\tau^2_{\theta 00} = 10$, $\alpha_\theta = 3$, $\beta_\theta = 1$, $\alpha_{\theta 0} = 3$, $\beta_{\theta 0} = 1$, $\alpha_{\tau\gamma} = 3$, $\beta_{\sigma\gamma} = 1$, $\alpha_{\sigma\gamma} = 3$, $\beta_{\sigma\gamma} = 1$, and $\lambda_\alpha = \lambda_\beta = 1$.

### 4.1 *Original Data*

Table 2 gives the posterior means of the $\theta$'s and the posterior probability that each $\theta > 0$, that is, that the AE rate on treatment is greater than on control.

The four AEs that were significant using Fisher's exact test do not have the highest $p(\theta > 0)$. The smallest $p$-value, for irritability, does have the largest probability that $\theta > 0$, but the second smallest $p$-value is associated with rash, which has the fourth largest probability that $\theta > 0$. This difference in the two approaches is largely due to the fact that rash is part of the largest body system (10), one in which there is not consistent evidence of a drug effect. Statistical significance levels take into account only individual AE data. The probability distribution of $\theta$ takes into account more than just the data within the individual AE, and thus the posterior probability that $\theta > 0$ is not monotone as a function of the $p$-value.

The posterior probability that the event rate on treatment is greater than on control is small to moderate (less than 50%) for 39 of the 40 types of AEs. The only type of AE with a high probability of being associated with treatment is irritability (in body system 8), with $p(\theta > 0) = 0.780$.

**Table 2**

*The three-level hierarchical model results for the example of Table* 1. $p(\theta = 0)$ *is the probability that the treatment and control have the same AE rates and* $p(\theta > 0)$ *is the probability that treatment has a higher AE rate. The entries in bold-faced type correspond to the asterisked entries in Table* 1.

| | | | Post probability | | | | | Post probability | |
|---|---|---|---|---|---|---|---|---|---|
| $b$ | $j$ | Type of AE | $\theta = 0$ | $\theta > 0$ | $b$ | $j$ | Type of AE | $\theta = 0$ | $\theta > 0$ |
| 1 | 1 | Astenia/fatigue | 0.762 | 0.211 | 9 | 4 | Cough | 0.906 | 0.062 |
| 1 | 2 | Fever | 0.827 | 0.122 | 9 | 5 | Infection, respiratory | 0.897 | 0.083 |
| 1 | 3 | Infection, fungal | 0.796 | 0.101 | 9 | 6 | Bronchitis | 0.898 | 0.047 |
| 1 | 4 | Infection, viral | 0.813 | 0.100 | 9 | 7 | Pharyngitis | 0.906 | 0.061 |
| 1 | 5 | Malaise | 0.826 | 0.116 | 9 | 8 | Rhinorrhea | 0.904 | 0.051 |
| 3 | 1 | Anorexia | 0.821 | 0.117 | 9 | 9 | Sinusitis | 0.903 | 0.051 |
| 3 | 2 | Cendisiasis, oral | 0.835 | 0.083 | 9 | 10 | Tonsillitis | 0.905 | 0.042 |
| 3 | 3 | Constipation | 0.812 | 0.101 | 9 | 11 | Wheezing | 0.907 | 0.050 |
| 3 | 4 | **Diarrhea** | 0.743 | **0.231** | 10 | 1 | Bite/sting | 0.859 | 0.087 |
| 3 | 5 | Gastroenteritis | 0.823 | 0.093 | 10 | 2 | Eczema | 0.860 | 0.070 |
| 3 | 6 | Nausea | 0.805 | 0.050 | 10 | 3 | Pruritis | 0.868 | 0.062 |
| 3 | 7 | Vomiting | 0.849 | 0.076 | 10 | 4 | **Rash** | 0.784 | **0.190** |
| 5 | 1 | Lymphadenopathy | 0.717 | 0.136 | 10 | 5 | Rash, diaper | 0.852 | 0.099 |
| 6 | 1 | Dehydration | 0.666 | 0.087 | 10 | 6 | **Rash, measles/rub-like** | 0.836 | **0.126** |
| 8 | 1 | Crying | 0.655 | 0.185 | 10 | 7 | Rash, varicella-like | 0.862 | 0.076 |
| 8 | 2 | Insomnia | 0.661 | 0.153 | 10 | 8 | Urticaria | 0.852 | 0.048 |
| 8 | 3 | **Irritability** | 0.214 | **0.780** | 10 | 9 | Viral exanthema | 0.855 | 0.055 |
| 9 | 1 | Bronchitis | 0.900 | 0.059 | 11 | 1 | Conjunctivitis | 0.721 | 0.079 |
| 9 | 2 | Congestion, nasal | 0.901 | 0.058 | 11 | 2 | Otitis media | 0.757 | 0.102 |
| 9 | 3 | Congestion, respiratory | 0.896 | 0.040 | 11 | 3 | Otorrhea | 0.749 | 0.121 |

**Table 3**
*Two-sided Fisher-exact-test p-values compared with the probability of a treatment effect $p(\theta > 0)$ from a "one-stage" solo Bayesian model (see text) and from the three-stage hierarchical model*

| Type of adverse event | Fisher $2p$ | Solo Bayesian | Hierarchical Bayesian |
|---|---|---|---|
| Diarrhea | 0.029 | 0.885 | 0.231 |
| Irritability | 0.003 | 0.984 | 0.780 |
| Rash | 0.021 | 0.923 | 0.190 |
| Rash, measles/ rubella-like | 0.039 | 0.889 | 0.126 |

Our hierarchical model addresses multiplicities. While the data for a particular type of AE may seem convincing when considered alone, it may be less convincing in the context of the other 39 types of AEs. In our example, because the other types of AEs in body system 8 (crying and insomnia) evince little treatment effect, the estimated treatment effect for irritability is "shrunk" toward 0. However, despite this shrinkage, the estimated treatment effect for irritability is moderately large. Therefore, our conclusion is that the difference is likely due to a treatment effect.

Table 3 gives the two-sided Fisher-exact-test $p$-values for the null hypothesis that the rates of AEs are equivalent. Table 3 also presents the conclusion of a "Solo Bayesian model." The purpose of this alternative Bayesian formulation is to show how the three-stage hierarchical model affects the conclusions. The Solo Bayesian model is simply the first stage of the three-stage hierarchical mixture model. In the Solo model the individual types of AEs are considered in isolation. Each $\gamma$ has a normal prior distribution with mean $-2$ and standard deviation 100. The prior distribution for $\theta$ is a mixture of a point mass at 0 and a continuous standard normal distribution. Each component of the mixture has prior probability 0.5.

The posterior probability that $\theta > 0$ for irritability is 0.780 in the full three-stage model. When considered in isolation this probability is 0.984. The difference is substantial, but not as large as in the other three AEs presented in Table 3. There are only two types of AEs in body system 8 except for irritability: crying and insomnia, with a total of only six occurrences between them. Therefore the effect of irritability is regressed toward no effect, but not as strongly as it is in the other three AEs. For diarrhea (in body system 3), in the Solo model, there is a 0.885 probability of being elevated by treatment, but in the full model this probability is only 0.231. This difference is because there are six other AEs in body system 3 and they show very little evidence of a treatment effect.

There are nine types of AEs in body system 10. For rash the probability of a treatment effect in the Solo Bayesian model is 0.923. However, when taken in the context of the other eight types of AEs in body system 10, and the remaining 31 types of AEs in the other body systems, this probability is only 0.190. For measles/rubella-like rash the Solo Bayesian model gives a probability of 0.889 while the full model probability is only 0.126. The full model probabilities better reflect the complete data. The full model allows for borrowing strength both from within and across body systems.

### 4.2 *Altered Data*

To further illustrate how modeling body systems affects conclusions, we present results for hypothetical changes in the example data. For the AE irritability we alter the number of occurrences in the treatment group. We also artificially remove it from body system 8 and moved into other body systems. In our full model the types of AEs within each body system are exchangeable, but the types of AEs in different body systems are not. Figure 1 presents the probability that $\theta > 0$ for irritability, under various hypothetical changes. In particular, we change the number of cases in the treatment group from the actual value of 75 to 65, 70, 80, and 85. The probability of a treatment effect ranges from 0.394 (for 65 cases) to 0.780 (actual number of 75 cases) to 0.965 (for 85 cases).

In the line "Body 9" of Figure 1, irritability has been moved to body system 9. There are 11 other types of AEs in body system 9 and none show much treatment effect. This information has a substantial impact, lowering the probability of a treatment effect as compared with the actual case of irritability in body system 8.

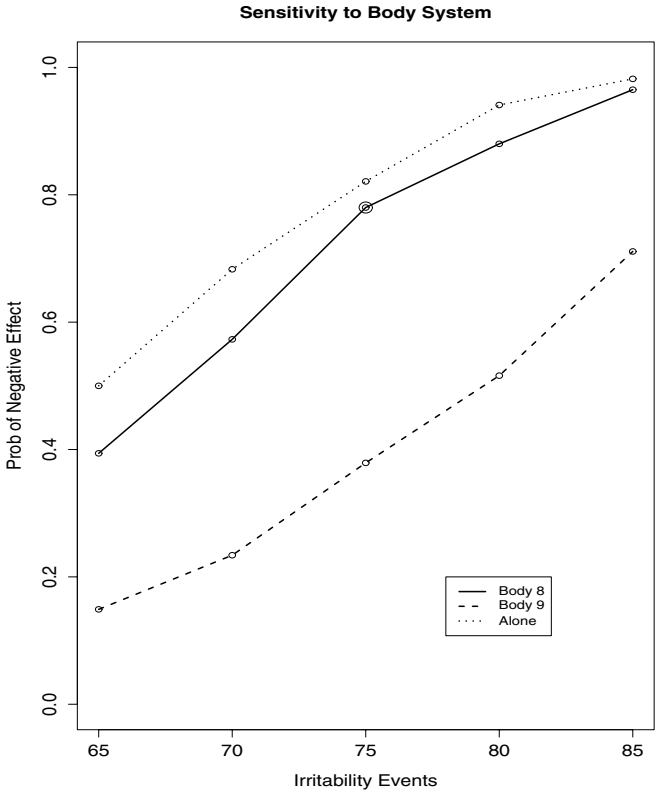


**Figure 1.** The probability that $\theta > 0$ for irritability is plotted as the $y$-axis. The $x$-axis gives the number of events in the treatment group and the different lines refer to the body system in which irritability belongs. The double circle shows the actual number of events and body system.

The third line of Figure 1 shows the result of moving irritability to its own body system, one with no other types of AEs from which to borrow information. Now the effect is the opposite of moving irritability to body system 9. Now the probability that treatment affects irritability is larger than when it is a member of body system 8.

These different hypothetical data sets and changes to the Bayesian model highlight the different forms of borrowing that are present in the three-stage model. For the original data set, the Solo Bayesian model results in a posterior probability of 0.984 of a treatment effect for irritability. In this scenario there is no borrowing—the irritability data are analyzed in isolation. In the hypothetical alone data set of Figure 1, there are no other types of AEs in the same body system with irritability, and thus the only borrowing is via the assumed exchangeability of the body systems themselves. With little information of a treatment effect in the remaining body systems the probability that treatment affects irritability in the alone data set is 0.821. When irritability is in body system 8, the actual data, there are two other types of AEs (crying and insomnia) in the same body system. Since there is little evidence of a treatment effect on crying or insomnia, the probability that treatment affects irritability drops slightly from 0.821 to 0.780. But when irritability is moved to body system 9 with 11 other types of AEs having little evidence of a treatment effect, the probability that treatment affects irritability drops markedly from 0.780 to 0.379.

Artificially changing the body system for irritability demonstrates the effect of the other AEs on the posterior distribution of the irritability parameters. Types of AEs within the same body system have a greater effect on each other because they are taken to be exchangeable. Types of AEs in different body systems are not exchangeable even though body systems are themselves assumed to be exchangeable at the next higher level of the hierarchy.

## 5. Simulations

In this section we present the results of simulations that are based on the example of Section 2. In particular, we use the same body systems. In setting 1 we assume that the true control and treatment rates of AEs are equal and are the same as the observed proportion of control events in the example (except that for those with 0 events, we assume a proportion of 1/132). We present results for three different sample sizes: 150, 300, and 450 per arm.

In setting 2 we change the theoretical proportion of events for lymphadenopathy, the only AE in body system 5. The probabilities of this AE are 0.10 and 0.20 for control and treatment, respectively.

In setting 3 the true rates for the first AE in body system 9 are 0.10 and 0.20 for control and treatment, respectively.

In setting 4 all 11 of the AEs in body system 9 have true rates 0.10 and 0.20 for control and treatment, respectively.

Figure 2 is a summary of the results for a subset of four AEs. We plot the average probability of a treatment effect for each of the three sample sizes and four settings. Each scenario was simulated 10,000 times.

In setting 1 there are no treatment effects. In none of the runs did a type of AE have a probability of a treatment effect larger than 0.90. In fewer than 1% of the runs there was a 50% chance of a treatment effect. Therefore, our method is quite conservative. Interestingly, although there are only 150 (or 300 or 450) patients within each AE/treatment group, there is borrowing across the AEs, and thus the results are stronger than the indicated number of patients.

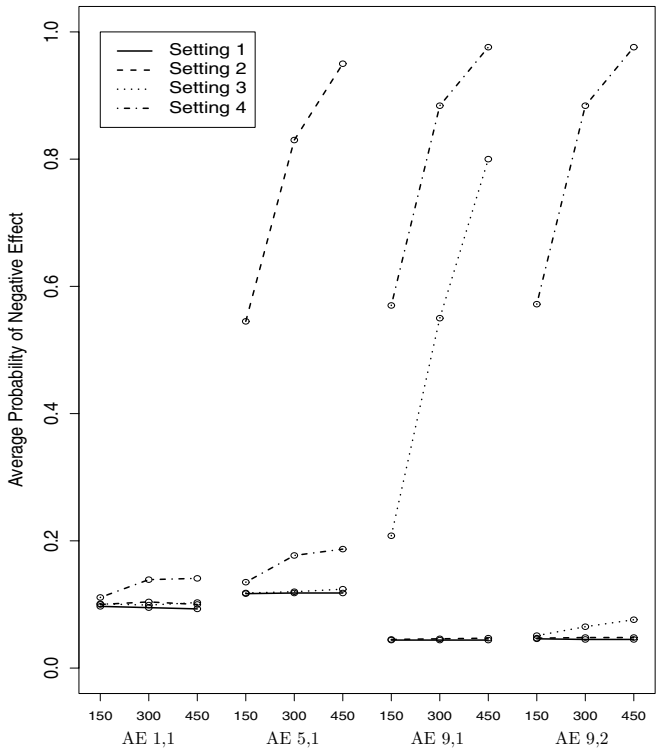


**Figure 2.** Operating characteristics based on 10,000 simulations. The entries are the average probabilities of a treatment effect. The results are shown for four AEs. The $x$-axis shows the AE and the sample size.

In setting 2 the true rate for the lone type of AE in body system 5 is elevated by the treatment. There is a clear sample size effect. The average probability of a treatment effect increases from 0.545 to 0.830 when the sample size is doubled from 150 per group and to 0.950 when it is tripled. In this simulation there are no other types of AEs in this body system and there is little information borrowed from the other types of AEs.

In setting 3 the true rate of the first type of AE in body system 9 is elevated by the treatment just as the AE in body system 5 in setting 2. In this example there are 10 other AEs which are unaffected by the treatment. There is a good deal of information borrowed from these other 10 types of AEs. Therefore, achieving a 0.90 probability of treatment effect is more difficult in setting 3 than in setting 2. In the case of 150 observations the average probability of a treatment effect is 0.208 in setting 3 as opposed to 0.545 in setting 2. This difference is entirely due to the other 10 types of AEs in body system 9. There is less reduction in the probability of a treatment effect in the 300 and 450 sample size scenarios in setting 3 than in setting 2. The data are stronger with larger sample size and so the other 10 types of AEs have less influence on the conclusions. Also, when the sample sizes are increased the

effect of the borrowing is less. Although there is still a significant amount of borrowing, its impact is less when the sample size is larger.

In setting 4 each of the types of AEs in body system 9 has the same rate as for the first type of AE in setting 3. In setting 4 our hierarchical model is more likely to identify a treatment effect as being real. For $N_T = N_C = 150$, the average probability of a treatment effect for the first AE in body system 9 is 0.570. The relevant comparisons are with settings 2 and 3. In setting 2 there are no other types of AEs in the same body system. In setting 3 there are other types of AEs in the same body system with no treatment effect. In setting 4 there are other types of AEs in the same body system, but they also have treatment effects. The average probability of a treatment effect reflects the different information available.

There are differences in the probability of a treatment effect in setting 4 as compared with setting 3. The additional information that the other AEs have apparent treatment effects markedly increases the probability of a treatment effect for individual AEs in setting 4. From this simulation it is clear that the other AEs within a body system have an important effect on the individual AEs within that body system—whether that evidence is for or against a treatment effect.

## 6. Discussion

As with all multiple comparison problems, handling AEs in a clinical trial is a challenging statistical problem. Modeling the existing structure and the available information is important for doing good science. There is valuable information in the patterns of AEs, especially when they are grouped by body system. In this article, we present a three-stage hierarchical model in which relationships among types of AEs are modeled explicitly depending on their body systems.

In our model, the conclusion that one type of AE is affected by treatment depends on the data from the other types of AEs, especially from those within the same body system. This is different from conclusions of more traditional multiple comparison methods in which only the number of types of AEs under consideration matters.

The use of a point-mass mixture prior for $\theta$ is an important aspect to our model. A mixture prior is reasonable for addressing Tier 2 AEs because many of them may be completely unaffected by treatment. Therefore for some types of AEs it may well be that $\theta \equiv 0$.

Body system plays an important role in our model and the conclusions are sensitive to the assignments of types of AEs to the various body systems. In particular, we showed in Subsection 4.1 that moving a type of AE to a different body system can have a dramatic effect on one's conclusions. This is both a positive and negative aspect of the model. It is positive in the sense that it allows for exploiting information across types of AEs that are related. It is negative in that it means that assigning types of AEs into body systems requires care. And it requires expert biological help. Assignments should be made separate from the data and on biological grounds. In particular, it would violate the spirit of the modeling to assign types of AEs to body systems based on their empirical correlations. If two AEs are known to be unrelated biologically then they should not be in the same body system. If there is some question as to which body system should contain a particular type of AE, we suggest multiple runs of the model with the different assignments considered. Such an approach provides a sensitivity analysis of the conclusions. Subsection 4.1 provides an example.

Suppose that a type of AE is assigned to the "wrong" body system. As we have indicated, this may have an impact on our model's conclusions. In this article we have not addressed how to make reparations. And indeed, any post-hoc corrections will be difficult to make without biasing the conclusions. This process represents another level of multiplicity and is on a par with data dredging: It is worthwhile, but difficult to carry out and potentially very misleading.

Our model is based on summary (marginal) data regarding numbers of occurrences of the various AEs. We did not have access to the raw data for the example we considered. Having raw data would enable modeling possible dependencies among the various AEs at the patient level. Conclusions about treatment effects could be made with greater precision.

Our model has applications beyond drug AEs. One that is potentially important is the analysis of cDNA microarray data (Zhang et al., 2001). The multiplicity problem is even more pronounced than when assessing AEs because tens of thousands of genes may be involved. A typical question is which genes are differentially expressed in diseased as compared with normal tissue. Our model finds the posterior probability that each of the genes is implicated. Categorizing genes into genetic pathways (the analog of body systems in the AE setting) is helpful. And such categorizations may be necessary in order to make progress in understanding the genetic basis and heterogeneity of a disease and its treatment. As in the AE application, having biologists identify pathways (or other groupings of genes) is essential.

## Acknowledgements

The authors would like to thank the referee and an associate editor for their very helpful suggestions.

## Résumé

Les problèmes de comparaisons multiples, et les problèmes de multiplicité en général, sont parmi les plus difficiles auxquels soient confrontés les statisticiens, fréquentistes comme bayésiens. Ainsi de l'analyse des nombreux types d'événements indésirables (EI) recueillis dans le cadre des essais cliniques. Pour une telle analyse, nous proposons un modèle mixte hiérarchique à trois niveaux: le niveau élémentaire est le type d'EI, le second niveau le système organe—chacun regroupant un certain nombre de types d'EI possiblement reliés—,et le niveau le plus élevé l'ensemble de tous les systèmes organes. Notre analyse permet d'utiliser, pour un EI donné d'un système organe donné, l'information recueillie dans les autres systèmes organes; mais elle peut présenter un intérêt encore plus grand (en fonction des données réelles), celui d'utiliser l'information recueillie dans le même système organe. En effet, la probabilité qu'un médicament ait induit un certain type d'EI est plus grande si la fréquence observée est élevée pour plusieurs types d'EI du même système organe que si les EI aux fréquences élevées sont dispersés dans différents systèmes organes. Nous illustrons notre méthode à travers des exemples et décrivons en quoi elle s'applique à d'autres types de problèmes.

## References


Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57,** 289–300.

Berry, D. A. (1988). Multiple comparisons, multiple tests, and data dredging: A Bayesian perspective (with discussion). In *Bayesian statistics*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (eds). Oxford: Oxford University Press.

Berry, S. M. (2000). Meta-analysis versus large trials: Resolving the controversy. In *Meta-analysis in Medicine and Health Policy*, D. K. Stangl and D. A. Berry (eds), 65–82. New York: Marcel Dekker.

Berry, D. A. and Hochberg, Y. (1999). Bayesian perspectives on multiple comparisons. *Journal of Statistical Planning and Inference* **82,** 215–227.

Berry, D. A. and Stangl, D. K. (1996). *Bayesian Biostatistics*. New York: Marcel Dekker.

Chen, M. H., Shao, Q. M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Analysis*. New York: Springer-Verlag.

Chi, G., Hung, H. M. J., and O'Neill, R. (2002). Some comments on "Adaptive Trials and Bayesian Statistics in Drug Development" by Donald A. Berry. In *Pharmaceutical Report*, Volume 9, 1–11. Washington, D.C.: American Statistical Association.

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

Gopalan, R. and Berry, D. A. (1998). Bayesian multiple comparisons using Dirichlet process priors. *Journal of the American Statistical Association* **93,** 1130–1139.

Hochberg, Y. and Tamhane, C. A. (1987). *Multiple Comparison Procedures*. New York: Wiley.

Mehrotra, D. V. and Heyse, J. F. (2004). Multiplicity considerations in clinical safety analyses. *Statistical Methods in Medical Research* **13,** 227–238.

Morris, C. and Normand, S. L. (1992). Hierarchical models for combining information and for meta-analysis (with discussion). In *Bayesian Statistics*, Volume 4, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 321–344. Oxford: Oxford University Press.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

Smith, T. C., Spiegelhalter, D. J., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine* **14,** 2685–2699.

Stangl, D. (1995). Prediction and decision making using Bayesian hierarchical models. *Statistics in Medicine* **14,** 2173–2190.

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science* **6,** 100–116.

Zhang, W., Laborde, P. M., Coombes, K., Berry, D. A., and Hamilton, S. (2001). Cancer genomics: Promises and complexities. *Clinical Cancer Research* **7,** 2159–2167.


*Received January* 2003. *Revised November* 2003.
*Accepted December* 2003.

## Appendix

We simulate from the posterior distributions using MCMC. The successive-substitution approach we use is standard in the MCMC literature. Additional details can be found in Robert and Casella (1999) and Chen, Shao, and Ibrahim (2000).

The joint posterior distribution of the parameters is

$$\begin{aligned}
&\left[\tilde{\theta},\tilde{\gamma},\tilde{\pi},\tilde{\sigma}_\theta,\tilde{\mu}_\theta,\sigma_\gamma,\tilde{\mu}_\gamma,\alpha_\pi,\beta_\pi,\tau_{\gamma 0},\mu_{\gamma 0},\tau_{\theta 0},\mu_{\theta 0}\,|\,\tilde{X},\tilde{Y}\right] \\
&\propto \prod_{b=1}^{B}\prod_{j=1}^{k_b}\frac{\exp(\theta_{bj}+\gamma_{bj})^{Y_{bj}}}{\{1+\exp(\theta_{bj}+\gamma_{bj})\}^{N_T}}\frac{\exp(\gamma_{bj})^{X_{bj}}}{\{1+\exp(\gamma_{bj})\}^{N_C}} \\
&\quad\times\prod_{b=1}^{B}\prod_{j=1}^{k_b}\left[\pi_b I_{[\theta_{bj}=0]}+(1-\pi_b)\left(\frac{1}{\sigma_{\theta b}}\right)\right. \\
&\qquad\left.\times\exp\left\{-\frac{1}{2\sigma_{\theta b}^2}\left(\theta_{bj}-\mu_{\theta b}\right)^2\right\}\right] \\
&\quad\times\prod_{b=1}^{B}\left\{\frac{\Gamma(\alpha_\pi+\beta_\pi)}{\Gamma(\alpha_\pi)\Gamma(\beta_\pi)}(\pi_b)^{\alpha_\pi-1}(1-\pi_b)^{\beta_\pi-1}\right. \\
&\qquad\left.\times\left(\sigma_{\theta b}^2\right)^{-(\alpha_\theta+1)}\exp\left(-\frac{1}{\beta_\theta\sigma_{\theta b}^2}\right)\right\} \\
&\quad\times\left[(\sigma_\gamma)^{-\sum_{b=1}^{B}k_b}\exp\left\{-\frac{1}{2\sigma_\gamma^2}\sum_{b=1}^{B}\sum_{j=1}^{k_b}\left(\gamma_{bj}-\mu_{\gamma b}\right)^2\right\}\right] \\
&\quad\times\left[(\tau_{\gamma 0})^{-B}\exp\left\{-\frac{1}{2\tau_{\gamma 0}^2}\sum_{b=1}^{B}\left(\mu_{\gamma b}-\mu_{\gamma 0}\right)^2\right\}\right] \\
&\quad\times\left[(\tau_{\theta 0})^{-B}\exp\left\{-\frac{1}{2\tau_{\theta 0}^2}\sum_{b=1}^{B}\left(\mu_{\theta b}-\mu_{\theta 0}\right)^2\right\}\right] \\
&\quad\times\left\{\exp(-\alpha_\pi\lambda_\alpha)I_{[\alpha_\pi>1]}\exp(-\beta_\pi\lambda_\beta)I_{[\beta_\pi>1]}\right\}\left(\sigma_\gamma^2\right)^{\alpha_{\sigma\gamma}+1} \\
&\quad\times\exp\left(-\frac{1}{\beta_{\sigma\gamma}\sigma_\gamma^2}\right)\left(\tau_{\gamma 0}^2\right)^{-(\alpha_{\tau\gamma}+1)}\exp\left(-\frac{1}{\beta_{\tau\gamma}\tau_{\gamma 0}^2}\right) \\
&\quad\times\left(\tau_{\theta 0}^2\right)^{-(\alpha_{\tau\theta}+1)}\exp\left(-\frac{1}{\beta_{\tau\theta}\tau_{\theta 0}^2}\right) \\
&\quad\times\exp\left\{-\frac{1}{2\tau_{\gamma 00}^2}\left(\mu_{\gamma 0}-\mu_{\gamma 00}\right)^2\right\}\exp\left\{-\frac{1}{2\tau_{\theta 00}^2}\left(\mu_{\theta 0}-\mu_{\theta 00}\right)^2\right\}
\end{aligned}$$

The complete conditional distributions are

$$\begin{aligned}
&[\theta_{bj}\,|\,\gamma_{bj},\pi_b,\sigma_{\theta b},\mu_{\theta b},Y_{bj}] \\
&\quad\propto\frac{\exp(\theta_{bj}+\gamma_{bj})^{Y_{bj}}}{\{1+\exp(\theta_{bj}+\gamma_{bj})\}^{N_T}}\left[\pi_b I_{[\theta_{bj}=0]}+(1-\pi_b)\left(\frac{1}{\sqrt{2\pi}\sigma_{\theta b}}\right)\right. \\
&\qquad\left.\times\exp\left\{-\frac{1}{2\sigma_{\theta b}^2}\left(\theta_{bj}-\mu_{\theta b}\right)^2\right\}\right]
\end{aligned}$$

$$\begin{aligned}
&[\gamma_{bj}|\theta_{bj},\sigma_\gamma,\mu_{\gamma b},Y_{bj},X_{bj}] \\
&\quad\propto\left[\frac{\exp(\theta_{bj}+\gamma_{bj})^{Y_{bj}}}{\{1+\exp(\theta_{bj}+\gamma_{bj})\}^{N_T}}\frac{\exp(\gamma_{bj})^{X_{bj}}}{\{1+\exp(\gamma_{bj})\}^{N_C}}\right] \\
&\qquad\times\exp\left\{-\frac{1}{2\sigma_\gamma^2}\left(\gamma_{bj}-\mu_{\gamma b}\right)^2\right\}
\end{aligned}$$

$$[\pi_b \mid \tilde{\theta}_b, \alpha_\pi, \beta_\pi] \sim \text{Beta}\left(\alpha_\pi + \sum_{j=1}^{k_b} I_{[\theta_{bj}=0]}, \beta_\pi + k_b - \sum_{j=1}^{k_b} I_{[\theta_{bj}=0]}\right)$$

$$\left[\sigma^2_{\theta b} \,\middle|\, \tilde{\theta}_b, \mu_{\theta b}\right] \sim \text{IG}\left[\frac{k_b}{2} + \alpha_\theta, \left\{\frac{1}{\beta_\theta} + \frac{1}{2}\sum_{j=1}^{k_b}(\theta_{bj} - \mu_{\theta b})^2\right\}^{-1}\right]$$

$$\left[\sigma^2_\gamma \,\middle|\, \tilde{\gamma}, \tilde{\mu}_{\gamma b}\right] \sim \text{IG}\left[\frac{1}{2}\sum_{b=1}^{B} k_b + \alpha_\gamma, \left\{\frac{1}{\beta_\gamma} + \frac{1}{2}\sum_{b=1}^{B}\sum_{j=1}^{k_b}(\gamma_{bj} - \mu_{\gamma b})^2\right\}^{-1}\right]$$

$$\left[\tau^2_{\gamma 0} \,\middle|\, \tilde{\mu}_{\gamma b}, \mu_{\gamma 0}\right] \sim \text{IG}\left[\frac{B}{2} + \alpha_{\tau\gamma}, \left\{\frac{1}{\beta_{\tau\gamma}} + \frac{1}{2}\sum_{b=1}^{B}(\mu_{\gamma b} - \mu_{\gamma 0})^2\right\}^{-1}\right]$$

$$\left[\tau^2_{\theta 0} \,\middle|\, \tilde{\mu}_{\theta b}, \mu_{\theta 0}\right] \sim \text{IG}\left[\frac{B}{2} + \alpha_{\tau\theta}, \left\{\frac{1}{\beta_{\tau\theta}} + \frac{1}{2}\sum_{b=1}^{B}(\mu_{\theta b} - \mu_{\theta 0})^2\right\}^{-1}\right]$$

$$[\mu_{\gamma b} \mid \mu_{\gamma 0}, \tilde{\gamma}_{bj}, \sigma_\gamma, \tau_{\gamma 0}] \sim N\left(\frac{\tau^2_\gamma \sum_{j=1}^{k_b}\gamma_{bj} + \sigma^2_\gamma \mu_{\gamma 0}}{\tau^2_\gamma k_b + \sigma^2_\gamma}, \frac{\tau^2_\gamma \sigma^2_\gamma}{\tau^2_\gamma k_b + \sigma^2_\gamma}\right)$$

$$[\mu_{\theta b} \mid \mu_{\theta 0}, \tilde{\theta}_{bj}, \sigma_\theta, \tau_{\theta 0}] \sim N\left(\frac{\tau^2_\theta \sum_{j=1}^{k_b}\theta_{bj} + \sigma^2_\theta \mu_{\theta 0}}{\tau^2_\theta k_b + \sigma^2_\theta}, \frac{\tau^2_\theta \sigma^2_\theta}{\tau^2_\theta k_b + \sigma^2_\theta}\right)$$

$$[\mu_{\gamma 0} \mid \tilde{\mu}_\gamma, \tau_{\gamma 0},] \sim N\left(\frac{\tau^2_{\gamma 00}\sum_{b=1}^{B}\mu_{\gamma b} + \tau^2_{\gamma 0}\mu_{\gamma 00}}{\tau^2_{\gamma 00}B + \tau^2_{\gamma 0}}, \frac{\tau^2_{\gamma 00}\tau^2_{\gamma 00}}{\tau^2_{\gamma 00}B + \tau^2_{\gamma 0}}\right)$$

$$[\mu_{\theta 0} \mid \tilde{\mu}_\theta, \tau_{\theta 0},] \sim N\left(\frac{\tau^2_{\theta 00}\sum_{b=1}^{B}\mu_{\theta b} + \tau^2_{\theta 0}\mu_{\theta 00}}{\tau^2_{\theta 00}B + \tau^2_{\theta 0}}, \frac{\tau^2_{\theta 00}\tau^2_{\theta 00}}{\tau^2_{\theta 00}B + \tau^2_{\theta 0}}\right)$$

$$[\alpha_\pi \mid \beta_\pi, \tilde{\pi}] \propto \left\{\frac{\Gamma(\alpha_\pi + \beta_\pi)}{\Gamma(\alpha_\pi)}\right\}^B \left(\prod_{b=1}^{B}\pi_b\right)^{(\alpha_\pi + 1)} \times \exp(-\alpha_\pi \lambda_\alpha) I_{[\alpha_\pi > 1]}$$

$$[\alpha_\pi, \beta_\pi \mid \tilde{\pi}] \propto \left\{\frac{\Gamma(\alpha_\pi + \beta_\pi)}{\Gamma(\beta_\pi)}\right\}^B \left\{\prod_{b=1}^{B}(1 - \pi_b)\right\}^{(\beta_\pi + 1)} \times \exp(-\beta_\pi \lambda_\beta) I_{[\beta_\pi > 1]}$$

We simulate successively from the above complete conditionals. Sampling from those not available in closed form uses a Metropolis–Hastings (MH) step. Simulations from all but $\theta$ are straightforward MH steps, using a normally distributed draw centered on the current value. Simulations from the mixture distributions for the values of $\theta$ are nonstandard. When sampling from the distribution of $\theta_{bj}$ we employ a mixture MH step. Namely, we simulate from a point mass at 0 (with probability 0.5), and a normal distribution centered on the current value of $\theta$. Let $\theta^C$ be the candidate draw and $\theta^O$ the "old" value. The ratio of the complete conditionals, $r(\theta^C, \theta^O)$ is defined as

$$r(\theta^C, \theta^O) = \frac{\left[\theta^C \mid \gamma_{bj}, \pi_b, \sigma_{\theta b}, \mu_{\theta b}, Y_{bj}\right]}{[\theta^O \mid \gamma_{bj}, \pi_b, \sigma_{\theta b}, \mu_{\theta b}, Y_{bj}]}.$$

For an MH step for $\theta$, with candidate draw, $\theta^C$, the acceptance probability is

$$\begin{cases} r(\theta^C, \theta^O) & \text{if } \theta^C = \theta^0 = 0 \\ \dfrac{1}{\sigma_{\text{MH}}\sqrt{2\pi}}\exp\left(-\dfrac{1}{2\sigma^2_{\text{MH}}}(\theta^O)^2\right) r(\theta^C, \theta^O) & \text{if } \theta^C = 0 \text{ and } \theta^O \neq 0 \\ \dfrac{r(\theta^C, \theta^O)}{\dfrac{1}{\sigma_{\text{MH}}\sqrt{2\pi}}\exp\left(-\dfrac{1}{2\sigma^2_{\text{MH}}}(\theta^O)^2\right)} & \text{if } \theta^C \neq 0 \text{ and } \theta^O = 0 \\ r(\theta^C, \theta^O) & \text{if } \theta^C \neq 0 \text{ and } \theta^O \neq 0. \end{cases}$$