Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v11n2p622

Approximate posterior inference for multiple testing using a hierarchical mixed-effect Poisson regression model

By Murua, Nembot-Simo

Published: 14 October 2018

Electronic Journal of Applied Statistical Analysis
Vol. 11, Issue 02, October 2018, 622-654
DOI: 10.1285/i20705948v11n2p622

Approximate posterior inference for multiple testing using a hierarchical mixed-effect Poisson regression model

Alejandro Murua*[a] and Annick Nembot-Simo[a]

[a] *Département de mathématiques et de statistique, Université de Montréal, CP 6128, succ. centre-ville, Montréal, Québec H3C 3J7 Canada*

Published: 14 October 2018

We present an approximate posterior inference methodology for a Bayesian hierarchical mixed-effect Poisson regression model. The model serves us to address the multiple testing problem in the presence of many group or cluster effects. This is carried out through a specialized Bayesian false discovery rate procedure. The likelihood is simplified by an approximation based on Laplace's approximation for integrals and a trace approximation for the determinants. The posterior marginals are estimated using this approximated likelihood. In particular, we obtain credible regions for the parameters, as well as probability estimates for the difference between risks (Poisson intensities) associated with different groups or clusters, or different levels of the fixed effects. The methodology is illustrated through an application to a vaccine trial.

keywords: Laplace's approximation, multiple comparison, pseudo-likelihood, Bayesian false discovery rate, vaccine trial.

1. Introduction

Consider the following multiple testing problem involving a series of adverse effects arising from a vaccine clinical trial. The data are taken from the work of Mehrotra and Heyse (2004). These are the count results of a clinical trial involving a quadrivalent vaccine against measles, mumps, rubella and varicella. The individuals participating

*Corresponding author: murua@dms.umontreal.ca

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

in the trial were 296 infants 12 to 18 month-old in good health. They were randomly assigned to the treatment (quadrivalent vaccine) or control. Infants in the control group were administered at day zero a vaccine for measles, mumps and rubella. Later on at day 42, they were given a varicella vaccine. Infants in the treatment group were administered the quadrivalent vaccine at day 0. One of the goals of the study is to evaluate the local and systematic reactions that infants face to the varicella component of the vaccine. Therefore, a comparison of secondary or adverse effects between the control group at days 42-84 and the treatment group at days 0-42 is of interest. Forty different adverse effects divided among eight human body systems are considered. These are displayed in Table 4. The allocation of the adverse effects into the eight body system groups was done from a purely biological point of view. The data was reproduced by Berry and Berry (2004) who also analyzed the data in the context of multiple comparisons. Individual Fisher exact tests comparing treatment and control groups for each adverse effect show that there are statistically significant differences at a level $\alpha = 0.05$ between the severity of four of the forty adverse effects. These are *3-Diarrhea, 8-Irritability, 10-Rash* and *10-Rash-measles-rubella-like* (the number refers to the body system group). However, false discovery rate adjustments (Benjamini and Hochberg, 1995; Mehrotra and Heyse, 2004) only points out to *3-Diarrhea* and *8-Irritability* as being more severe for the treatment group. Berry and Berry (2004) also arrive to the same conclusion using a Bayesian hierarchical model for the log-odds and the difference between the control and treatment log-odds. In the present work we analyze these data by performing a Bayesian false discovery rate procedure (Whittemore, 2007) with a hierarchical multilevel mixed-effect Poisson regression model. Intuitively, adverse effects within any body system group may be correlated. One way to account for this is to consider that the Body systems contribute to random effects in the counts. This is our assumption. Also, similar adverse effects may be presented within different body systems (see Table 4). Therefore, the information gathered from all body system groups about the same adverse effect should be incorporated in all groups concerned. This is achieved by considering a hierarchical model where at a second level of the model, information is shared among the groups.

The model introduced in the present work is an extension of the hierarchical Poisson regression model considered by Christiansen and Morris (1997). In the last decades, Poisson models have been extensively used to model regression involving count data. Due to the complexity of Poisson regression, the methods used to fit the data to the models are almost as important as the models themselves. For example, Holford (1980) fits log-linear models for the analysis of proportions and hazard rates using the proportional fitting (IPF) algorithm. Frome (1983) fits Poisson regression models that entail log-linear, quasi-linear and non-linear models using iterative re-weighted least-squares (IRLS). Usually, though, count data show overdispersion. More realistic models have been conceived to deal with this extra-Poisson variation. For example, by embedding Poisson regression into a Bayesian framework, Albert (1992) is able to analyze sports data using a Bayesian hierarchical Gamma-Poisson mixture model. The estimation is achieved through Markov chain Monte Carlo (MCMC) techniques. There are also models that incorporate random effects. Vonesh (1990) shows a mixed-effects Poisson regres-

sion model to study the risk factors associated with peritonitis, a bacterial infection of the peritoneum membrane. The model considers fixed effects corresponding to the collected information from the patients, and random effects due to the patients themselves. Tempelman and Gianola (1996) develop models for applications to animal reproduction that take into account the overdispersion not accounted for by the mixed-effect models. This is modeled by fitting a Gamma distribution to the Poisson parameters, yielding in a Negative Binomial model for the counts. The Maximum A Posteriori (MAP) estimate of the parameters is estimated via the Newton-Raphson algorithm. Christiansen and Morris (1997) present a Bayesian hierarchical fixed-effect Gamma-Poisson regression model that leads to a Negative Binomial model for count data. They show the superiority of their approach with respect to other methods such as the generalized linear model in the presence of overdispersion. They also show how to obtain estimates of the parameters posteriors without having to resort to MCMC computations. Basically, they used a first-order (i.e., asymptotic normality) approximation of the posteriors. We pursue further their ideas to extend their model to a Bayesian multi-level mixed-effect hierarchical Poisson regression model. However, in contrast to their work, we estimate the parameter posteriors using higher-order approximations (i.e., beyond asymptotic normality which is appropriate only when the data size is very large) that are derived from Laplace's approximation (Tierney and Kadane, 1986; Davison, 1986) and a matrix trace approximation to determinants. Furthermore, we are able to simplify some calculations using the orthogonality between some of the model parameters. Recently, Guihenneuc-Jouyaux and Rousseau (2007) introduced a somewhat hybrid approach that exploits similar ideas to improve the performance of an MCMC sampling. Even though their approach could be combined with our approximate posterior density estimates for our model, we prefer to pursue here a complete analytic approach, that is, without the need to resort to MCMC sampling. A simulation study presented in Section 4 corroborates the validity of our approach.

The paper is organized as follows. Section 2 introduces the hierarchical mixed effect Poisson model. The approximate posterior is deduced in Section 3. This section also described the approximate posterior inference. In particular, Section 3.1 shows an approximation of the marginal posterior of the main parameters in the presence of nuisance parameters. The results of a simulation study conceived to compare the performance of the procedure introduced in this work with that of an MCMC sampler are presented in Section 4. The posterior density associated with the difference between two Poisson intensities is presented in Section 5. This result is key for carrying over multiple testing based on a version of the Bayesian false discovery rate suggested by Whittemore (2007). Section 6 shows the application of our methodology to a few data sets, including the vaccine trial describe above. There are three appendices: the first one shows the calculation leading to the approximate likelihood. The second one, describes in detail the computations for the special case of orthogonal parameters. The third one displays the first, second and third derivatives of the approximate likelihood. These are used to build our approximation to the marginal posterior densities.

2. The model

Our model is an extension of the hierarchical Poisson regression model considered by Christiansen and Morris (1997) to a multi-level mixed-effect Poisson regression model. Our approximate posterior inference for this latter model is inspired by the work of Christiansen and Morris (1997) on what they named the Poisson Regression Interactive Multilevel Modeling (PRIMM). In our framework, the population of counts is stratified in p strata or groups. Let $\mathcal{D} = \{Z_{ij}\}$ ($i = 1, \ldots, n_j$, and $j = 1, \ldots, p$) be a sample from this population. We assume that the distribution of Z_{ij} is Poisson with intensity λ_{ij}. That is, $Z_{ij}|\lambda_{ij} \sim \text{Pois}(e_{ij}\lambda_{ij})$, where the $\{e_{ij}\}$ are known expositions. We will denote the observed counts and rates by z_{ij} and $y_{ij} = z_{ij}/e_{ij}$, respectively. We also assume that the intensity λ_{ij} follows a Gamma distribution with shape and rate parameters ξ_j, and ξ_j/μ_{ij}. The logarithm of its mean is linearly linked to an r-dimensional vector of fixed-effects covariates $X_{ij} = (1, x_{ij,1}, \ldots, x_{ij,r-1})'$ and a q-dimensional vector of group (or "random") effects covariates $W_{ij} = (1, w_{ij,1}, \ldots, w_{ij,q-1})'$ through the parameters $\beta = (\beta_0, \ldots, \beta_{r-1}) \in \mathbb{R}^r$ and $b_j = (b_{j0}, b_{j1}, \ldots, b_{j,q-1})' \in \mathbb{R}^q$, $j = 1, \ldots, p$, respectively. That is,

$$E(\lambda_{ij}) = \mu_{ij} = \exp\{X'_{ij}\beta + W'_{ij}b_j\}.$$

Note that since $\text{Var}(\lambda_{ij}) = \mu_{ij}^2/\xi_j$, the parameters ξ_j control the variance of the Poisson intensities. Therefore we will refer to them as the *precision parameters.* From now on we will denote the set of log-transformed precision parameters as $\vec{\tau} = (\tau_1, \ldots, \tau_p)$, where $\tau_j = \log(\xi_j)$. We will also refer to the vector $\vec{\tau}$ as the vector of precision parameters. The set of group effects will be denoted by $b \in \mathbb{M}_{q\times p}$ (here $\mathbb{M}_{q\times p}$ denotes the set of matrices of q rows and p columns). We will denote by X_j and W_j the $n_j \times r$ and $n_j \times q$ matrices of covariables associated with the j-th group, $j = 1, \ldots, p$.

Let $B_{ij} = \xi_j/(\xi_j + e_{ij}\mu_{ij})$. It is straightforward to verify that the distribution of $Z_{ij}|(\beta, \vec{\tau}, b)$ is the Polya distribution $\text{Polya}(\xi_j, B_{ij})$, i.e.

$$p(z_{ij}|(\beta, \vec{\tau}, b)) = p(z_{ij}|\xi_j, \mu_{ij}) = \frac{\Gamma(z_{ij} + \xi_j)}{z_{ij}!\Gamma(\xi_j)} B_{ij}^{\xi_j}(1 - B_{ij})^{z_{ij}}.$$

Hence, the ξ_j's account for possible overdispersion of the Poisson counts Z_{ij}, for example, if burst of events are likely to occur. The overdispersion is more severe for small values of ξ_j, and is nonexistent for large values of ξ_j.

Since the counts are supposed to be independent, the likelihood of the model is

$$L(\beta, \vec{\tau}, b|\mathcal{D}) = \prod_{j=1}^{p}\prod_{i=1}^{n_j} \frac{\Gamma(z_{ij} + \xi_j)}{z_{ij}!\Gamma(\xi_j)} B_{ij}^{\xi_j}(1 - B_{ij})^{z_{ij}}.$$

Let us define for each $j \in \{1, 2, \ldots, p\}$, $M_j = \max\{z_{ij} : i = 1, \ldots, n_j\}$, $N_{sj} = \text{cardinality}(\{i : z_{ij} \geq s\})$, for $s = 1, \ldots, M_j$, and $\bar{z}_j = \sum_{i=1}^{n_j} z_{ij}/n_j$. In what follows, z_j will denote the n_j-dimensional vector $(z_{1j}, z_{2j}, \ldots, z_{n_j,j})$. All vectors are assumed to be represented as column matrices. A direct generalization of the formula derived in Lemma 1 in (Christiansen and Morris, 1997) yields the following expression for the log-likelihood $\ell(\beta, \vec{\tau}, b)$

of our model

$$\ell(\beta,\vec{\tau},b)=\sum_{j=1}^{p}\Bigg\{\sum_{s=1}^{M_j}N_{sj}\log(e^{\tau_j}+s-1)-n_j\bar{z}_j\tau_j+z_j'X_j\beta+z_j'W_jb_j \\ -\sum_{i=1}^{n_j}(e^{\tau_j}+z_{ij})\log(1+e_{ij}\mu_{ij}e^{-\tau_j})\Bigg\}. \quad (1)$$

The hierarchical model.
As it is usually done with random-effects, we impose a common prior for all the group effects b_j's. This is a q-dimensional zero-mean Normal distribution with variance-covariance matrix Ψ, $j=1,\dots,p$. Ψ itself is assumed to follow (a priori) an Inverse-Wishart distribution with m_0 degrees of freedom and scale matrix $\sigma_0^2\Lambda_0$, for a scalar σ_0^2. That is, $b_j\sim\text{Normal}_q(0,\Psi)$, and $\Psi\sim\text{Inverse-Wishart}(m_0,\sigma_0^2\Lambda_0)$. We also assume an improper uniform prior for the fixed effects β, and the following priors for the precision parameters:

$$\begin{aligned}\tau_j=\log(\xi_j)&\sim\text{Normal}(\tau,\sigma_\tau^2),\\ \sigma_\tau^2&\sim\text{Inverse-Chi-squared}(\nu_0,s_0^2),\ \text{and}\\ \tau&\sim\text{Normal}(\tau_0,v_0^2).\end{aligned}$$

In what follows η will stand for the set of hyper-parameters $(\tau,\sigma_\tau^2,\Psi)$. It is common to set $\Lambda_0=I_q$, the identity matrix of dimension $q\times q$. In what follows we will suppose that σ_0^2 is rather large. Besides providing a flatter prior on the variance of the group effects, this choice will allow us to simplify the computations when deriving an approximation to the posterior distribution of the parameters.

3. The posterior of $(\beta,\vec{\tau})$

In principle we would like to make inference only on the set of parameters $(\beta,\vec{\tau})$. That is, (b,η) may be seen as nuisance parameters. Therefore we are particularly interested in estimating

$$\pi(\beta,\vec{\tau}|\mathcal{D})=\frac{\int L(\beta,\vec{\tau},b|\mathcal{D})\,\pi(\eta)\pi(b|\eta)\pi(\beta,\vec{\tau}|\eta)\,d\eta\,db}{\int L(\beta,\vec{\tau},b|\mathcal{D})\,\pi(\eta)\pi(b|\eta)\pi(\beta,\vec{\tau}|\eta)\,d\eta\,db\,d\beta\,d\vec{\tau}.} \quad (2)$$

There are several alternatives to estimate this posterior. The main problem is to integrate out the group effects b. There are techniques based on Gaussian quadrature (Pinheiro and Bates, 1995), Bayesian quadrature, and quasi-Monte-Carlo (Morokoff and Caflisch, 1995). We are going to approximate this posterior using Laplace's approximation to the integral. Already Tierney and Kadane (1986) suggested a similar technique to approximate the means and variances of posterior densities, showing that the error in the approximation of the integrals is $O(n^{-2})$, where n is the total number of observa-

tions. Joe (2008) also shows the adequacy and accuracy of Laplace's approximation for Poisson count response mixed models. There are several works that try to exploit these results even within the frequentist approach (Wolfinger, 1993; Breslow and Lin, 1995; Vonesh, 1996; Raudenbush et al., 2000) and extend them to facilitate their computation (Nott et al., 2009) or to use them in combination with important sampling or Gibbs sampling techniques (Kuk, 1999; Skaug and Fournier, 2006). Our approach is in two stages. As in the work of Sutradhar and Zhende (1998), we first approximate the full likelihood (or full log-posterior) by a function that is simpler to work with than the full likelihood. We refer to the resulting approximation as the q-likelihood. This is obtained by using Laplace's approximation and an approximation of the determinant involving the group effects. Then, we obtain a series of approximations to the parameter posteriors using the q-likelihood. One of the main drawbacks cited in the literature for using higher order approximations to draw inference, such as Laplace's approximation, is the difficulty in finding the second and third order derivatives of the log-likelihood. Here, we have done the work for our hierarchical mixed effect Poisson regression model using the q-likelihood. The derivatives are shown in Appendix C.

Let $\mathcal{L}(\beta, \vec{\tau}, b|\mathcal{D})$ denote the integrand in the numerator of (2). This is the likelihood of $(\beta, \vec{\tau}, b)$. Our approximation is obtained as follows. Let us denote by $a = \sum_{j=1}^{p} \tau_j/p$, and by $A = p^{-1}(\nu_0 s_0^2 + \sum_{j=1}^{p}(\tau_j - a)^2)$. A first approximation to $\mathcal{L}(\beta, \vec{\tau}, b|\mathcal{D})$, derived from a Laplace's approximation of the integral with respect to η of the integrand in (2) is

$$L(\beta, \vec{\tau}, b|\mathcal{D}) \exp\left\{ -\frac{(a-\tau_0)^2}{2v_0^2} - \frac{\nu_0 + p - 1}{2} \log(A) - \frac{m_0 + p}{2} \log(|\sigma_0^2 \Lambda_0 + bb'|) \right\}. \quad (3)$$

The derivation is postponed to Appendix A. The following matrix determinant approximation is very useful in the derivation of a final approximation to $\mathcal{L}(\beta, \vec{\tau}, b|\mathcal{D})$ and to the posteriors. Let C be a $k \times k$ real symmetric matrix, and ϵ be a small real number. Then

$$|I_k + \epsilon C| = 1 + \epsilon \,\mathrm{trace}(C) + \mathcal{O}(\epsilon^2). \quad (4)$$

This approximation is easily derived from Jacobi's formula for the derivative of a matrix. Using this expression we obtain, for large σ_0^2

$$\begin{aligned} \log|\sigma_0^2\Lambda_0 + bb'| &= q\log\sigma_0^2 + \log|\Lambda_0| + \log|I_q + \sigma_0^{-2}\Lambda_0^{-1}bb'| \\ &\approx q\log\sigma_0^2 + \log|\Lambda_0| + \log(1 + \sigma_0^{-2}\mathrm{trace}(\Lambda_0^{-1}bb')), \end{aligned}$$

where the symbol $\approx$ stands for *approximately equals to.* Consequently, from now on we will work with the pseudo-likelihood given by the q-likelihood

$$\begin{aligned} q(\beta, \vec{\tau}, b|\mathcal{D}) \doteq \exp\Bigg\{ &\ell(\beta, \vec{\tau}, b|\mathcal{D}) - \frac{1}{2v_0^2}(a-\tau_0)^2 - \frac{1}{2}(\nu_0 + p - 1)\log(A) \\ &- \frac{1}{2}(m_0 + p)\left(q\log\sigma_0^2 + \log|\Lambda_0| + \log(1 + \sigma_0^{-2}\sum_{j=1}^{p} b_j'\Lambda_0^{-1}b_j) \right) \Bigg\}, \end{aligned} \quad (5)$$

where $\ell(\beta, \vec{\tau}, b|\mathcal{D})$ denotes the log-likelihood $\log L(\beta, \vec{\tau}, b|\mathcal{D})$. Let $j_q(\beta, \vec{\tau}, b)$ be the observed Fisher information associated with $(\beta, \vec{\tau}, b)$ derived from the q-likelihood $q(\beta, \vec{\tau}, b|\{z_{ij}\})$. That is,

$$j_q(\beta, \vec{\tau}, b) = -\frac{\partial^2 \log(q)}{\partial(\beta, \vec{\tau}, b)^2}(\beta, \vec{\tau}, b).$$

Also, we will use the notation $j_{q,b}$ for the submatrix of j_q associated with the second derivatives of b. Similarly, we will denote by $j_{q,\beta,b}$ and $j_{q,\vec{\tau},b}$ the submatrices of j_q associated with the second derivatives of the vectors (β, b) and $(\vec{\tau}, b)$, respectively.

Denote by $\hat{\beta}$, $\hat{\vec{\tau}}$ and $\hat{b}$ the maximizers of $q(\beta, \vec{\tau}, b)$. We will refer to these variables as the q-likelihood maximizers (note that they could also be referred to as the q-MAP estimates). Let $b_{\beta,\vec{\tau}}$ be the maximizer of $q(\beta, \vec{\tau}, b)$ for fixed $(\beta, \vec{\tau})$. Then

$$\pi(\beta, \vec{\tau}|\mathcal{D}) \approx p_L(\beta, \vec{\tau}|\mathcal{D}) \doteq c(2\pi)^{-(p+r)/2} \left(\frac{|j_q(\hat{\beta}, \hat{\vec{\tau}}, \hat{b})|}{|j_{q,b}(\beta, \vec{\tau}, b_{\beta,\vec{\tau}})|}\right)^{1/2} \frac{q(\beta, \vec{\tau}, b_{\beta,\vec{\tau}})}{q(\hat{\beta}, \hat{\vec{\tau}}, \hat{b})}, \tag{6}$$

where the symbol $\doteq$ stands for *equality by definition,* and where c is the normalizing constant so that the integral with respect to $(\beta, \vec{\tau})$ of the right-hand-side is one. In what follows, we are not going to write these normalizing constants explicitly, because the Laplace density approximation does not include them. Here, we have added it for technical reasons, to stress that the approximation does not necessarily integrates to one. Let $(\beta_{\vec{\tau}}, b_{\vec{\tau}})$ be the maximizer of $q(\beta, \vec{\tau}, b)$ for fixed $\vec{\tau}$, and let $(\vec{\tau}_\beta, b_\beta)$ be the maximizer of $q(\beta, \vec{\tau}, b)$ for fixed β. The marginal posteriors are approximately given by

$$\pi(\vec{\tau}|\mathcal{D}) \approx p_L(\vec{\tau}|\mathcal{D}) \doteq (2\pi)^{-p/2} \left(\frac{|j_q(\hat{\beta}, \hat{\vec{\tau}}, \hat{b})|}{|j_{q,\beta,b}(\beta_{\vec{\tau}}, \vec{\tau}, b_{\vec{\tau}})|}\right)^{1/2} \frac{q(\beta_{\vec{\tau}}, \vec{\tau}, b_{\vec{\tau}})}{q(\hat{\beta}, \hat{\vec{\tau}}, \hat{b})}, \tag{7}$$

$$\pi(\beta|\mathcal{D}) \approx p_L(\beta|\mathcal{D}) \doteq (2\pi)^{-r/2} \left(\frac{|j_q(\hat{\beta}, \hat{\vec{\tau}}, \hat{b})|}{|j_{q,\vec{\tau},b}(\beta, \vec{\tau}_\beta, b_\beta)|}\right)^{1/2} \frac{q(\beta, \vec{\tau}_\beta, b_\beta)}{q(\hat{\beta}, \hat{\vec{\tau}}, \hat{b})}. \tag{8}$$

These well-known approximations are derived directly from the application of Laplace's approximation to the integrals of $q(\beta, \vec{\tau}, b)$ with respect to the group effects b, the precision parameters $\vec{\tau}$ and the fixed effects β, as well as the set of parameters to $(\beta, \vec{\tau}, b)$. Usually, they are written in terms of the profile likelihood (Tierney and Kadane, 1986; Davison, 1986) (Brazzale et al., 2007, p. 162). Higher-order approximations, such as the p^*-density approximation of Barndorff-Nielsen and Cox, may be derived from these approximations for the scalar β case (Severini, 2000; Barndorff-Nielsen and Cox, 1994). But the expression and computations involved are complex. We will derive simpler approximations to these quantities in order to facilitate their computation.

3.1. An approximation of the marginal posterior in the presence of nuisance parameters

In order to find approximations for the marginal posteriors of β and $\vec{\tau}$, we are going to derive a couple of general approximations to the densities when the parameters, say (ψ, θ), can be decomposed in a parameter of interest $\psi \in \mathbf{R}^d$, and a nuisance parameter θ. For example, in our case, if $\psi = \beta$ is the parameter of interest, then $\theta = (\vec{\tau}, b)$ corresponds to the nuisance parameter. When $\psi = \vec{\tau}$ is of interest, $\theta = (\beta, b)$ is the nuisance parameter. As before, Laplace's approximation of the marginal posterior of ψ is given by

$$\pi(\psi|\mathcal{D}) \approx p_L(\psi|\mathcal{D}) \doteq (2\pi)^{-d/2}\left(\frac{|j(\hat{\psi},\hat{\theta})|}{|j_\theta(\psi,\theta_\psi)|}\right)^{1/2}\exp\{\ell_q(\psi,\theta_\psi) - \ell_q(\hat{\psi},\hat{\theta})\}, \tag{9}$$

where $\ell_q(\psi,\theta)$ and $j(\psi,\theta)$ denote the log-likelihood and minus the Hessian of the log-q-likelihood evaluated at (ψ,θ), respectively; $j_\theta(\psi,\theta)$ is the portion of $j(\psi,\theta)$ corresponding to the second derivatives with respect to θ; and $(\hat{\psi},\hat{\theta})$ is the maximum q-likelihood estimator. From now on, to ease the notation, a hat $\hat{}$ over a quantity will indicate that quantity evaluated at the maximum q-likelihood estimate; for example, $\hat{\jmath} = j(\hat{\psi},\hat{\theta})$.

3.2. An approximation for moderate deviation range values.

Next, we will simplify a bit further the density by deriving an Edgeworth-like type of expansion for the posterior from (9). That is, we will write the density as a deviation from the Normal density. This approximation will be valid for the moderate deviation range of values of ψ, that is, for values of ψ near $\hat{\psi}$, $(\psi = \hat{\psi} + O(n^{-1/2}))$. In this latter case, by smoothness (i.e., differentiability of the derivatives), the matrices $j_\theta(\psi,\theta_\psi)$ and $\hat{\jmath}_\theta$ are close to each other. That is, their Frobenious norm $\epsilon_\theta = ||j_\theta(\psi,\theta_\psi) - \hat{\jmath}_\theta||$ is small. Therefore, using (4) again, we have

$$\begin{aligned}|j_\theta(\psi,\theta_\psi)| &= |\hat{\jmath}_\theta| \times \left|I + \epsilon_\theta\, \hat{\jmath}_\theta^{-1}(j_\theta(\psi,\theta_\psi) - \hat{\jmath}_\theta)/\epsilon_\theta\right| \\ &\approx |\hat{\jmath}_\theta| \times (1 + \operatorname{trace}\{\hat{\jmath}_\theta^{-1}(j_\theta(\psi,\theta_\psi) - \hat{\jmath}_\theta)\}).\end{aligned}$$

Since this trace's approximation is of order $O\big(||\psi - \hat{\psi}||_2^2\big)$, and Laplace's approximation is of order $O(n^{-1})$, the approximations obtained in this section are of order $O\Big(\max\big\{n^{-1},\, ||\psi - \hat{\psi}||_2^2\big\}\Big)$, where n is the total number of observations. Therefore, for values of ψ in the moderate deviation range, the approximations are still of order $O(n^{-1})$.

Recall the determinant identity $|\hat{\jmath}| = |\hat{\jmath}_\theta| \times |\hat{\jmath}_\psi - \hat{\jmath}_{\psi\times\theta}\hat{\jmath}_\theta^{-1}\hat{\jmath}_{\theta\times\psi}|$, where $\hat{\jmath}_{\psi\times\theta}$ denotes the entries of $\hat{\jmath}$ corresponding to the cross-derivatives between ψ and θ, and the matrix $\hat{\jmath}_{\theta\times\psi} = \hat{\jmath}'_{\psi\times\theta}$. Let $\hat{D}_{\psi,\theta} \doteq \hat{\jmath}_\psi^{-1}\hat{\jmath}_{\psi\times\theta}\hat{\jmath}_\theta^{-1}\hat{\jmath}_{\theta\times\psi}$. Note that $\hat{\jmath}_{p,\psi} \doteq \hat{\jmath}_\psi(I - \hat{D}_{\psi,\theta})$ corresponds to the profile q-likelihood Hessian for ψ with θ as nuisance parameter. These latter

observations yield a second approximation to $\pi(\psi|\mathcal{D})$ given by

$$p_{L.2}(\psi|\mathcal{D}) \doteq \left(1+\text{trace}\{\hat{\jmath}_\theta^{-1}(j_\theta(\psi,\theta_\psi)-\hat{\jmath}_\theta)\}\right)^{-\frac{1}{2}}(2\pi)^{-\frac{d}{2}}|\hat{\jmath}_{p,\psi}|^{\frac{1}{2}}\,\exp\{\ell_q(\psi,\theta_\psi)-\hat{\ell}_q\}, \quad (10)$$

Note that the approximation $p_{L.2}(\psi|\mathcal{D})$ also has the form of the p^*-density approximation of Barndorff-Nielsen and Cox (Severini, 2000; Barndorff-Nielsen and Cox, 1994). Next, we use a Taylor series expansion to approximate both the numerator and denominator of $p_{L.2}(\psi|\mathcal{D})$. The series is expanded at $\psi=\hat{\psi}$. Let $T_\theta = \text{trace}\{\hat{\jmath}_\theta^{-1}(j_\theta(\psi,\theta_\psi)-\hat{\jmath}_\theta)\}$. Since $\hat{T}_\theta = 0$, we have $T_\theta(\psi,\theta_\psi) \approx \dot{\hat{T}}'_\theta(\psi-\hat{\psi})$, where $\dot{T}_\theta = \frac{dT_\theta}{d\psi}$. The exact form of the derivatives of $T_{(\vec{\tau},b)}$ and $T_{(\beta,b)}$ for our model are easily derived from Proposition 1 of the Appendix. A simpler, though still reasonable, approximation is obtained by writing the log-likelihood as

$$\ell_q(\psi,\theta_\psi)-\hat{\ell}_q \approx -\frac{1}{2}(\psi-\hat{\psi})'\,\hat{\jmath}_{p,\psi}\,(\psi-\hat{\psi}).$$

Using the fact that $(1+x)^{-1/2} \approx 1-x/2$ for small $|x|$, one obtains the following expansion approximation to the density

$$p_{L.3}(\psi|\mathcal{D}) \doteq \big(1-\frac{1}{2}\dot{\hat{T}}'_\theta(\psi-\hat{\psi})\big)\frac{|\hat{\jmath}_{p,\psi}|^{\frac{1}{2}}}{(2\pi)^{\frac{d}{2}}}\exp\left\{-\frac{1}{2}(\psi-\hat{\psi})'\,\hat{\jmath}_{p,\psi}\,(\psi-\hat{\psi})\right\}. \quad (11)$$

Equation (11) can also be used to find the component-wise marginal densities. Let $\hat{\imath}_{p,\psi} = (\hat{\jmath}_{p,\psi})^{-1}$, and $\hat{\imath}_{p,\psi;k}$, $\hat{\imath}_{p,\psi;kk}$, be the k-th column and the (k,k)-th element of $\hat{\imath}_{p,\psi}$, respectively. Using the formulas of the marginal and conditional normal distribution, it is straightforward to verify that

$$p_{L.3}(\psi_k|\mathcal{D}) = \left(1-\frac{1}{2}\dot{\hat{T}}'_\theta\frac{\hat{\imath}_{p,\psi;k}}{\hat{\imath}_{p,\psi;kk}}\,(\psi_k-\hat{\psi}_k)\right)\frac{1}{\sqrt{\hat{\imath}_{p,\psi;kk}}}\phi\left(\frac{\psi_k-\hat{\psi}_k}{\sqrt{\hat{\imath}_{p,\psi;kk}}}\right), \quad (12)$$

where ϕ denotes the density of the standard Normal distribution.

3.2.1. Approximate credible regions

In order to obtain credible regions and/or intervals for the parameters, one needs to evaluate the tail probabilities of the posterior. Let $\mathcal{B} \subset \mathbf{R}^d$ be a d-dimensional ball centered at the origin (i.e., $0 \in \mathbf{R}^d$). Using the approximation $p_{L.3}(\psi|\mathcal{D})$ given by

equation (11) for sets in the moderate deviation range $\hat{\psi} + \mathcal{B}$, one obtains

$$\begin{aligned}\text{Prob}(\psi \in \hat{\psi} + \mathcal{B}) \approx \\ \Phi_d(\hat{\jmath}_{p,\psi}^{1/2}\,\mathcal{B}) - \frac{1}{2}\hat{\dot{T}}'_\theta \int_{\hat{\psi}+\mathcal{B}} \frac{|\hat{\jmath}_{p,\psi}|^{1/2}}{(2\pi)^{d/2}}(\psi - \hat{\psi}) \exp\{-\frac{1}{2}(\psi - \hat{\psi})'\,\hat{\jmath}_{p,\psi}\,(\psi - \hat{\psi})\}\,d\psi \\ = \Phi_d(\hat{\jmath}_{p,\psi}^{1/2}\,\mathcal{B}) + \frac{1}{2}\hat{\dot{T}}'_\theta\,\hat{\jmath}_{p,\psi}^{-1/2} \int_{\hat{\jmath}_{p,\psi}^{1/2}\mathcal{B}} \left[\frac{d}{du}\frac{1}{(2\pi)^{d/2}} e^{-u'u/2}\right] du, \quad (13)\end{aligned}$$

where Φ_d denotes the multivariate standard Normal distribution function in $\mathbf{R}^d$. For $d = 1$, the above formula for $\psi_1, \psi_2 \geq 0$ reduces to

$$\begin{aligned}\text{Prob}(\hat{\psi} + \psi_1 \leq \psi \leq \hat{\psi} + \psi_2) \approx \\ \Phi_1(\hat{\jmath}_{p,\psi}^{1/2}\psi_2) - \Phi_1(\hat{\jmath}_{p,\psi}^{1/2}\psi_1) + \frac{1}{2}\hat{\dot{T}}_\theta\,\hat{\jmath}_{p,\psi}^{-1/2}\left[\phi(\hat{\jmath}_{p,\psi}^{1/2}\psi_2) - \phi(\hat{\jmath}_{p,\psi}^{1/2}\psi_1)\right].\end{aligned}$$

For $d \in \{2, 3\}$, one can use the divergence theorem to evaluate the above probability in (13).

3.3. The case of orthogonal parameters

Note that in our case, the parameter $\vec{\tau}$ is orthogonal to (β, b). This is easily seen from Proposition 1 of the Appendix. Therefore, the formulas may be simplified by assuming that $\partial^2\ell_q/(\partial\beta\partial\vec{\tau}) \approx 0$ and $\partial^2\ell_q/(\partial b\partial\vec{\tau}) \approx 0$. The explicit simplifications are shown in this section and in Section B of the Appendix. Consider again $T_\theta = T_\theta(\psi, \theta_\psi) = \sum_u j_\theta^{-u} j_{\theta,u}(\psi, \theta_\psi)$, where j_θ^{-u} and $j_{\theta,u}$ denote the u-row of j_θ^{-1} and the u-column of j_θ, respectively. We have

$$\dot{T}_\theta = -\sum_u \sum_v j_\theta^{-uv} \frac{d}{d\psi}\left\{\frac{\partial^2\ell_q}{\partial\theta_u\partial\theta_v}(\psi, \theta_\psi)\right\}(\psi, \theta_\psi),$$

where (θ_u, θ_v) denotes a pair of components of θ. Now let us write $\theta = (\zeta, \xi)$ and suppose that $\xi \in \mathbf{R}^{d_2}$ is orthogonal to both $\zeta \in \mathbf{R}^{d_1}$ and ψ. In this latter case, the matrix $\partial^2\ell_q/\partial\theta^2$ may be approximated by the block diagonal matrix $\text{diag}(\partial^2\ell_q/\partial\zeta^2, \partial^2\ell_q/\partial\xi^2)$, and

$$\frac{\partial^2\ell_q}{\partial\theta\partial\psi} \approx (\frac{\partial^2\ell_q}{\partial\zeta\partial\psi}, 0)', \quad \frac{\partial\zeta_\psi}{\partial\psi} \approx -[\frac{\partial^2\ell_q}{\partial\zeta^2}(\psi, \theta_\psi)]^{-1}\frac{\partial^2\ell_q}{\partial\zeta\partial\psi}(\psi, \theta_\psi), \quad \frac{\partial\xi_\psi}{\partial\psi} \approx 0. \quad (14)$$

These approximations yield

$$\begin{aligned}T_\theta(\psi, \theta_\psi) \approx \text{trace}\Bigg\{[\frac{\partial^2\ell_q}{\partial\zeta^2}(\hat{\psi}, \hat{\theta})]^{-1}\left(\frac{\partial^2\ell_q}{\partial\zeta^2}(\psi, \theta_\psi) - \frac{\partial^2\ell_q}{\partial\zeta^2}(\hat{\psi}, \hat{\theta})\right) \\ + [\frac{\partial^2\ell_q}{\partial\xi^2}(\hat{\psi}, \hat{\theta})]^{-1}\left(\frac{\partial^2\ell_q}{\partial\xi^2}(\psi, \theta_\psi) - \frac{\partial^2\ell_q}{\partial\xi^2}(\hat{\psi}, \hat{\theta})\right)\Bigg\}, \quad (15)\end{aligned}$$

$$\frac{d}{d\psi}\frac{\partial^2 \ell_q}{\partial\theta_u\partial\theta_v}(\hat\psi,\hat\theta)$$
$$\approx \frac{\partial^3 \ell_q}{\partial\theta_u\partial\theta_v\partial\psi}(\hat\psi,\hat\theta) - \frac{\partial^3 \ell_q}{\partial\theta_u\partial\theta_v\partial\zeta}(\hat\psi,\hat\theta)[\frac{\partial^2 \ell_q}{\partial\zeta^2}(\hat\psi,\hat\theta)]^{-1}\frac{\partial^2 \ell_q}{\partial\zeta\partial\psi}(\hat\psi,\hat\theta). \quad (16)$$

Further computations detailed in Section B of the Appendix allows to simplify these expressions even further. For example, for the most interesting case $\psi=\beta$, $\zeta=b$, $\xi=\vec\tau$, we have,

$$\hat{\dot T}_{b,\vec\tau}(\hat\beta,\hat b,\hat{\vec\tau}) \approx$$
$$-\sum_{j=1}^{p}\sum_{u=1}^{q}\sum_{v=1}^{q}\hat\jmath_b^{-(j,u)(j,v)}\frac{d}{d\beta}\frac{\partial^2 \ell_q}{\partial b_{ju}\partial b_{jv}}(\hat\beta,\hat b,\hat{\vec\tau}) - \sum_{j=1}^{p}\hat\jmath_{\vec\tau}^{-jj}\frac{d}{d\beta}\frac{\partial^2 \ell_q}{\partial\tau_j^2}(\hat\beta,\hat b,\hat{\vec\tau})$$
$$= -\sum_{j=1}^{p}\sum_{u=1}^{q}\sum_{v=1}^{q}\hat\jmath_b^{-(j,u)(j,v)}\left(\frac{\partial^3 \ell_q}{\partial b_{ju}\partial b_{jv}\partial\beta}(\hat\beta,\hat b,\hat{\vec\tau}) - \frac{\partial^3 \ell_q}{\partial b_{ju}\partial b_{jv}\partial b}(\hat\beta,\hat b,\hat{\vec\tau})\,[\hat\jmath_b]^{-1}\hat\jmath_{b\times\beta}\right)$$
$$-\sum_{j=1}^{p}\hat\jmath_{\vec\tau}^{-jj}\left(\frac{\partial^3 \ell_q}{\partial\tau_j^2\partial\beta}(\hat\beta,\hat b,\hat{\vec\tau}) - \frac{\partial^3 \ell_q}{\partial\tau_j^2\partial b}(\hat\beta,\hat b,\hat{\vec\tau})\,[\hat\jmath_b]^{-1}\hat\jmath_{b\times\beta}\right).$$

4. Simulation study

We show in this section the results of a comparison study between the approximate posterior inference developed in the previous sections and the posterior inference derived from Markov chain Monte-Carlo (MCMC) samples. The MCMC samples were obtained using the package JAGS (Plummer, 2013; Lunn et al., 2009). Both methods are implemented in the **R** statistical software package. The method introduced in this paper is available from the author's web-site. JAGS is available as the `rjags` package from the usual **R** repository `cran.r-project.org`.

For the generation of the data sets, we mimicked the characteristics of the vaccine trial data. All data sets consisted of ten groups with a random number of levels (categories) in each group. The number of levels was generated uniformly with values between two and ten. The number of groups was kept fixed to ten so as to ease the coding of the associated JAGS file. JAGS does not allow a variable number of levels (columns) in two-dimensional arrays. The fixed effects $X_{ij}=(1,X_t)$ consisted of two-dimensional arrays equal to $(1,0)$ or $(1,1)$ depending on the value of the "treatment" variable X_t. The first 1 in X_{ij} is associated with the constant β_0 of the model. The parameters $\beta=(\beta_0,\beta_1)$ were generated as independent Normal deviates with means μ and variances equal to one. The means μ were generated from a Normal distribution with mean $(-3,3)$ and variance-covariance matrix equals to the identity matrix. The group effects $b_j\in\mathbf{R}$ were generated from a Normal distribution with zero-mean and variance equals to $\sigma_b^2=0.25$. The log-precision parameters τ_j were also generated from a Normal distribution with mean $\tau=2$ and variance σ_τ^2. This latter was generated form an inverse-chi-squared distribution with five degrees of freedom and scale parameter equals to one. Finally,

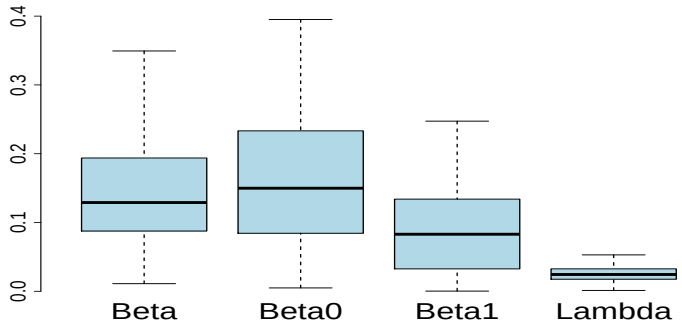

Figure 1: Boxplots of the absolute difference between the approximate posterior inference procedure and the MCMC sampler estimates of the parameters β_0, β_1, the vector parameter $\beta = (\beta_0, \beta_1)$ and the intensities λ_{ij}.

the counts z_{ij} and intensitiy parameters λ_{ij} were generated as in the original model described in Section 2.

The results of our simulation are based on a hundred data sets generated as explained above. The hyper-parameters were set as follows: $m_0 = 2$, $\Lambda_0 = 1$, $\sigma_0^2 = s_0^2 = v_0^2 = 10$, $\nu_0 = 3$, and $\tau_0 = 1$. We initialized the parameters with easy to obtain sensible estimates. For example, the parameters β were initialized using the coefficients of the robust regression of log(observed intensity) as function of the treatment variable, given by Huber's M-estimation as implemented in the method *rlm()* from the package *MASS* of **R** (Venables and Ripley, 1994, p. 216).

For each data set, we computed the difference between the MAP estimates obtained from the approximate posterior inference (MAP-API) and the posterior mean estimates obtained from the MCMC samples (JAGS). Boxplots of the absolute differences associated with the parameters β_0, and β_1, the vector parameter $\beta = (\beta_0, \beta_1)$, and the intensities λ_{ij} are displayed in Figure 1. Observe that despite some difference in the β estimates, the intensity parameters estimates are very close between the two methods. This is particularly important, since the multiple testing procedure proposed in the next section is based on the intensities λ_{ij}. We have also computed the differences between the MAP-API estimates and the true means μ for the parameters β_0 and β_1. Boxplots of these quantities are displayed in Figure 2. The corresponding differences associated with the JAGS estimates are also displayed in the figure. We can see that the difference to the true mean statistics are similar for both methods; however, the MAP-API estimates appear less biased (centered at the true values) than the JAGS estimates.

We also compared the similarity between the 95% credible intervals (CI) for β_0 and β_1 obtained from each one of the two methods. The upper right plot of Figure 3 displays density estimates of the length of the intervals for the API CI estimates and the JAGS CI estimates. The upper left plot of the figure displays a density estimate of the ratio between the length of the API CI and the corresponding JAGS CI. We note the similarity

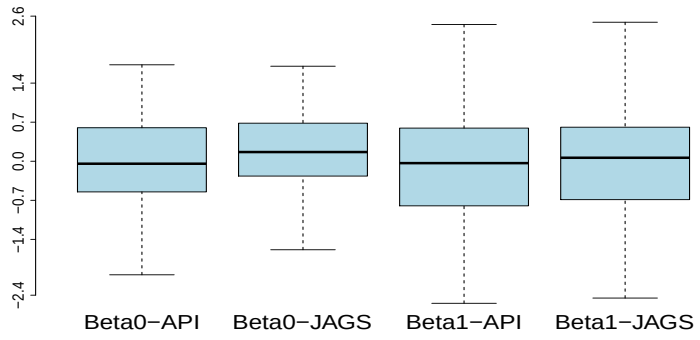

Figure 2: Boxplots of (a) the differences between the approximate posterior inference procedure estimates of the parameters β_0 and β_1 and the true mean of these parameters μ; and (b) the difference between the MCMC sampler estimates of the parameters β_0 and β_1 and the true mean of these parameters μ. Here *API* stands for the estimates obtained with the approximate posterior inference procedure, and *JAGS*, for the estimates obtained with the MCMC sampler.

in the distributions of the interval lengths. However, the left plot also appears to indicate that the API CI estimates have a slight tendency to be larger than those obtained with JAGS.

In order to further compare the CI estimates, we computed the F_1-measure of agreement (Allan et al., 1998) between the intervals. Let I_1, I_2 be two CI, and let $|I_1|, |I_2|$ be their lengths. Also, let $|I_1 \cap I_2|$ be the length of the intersection between the credible regions. The F_1-measure of agreement between these two CI is given by $F_1 = (\frac{1}{2}\{\frac{1}{R}+\frac{1}{P}\})^{-1}$ where $R = |I_1 \cap I_2|/|I_1|$ and $P = |I_1 \cap I_2|/|I_2|$, are the so-called, recall and precision measures. The F_1 measure has been extensively used in the text mining literature and more recently in the clustering and biclustering literature (Santamaria et al., 2007; Turner et al., 2005) Note that $0 \leq F_1 \leq 1$, and $F_1 = 1$ when there is a perfect match. In general, the larger the F_1-measure, the better the agreement between the credible intervals. The bottom plot of Figure 3 displays a density estimate of the F_1-measure of agreement between the CI estimates. In general, we see a good agreement between the credible regions. We would like to note that the simulation data are not very large. The median size (and average size) of the data in the simulations was $n = 60$. Therefore, the approximate posterior inference procedure did a remarkable job even though these datasets were of moderate size.

5. Multiple testing

In several studies it is necessary to test whether two intensities are the same or not. For example, in a clinical study, one may want to compare the intensities associated

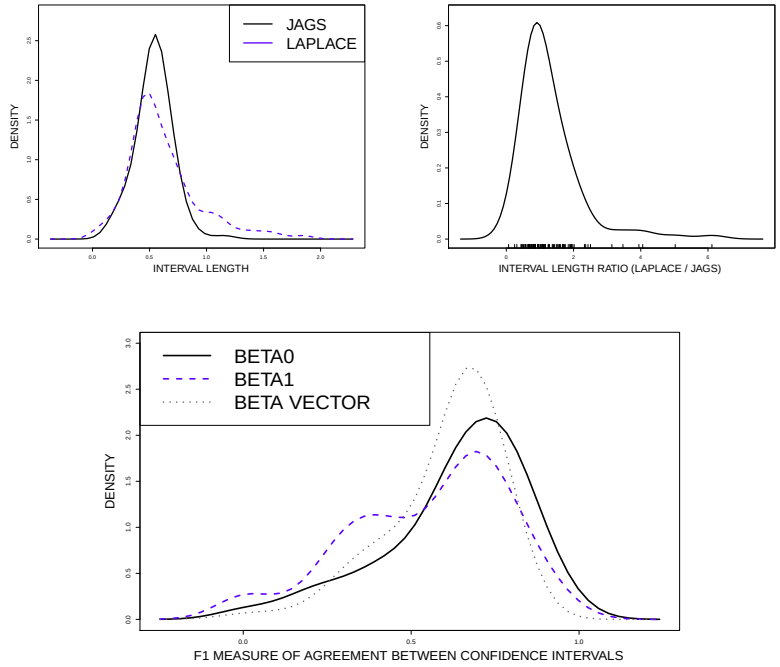

Figure 3: The upper left plot displays the distribution of the credible interval lengths generated by the approximate posterior inference (API) procedure and the MCMC sampler. Here *JAGS* stands for the lengths associated with the MCMC sampler, while *LAPLACE* stands for the lengths associated with the API procedure. The upper right plot displays an estimate of the density of the ratio between the lengths of the credible regions obtained with the API procedure and the MCMC sampler. The lower row displays estimates of the distribution of the F_1-measure of agreement between the API and MCMC sampler estimates of the credible intervals of the parameters β_0 and β_1, and the vector parameter $\beta = (\beta_0, \beta_1)$.

with the treatment λ_1 and the control λ_2 groups. In this case, one needs to estimate $p_\delta \doteq P(\lambda_1 - \lambda_2 < 0|\mathcal{D})$. In the case of multiple testing, the difference $\lambda_1 - \lambda_2$ may be used as a quantity whose extremely small values favor the hypothesis of a true (positive) treatment effect. Therefore, following the ideas of Whittemore (2007), we use p_δ as the base of a Bayes false discovery rate (Bayes FDR) procedure for Poisson regression. In the next three sections we proceed to find an estimate of p_δ, and then show how to use these probabilities in a multiple testing setup.

5.1. The marginal posterior of the intensities λ

Let $g(\lambda|\alpha,\gamma)$ denote the gamma distribution with shape and rate parameters α and γ, respectively, i.e. $g(\lambda|\alpha,\gamma) \propto \lambda^{\alpha-1}\exp\{-\gamma\lambda\}$. the posterior of λ_{ij} is given by

$$\pi(\lambda_{ij}|\mathcal{D}) = \int g(\lambda_{ij}|z_{ij}+e^{\tau_j},\, e_{ij}+e^{\tau_j}/\mu_{ij})\,\pi(\beta,\vec{\tau},b|\mathcal{D})d\beta\,d\vec{\tau}\,db.$$

A quick estimate, obtained through Laplace's approximation, yields

$$\pi(\lambda_{ij}|\mathcal{D}) \approx g(\lambda_{ij}|z_{ij}+e^{\hat{\tau}_j},\, e_{ij}+e^{\hat{\tau}_j}/\hat{\mu}_{ij}).$$

That is, the posterior is close to a Gamma distribution. Christiansen and Morris (1997) supposed that the posterior of the intensities in their model may be approximated by a Gamma distribution but gave no formal argument to support this assumption. However, their estimates of α and γ do not coincide with those derived from Laplace's approximation. They used instead the so-called adjusted density method (ADM) (Morris, 1983) to estimate them. These estimates were derived from a Beta distribution approximation to the variables B_{ij}. Here, we will still use Laplace's approximation derived in Section 3.1. For $j \in \{1,\ldots,p\}$, let $\vec{\tau}_{-j} = \{\tau_k : k \neq j\}$ and $b_{-j} = \{b_k : k \neq j\}$. Similarly, $\hat{\vec{\tau}}_{-j} = \{\hat{\tau}_k : k \neq j\}$ and $\hat{b}_{-j} = \{\hat{b}_k : k \neq j\}$ will denote the maximum q-likelihood estimators. From now on, for any quantity Q, the expression $\hat{Q}_{-j}$ will denote the quantity Q evaluated at $(\tau_j, \hat{\vec{\tau}}_{-j})$ and $(b_j, \hat{b}_{-j})$. Let $\ell_j \doteq \sum_{s=1}^{M_j} N_{js}\log(e^{\tau_j}+s-1) - n_j\bar{z}_{\cdot j}\tau_j + z_j'W_jb_j + \sum_{k=1}^{p}\{z_k'X_k\beta + \sum_{i=1}^{n_k}(e^{\tau_k}+z_{ik})\log(B_{ik})\}$, where $\bar{z}_{\cdot j}$ is the mean of the counts over the j-th group. Consider

$$\log q_j = \log q_j(\beta,\tau_j,b_j) \doteq \ell_j(\beta,\tau_j,b_j,\hat{\vec{\tau}}_{-j},\hat{b}_{-j}) - \frac{1}{2v_0^2}(\hat{a}_{-j}-\tau_0)^2$$
$$- \frac{1}{2}(\nu_0+p-1)\log(\hat{A}_{-j}) - \frac{1}{2}(m_0+p)\log(1+\sigma_0^{-2}(b_j'\Lambda_0^{-1}b_j + \sum_{k\neq j}^{p}\hat{b}_k'\Lambda_0^{-1}\hat{b}_k)).$$

This latter quantity contains all the dependency in the q-likelihood $q(\beta,\vec{\tau},b)$ that depends only on (β,τ_j,b_j). Let $h = \log g(\lambda_{ij}|z_{ij}+e^{\tau_j},\, e_{ij}+e^{\tau_j}/\mu_{ij}) + \log q(\beta,\vec{\tau},b)$, and j_h denote minus the Hessian of h. Applying the approximation (10) to h instead of ℓ_q, with

$\psi = \lambda \doteq \lambda_{ij}$ and $\theta = (\beta, \vec{\tau}, b)$ yields

$$p_{L.2}(\lambda|\mathcal{D}) = (\hat{\jmath}_{h,p,\lambda}/(2\pi))^{1/2}(1 + T_\theta(\lambda, \beta_\lambda, \vec{\tau}_\lambda, b_\lambda))^{-1/2} h(\lambda, \beta_\lambda, \vec{\tau}_\lambda, b_\lambda)/h(\hat{\lambda}, \theta_{\hat{\lambda}}),$$

where $\hat{\theta}_{\hat{\lambda}} = (\beta_{\hat{\lambda}}, \tau_{\hat{\lambda}}, b_{\hat{\lambda}})$ is not necessarily $\hat{\theta} = (\hat{\beta}, \hat{\vec{\tau}}, \hat{b})$, since the former value satisfies $(\partial \log q_j/\partial\beta)(\hat{\theta}_{\hat{\lambda}}) = -(\partial \log g/\partial\beta)(\hat{\lambda}, \hat{\theta}_{\hat{\lambda}})$, $(\partial \log q_j/\partial\tau_j)(\hat{\theta}_{\hat{\lambda}}) = -(\partial \log g/\partial\tau_j)(\hat{\lambda}, \hat{\theta}_{\hat{\lambda}})$, and $(\partial \log q_j/\partial b_j)(\hat{\theta}_{\hat{\lambda}}) = -(\partial \log g/\partial b_j)(\hat{\lambda}, \hat{\theta}_{\hat{\lambda}})$. However, the maximizer $\hat{\lambda} = \hat{\mu}_{ij,\hat{\lambda}}(z_{ij} + e^{\hat{\tau}_{j,\hat{\lambda}}} - 1)/(e_{ij}\hat{\mu}_{ij,\hat{\lambda}} + e^{\hat{\tau}_{j,\hat{\lambda}}}) \approx \bar{\lambda} \doteq \hat{\mu}_{ij,\hat{\lambda}}(z_{ij} + e^{\hat{\tau}_{j,\hat{\lambda}}})/(e_{ij}\hat{\mu}_{ij,\hat{\lambda}} + e^{\hat{\tau}_{j,\hat{\lambda}}})$ for large z_{ij}. It is easy to verify that $(\partial \log g/\partial\beta)(\bar{\lambda}, \theta_{\bar{\lambda}}) = 0$, $(\partial \log g/\partial b_j)(\bar{\lambda}, \theta_{\bar{\lambda}}) = 0$, and

$$\frac{\partial \log g}{\partial \tau_j}(\theta_{\bar{\lambda}}) = e^{\tau_{j,\bar{\lambda}}}[\log(z_{ij} + e^{\tau_{j,\bar{\lambda}}}) - \Psi(z_{ij} + e^{\tau_{j,\bar{\lambda}}})] \approx \frac{e^{\tau_{j,\bar{\lambda}}}}{2(z_{ij} + e^{\tau_{j,\bar{\lambda}}})},$$

where we have used the approximation for the digamma function $\Psi(x) \approx \log(x) - 1/(2x)$. Note that the last derivative with respect to τ_j is small for large z_{ij}. Therefore, we will suppose that $p_{L.2}(\lambda|\mathcal{D})$ is computed using the expected value $\bar{\lambda}$ instead of the mode $\hat{\lambda}$. In this latter case, $\theta_{\bar{\lambda}} \approx \hat{\theta}$, and $\hat{h} = h(\bar{\lambda}, \hat{\beta}, \hat{\vec{\tau}}, \hat{b})$. This simplifies the calculation that follows as well as the computations in practice, since there will be no need to evaluate $\hat{\theta}_{\hat{\lambda}}$ for each individual λ_{ij}.

Also note that since for all $k \neq j$, $\partial^2 h/\partial\lambda\partial\tau_k = \partial^2 h/\partial\lambda\partial b_k = 0$, we have $(d/d\lambda)\tau_{k,\lambda} \approx 0$ and $(d/d\lambda)b_{k,\lambda} \approx 0$. Therefore, $\vec{\tau}_\lambda \approx (\tau_{j,\lambda}, \hat{\vec{\tau}}_{-j})$ and $b_\lambda \approx (b_{j,\lambda}, \hat{b}_{-j})$. This implies that we may consider the approximation to the posterior of λ:

$$\begin{aligned} p_{L.2}(\lambda|\mathcal{D}) = (\hat{\jmath}_{h,p,\lambda}/(2\pi))^{1/2}\big(1 + T_\theta(\lambda, \beta_\lambda, \tau_{j,\lambda}, b_{j,\lambda}, \hat{\vec{\tau}}_{-j}, \hat{b}_{-j})\big)^{-1/2} \\ \times \big(q_j(\beta_\lambda, \tau_{j,\lambda}, b_{j,\lambda})/\hat{q}_j\big)\big(g(\lambda|z_{ij} + e^{\tau_{j,\lambda}},\, e_{ij} + e^{\tau_{j,\lambda}}/\mu_{ij,\lambda})/\hat{g}\big), \quad (17) \end{aligned}$$

where $\mu_{ij,\lambda}$ denotes μ_{ij} evaluated at $(\beta_\lambda, \tau_{j,\lambda}, b_{j,\lambda})$, and $\hat{q}_j$, and $\hat{g}$ denote q_j, and g evaluated at $(\bar{\lambda}, \hat{\theta})$.

5.2. Comparing two intensities

Suppose that λ_i follows a distribution $g_i(\lambda)$, $i = 1, 2$, and that these two variables are independent. Let $U_1 = \lambda_1 - \lambda_2$, and $U_2 = \lambda_1$. It is easily shown that $p(u_1) = \int_{u_2 > \max\{u_1, 0\}} g_1(u_2)g_2(u_2 - u_1)\, du_2$. We will approximate this integral using a hybrid between a numerical integration and Laplace's approximation. The resulting expression may be integrated numerically to find p_δ. We use the form of the approximation (17) for $g_i(\lambda) = f_i(\lambda)g(\lambda|\alpha_i, \gamma_i)$, $i = 1, 2$, where $g(\cdot|\alpha_i, \gamma_i)$ denotes a gamma density with shape parameter α_i, and rate γ_i. We have

$$p(u_1, u_2) \propto f_1(u_2)f_2(u_2 - u_1) \times (u_2 - u_1)^{\alpha_2 - 1}u_2^{\alpha_1 - 1} \exp\{-(\gamma_1 + \gamma_2)u_2 + \gamma_2 u_1\},$$

$u_2 \geq \max\{u_1, 0\}$. Let $I_2(u_2) = \int_{u_2}^{\infty} \frac{\gamma_2^{\alpha_2}}{\Gamma(\alpha_2)} f_2(v)v^{\alpha_2 - 1}e^{-\gamma_2 v}\, dv$. And let $v_0 = \arg\max_{v \geq u_2}(f_2(v)v^{\alpha_2 - 1}e^{-\gamma_2 v})$. Note that v_0 may be zero if $\alpha_2 < 1$. Using Laplace's approximation and the fact that,

besides f_2, the integrand is a gamma density, we have

$$I_2(u_2) \approx \begin{cases} f_2(v_0)(1 - G(u_2|\alpha_2, \gamma_2)), & \text{if } u_2 < v_0, \\ f_2(u_2)(1 - G(u_2|\alpha_2, \gamma_2)), & \text{otherwise.} \end{cases}$$

where $G(|)$ denotes the gamma cdf. Therefore

$$\begin{aligned} p_\delta \propto P^- \doteq f_2(v_0) \int_0^{v0} \frac{\gamma_1^{\alpha_1}}{\Gamma(\alpha_1)} f_1(u_2) u_2^{\alpha_1 - 1} e^{-\gamma_1 u_2} \left(1 - G(u_2|\alpha_2, \gamma_2)\right) du_2 \\ + \int_{v_0}^{\infty} \frac{\gamma_1^{\alpha_1}}{\Gamma(\alpha_1)} f_1(u_2) f_2(u_2) u_2^{\alpha_1 - 1} e^{-\gamma_1 u_2} \left(1 - G(u_2|\alpha_2, \gamma_2)\right) du_2 \\ = f_2(v_0) \int_0^{v0} f_1(u_2) g(u_2|\alpha_1, \gamma_1)(1 - G(u_2|\alpha_2, \gamma_2)) \, du_2 \\ + \int_{v_0}^{\infty} f_1(u_2) f_2(u_2) g(u_2|\alpha_1, \gamma_1)(1 - G(u_2|\alpha_2, \gamma_2)) \, du_2. \end{aligned}$$

In practice, we calculate the maximum of $g(u_2|\alpha_1, \gamma_1)(1 - G(u_2|\alpha_2, \gamma_2))$, say u_*, and divide the real line into a few bins with end-points $u_1 < u_2 < \ldots < u_m$ so that, if $S_i = \int_{u_i}^{u_{i+1}} g(u_2|\alpha_1, \gamma_1)(1 - G(u_2|\alpha_2, \gamma_2)) \, du_2$,, then

$$\int_{u_i}^{u_{i+1}} f_1(u_2) f_2(u_2) g(u_2|\alpha_1, \gamma_1)(1 - G(u_2|\alpha_2, \gamma_2)) \, du_2 \approx \begin{cases} f_1(u_*) f_2(u_*) S_i, & \text{if } u_i \le u_* < u_{i+1}, \\ f_1(u_i) f_2(u_i) S_i, & \text{otherwise.} \end{cases}$$

In order to get an estimate of p_δ, we need to normalize the estimate of P^- by $(P^- + P^+)$, where

$$P^+ = f_2(v_0) \int_0^{v0} f_1(u_2) g(u_2|\alpha_1, \gamma_1) G(u_2|\alpha_2, \gamma_2) \, du_2 + \int_{v_0}^{\infty} f_1(u_2) f_2(u_2) g(u_2|\alpha_1, \gamma_1) G(u_2|\alpha_2, \gamma_2) \, du_2,$$

is proportional to $1 - p_\delta$. We approximate P^+ in a similar fashion as P^-.

In many situations we would like to compare the intensities associated with different values of the fixed effect variable X (for example, for control and treatment patients). Let us denote these two values by x_1 and x_2. There are two different cases. The first one is given by $\lambda_{ij}(x_1) - \lambda_{i'j}(x_2)$. The second one is given by $\lambda_{ij}(x_1) - \lambda_{lk}(x_2)$ with $j \neq k$. Note that in the latter case, it may be interesting to test if these intensities are the same for the same value of the fixed effects, i.e. we may consider here $x_1 = x_2$. However, the most important scenario is to compare the same intensities associated with different values of X. This corresponds to the first case with $i' = i$. The conditional intensities $\lambda_{ij}|\beta, \tau_j, b_j, x_1, \mathcal{D}$ and $\lambda_{lk}|\beta, \tau_k, b_k, x_2, \mathcal{D}$ are independent if either, $i \neq l$, $j \neq k$, or $x_1 \neq$

x_2. Therefore, we can approximate $\pi(\lambda_{ij}(x_1) - \lambda_{lk}(x_2) < 0|\beta, \tau_j, \tau_k, b_j, b_k, x_1, x_2, \mathcal{D})$ as above. The resulting expression, say, $\hat{p}_L(\lambda_{ij}(x_1) - \lambda_{lk}(x_2) < 0|\beta, \tau_j, \tau_k, b_j, b_k, x_1, x_2, \mathcal{D})$, need to be integrated with respect to the measure $\pi(\beta, \tau_j, \tau_k, b_j, b_k|\mathcal{D})$. Consequently, a simple approximation to $\pi(\lambda_{ij}(x_1) - \lambda_{lk}(x_2) < 0|\mathcal{D})$ is given by $\hat{p}_L(\lambda_{ij}(x_1) - \lambda_{lk}(x_2) < 0 \,|\, \hat{\beta}, \hat{\tau}_j, \hat{\tau}_k, \hat{b}_j, \hat{b}_k, x_1, x_2, \mathcal{D})$.

5.3. A Bayes false discovery rate

Since we are considering multiple tests simultaneously, we need to apply a multiple comparison procedure to evaluate if these effects are likely to be real effects. We follow the Bayesian FDR suggested by Whittemore (2007). In this framework, we compare the results yielded by the full model with those yielded by the model that supposes that the null hypothesis of no treatment effect is true, that is, the model that imposes the constraint $\beta = 0$. Since this latter model, which we are going to call, the null-hypothesis model, is a special case of the full model, we can easily obtain the results under the null-hypothesis by fixing β to zero in the above equations. Whittemore (2007) considers comparing the probabilities of having an extreme value of an associated test statistics under both the null and the alternative (full model) hypothesis. This idea is easily generalized to our case, by considering the difference $\lambda_1 - \lambda_2$ as our "test statistics". This entails computing the probability p_δ under both hypothesis. Let π_i be the (prior) probability that the i-th null-hypothesis is false. Let $p_{i,\delta}$ be the associated p_δ for the i-hypothesis, and let $p_{i,\delta,o}$ be the corresponding probability computed from the null-hypothesis model, $i = 1, \ldots, N$, where N is the number of tests. Consider the posterior odds against the null-hypothesis

$$o_i = \frac{\pi_i}{1 - \pi_i} \frac{p_{i,\delta}}{p_{i,\delta,o}}.$$

The Bayes FDR is based on the so-called b-values: $b_i = 1/(1 + o_i)$. For a level $0 < \alpha < 1$, Whittemore (2007) shows that the Bayes FDR is controlled at a level α if all hypothesis satisfying $b_i \leq \alpha$ are rejected.

6. Applications

We start by applying our methodology to two problems that do not involve random effects. These are the pump and heart data sets already studied by Christiansen and Morris (1997). The goal here is not to do a deep analysis of these data, but to compare our results with those already published. In so doing, we aim at showing that our methodology works well in these simple cases. We note that the results of Christiansen and Morris (1997) have been calibrated by extensive simulations. They have also been compared to full Markov chain Monte Carlo techniques via WinBugs (The BUGS Project, 2012). The next section shows a more complex application involving group effects, namely the quadrivalent vaccine trial described in the Introduction.

6.1. The pump and heart rate data sets

The pump data (Gaver and O'Muircheartaigh, 1987) refers to pump failure rates at pressurized water in nuclear reactor power plants. These data have been analyzed using several different models (Carlin and Gelfand, 1991; Johnson, 1992; Christiansen and Morris, 1997). Each observation z_i ($i = 1, \ldots, 10$) represents the number of failures in the i-th pump. There is only one covariable that takes the values $x_{i1} = 1$ for four pumps that operates continually, and $x_{i1} = 0$ for the other six pumps that operate intermittently. The exposure e_i of the i-th pump represents the number of pump hours of operation. Thus, $y_i = z_i/e_i$ is the observed failure rate associated with the i-th pump.

The heart transplant data (Christiansen and Morris, 1996) are mortality rates from patients that had undergone a heart transplant. The data was gathered from fifteen heart transplant centers in the USA between October-1987 and December-1989. For each center i ($i = 1, \ldots, 15$), z_i represents the number of deaths during the first month after the transplant. The exposure e_i is the number of heart transplants performed in the corresponding center. As in the analysis of Christiansen and Morris (1997), we use the variable $X =$*severity* as a covariable. This refers to the expected number of deaths in each center. The severity index is based on several demographic and health variables considered as risk factors for the patients treated in the different centers. For these two data sets there is only one strata, hence $p = 1$.

As in the simulation study of Section 4, for all the applications of this section, we set the hyper-parameters values to $s_0^2 = v_0^2 = 10$, $\nu_0 = 3$, and $\tau_0 = 1$. We initialized the parameters with sensible estimates. For example, the parameters β were initialized using an appropriate robust regression.

The fixed effects MAP estimates together with 95% credible regions are shown in Table 1. These credible regions were obtained by integration of the approximate density given by expression (12), as in Section 3.2.1. The posterior standard deviations were also estimated from the corresponding approximate marginal posterior. The MAP intensity estimates are shown in Table 2. Our results are very similar to those obtained by Christiansen and Morris (1997). Note that at first sight the pump data estimates of β, are quite different. This is simply due to the different coding of the type of pump variable (we use a 0/1 coding, whilst Christiansen and Morris (1996) use a $-1.0/1.5$ coding). We also note that there seems to be a typo in their reported fixed effects estimates for the heart transplant data: the coefficient associated with the intercept is reported as -0.295. We believe that the right figure is -2.95, otherwise, their estimates of the shrinkage weights B_i and the intensities λ_i would be far off those reported in their paper. We would like to stress, that our method yields credible intervals that are not necessarily symmetric about the estimated value of the parameters. For the heart transplant data, the center severity does not seem to explain all the variability observed in the death rates after the heart transplants. Note that Center 1 presents the highest observed death rate among the fifteen centers. A one-to-one comparison between the center intensities $\lambda_1 - \lambda_j$ yield significant death rate differences between the first center and the all the other centers ($p_\delta < 0.15$), except center 3 ($p_\delta = 0.23$). On the opposite death count (zero), center 10 seems to be the most different center with $1 - p_\delta < 0.04$

for center 1, 3, and 9, and $1 - p_\delta < 0.15$ for centers 5,7, and 8. We also note that despite the fact that the observed death rate in center 5 is larger than that in center 7, $P(\lambda_5 - \lambda_7 < 0|$ data $) = 0.32$ is not small. This is explained by the smaller severity for patients in center 5 than for patients in center 7. This fact was already noted by Christiansen and Morris (1997) from the fact that the MAP estimate of λ_5 (0.044) is smaller than that of λ_7 (0.046). However, they could not produce a posterior probability statement like we do with our methodology.

Data	$\boldsymbol{\beta}$		95% credible region		Standard
			Lower	Upper	Deviation
Pump	Intercept	0.16	-0.42	0.72	0.31
Failure	Treatment	-1.78	-2.77	-0.79	0.54
Heart	Intercept	-2.96	-3.39	-2.53	0.24
Transplant	Treatment	1.40	-1.64	4.39	1.66

Table 1: The MAP regression parameter estimates for the Pump failure and Heart transplant data sets. The posterior standard deviations are estimated from the approximate marginal posterior of the parameters.

	Intensities $\boldsymbol{\lambda}_i$				
Pump	1	2	3	4	5
Failure	1.08	1.08	1.76	0.68	2.12
Data	6	7	8	9	10
	0.12	0.57	0.09	0.02	0.12

	1	2	3	4	5
Heart	0.144	0.046	0.091	0.038	0.044
Transplant	6	7	8	9	10
Data	0.035	0.046	0.055	0.084	0.015
	11	12	13	14	15
	0.092	0.043	0.039	0.084	0.068

Table 2: The MAP intensity parameter estimates for the Pump failure and Heart transplant data sets.

6.2. A quadrivalent vaccine trial

We applied our model to the human body system data introduced earlier in the paper. The covariable used was $x = 1$ for the treatment group, and $x = 0$ for the control group. The q-likelihood maximizers were estimated with a greedy gradient ascent algorithm. We initialized the parameters with easy to obtain sensible estimates. For example, the parameters β were initialized using the coefficients of the robust regression of log(observed intensity) as function of Treatment/Control. The hyper-parameters were set as follows: $m_0 = 2$, $\Lambda_0 = 1$, $\sigma_0^2 = s_0^2 = v_0^2 = 10$, $\nu_0 = 3$, and $\tau_0 = 1$.

The MAP estimates of the fixed effects (β_0, β_1) and their 95% credible regions, computed by integration of the approximating density given by the expression (12) as in Section 3.2.1, together with their posterior standard deviations (estimated from the observed Fisher information matrix) are displayed in Table 3. Note that the model says that in general there is a slight treatment effect. The MAP intensity estimates are shown in the Table 4. The probabilities $\pi(\lambda_{ij}(x = 1) - \lambda_{ik}(x = 0) < 0|\mathcal{D})$ were estimated with the procedure described in Section 5.2. One can observe from Table 4 that three adverse effect, *3-Diarrhea*, *8-Irritability* and to a minor degree *10-Rash*, are likely to be more exacerbated in the treatment group (the one receiving the quadrivalent vaccine) than in the control group. Also, due to the large differences between *Irritability* and the other two adverse effects, the body system group *8* presents the largest overdispersion ($\xi_8 = 0.43$, against a median of 0.82) among the eight body systems.

We applied the Bayes FDR procedure of Section 5.3 to the vaccine data. The last five columns of Table 4 shows the corresponding b-values for different values of the prior probabilities π_i against the null hypothesis of no treatment effect. Only *8-Irritability* present evidence against the null hypothesis of no treatment effect. These findings corroborate the findings of Mehrotra and Heyse (2004) and Berry and Berry (2004). It also shows that our method is a reliable method to handle multiple comparisons. In conclusion, this study did offer enough evidence to conclude that the quadrivalent vaccine do exacerbate some adverse effect in the treatment group. The study strongly indicates that the mood of the patients might be altered more often with the quadrivalent vaccine. Although, it was suspected that the digestive system and the skin may suffer some complications, the study failed to offer credible evidence that these and other areas of the body are affected by the quadrivalent vaccine.

$\boldsymbol{\beta}$		95% credible region		Standard
		Lower	Upper	Deviation
Intercept	-4.10	-4.30	-3.92	0.10
Treatment	0.57	0.24	0.90	0.18

Table 3: The vaccine data: The MAP regression parameter estimates. The posterior standard deviations are estimated from the approximate marginal posterior of the parameters.

7. Conclusions

We have presented both a Bayesian hierarchical mixed-effect Poisson regression model and a Bayes FDR methodology in order to address the problem of multiple testing. In this framework, the problem consists of comparing the risks or intensities associated with the different groups of interest under both the null (reduced model) hypothesis of no treatment effect, and the alternative (full model) one.

We note that in addition to the different treatment groups considered in a given study, every level L of a given discrete variable X give rise to a level-dependent subgroup formed by those individuals for which the corresponding value of X is L. We derived an approximation to the posterior density of the difference in risk between two treatment groups or two level-dependent subgroups. This density is then integrated numerically in order to obtain estimates of the probability that one of the groups or level-dependent subgroups shows a higher risk. Simplifications to the estimation of the joint or marginal posterior were achieved by using among other techniques, Laplace's approximation, and a trace approximation to the determinant involving the group-effect parameters. Orthogonality between the regression coefficients and the precision parameters allowed us to simplify even further the approximations. We note that Laplace's approximation has been widely used to approximate mixed-effect generalized linear models and has been proven to work both accurately and efficiently (as opposed to quadrature or Monte Carlo integration). We showed that our methodology generalizes the PRIMM methodology introduced by Christiansen and Morris (1997). In a sense, our method is more general, since the formulas derived in the present work are very general and may be applied to other hierarchical models as well. The application of our methodology to the quadrivalent vaccine trial (Mehrotra and Heyse, 2004; Berry and Berry, 2004) shows that there are significant differences between the standard two-stage vaccine and the quadrivalent one. In general, patients given the quadrivalent vaccine show a higher risk of suffering an adverse or secondary effect: *irritability* presented the most pronounced difference.

Acknowledgement

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) through grant number 327689-06.

A. Derivation of the q-likelihood $q(\beta, \vec{\tau}, b|\mathcal{D})$

Integrating $\pi(\beta, \vec{\tau}, b, \eta|\mathcal{D})$ with respect to Ψ one obtains

$$\pi_1(\beta, \vec{\tau}, b, (\tau, \sigma_\tau^2)|\mathcal{D}) = C_1 \int_\Psi \prod_{j=1}^{p} \pi(b_j|\eta)\, \pi(\Psi) d\Psi,$$

Body System	Adverse Effect	Counts T	Counts C	$\boldsymbol{\lambda}$ T	$\boldsymbol{\lambda}$ C	Diff
1	Fatigue	57	40	0.377	0.288	0.089
1	Fever	34	26	0.228	0.190	0.038
1	Fungal infection	2	0	0.021	0.008	0.012
1	Viral infection	3	1	0.027	0.015	0.012
1	Malaise	27	20	0.183	0.148	0.035
3	Anorexia	7	2	0.049	0.018	0.030
3	Oral candidiasis	2	0	0.017	0.005	0.012
3	Constipation	2	0	0.017	0.005	0.012
3	Diarrhea	24	10	0.154	0.070	0.084
3	Gastroenteritis	3	1	0.024	0.012	0.012
3	Nausea	2	7	0.017	0.051	-0.033
3	Vomiting	19	19	0.123	0.128	-0.005
5	Lymphadenopathy	3	2	0.043	0.028	0.015
6	Dehydration	0	2	0.003	0.014	-0.011
8	Crying	2	0	0.016	0.003	0.013
8	Insomnia	2	2	0.016	0.018	-0.002
8	Irritability	75	43	0.499	0.316	0.183
9	Bronchitis	4	1	0.031	0.012	0.018
9	Nasal congestion	4	2	0.031	0.019	0.012
9	Respiratory congestion	1	2	0.012	0.019	-0.007
9	Cough	13	8	0.087	0.059	0.029
9	Respiratory infection	28	20	0.182	0.139	0.044
9	Laryngotracheobronchitis	2	1	0.018	0.012	0.006
9	Pharyngitis	13	8	0.087	0.059	0.029
9	Rhinorrhea	15	14	0.100	0.099	0.001
9	Sinusitis	3	1	0.024	0.012	0.012
9	Tonsillitis	2	1	0.018	0.012	0.006
9	Wheezing	3	1	0.024	0.012	0.012
10	Bite/sting	4	0	0.027	0.005	0.022
10	Eczema	2	0	0.016	0.005	0.011
10	Pruritus	2	1	0.016	0.010	0.006
10	Rash	13	3	0.076	0.020	0.056
10	Diaper rash	6	2	0.038	0.015	0.023
10	Measles/rubella-like rash	8	1	0.048	0.010	0.039
10	Varicella-like rush	4	2	0.027	0.015	0.012
10	Urticaria	0	2	0.005	0.015	-0.010
10	Viral exanthema	1	2	0.010	0.015	-0.005
11	Conjunctivitis	0	2	0.004	0.017	-0.013
11	Otitis media	18	14	0.117	0.096	0.021
11	Otorrhea	2	1	0.017	0.011	0.006

Table 4: The adverse effects associated with eight body systems. The exposures for the treatment and control groups are 148 and 132, respectively. The symbols "T" and "C" stand for the treatment and control groups, respectively. The columns under $\boldsymbol{\lambda}$ are the MAP estimates of the corresponding group intensities. The column "Diff" shows the difference between the corresponding groups intensities.

where C_1 is a function independent of Ψ. Recall that $b_j \sim N_q(0, \Psi)$ and $\Psi \sim W^{-1}(m_0, \Lambda_0)$, i.e., $\pi(b_j|\eta) = \pi(b_j|\Psi) = (2\pi)^{-q/2}|\Psi|^{-1/2}\exp[-b_j'\Psi^{-1}b_j/2]$ and,

$$\pi(\Psi) = |\sigma_0^2\Lambda_0|^{m_0/2}|\Psi|^{-(m_0+q+1)/2}\exp[-\text{trace}(\sigma_0^2\Lambda_0\Psi^{-1})/2]/(2^{m_0q/2}\Gamma_q(m_0/2)).$$

These yield

$$\pi_1(\beta, \vec{\tau}, b, (\tau, \sigma_\tau^2)|\mathcal{D}) = \\ C_1\pi^{-pq/2}(\Gamma_q((m_0+p)/2)/\Gamma_q(m_0/2)) \times |\sigma_0^2\Lambda_0|^{m_0}\,|\sigma_0^2\Lambda_0 + b'b|^{-(m_0+p)/2}.$$

The integral of π_1 with respect to σ_τ^2 is

$$\pi_2(\beta, \vec{\tau}, b, \tau|\mathcal{D}) = C_2\int_{\sigma_\tau^2}\prod_{j=1}^{p}\pi(\tau_j|\tau, \sigma_\tau^2)\,\pi(\sigma_\tau^2)d\sigma_\tau^2.$$

where C_2 is a function independent of (σ_τ^2, Ψ). Recall again that $\tau_j \sim \text{Normal}(\tau, \sigma_\tau^2)$ and $\sigma_\tau^2 \sim$Inverse-Chi-squared(ν_0, s_0^2), i.e.,

$$\pi(\tau_j|\tau, \sigma_\tau^2) = (2\pi\sigma_\tau^2)^{-1/2}\exp[-(\tau_j-\tau)^2/2\sigma_\tau^2] \quad \text{and}$$

$$\pi(\sigma_\tau^2) = ((s_0^2\nu_0/2)^{\nu_0/2}/\Gamma(\nu_0/2)) \times (\sigma_\tau^2)^{-(\nu_0/2+1)}\exp(-\nu_0 s_0^2/2\sigma_\tau^2).$$

Hence,

$$\pi_2(\beta, \vec{\tau}, b, \tau|\mathcal{D}) = C_2\frac{(s_0^2\nu_0/2)^{\nu_0/2}\Gamma((\nu_0+p)/2)2^{(\nu_0+p)/2}}{(2\pi)^{p/2}\Gamma(\nu_0/2)}[pA(1+\frac{(\tau-a)^2}{A})]^{-(\nu_0+p)/2},$$

where $A = \frac{\nu_0 s_0^2}{p} + \frac{1}{p}\sum_{j=1}^{p}(\tau_j - a)^2$ and $a = \frac{1}{p}\sum_{j=1}^{p}\tau_j$. Finally, the integral of π_2 with respect to τ is

$$\pi_3(\beta, \vec{\tau}, b|\mathcal{D}) = C_3\int_\tau(1+\frac{(\tau-a)^2}{A})^{-(\nu_0+p)/2}\frac{1}{\sqrt{2\pi}\,v_0}\exp\{-\frac{(\tau-\tau_0)^2}{2v_0^2}\}\,d\tau,$$

where C_3 is a function independent of $(\tau, \sigma_\tau^2, \Psi)$. We are going to estimate this integral using Laplace's approximation. Let $g(\tau) = -((\nu_0+p)/2)\log(1+\frac{(\tau-a)^2}{A})$. Its first and second derivatives are $g'(\tau) = -(\nu_0+p)(\tau-a)/(A+(\tau-a)^2)$ and $g''(\tau) = -(\nu_0+p)[(A-(\tau-a)^2)/(A+(\tau-a)^2)^2]$. Observe that $\tau = a$ is the global maximum of $g(\tau)$. Hence, Laplace's approximation yields

$$\int_\tau \exp\{g(\tau)\}\,\pi(\tau)d\tau \approx \exp\{g(a)\}\sqrt{2\pi}\,|g''(a)|^{-1/2}\,\pi(a) = \\ \frac{1}{v_0\sqrt{\nu_0+p}}A^{1/2}\,\exp\{-\frac{(a-\tau_0)^2}{2v_0^2}\}.$$

So finally, $\pi_3(\beta, \vec{\tau}, b|\mathcal{D}) \approx$

$$c_0 L(\beta, \vec{\tau}, b|\mathcal{D})|\sigma_0^2 \Lambda_0 + b'b|^{-(m_0+p)/2}(pA)^{-(\nu_0+p-1)/2} \exp\{-\frac{(a-\tau_0)^2}{2v_0^2}\},$$

where c_0 is a constant independent of the parameters of the model. The right-hand-side of the above expression is, up to a constant, exactly $q(\beta, \vec{\tau}, b|\mathcal{D})$.

B. The case of orthogonal parameters

Consider $T_\theta = T_\theta(\psi, \theta_\psi) = \sum_u j_\theta^{-u} j_{\theta,u}(\psi, \theta_\psi)$, where j_θ^{-u} and $j_{\theta,u}$ denote the u-row of j_θ^{-1} and the u-column of j_θ, respectively. We have

$$\dot{T}_\theta = -\sum_u \sum_v j_\theta^{-uv} \frac{d}{d\psi}\left\{\frac{\partial^2 \ell_q}{\partial\theta_u \partial\theta_v}(\psi, \theta_\psi)\right\}(\psi, \theta_\psi),$$

where (θ_u, θ_v) denotes a pair of components of θ. In addition,

$$-\frac{d}{d\psi}(j_\theta(\psi, \theta_\psi))_{uv} = \frac{d}{d\psi}\frac{\partial^2 \ell_q(\psi, \theta_\psi)}{\partial\theta_u \partial\theta_v} = \frac{\partial^3 \ell_q(\psi, \theta_\psi)}{\partial\theta_u \partial\theta_v \partial\psi} + \frac{\partial^3 \ell_q(\psi, \theta_\psi)}{\partial\theta_u \partial\theta_v \partial\theta}\frac{\partial\theta_\psi}{\partial\psi}.$$

The last partial derivative may be obtained from the fact that by definition $(\partial\ell_q/\partial\theta)(\psi, \theta_\psi) = 0$ for all ψ. This implies that

$$\frac{\partial\theta_\psi}{\partial\psi} = -[\frac{\partial^2 \ell_q}{\partial\theta^2}(\psi, \theta_\psi)]^{-1}\frac{\partial^2 \ell_q}{\partial\theta\partial\psi}(\psi, \theta_\psi).$$

Now let us write $\theta = (\zeta, \xi)$ and suppose that $\xi \in \mathbf{R}^{d_2}$ is orthogonal to both $\zeta \in \mathbf{R}^{d_1}$ and ψ. As mentioned in Section 3.3, the matrix $\partial^2\ell_q/\partial\theta^2$ may be approximated by the block diagonal matrix, $\mathrm{diag}(\partial^2\ell_q/\partial\zeta^2, \partial^2\ell_q/\partial\xi^2)$, with blocks given by equation (14). These approximations yield equations (15) and (16) of Section 3.3, and formally,

$$\begin{aligned}\hat{\dot{T}}_\theta(\hat\psi, \hat\theta) \approx \\ \text{''trace''}\Big\{&[\frac{\partial^2 \ell_q}{\partial\zeta^2}(\hat\psi, \hat\theta)]^{-1}(\frac{\partial^3 \ell_q}{\partial\zeta^2\partial\psi}(\hat\psi, \hat\theta) - \frac{\partial^3 \ell_q}{\partial\zeta^3}(\hat\psi, \hat\theta)[\frac{\partial^2 \ell_q}{\partial\zeta^2}(\hat\psi, \hat\theta)]^{-1}\frac{\partial^2 \ell_q}{\partial\zeta\partial\psi}(\hat\psi, \hat\theta)) \\ &+ [\frac{\partial^2 \ell_q}{\partial\xi^2}(\hat\psi, \hat\theta)]^{-1}(\frac{\partial^3 \ell_q}{\partial\xi^2\partial\psi}(\hat\psi, \hat\theta) - \frac{\partial^3 \ell_q}{\partial\xi^2\partial\zeta}(\hat\psi, \hat\theta)[\frac{\partial^2 \ell_q}{\partial\zeta^2}(\hat\psi, \hat\theta)]^{-1}\frac{\partial^2 \ell_q}{\partial\zeta\partial\psi}(\hat\psi, \hat\theta))\Big\}.\end{aligned}$$

More explicitly, let j_ζ, j_ξ be the submatrices of j_θ corresponding to the second derivatives with respect to ζ and ξ, respectively. Consider the matrices of complete derivatives with respect to ψ evaluated at $(\hat\psi, \hat\theta)$,

$$\nabla_\psi \hat{j}_{\zeta,u} = -\left(\frac{d}{d\psi}\frac{\partial^2 \ell_q}{\partial\zeta_1\partial\zeta_u}\middle|\frac{d}{d\psi}\frac{\partial^2 \ell_q}{\partial\zeta_2\partial\zeta_u}\middle|\cdots\middle|\frac{d}{d\psi}\frac{\partial^2 \ell_q}{\partial\zeta_{d_1}\partial\zeta_u}\right)'(\hat\psi, \hat\theta), \qquad u = 1, \ldots, d_1,$$

$$\nabla_\psi \hat{\jmath}_{\xi,v} = -\left(\frac{d}{d\psi}\frac{\partial^2 \ell_q}{\partial \xi_1 \partial \xi_v} \middle| \frac{d}{d\psi}\frac{\partial^2 \ell_q}{\partial \xi_2 \partial \xi_v} \middle| \cdots \middle| \frac{d}{d\psi}\frac{\partial^2 \ell_q}{\partial \xi_{d_2} \partial \xi_v} \right)'(\hat{\psi}, \hat{\theta}), \qquad v = 1, \ldots, d_2.$$

Then

$$\hat{\hat{T}}_\theta(\hat{\psi}, \hat{\theta}) \approx \sum_{u=1}^{d_1} \hat{\jmath}_\zeta^{-u} \nabla_\psi \hat{\jmath}_{\zeta,u} + \sum_{v=1}^{d_2} \hat{\jmath}_\xi^{-v} \nabla_\psi \hat{\jmath}_{\xi,v},$$

where $\hat{\jmath}_\zeta^{-u}$ and $\hat{\jmath}_\xi^{-v}$ denotes the u-th and v-th rows of the inverses of $\hat{\jmath}_\zeta$ and $\hat{\jmath}_\xi$, respectively. We can also find approximations for the determinants. We have $j_{\psi\times\theta} \approx (j_{\psi\times\zeta} \,|\, 0)$, and therefore $|\hat{\jmath}| \approx |\hat{\jmath}_\zeta| \times |\hat{\jmath}_\xi| \times |\hat{\jmath}_\psi - \hat{\jmath}_{\psi\times\zeta}\hat{\jmath}_\zeta^{-1}\hat{\jmath}_{\zeta\times\psi}|$. This yields $\hat{\jmath}_{p,\psi} \approx \hat{\jmath}_\psi \times (I - \hat{\jmath}_\psi^{-1}\hat{\jmath}_{\psi\times\zeta}\hat{\jmath}_\zeta^{-1}\hat{\jmath}_{\zeta\times\psi})$. Note that if ζ is also orthogonal to ψ, then we simply have $\hat{\jmath}_{p,\psi} \approx \hat{\jmath}_\psi$. This is the case of the parameter $\vec{\tau}$ which is also orthogonal to b.

C. Some useful derivatives

Proposition 1

$$\frac{\partial \log q}{\partial \beta} = \frac{\partial \ell}{\partial \beta} = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(z_{ij} - e_{ij}\mu_{ij})B_{ij}X_{ij}$$

$$\frac{\partial^2 \log q}{\partial \beta^2} = \frac{\partial^2 \ell}{\partial \beta^2} = -\sum_{j=1}^{p} X_j' D_j X_j,$$

where $D_j = \mathrm{diag}(e_{ij}B_{ij}\lambda_{ij}^*)$, *with* $\lambda_{ij}^* = (1 - B_{ij})y_{ij} + B_{ij}\mu_{ij}$. *In particular,*

$$\frac{\partial^2 \log q}{\partial \beta_k \partial \beta_\ell} = -\sum_{j=1}^{p}\sum_{i=1}^{n_j} x_{ij,k}x_{ij,\ell}e_{ij}B_{ij}\lambda_{ij}^*.$$

$$\frac{\partial \log q}{\partial \tau_j} = \frac{\partial \ell}{\partial \tau_j} - \frac{a - \tau_0}{pv_0^2} - \frac{(\nu_0 + p - 1)(p-1)(\tau_j - a)}{p^2 A},$$

$$\frac{\partial^2 \log q}{\partial \tau_j \partial \tau_k} = \delta_{jk}\frac{\partial^2 \ell}{\partial \tau_j \partial \tau_k} - \frac{1}{(pv_0)^2} - \frac{(\nu_0 + p - 1)(p-1)}{p^4 A^2}\left((\delta_{jk}p - 1)pA - 2(p-1)(\tau_j - a)(\tau_k - a)\right),$$

where δ_{jk} denotes the Kronecker's delta, that is, $\delta_{jk} = 1$ if and only if $j = k$, and $\delta_{jk} = 0$, otherwise, and

$$\frac{\partial \ell}{\partial \tau_j} = \sum_{s=1}^{M_j} N_{js} \frac{e^{\tau_j}}{e^{\tau_j} + s - 1} + e^{\tau_j} \sum_{i=1}^{n_j} \log B_{ij} - \sum_{i=1}^{n_j} (z_{ij} - e_{ij}\mu_{ij}) B_{ij},$$

$$\frac{\partial^2 \ell}{\partial \tau_j^2} = \sum_{s=1}^{M_j} N_{js} \frac{e^{\tau_j}(s-1)}{(e^{\tau_j} + s - 1)^2} + \sum_{i=1}^{n_j} \Big\{ e^{\tau_j} [\log(B_{ij}) + (1 - B_{ij})] - (z_{ij} - e_{ij}\mu_{ij}) B_{ij}(1 - B_{ij}) \Big\}.$$

$$\frac{\partial \log q}{\partial b_j} = \sum_{i=1}^{n_j} (z_{ij} - e_{ij}\mu_{ij}) B_{ij} W_{ij} - \frac{(m_0 + p)\sigma_0^{-2}}{1 + \sigma_0^{-2} \sum_{\ell=1}^{p} b'_\ell \Lambda_0^{-1} b_\ell} \Lambda_0^{-1} b_j,$$

$$\frac{\partial^2 \log q}{\partial b_j \partial b_k} = \begin{cases} -W'_j D_j W_j - \sigma_0^{-2}(m_0 + p)\Lambda_0^{-1} \left(\frac{(1+\sigma_0^{-2} \sum_{\ell=1}^{p} b'_\ell \Lambda_0^{-1} b_\ell) I_q - 2\sigma_0^{-2} b_j b'_j \Lambda_0^{-1}}{(1+\sigma_0^{-2} \sum_{\ell=1}^{p} b'_\ell \Lambda_0^{-1} b_\ell)^2} \right), & k = j \\ 2\sigma_0^{-4}(m_0 + p) \frac{\Lambda_0^{-1} b_j b'_k \Lambda_0^{-1}}{(1+\sigma_0^{-2} \sum_{\ell=1}^{p} b'_\ell \Lambda_0^{-1} b_\ell)^2}, & k \neq j \end{cases}$$

In particular,

$$\frac{\partial^2 \log q}{\partial b_{ju} \partial b_{jv}} = -\sum_{i=1}^{n_j} w_{ij,u} w_{ij,v} e_{ij} B_{ij} \lambda^*_{ij} - \sigma_0^{-2}(m_0 + p) \left(\frac{\Lambda_0^{uv} (1 + \sigma_0^{-2} \sum_{\ell=1}^{p} b'_\ell \Lambda_0^{-1} b_\ell) - 2\sigma_0^{-2} \Lambda_0^u b_j \, \Lambda_0^v b_j}{(1 + \sigma_0^{-2} \sum_{\ell=1}^{p} b'_\ell \Lambda_0^{-1} b_\ell)^2} \right),$$

$$\frac{\partial^2 \log q}{\partial b_{ju} \partial b_{kv}} = 2\sigma_0^{-4}(m_0 + p) \frac{\Lambda_0^u b_j \Lambda_0^v b_k}{(1 + \sigma_0^{-2} \sum_{\ell=1}^{p} b'_\ell \Lambda_0^{-1} b_\ell)^2}, \quad j \neq k.$$

where Λ_0^u denotes the u-th row of Λ_0^{-1}, and Λ_0^{uv}, the v-th component of the u-th row of Λ_0^{-1} (recall that Λ_0^{-1} is symmetric). Also, for every $j \in \{1, \ldots, p\}$

$$\frac{\partial^2 \log q}{\partial \beta \partial \tau_j} = \sum_{i=1}^{n_j} (z_{ij} - e_{ij}\mu_{ij}) X_{ij} B_{ij}(1 - B_{ij})$$

$$\frac{\partial^2 \log q}{\partial \beta \partial b_j} = -X'_j D_j W_j = (-\sum_{i=1}^{n_j} \lambda^*_{ij} e_{ij} B_{ij} x_{ij,u} w_{ij,v})$$

$$\frac{\partial^2 \log q}{\partial b_j \partial \tau_k} = \left(\sum_{i=1}^{n_j} (z_{ij} - e_{ij}\mu_{ij}) B_{ij}(1 - B_{ij}) W_{ij} \right) \delta_{kj}.$$

Moreover,

$$\frac{\partial^3 \log q}{\partial \beta_u \partial \beta_v \partial \beta_t} = \sum_{j=1}^{p} \sum_{i=1}^{n_j} e_{ij} B_{ij}(1 - 2B_{ij}) \lambda^*_{ij} x_{ij,u} x_{ij,v} x_{ij,t}.$$

$$\frac{\partial^3 \log q}{\partial b_{ju}\partial b_{jv}\partial b_{jt}} = \sum_{i=1}^{n_j} e_{ij}B_{ij}(1-2B_{ij})\lambda^*_{ij}w_{ij,u}w_{ij,v}w_{ij,t} - \frac{8\sigma^{-6}(m_0+p)}{(1+\sigma_0^{-2}\sum_{\ell=1}^p b'_\ell\Lambda_0^{-1}b_\ell)^2}\times$$
$$\left(\frac{\Lambda_0^u b_j\Lambda_0^v b_j\Lambda_0^t b_j}{1+\sigma_0^{-2}\sum_{\ell=1}^p b'_\ell\Lambda_0^{-1}b_\ell} - \frac{1}{4\sigma_0^{-2}}(\Lambda_0^{uv}\Lambda_0^t b_j + \Lambda_0^{ut}\Lambda_0^v b_j + \Lambda_0^{vt}\Lambda_0^u b_j)\right),$$

$$\frac{\partial^3 \log q}{\partial b_{ju}\partial b_{ks}\partial b_{ht}} = -8\sigma^{-6}(m_0+p)\frac{\Lambda_0^u b_j\Lambda_0^v b_k\Lambda_0^t b_h}{(1+\sigma_0^{-2}\sum_{\ell=1}^p b'_\ell\Lambda_0^{-1}b_\ell)^3},\ \textit{for}\ \ \delta_{jk}\delta_{jh} = 0.$$

$$\frac{\partial^3 \log q}{\partial\tau_j\partial\tau_k\partial\tau_\ell} = \frac{\partial^3\ell}{\partial\tau_j^3}\delta_{jk}\delta_{j\ell} - \frac{2(\nu_0+p-1)(p-1)^2}{p^6A^3}\times$$
$$\Big(4(p-1)(\tau_\ell-a)(\tau_k-a)(\tau_j-a) - pA[(\tau_\ell-a)(\delta_{jk}p-1) + (\tau_k-a)(\delta_{j\ell}p-1) + (\tau_j-a)(\delta_{k\ell}p-1)]\Big),$$

where

$$\frac{\partial^3\ell}{\partial\tau_j^3} = \sum_{s=1}^{M_j} N_{js}(s-1)\frac{(s-1-e^{\tau_j})e^{\tau_j}}{(s-1+e^{\tau_j})^3} +$$
$$\sum_{i=1}^{n_j}\Big\{e^{\tau_j}[B_{ij}(1-B_{ij}) - \log B_{ij}] - (z_{ij}-e_{ij}\mu_{ij})B_{ij}(1-B_{ij})(1-2B_{ij})\Big\}.$$

$$\frac{\partial^3 \log q}{\partial\beta_u\partial\beta_v\partial\tau_j} = -\sum_{i=1}^{n_j} B_{ij}(1-B_{ij})(2\lambda^*_{ij}-y_{ij})e_{ij}x_{ij,u}x_{ij,v},$$
$$\frac{\partial^3 \log q}{\partial b_{ju}\partial b_{kv}\partial\tau_\ell} = -\left(\sum_{i=1}^{n_j} B_{ij}(1-B_{ij})(2\lambda^*_{ij}-y_{ij})e_{ij}w_{ij,u}w_{ij,v}\right)\delta_{jk}\delta_{j\ell},$$
$$\frac{\partial^3 \log q}{\partial b_{ju}\partial b_{kv}\partial\beta_h} = \left(\sum_{i=1}^{n_j} B_{ij}(1-2B_{ij})\lambda^*_{ij}e_{ij}w_{ij,u}w_{ij,v}x_{ij,h}\right)\delta_{kj}$$
$$\frac{\partial^3 \log q}{\partial b_{ju}\partial\beta_v\partial\beta_t} = \sum_{i}^{n_j} B_{ij}(1-2B_{ij})\lambda^*_{ij}e_{ij}w_{ij,u}x_{ij,v}x_{ij,t},$$
$$\frac{\partial^3 \log q}{\partial b_{ju}\partial\beta_v\partial\tau_\ell} = -\left(\sum_{i=1}^{n_j} B_{ij}(1-B_{ij})(2\lambda^*_{ij}-y_{ij})e_{ij}w_{ij,u}x_{ij,v}\right)\delta_{j\ell},$$
$$\frac{\partial^3 \log q}{\partial\tau_j\partial\tau_k\partial\beta_h} = -\left(\sum_{i=1}^{n_j}(z_{ij}-e_{ij}\mu_{ij})B_{ij}(1-B_{ij})(1-2B_{ij})x_{ij,h}\right)\delta_{kj},$$
$$\frac{\partial^3 \log q}{\partial\tau_j\partial\tau_k\partial b_{\ell,u}} = -\left(\sum_{i=1}^{n_j}(z_{ij}-e_{ij}\mu_{ij})B_{ij}(1-B_{ij})(1-2B_{ij})w_{ij,u}\right)\delta_{jk}\delta_{j\ell}.$$

The proof of the proposition is straightforward. One only needs to verify that $(\partial B_{ij}/\partial\beta_u) = -B_{ij}(1-B_{ij})x_{ij,u}$, $(\partial B_{ij}/\partial\tau_k) = B_{ij}(1-B_{ij})\delta_{jk}$, $(\partial B_{ij}/\partial b_{ku}) = -B_{ij}(1-B_{ij})w_{ij,u}\delta_{jk}$,

$(\partial\mu_{ij}/\partial\beta_u) = \mu_{ij}x_{ij,u}$, and $(\partial\mu_{ij}/\partial b_{ku}) = \mu_{ij}w_{ij,u}\delta_{jk}$.

Proposition 2 *Let* $\alpha_{ij} = z_{ij} + e^{\tau_j}$, *and* $\eta_{ij} = e_{ij} + e^{\tau_j}/\mu_{ij}$.

$$
\begin{aligned}
\frac{\partial \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \beta_u} &= \{\frac{\lambda}{\mu_{ij}}e^{\tau_j} - B_{ij}(z_{ij}+e^{\tau_j})\}X_{ij,u},\\
\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \beta_u \partial \beta_v} &= \{-\frac{\lambda}{\mu_{ij}}e^{\tau_j} + B_{ij}(1-B_{ij})(z_{ij}+e^{\tau_j})\}X_{ij,u}X_{ij,v},\\
\frac{\partial \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial b_{k,u}} &= \{\frac{\lambda}{\mu_{ij}}e^{\tau_j} - B_{ij}(z_{ij}+e^{\tau_j})\}W_{ij,u}\,\delta_{kj},\\
\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial b_{j,u} \partial b_{j,v}} &= \{-\frac{\lambda}{\mu_{ij}}e^{\tau_j} + B_{ij}(1-B_{ij})(z_{ij}+e^{\tau_j})\}W_{ij,u}W_{ij,v},\\
\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \beta_u \partial b_{k,v}} &= \{-\frac{\lambda}{\mu_{ij}}e^{\tau_j} + B_{ij}(1-B_{ij})(z_{ij}+e^{\tau_j})\}X_{ij,u}W_{ij,v}\,\delta_{kj},\\
\frac{\partial \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \tau_k} &= \Big(-e^{\tau_j}\{\log(1-B_{ij}) - \log(e_{ij}\lambda)\\
&\qquad + \Psi(z_{ij}+e^{\tau_j})\} + B_{ij}(z_{ij}+e^{\tau_j}) - \frac{\lambda}{\mu_{ij}}e^{\tau_j}\Big)\delta_{kj},\\
\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \tau_j^2} &= \frac{\partial \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \tau_j} + B_{ij}^2(e_{ij}\mu_{ij} - z_{ij}) - e^{\tau_j}\Psi'(z_{ij}+e^{\tau_j}),\\
\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \tau_j \partial \beta_u} &= (1-B_{ij})B_{ij}(e_{ij}\mu_{ij} - z_{ij})X_{ij,u} + \frac{\lambda-\mu_{ij}}{\mu_{ij}}e^{\tau_j}X_{ij,u},\\
\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \tau_j \partial b_{k,u}} &= \Big((1-B_{ij})B_{ij}(e_{ij}\mu_{ij} - z_{ij})W_{ij,u} + \frac{\lambda-\mu_{ij}}{\mu_{ij}}e^{\tau_j}W_{ij,u}\Big)\delta_{kj}.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \beta_u \partial \beta_v \partial \beta_t} &= \{\frac{\lambda}{\mu_{ij}}e^{\tau_j} - B_{ij}(1-B_{ij})(1-2B_{ij})(z_{ij}+e^{\tau_j})\}X_{ij,u}X_{ij,v}X_{ij,t},\\
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial b_{j,u} \partial b_{j,v} \partial b_{j,t}} &= \{\frac{\lambda}{\mu_{ij}}e^{\tau_j} - B_{ij}(1-B_{ij})(1-2B_{ij})(z_{ij}+e^{\tau_j})\}W_{ij,u}W_{ij,v}W_{ij,t},\\
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \beta_u \partial \beta_v \partial b_{j,t}} &= \{\frac{\lambda}{\mu_{ij}}e^{\tau_j} - B_{ij}(1-B_{ij})(1-2B_{ij})(z_{ij}+e^{\tau_j})\}X_{ij,u}X_{ij,v}W_{ij,t},\\
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial b_{j,u} \partial b_{j,v} \partial \beta_t} &= \{\frac{\lambda}{\mu_{ij}}e^{\tau_j} - B_{ij}(1-B_{ij})(1-2B_{ij})(z_{ij}+e^{\tau_j})\}W_{ij,u}W_{ij,v}X_{ij,t},
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial \beta_u \partial \beta_v \partial \tau_j} =\\
&\qquad \{-\frac{\lambda}{\mu_{ij}}e^{\tau_j} + B_{ij}(1-B_{ij})(1-2B_{ij})(z_{ij}+e^{\tau_j}) + B_{ij}(1-B_{ij})e^{\tau_j}\}X_{ij,u}X_{ij,v},
\end{aligned}
$$

$$\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\beta_u \partial b_{j,v} \partial\tau_j} = \{-\frac{\lambda}{\mu_{ij}} e^{\tau_j} + B_{ij}(1-B_{ij})(1-2B_{ij})(z_{ij}+e^{\tau_j}) + B_{ij}(1-B_{ij})e^{\tau_j}\} X_{ij,u} W_{ij,v},$$

$$\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial b_{j,u} \partial b_{j,v} \partial\tau_j} = \{-\frac{\lambda}{\mu_{ij}} e^{\tau_j} + B_{ij}(1-B_{ij})(1-2B_{ij})(z_{ij}+e^{\tau_j}) + B_{ij}(1-B_{ij})e^{\tau_j}\} W_{ij,u} W_{ij,v},$$

$$\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\beta_u \partial\tau_j^2} = \frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\beta_u \partial\tau_j} + B_{ij}^2[e_{ij}\mu_{ij} - 2(1-B_{ij})(e_{ij}\mu_{ij} - z_{ij})] X_{ij,u},$$

$$\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial b_{j,u} \partial\tau_j^2} = \frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial b_{j,u} \partial\tau_j} + B_{ij}^2[e_{ij}\mu_{ij} - 2(1-B_{ij})(e_{ij}\mu_{ij} - z_{ij})] W_{ij,u},$$

$$\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\tau_j^3} = \frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\tau_j^2} + 2B_{ij}^2(1-B_{ij})(e_{ij}\mu_{ij} - z_{ij}) - e^{\tau_j}\Psi'(z_{ij}+e^{\tau_j}) - e^{2\tau_j}\Psi''(z_{ij}+e^{\tau_j}).$$

We also have

$$\frac{\partial \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda} = \frac{\alpha_{ij}-1}{\lambda} - \eta_{ij},$$

$$\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda^2} = -\frac{\alpha_{ij}-1}{\lambda^2},$$

$$\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial\beta} = -\frac{e^{\tau_{ij}}}{\mu_{ij}} X_{ij},$$

$$\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial b_k} = -\frac{e^{\tau_{ij}}}{\mu_{ij}} W_{ij}\,\delta_{kj},$$

$$\frac{\partial^2 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial\tau_k} = \left(\frac{e^{\tau_{ij}}}{\lambda} - \frac{e^{\tau_{ij}}}{\mu_{ij}}\right)\delta_{kj}.$$

$$\begin{aligned}
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial\beta_u\partial\beta_v} &= -\frac{e^{\tau_j}}{\mu_{ij}} X_{ij,u}X_{ij,v},\\
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial\beta_u\partial b_{k,v}} &= -\frac{e^{\tau_j}}{\mu_{ij}} X_{ij,u}W_{ij,v}\delta_{kj},\\
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial b_{k,u}\partial b_{\ell,v}} &= -\frac{e^{\tau_j}}{\mu_{ij}} W_{ij,u}W_{ij,v}\delta_{kj}\delta_{\ell j},\\
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial\tau_k\partial\tau_\ell} &= e^{\tau_j}\left(\frac{1}{\lambda}-\frac{1}{\mu_{ij}}\right)\delta_{kj}\delta_{\ell j},\\
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial\tau_k\partial\beta_u} &= \frac{e^{\tau_j}}{\mu_{ij}} X_{ij,u}\delta_{kj},\\
\frac{\partial^3 \log g(\lambda|\alpha_{ij},\eta_{ij})}{\partial\lambda\partial\tau_k\partial b_{\ell,u}} &= \frac{e^{\tau_j}}{\mu_{ij}} W_{ij,u}\delta_{kj}\delta_{\ell j}.
\end{aligned}$$

Also note that since $\partial \log q/\partial\lambda = 0$*, we have*

$$\begin{aligned}
\frac{d}{d\,\lambda}\frac{\partial^2 \log h}{\partial\zeta_1\partial\zeta_2} &= \frac{d}{d\,\lambda}\frac{\partial^2(\log q+\log g)}{\partial\zeta_1\partial\zeta_2}\\
&= -\left(\frac{\partial^3 \log q}{\partial\zeta_1\partial\zeta_2\partial\zeta}+\frac{\partial^3 \log g}{\partial\zeta_1\partial\zeta_2\partial\zeta}\right)\left(\frac{\partial^2 \log q}{\partial\zeta^2}+\frac{\partial^2 \log g}{\partial\zeta^2}\right)^{-1}\frac{\partial^2 \log g}{\partial\zeta\partial\lambda}+\frac{\partial^3 \log g}{\partial\lambda\partial\zeta_1\partial\zeta_2}
\end{aligned}$$

where $\zeta = (\beta, \vec{\tau}, b)$*, and* ζ_1, ζ_2 *denote any pair of single components of* ζ*.*

References

Albert, J. (1992). A Bayesian analysis of a Poisson random effects model for home run hitters. *The American Statistician*, 46:246–253.

Allan, J., Carbonell, J., Doddington, G., Yamron, J., and Yang, Y. (1998). Topic detection and tracking pilot study: Final report. *In Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, pages 194–218.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman & Hall.

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, S. M. and Berry, D. A. (2004). Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics*, 60:418–426.

Brazzale, A. R., Davison, A. C., and Reid, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge University Press.

Breslow, N. E. and Lin, X. (1995). Bias-correction in generalized linear mixed models with a single component of dispersion. *Biometrika*, 82:81–91.

Carlin, B. and Gelfand, A. (1991). A sample reuse method for accurate parametric

empirical Bayes confidence intervals. *Journal of the Royal Statistical Society, Series B*, 53:189–200.

Christiansen, C. L. and Morris, C. (1996). *Fitting and checking a two-level Poisson model: modeling patient mortality rates in heart transplant patients*, pages 467–501. Marcel Dekker, New York. Berry, D. and Stangl, D., editors.

Christiansen, C. L. and Morris, C. (1997). Hierarchical Poisson regression modeling. *Journal of the American Statistical Association*, 92:618–632.

Davison, A. C. (1986). Approximate predictive likelihood. *Biometrika*, 73(2):323–332.

Frome, E. L. (1983). The analysis of rates using Poisson regression models. *Biometrics*, 39:665–674.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analysis of event rates. *Technometrics*, 29:1–15.

Guihenneuc-Jouyaux, C. and Rousseau, J. (2007). Laplace expansions in Markov chain Monte Carlo algorithms. *Journal of Computational and Graphical Statistics*, 14(1):75–94.

Holford, T. R. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics*, 36:299–305.

Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*, 52:5066–5074.

Johnson, V. E. (1992). A technique for estimating marginal posterior densities in hierarchical models using mixtures of conditional densities. *Journal of the American Statistical Association*, 87:852–860.

Kuk, A. Y. C. (1999). Laplace Important Sampling for generalized linear mixed models. *Journal of Statistical Computation and Simulation*, 63:143–158.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The bugs project: evolution, critique and future directions. *Statistics in Medicine*, 28:3049–67.

Mehrotra, D. V. and Heyse, J. F. (2004). Use of the false discovery rate for evaluating clinical safety data. *Statistical Methods in Medical Research*, 13:227–238.

Morokoff, W. J. and Caflisch, R. E. (1995). Quasi-Monte Carlo integration. *Journal of Computational Physics*, pages 218–230.

Morris, C. N. (1983). *Parametric empirical Bayes confidence intervals*, pages 25–50. Academic Press, New York.

Nott, J. D., Fielding, M., and Leonte, D. (2009). On a generalization of the Laplace approximation. *Statistics and Probability Letters*, 79:1397–1403.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the non-linear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4:12–35.

Plummer, M. (2013). *rjags: Bayesian graphical models using MCMC*. R package version 3-10.

Raudenbush, S. W., Yang, M. L., and Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace

approximation. *Journal of Computational and Graphical Statistics*, 9:141–157.

Santamaria, R., Quintales, L., and Theron, R. (2007). Methods to bicluster validation and comparison in microarray data. *IDEAL 07 Proceedings of the 8th international conference on Intelligent data engineering and automated learning*, pages 780–789.

Severini, T. (2000). *Likelihood Methods in Statistics*. Oxford University Press.

Skaug, H. J. and Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics and Data Analysis*, 51:699–709.

Sutradhar, B. C. and Zhende, Q. (1998). On approximate likelihood inference in a Poisson mixed model. *The Canadian Journal of Statistics*, 26:169–186.

Tempelman, R. J. and Gianola, D. (1996). A mixed effects model for overdispersed count data in animal breeding. *Biometrics*, 52:265–279.

The BUGS Project (2012). Hosted by the MRC Biostatistics Unit, Cambridge, UK.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81:82–86.

Turner, H., Bailey, T., and Krzanowski, W. (2005). Improved biclustering of microarray data demonstrated through systematic performance tests. *Computational Statistics & Data Analysis*, 48:235–254.

Venables, W. N. and Ripley, B. D. (1994). *Nodern Applied Statisitcs with S-Plus*. Springer-Verlag, New York.

Vonesh, E. F. (1990). Modelling peritonitis rates and associated risk factors for individuals on continuous ambulatory peritoneal dialysis. *Statistics in Medicine*, 9:263–271.

Vonesh, E. F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika*, 83:447–452.

Whittemore, A. S. (2007). A Bayesian false discovery rate for multiple testing. *Journal of Applied Statistics*, 34(1):1–9.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80:791–795.

Body System	Adverse Effect	$\boldsymbol{p}_\delta$	$\boldsymbol{p}_{\delta,o}$	$\boldsymbol{b}$-values **0**.10	**0**.25	**0**.50	**0**.75	**0**.90
1	Fatigue	0.10	0.12	0.915	0.783	0.546	0.286	0.118
1	Fever	0.25	0.32	0.920	0.792	0.560	0.298	0.124
1	Fungal infection	0.36	0.36	0.901	0.751	0.502	0.251	0.101
1	Viral infection	0.41	0.41	0.901	0.751	0.502	0.251	0.101
1	Malaise	0.25	0.26	0.905	0.761	0.514	0.261	0.105
3	Anorexia	0.23	0.24	0.905	0.760	0.514	0.261	0.105
3	Oral candidiasis	0.26	0.27	0.902	0.754	0.506	0.254	0.102
3	Constipation	0.26	0.27	0.902	0.754	0.506	0.254	0.102
3	Diarrhea	**0.001**	0.01	0.984	0.955	0.876	0.701	0.439
3	Gastroenteritis	0.50	0.50	0.900	0.750	0.500	0.250	0.100
3	Nausea	0.79	0.80	0.901	0.753	0.503	0.253	0.101
3	Vomiting	0.65	0.71	0.908	0.766	0.522	0.267	0.108
5	Lymphadenopathy	0.11	0.33	0.965	0.902	0.754	0.505	0.254
6	Dehydration	0.53	0.54	0.901	0.751	0.502	0.251	0.101
8	Crying	0.46	0.45	0.900	0.750	0.500	0.250	0.100
8	Insomnia	0.50	0.51	0.900	0.750	0.500	0.250	0.100
8	Irritability	**0.04**	**0.0002**	**0.039**	**0.013**	**0.004**	**0.001**	**0.0005**
9	Bronchitis	0.35	0.35	0.900	0.750	0.501	0.250	0.100
9	Nasal congestion	0.40	0.40	0.900	0.751	0.501	0.251	0.100
9	Respiratory congestion	0.57	0.57	0.901	0.752	0.502	0.252	0.101
9	Cough	0.17	0.22	0.920	0.793	0.560	0.298	0.124
9	Respiratory infection	0.15	0.19	0.918	0.788	0.554	0.293	0.121
9	Laryngotracheobronchitis	0.44	0.44	0.900	0.750	0.500	0.250	0.100
9	Pharyngitis	0.17	0.22	0.920	0.793	0.560	0.298	0.124
9	Rhinorrhea	0.48	0.56	0.913	0.777	0.537	0.279	0.114
9	Sinusitis	0.37	0.37	0.900	0.751	0.501	0.251	0.100
9	Tonsillitis	0.44	0.44	0.900	0.750	0.500	0.250	0.100
9	Wheezing	0.37	0.37	0.900	0.751	0.501	0.251	0.100
10	Bite/sting	0.13	0.15	0.912	0.776	0.536	0.278	0.114
10	Eczema	0.49	0.49	0.900	0.750	0.500	0.250	0.100
10	Pruritus	0.42	0.43	0.903	0.755	0.507	0.255	0.103
10	Rash	**0.07**	0.10	0.930	0.816	0.597	0.331	0.141
10	Diaper rash	0.32	0.35	0.908	0.766	0.522	0.267	0.108
10	Measles/rubella-like rash	0.10	0.13	0.919	0.792	0.559	0.297	0.123
10	Varicella-like rush	0.41	0.43	0.903	0.757	0.510	0.257	0.104
10	Urticaria	0.51	0.51	0.900	0.750	0.500	0.250	0.100
10	Viral exanthema	0.58	0.58	0.901	0.752	0.503	0.252	0.101
11	Conjunctivitis	0.52	0.52	0.900	0.750	0.500	0.250	0.100
11	Otitis media	0.43	0.16	0.767	0.523	0.267	0.108	0.039
11	Otorrhea	0.42	0.42	0.901	0.752	0.502	0.252	0.101

Table 5: The adverse effects associated with eight body systems. The columns $\boldsymbol{p}_\delta$ and $\boldsymbol{p}_{\delta,0}$ respectively show the estimate of the probability that the difference between the treatment and control intensities is negative under the model, and under the null hypothesis of no treatment-effects. The last five columns show the Bayesian false discovery rate b-values for different values of the prior probabilities π_i against the null hypothesis of no treatment effect.