

A Bayesian Approach for Assessing Heterogeneity in Generalized Linear Models

Zhen Chen and David B. Dunson*

Biostatistics Branch, MD A3-03,
National Institute of Environmental Health Sciences,
P.O. Box 12233, Research Triangle Park, NC 27709

**email*: dunson1@niehs.nih.gov

SUMMARY. Generalized linear mixed models (GLMMs) are used routinely for analyzing clustered data arising in a broad variety of applications. In Bayesian analyses, inverse Wishart or inverse gamma priors are almost always used for the covariance of the random effects, for computational convenience and to enforce the positive definite constraint on the covariance matrix. In this article, we propose a new class of prior distributions based on a Gaussian structure for variance component parameters underlying the random effects covariance. The proposed prior assigns positive probability not only to the full model but also to reduced models that exclude one or more of the random effects. This structure facilitates Bayesian inferences about the covariance structure, while also accounting for uncertainty in the random effects model in estimating the population parameters. A Markov chain Monte Carlo algorithm is proposed for posterior computation, and the approach is illustrated using data on prenatal exposure to PCBs and psychomotor development.

KEY WORDS: Latent variables; Longitudinal data; Mixed model; Markov chain Monte Carlo; Model averaging; Prior elicitation; Random effects; Bayes factor.

1. Introduction

It is common practice in many applications to collect multiple measurements on a subject. Mixed effects models attempt to account for within-subject dependency in the multiple measurements by including one or more subject-specific latent variables (i.e., random effects) in the regression model (c.f., Laird and Ware, 1982; Gilks et al., 1993). Typically, an additional random effect is included for each regression coefficient that is expected to vary among subjects. For example, in dose response settings, one may account for baseline heterogeneity through a random intercept and for heterogeneity in susceptibility through a random slope, with these two factors potentially correlated. Often, it is not known a priori which coefficients vary among subjects, and there may be interest in investigating the covariance structure. In addition, when interest focuses on the population parameters, one may wish to account for uncertainty in the random effects model to avoid biased inferences.

To motivate the problem, we focus on data from an epidemiologic study relating prenatal exposure to Polychlorinated biphenyls (PCBs) and children’s psychomotor development (Daniels et al., 2002). Data were drawn from 12 prenatal care centers participating in the Collaborative Perinatal Project. Demographic data were obtained from women participating in the study and blood samples were collected, which were later analyzed to measure prenatal PCB exposure. Psychomotor development was measured for a sample of children at 8 months of age using the Bayley Psychomotor Index. Based on a classical linear regression analysis of the data from 1124 children, with eight possible confounders included and with center-specific intercepts and PCB slopes treated as fixed effects, Daniels et al. (2002) concluded that there was significant heterogeneity among centers in the PCB effect but no overall effect across the centers. An association between prenatal PCB exposure and impaired psychomotor development would be an important finding, even if it occurred only in certain sub-populations or study centers. However, by assuming the confounding coefficients are fixed for all study centers, the estimates for the center-specific PCB slopes and the mag-

nitude of heterogeneity in the slopes may be biased if the confounding coefficients actually vary. To properly adjust for confounding in assessing the PCB effect and to additionally assess heterogeneity in the confounding coefficients, it is of interest to consider models that allow each of the regression coefficients to vary across centers.

Potentially, we could just fit a mixed effects model that has random coefficients for the intercept, the PCB slope, and each of the confounders. A frequentist analysis could be implemented in SAS PROC MIXED (Littel et al. 1996), while a Bayesian analysis could be implemented in WinBUGS with an inverse Wishart prior used for the random-effects covariance (Spiegelhalter, Thomas, and Best, 1999). However, such analyses implicitly assume that there is some level of heterogeneity among centers in each of the regression coefficients. In addition, it can be difficult to fit random effects models that have more than a few random coefficients, due to computational problems. For the PCB data, a mixed effects model with random coefficients for each of the covariates failed to converge when SAS PROC MIXED was applied. When using WinBUGS with an inverse Wishart prior, extremely high autocorrelation was present for most of the population-level parameters, indicating problems with slow mixing and hence inefficiency in posterior computation.

From a Bayesian perspective, a priori uncertainty in the covariance structure should be accommodated through a prior that allocates probability not only to the full model, which includes a random effect for each coefficient that possibly varies among subjects, but also to reduced models that exclude one or more of these random effects. In the Bayesian literature on generalized linear mixed models (GLMMs), inverse Wishart or inverse gamma priors are routinely used for the random effects covariance, with independent inverse gamma priors typically used when the response is non-Gaussian (c.f., Dunson, 2000; Dellaportas and Smith, 1993; Gilks and Wild, 1992; Zeger and Karim, 1991, among others). Such priors do not accommodate uncertainty in the random effects structure.

There has been some research on accommodating uncertainty in other aspects of a

GLMM, including the shape of the density function (Mukhopadhyay and Gelfand, 1997) and the assumption of normality of the random effects (Kleinman and Ibrahim, 1998). However, these approaches focus on the case where the basic random effects structure (i.e., the specific covariates that have random coefficients) is known. In addition, some methods have been described for model selection in mixed effects models based on Bayes factors (c.f., Albert and Chib, 1997; Weiss, Wang, and Ibrahim, 1997). The approach of Weiss, Wang and Ibrahim (1997) focuses on selection of the fixed effects part of a normal linear mixed model (Laird and Ware, 1982), and requires separate fitting of each model under consideration. In contrast, the approach of Albert and Chib (1997) allows comparison of a random intercept GLMM with the corresponding fixed effects generalized linear model by using a prior on the random effect variance with a point mass at 0. The extension of the Albert and Chib (1997) approach to GLMMs with multiple random effects is not straightforward, particularly if the random effects are not assumed to be a priori independent. Frequentist approach for testing random effects structure usually involves the use of likelihood ratio methods (c.f., Stram and Lee, 1994; Morrell, 1998).

This article proposes an alternative approach for accommodating a priori uncertainty in the random effects component of a GLMM. We first reparameterize the random effects model in terms of independent standard normal latent variables and variance component parameters measuring heterogeneity among subjects in the different regression parameters. The reexpressed model contains parameter vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$, which measure overall heterogeneity in the different regression coefficients and within-subject dependency in the coefficients, respectively. A value of $\lambda_l = 0$ implies that the l th regression coefficient is constant for all subjects, and hence that the l th random effect can be excluded from the model. Therefore, by choosing prior densities for the elements of $\boldsymbol{\lambda}$ that allocate non-zero probability to $\lambda_l = 0$, a priori uncertainty in the random effects structure can be accommodated. We use a simple and efficient Markov chain Monte Carlo (MCMC) algorithm (c.f., Chen, Shao and Ibrahim,

2000 for a recent book on MCMC) to conduct the posterior computation. The proposed approach can be used for inferences on the random effects structure, and to account for uncertainty in the random effects model in estimating population parameters.

Section 2 proposes the model. Section 3 consider the prior structure and inference. Section 4 outlines the MCMC approach. Section 5 illustrates the methodology through application to an epidemiology study of prenatal exposure to PCBs and children’s psychomotor development, and Section 6 discusses the results.

2. The Method

2.1 Generalized Linear Mixed Model

Suppose there are n subjects under study, with subject i contributing n_i observations, for $i = 1, \dots, n$. For subject i at observation j , let y_{ij} denote a response variable, let \mathbf{x}_{ij} denote a $p \times 1$ vector of covariates, and let \mathbf{z}_{ij} denote a $q \times 1$ vector of covariates (typically chosen to be a subvector of \mathbf{x}_{ij} with $q \leq p$). Following McCullagh and Nelder (1989), we assume that y_{ij} is a random variable in the exponential family,

$$f(y_{ij} | \theta_{ij}, \tau) = \exp\{\tau[y_{ij}\theta_{ij} - a(\theta_{ij})] + c(y_{ij}, \tau)\},$$

where θ_{ij} is the canonical parameter, $E(y_{ij} | \theta_{ij}, \tau) = a'(\theta_{ij})$, and $\text{var}(y_{ij} | \theta_{ij}, \tau) = a''(\theta_{ij})/\tau$. Letting $\eta_{ij} = h(\theta_{ij})$ and $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{in_i})^T$, with $h(\cdot)$ a known monotone link function, we assume that

$$\boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\boldsymbol{\beta}_i, \tag{1}$$

where $\mathbf{X}_i = (\mathbf{x}_{i1}^T, \dots, \mathbf{x}_{in_i}^T)^T$, $\mathbf{Z}_i = (\mathbf{z}_{i1}^T, \dots, \mathbf{z}_{in_i}^T)^T$, $\boldsymbol{\alpha}$ is a $p \times 1$ vector of unknown population parameters, and $\boldsymbol{\beta}_i$ is a $q \times 1$ vector of unknown subject-specific deviations (random effects) with $\boldsymbol{\beta}_i \sim N_q(\mathbf{0}, \boldsymbol{\Omega})$.

Expression (1) is in the form of a generalized linear mixed model (c.f., Zeger and Karim, 1991). Integrating out the random-effects $\boldsymbol{\beta}_i$, the model implies that $\boldsymbol{\eta}_i \sim N_{n_i}(\mathbf{X}_i\boldsymbol{\alpha}, \mathbf{Z}_i\boldsymbol{\Omega}\mathbf{Z}_i^T)$.

Heterogeneity among subjects is accommodated by allowing the linear predictor conditional on the covariates to vary. When \mathbf{z}_{ij} is a subvector of \mathbf{x}_{ij} , the model allows the regression coefficients for the covariates included in \mathbf{z}_{ij} to vary among subjects, while assuming that the remaining coefficients are fixed for all subjects. The most commonly used model in practice is the random intercept model, which takes $\mathbf{Z}_i = \mathbf{1}_{n_i}$. Typically, in Bayesian analyses, one either accommodates dependency in the different random effects through an inverse Wishart prior for $\mathbf{\Omega}$ or assumes a priori independence by choosing $\mathbf{\Omega}$ to be diagonal, with independent inverse gamma priors specified for the diagonal elements. These priors are computationally convenient, while also constraining $\mathbf{\Omega}$ to be positive definite. Unfortunately, inverse Wishart and inverse gamma priors implicitly assume that there is at least a minimal degree of heterogeneity among subjects in each of the regression coefficients that have random effects (i.e., all the coefficients if $\mathbf{z}_{ij} = \mathbf{x}_{ij}$). In order to perform inferences on the random effects structure and to accommodate uncertainty in the random effects model in estimating population parameters, it is necessary to consider alternative priors.

2.2 Reparameterization

Note that β_i in (1) can be reexpressed as $\beta_i = \mathbf{\Psi}\mathbf{b}_i$, where $\mathbf{\Psi}$ is the Cholesky decomposition of $\mathbf{\Omega}$ and $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^T$ is a vector of independent standard normal latent variables, $b_{il} \stackrel{iid}{\sim} N(0, 1)$. Let $\mathbf{\Psi} = \mathbf{\Lambda}\mathbf{\Gamma}$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_q)$ and $\mathbf{\Gamma}$ is a $q \times q$ matrix with the (l, m) th element denoted by γ_{lm} . In order for $\mathbf{\Omega}$ to be uniquely defined from $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$, we assume that

$$\lambda_l \geq 0, \quad \gamma_{ll} = 1, \quad \text{and} \quad \gamma_{lm} = 0, \quad \text{for } l = 1, \dots, q; \quad m = l + 1, \dots, q. \quad (2)$$

We then focus on the reparameterized GLMM,

$$\boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{\Lambda}\mathbf{\Gamma}\mathbf{b}_i, \quad (3)$$

where $\mathbf{\Omega} = \mathbf{\Lambda}\mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{\Lambda}$ and $\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{I}_{q \times q})$.

The covariance matrix of β_i can be expressed as a function of $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_q)^T$ and the free elements of $\boldsymbol{\Gamma}$, $\boldsymbol{\gamma}$. Letting ω_{lm} denote the element in the l th row and m th column of $\boldsymbol{\Omega}$, the diagonal elements of $\boldsymbol{\Omega}$ are

$$\omega_{ll} = \lambda_l^2 \left(1 + \sum_{m=1}^{l-1} \gamma_{lm}^2 \right) \quad \text{for } l = 1, \dots, q \quad (4)$$

and the off-diagonal elements are

$$\omega_{lm} = \omega_{ml} = \lambda_l \lambda_m \left(\gamma_{ml} + \sum_{r=1}^{l-1} \gamma_{lr} \gamma_{mr} \right) \quad \text{for } l = 1, \dots, q; m = l+1, \dots, q, \quad (5)$$

which follows from straightforward matrix algebra. Note that in the case where $\lambda_l = 0$, which is admitted by constraint (2), the l th row and l th column of $\boldsymbol{\Omega}$ consist of all 0s. In this case, $\text{var}(\beta_l) = 0$ and the l th random effect is effectively dropped from the model. By allowing λ_l to equal 0, we expand the support of $\boldsymbol{\Omega}$ to include not just positive definite but also non-negative definite matrices.

The parameters $\boldsymbol{\gamma} = (\gamma_{ml} : m = 2, \dots, q; l = 1, \dots, m-1)^T \in \mathfrak{R}^{q(q-1)/2}$ measure the degree of within-subject dependency in the random-effects β_i , as is clear from the expression for the correlation coefficient between β_{il} and β_{im} , for $l \neq m$,

$$\rho(\beta_{im}, \beta_{il}) = \frac{\gamma_{ml} + \sum_{r=1}^{l-1} \gamma_{lr} \gamma_{mr}}{\sqrt{\left(1 + \sum_{r=1}^{l-1} \gamma_{lr}^2\right) \left(1 + \sum_{r=1}^{l-1} \gamma_{mr}^2\right)}}, \quad (6)$$

which does not depend on $\boldsymbol{\lambda}$. In the case where $\gamma_{l1} \gamma_{m1} = \dots = \gamma_{l,l-1} \gamma_{m,l-1} = \gamma_{ml} = 0$, $\rho(\beta_{il}, \beta_{im}) = 0$ and the random effects β_{il} and β_{im} are independent. In addition, for $\boldsymbol{\gamma} = \mathbf{0}_{q(q-1)/2}$, the elements of β_i are mutually independent with the random effects covariance reducing to the simple diagonal form: $\boldsymbol{\Omega} = \text{diag}(\lambda_1^2, \dots, \lambda_q^2)$.

It is interesting to consider the important special case where $h(\cdot)$ is the identity link, the outcomes are normally distributed, and the covariates are subject-specific (i.e., $\mathbf{x}_{ij} = \mathbf{x}_i$ for all i, j). In this case, the correlation coefficient between two observations on the same

subject, y_{ij} and $y_{ij'}$, conditional on $\mathbf{z}_{ij} = \mathbf{z}_{ij'} = \mathbf{z}_i$ is

$$\rho(y_{ij}, y_{ij'}; \mathbf{z}_i) = \frac{\sum_{m=1}^q (\lambda_m z_{im} + \sum_{l=m+1}^q \lambda_l z_{il} \gamma_{lm})^2}{\tau^{-1} + \sum_{m=1}^q (\lambda_m z_{im} + \sum_{l=m+1}^q \lambda_l z_{il} \gamma_{lm})^2}, \quad (7)$$

which equals 0 for $\lambda_1 = \dots = \lambda_q = 0$.

3. Prior Specification and Inference

By choosing a prior distribution for λ_l with a point mass at 0, prior probability is allocated to models that exclude the l th random effect (β_{il}), and hence assume homogeneity in the corresponding regression coefficient. A convenient choice for the prior, for $l = 1, \dots, q$, is

$$\pi(\lambda_l) = \Phi\left(-\frac{m_{l0}}{s_{l0}}\right)1(\lambda_l = 0) + \phi\left(\frac{\lambda_l - m_{l0}}{s_{l0}}\right)1(\lambda_l > 0), \quad (8)$$

where $\Phi(\cdot)$ is the standard normal distribution function, $\phi(\cdot)$ is the standard normal density function, and m_{l0}, s_{l0} are investigator-specified hyperparameters. This prior restricts $\lambda_l \geq 0$, assigns $\lambda_l = 0$ with probability $\rho_{l0} = \Phi(-m_{l0}/s_{l0})$, and has expectation $s_{l0}^2 \phi(m_{l0}/s_{l0}) + m_{l0} \Phi(m_{l0}/s_{l0})$. We assume apriori independence so that $\pi(\boldsymbol{\lambda}) = \prod_{l=1}^q \pi(\lambda_l)$. A Bayesian specification of the model is completed with Gaussian priors for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, and a prior for the scale parameter τ , if τ is unknown. For normally distributed outcomes, $\tau \sim \text{gamma}(c_0, d_0)$ corresponds to a conditionally conjugate prior.

Inference about the model parameters $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$, and τ proceeds as usual once the posterior distributions or simulated samples from the posterior distributions are available. In particular, one can report posterior means, posterior standard deviations, and 95% highest posterior density intervals.

The question of whether the l th random effect is necessary in the model ($l = 1, \dots, q$) can be addressed by testing the hypothesis $H_{1l} : \lambda_l > 0$ against $H_{0l} : \lambda_l = 0$ using the Bayes factor (c.f., Kass and Raftery, 1995). In particular, the Bayes factor in support of H_{1l} over H_{0l} is given by

$$\text{BF}_{10}^{(l)} = \frac{p(\mathbf{y}|H_{1l})}{p(\mathbf{y}|H_{0l})},$$

the ratio of the marginal likelihood of the data under H_{1l} and H_{0l} , respectively. In practice, values of the Bayes factor are usually computed using the posterior odds and the prior odds: $\widehat{\text{BF}}_{10} = [(1 - \hat{\rho}_l)/\hat{\rho}_l]/[(1 - \rho_{l0})/\rho_{l0}]$, where $\hat{\rho}_l$ is the estimated posterior probability that $\lambda_l = 0$. With a simulated sample $\lambda_l^{(1)}, \dots, \lambda_l^{(G)}$ from the posterior distribution of λ_l (e.g., based on MCMC), this posterior probability can be estimated by $\hat{\rho}_l = \sum_{g=1}^G 1(\lambda_l^{(g)} = 0)/G$.

4. Posterior Computation

Under constraint (2), expression (3) implies that

$$\eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\alpha} + \sum_{l=1}^q b_{il} \left(\lambda_l z_{ijl} + \sum_{m=l+1}^q \lambda_m z_{ijm} \gamma_{ml} \right). \quad (9)$$

Since η_{ij} is non-linear in the parameters $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T)^T$, $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$, depending on multiplicative terms, our reparameterized model is no longer in standard GLMM form. However, posterior computation can proceed via a Markov chain Monte Carlo algorithm, which is a minor modification of Gibbs sampling algorithms for GLMMs, due to the conditionally linear structure of the model.

4.1 Linear Mixed Models

We first consider the case where y_{ij} is normally distributed and $h(\cdot)$ is the identity link. In this case, the GLMM is in the form of a linear mixed effects model (Laird and Ware, 1982), and a simple modification of the Gibbs sampling approach of Gilks et al. (1993) can be used for posterior computation. After specifying initial values for the parameters and latent variables, our proposed MCMC algorithm proceeds as follows:

1. Update \mathbf{b}_i , for $i = 1, \dots, n$, by sampling from the conjugate Gaussian full conditional distribution for \mathbf{b}_i given the data for subject i and the current values for $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{\tau}$.
2. Update $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ by sampling from the conjugate Gaussian full conditional distribution for $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ given the data and the current values for $\boldsymbol{\lambda}, \boldsymbol{\tau}$, and \mathbf{b} . In particular, set $\gamma_{lm} = \gamma_{ml} = 0$ for $m = 1, \dots, q$ when $\lambda_l = 0$.

3. Update τ by sampling from the conjugate gamma full conditional distribution for τ given the data and the current values for $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, \mathbf{b} , and $\boldsymbol{\lambda}$.
4. Update λ_l , for $l = 1, \dots, q$, by sampling from the conjugate full conditional distribution of λ_l given the data and the current values for \mathbf{b} , $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, τ , and $\boldsymbol{\lambda}_{(l)}$, where $\boldsymbol{\lambda}_{(l)}$ is the vector of $\boldsymbol{\lambda}$ excluding the l th element.
5. Repeat steps 1-4 for a large number of iterations, and calculate posterior summaries based on samples collected after apparent convergence.

The conditional densities are provided in the Appendix.

4.2 Probit Models with Random Effects

When y_{ij} is binary, choosing the link function $h(\cdot)$ to be the standard normal cumulative distribution function, $\Phi(\cdot)$, results in a probit mixed effects model,

$$\Pr(y_{ij} = 1 | \mathbf{b}_i, \mathbf{x}_{ij}, \mathbf{z}_{ij}) = \Phi(\eta_{ij}) = \Phi\left(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + \sum_{l=1}^q b_{il}(\lambda_l z_{ijl} + \sum_{m=l+1}^q \lambda_m z_{ijm} \gamma_{ml})\right), \quad (10)$$

which is a reparameterization of standard probit mixed models (c.f., Chib and Greenberg, 1998). In this case, the data augmentation approach of Albert and Chib (1993) can be used to simplify posterior computation. In particular, an unobserved continuous variable w_{ij} is introduced which relates to y_{ij} via $y_{ij} = 1(w_{ij} > 0)$ and to η_{ij} via

$$w_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\alpha} + \sum_{l=1}^q b_{il} \left(\lambda_l z_{ijl} + \sum_{m=l+1}^q \lambda_m z_{ijm} \gamma_{ml} \right) + \delta_{ij}, \quad (11)$$

where $\delta_{ij} \sim N(0, 1)$.

Letting $\mathbf{w} = (w_{ij}, i = 1, \dots, n, j = 1, \dots, n_i)^T$, posterior computation for model (10) can proceed via a simple modification of the MCMC algorithm described in Section 4.1.

1. Update \mathbf{b}_i , for $i = 1, \dots, n$, by sampling from the conjugate Gaussian full conditional distribution for \mathbf{b}_i given the current values for \mathbf{w} , $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$, and τ .

2. Update $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ by sampling from the conjugate Gaussian full conditional distribution for $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ given the current values for \mathbf{w} , $\boldsymbol{\lambda}$, τ , and \mathbf{b} . In particular, set $\gamma_{lm} = \gamma_{ml} = 0$ for $m = 1, \dots, q$ when $\lambda_l = 0$.
3. Update w_{ij} , for $i = 1, \dots, n$, $j = 1, \dots, n_i$, by sampling from the its full conditional distribution given y_{ij} and the current values for $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, \mathbf{b} , and $\boldsymbol{\lambda}$. This full conditional distribution is a $N(\eta_{ij}, 1)$ truncated above (below) by 0 for $y_{ij} = 0$ ($y_{ij} = 1$).
4. Update λ_l , for $l = 1, \dots, q$, by sampling from the conjugate full conditional distribution of λ_l given the current values for \mathbf{w} , \mathbf{b} , $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, τ , and $\boldsymbol{\lambda}_{(l)}$, where $\boldsymbol{\lambda}_{(l)}$ is the vector of $\boldsymbol{\lambda}$ excluding the l th element.
5. Repeat steps 1-4 for a large number of iterations, and calculate posterior summaries based on samples collected after apparent convergence.

4.3 General Case

In the general case, the full conditional distributions may have non-conjugate forms. However, in most cases, the full conditional densities of \mathbf{b}_i , $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$, and τ are log-concave, and the adaptive rejection sampling algorithm can be used for direct sampling (c.f., Dellaportas and Smith, 1993). In addition, a Metropolis-Hastings algorithm can be used to update $\boldsymbol{\lambda}$.

5. Epidemiology Application

5.1 Data and Model

To illustrate the methodology, we considered an epidemiologic study on the relationship between prenatal exposure to polychlorinated biphenyls and motor development in young children. The data were drawn from twelve centers participating in the Collaborative Perinatal Project between 1959 and 1966. The outcome of interest was a child's psychomotor development at 8 months of age, as measured by the Bayley Psychomotor Index. Data were available for a total of 1124 children. The risk factor of primary interest in this study was the

level of prenatal exposure to PCBs, measured in maternal serum taken in the third trimester of pregnancy. Also available were the potential confounding variables: third trimester triglyceride and serum cholesterol level, maternal race (other, white, black), maternal education (below high school, high school, above high school), child’s birth order (first or not), and an indicator of whether the child had ever been breastfed.

Possible heterogeneity among the twelve centers in children’s level of psychomotor development and in the association between this outcome and the predictors was a major concern to the study investigators (Daniel et al, 2002). To accommodate such heterogeneity and to assess its sources, we considered a linear mixed model with $\mathbf{z}_{ij} = \mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ij,10})^T$, where $x_{ij1} = 1$, x_{ij2} was the level of prenatal exposure to PCBs for the j th child in center i , and $x_{ij3}, \dots, x_{ij,10}$ consisted of the confounders. This parameterization allows each regression coefficient, including the intercept, the PCB slope, and the eight confounding coefficients, to possibly vary for the different study centers. However, we did not assume a priori that these coefficients vary, and one of the primary goals of our analysis (in addition to assessing a PCB population-level effect) was to identify factors contribution to heterogeneity among centers.

5.2 Priors and Analysis

We chose $m_{l0} = 0$ and $s_{l0}^2 = 10$ to place 0.5 prior probability on the set $\lambda_l = 0$ for $l = 1, \dots, 10$. For this choice of prior, the conditional prior density of λ_l given $\lambda_l > 0$ is $N(0, 10)$ truncated below by 0, with mean and standard deviation at about 2.5 and 1.9, respectively. The prior for $\boldsymbol{\alpha}$ was $\pi(\boldsymbol{\alpha}) \propto 1$, reflecting no information about the population-level effects. For $\boldsymbol{\gamma}$, we chose a $N(\mathbf{0}_{q(q-1)/2}, \mathbf{I}_{q(q-1)/2 \times q(q-1)/2})$ prior to reflect our belief that the level of correlation between the different center-specific parameters is low to moderate. A gamma distribution with mean 1 and variance 1000 was used as the prior for τ (corresponding to $c_0 = 0.001$ and $d_0 = 0.001$).

Posterior summaries were calculated based on 100,000 samples collected after discard-

ing a 6,000 sample burn-in. There was no evidence of lack of convergence based on an examination of trace plots and convergence diagnostic procedures recommended by Cowles and Carlin (1996). In addition, autocorrelation tended to be low, suggesting high sampling efficiency.

We first investigated heterogeneity among centers in the regression coefficients by testing the hypotheses:

$$H_{1l} : \lambda_l > 0 \text{ vs } H_{0l} : \lambda_l = 0,$$

separately for $l = 1, \dots, q$ (note that H_{0l} implies $\beta_{1l} = \dots = \beta_{12,l}$). Table 1 shows the corresponding Bayes factors obtained using the approach of Section 3. Based on a Bayes factor of $\widehat{\text{BF}}_{10}^{(1)} = 2.922$, there appeared to be clear heterogeneity among the centers in the baseline mean psychomotor scores. However, we found no evidence that the effect of PCB exposure or of any of the confounders varied from center to center, since $\widehat{\text{BF}}_{10}^{(l)} < 1$ for $l = 2, \dots, 10$. Hence, a random intercept model appeared to adequately capture the heterogeneity structure. These results were robust to the prior choice, and our conclusions were the same for each of the priors considered in our sensitivity analyses (in particular, using 3, 4, 5, 7, 10, and 50 as the values for s_{l0} , and 0.5, 2, and 4 as the values for the prior variance for the elements of $\boldsymbol{\gamma}$). The range of the estimated Bayes factors for different choices of s_{l0} are provided in the final column of Table 1.

In addition to investigating differences among centers, investigators were interested in assessing evidence of an overall population-level effect of PCBs. The posterior mean of the PCB slope was $\hat{\alpha}_2 = 0.398$, with a standard deviation of 0.350; the 95% highest posterior density interval for α_2 was $(-0.304, 1.079)$. Hence there was no evidence of an overall dose-response relationship between PCB exposure and children's psychomotor development. This finding was also robust to the different prior choices for $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$.

5.3 Comparison to Previous Approaches

As mentioned in Section 1, an alternative to our analysis would be to fit a conventional mixed effects model with a random coefficient for each of the covariates, using PROC MIXED in the frequentist case and WinBUGS with an inverse Wishart prior for $\mathbf{\Omega}$ in the Bayesian case. We were unable to obtain convergence using SAS PROC MIXED, even after trying a variety of reasonable initial values. We also noted problems with the use of WinBUGS to implement the Bayesian analysis, related not to convergence, but to slow mixing and inefficient computation. In particular, to mirror the analysis of Section 5.2, we chose the same prior for τ , a diffuse $N(\mathbf{0}, 10^4 \mathbf{I}_{10 \times 10})$ prior for $\boldsymbol{\alpha}$, and a $\text{Wishart}(10, \mathbf{I}_{10 \times 10})$ prior for $\mathbf{\Omega}^{-1}$. The resulting Gibbs iterates were highly autocorrelated, compared to the sample from our method. Figure 1 presents a side by side comparison of the autocorrelation in samples of α_1 to α_4 . The autocorrelation in the α_1 sample from WinBUGS is still about 0.6 at lag 50, but that from our method almost reaches 0 at lag 10. The autocorrelation drops to 0 even faster for α_5 to α_{10} for our procedure. Similar improvements in sampling efficiency were also observed for the variance component parameters.

Since our approach allows random coefficients to have 0 variances while the conventional approach does not, one would expect the results from the two approaches to differ, especially regarding random effects estimates of the covariance $\mathbf{\Omega}$. Figure 2 presents the posterior densities of $\sqrt{\omega_{11}}$ and $\sqrt{\omega_{22}}$, based on samples from WinBUGS and from our procedure, respectively. It appears that by constraining the variance of the PCB slope parameter to be positive, the WinBUGS analysis based on a Wishart prior may have underestimated the magnitude of baseline heterogeneity among centers (as is apparent from the estimated marginal posterior densities of $\sqrt{\omega_{11}}$). In spite of differences in the estimates of $\mathbf{\Omega}$ between the two procedures, estimates for the population-level parameters, $\alpha_1, \dots, \alpha_{10}$, were similar. In particular, the WinBUGS analysis had $\hat{\alpha}_2 = 0.245$ (standard deviation = 0.459, 95% interval = $(-0.695, 1.130)$), which is similar to the estimate based on our procedure.

5. Discussion

In extending generalized linear models to allow multiple observations on a study subject, it is important to account for within-subject dependency that results from heterogeneity among subjects in the regression coefficients. Typically, this is done by incorporating subject-specific random effects for each regression coefficient that is expected to vary among subjects. By restricting the random effects covariance matrix to be positive definite, it is implicitly assumed that there is some degree of heterogeneity in each of the regression coefficients that have a random effect, while regression coefficients without random effects are fixed for all subjects.

Although Bayesian inferences on the covariance structure can potentially be implemented based on Bayes factors or other model selection strategies, such approaches typically require separate fitting of each model under consideration, which can be infeasible when the number of models is large. In addition, when interest focuses on the population parameters, a model averaging approach that properly accounts for uncertainty in the random effects structure seems preferable to basing inferences on a single model selected to optimize some criterion. An additional factor to consider is that current standard methods of fitting mixed effects models often face computational problems when the model has more than a few random effects. For example, in the PCB application, we were unable to use SAS PROC MIXED to obtain frequentist estimates due to lack of convergence. In addition, although we were able to implement a Bayesian Gibbs sampling analysis with a conjugate Wishart prior, we noticed problems with slow mixing.

Motivated by these drawbacks of current standard approaches, this article has proposed a class of priors for the variance components in a reparameterized version of the generalized linear mixed model. By allocating positive probability to non-negative definite covariance matrices for the random effects, the proposed approach allows for uncertainty in the regression coefficients that vary among subjects. In addition, the prior has an appealing con-

ditionally conjugate structure, which results in easy posterior computation using MCMC. In the PCB application, our MCMC algorithm had substantially higher sampling efficiency compared with standard approaches.

ACKNOWLEDGEMENTS

The authors thank Shyamal Peddada for his critical reading of the manuscript.

REFERENCES

- Albert, J. and Chib, S. (1997). Bayesian tests and model diagnostics in conditionally independent hierarchical models. *Journal of the American Statistical Association* **92**, 916-925.
- Chen, M.-H., Shao, Q.-M., and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.
- Chib, S. and Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika* **85**, 347-361.
- Dellaportas, P. and Smith, A.F.M. (1993). Bayesian-inference for generalized linear and proportional hazards models via Gibbs sampling. *Applied Statistics* **42**, 443-459.
- Cowles, M.K. and Carlin, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistician Association* **91**, 883-904.
- Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* **7**, 57-68.
- Gilks, W.R., Wang, C.C., Yvonnet, B., and Coursaget, P. (1993). Random-effects models for longitudinal data using Gibbs sampling. *Biometrics* **49**, 441-453.

- Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337-348.
- Hall, D.B. and Praestgaard, J.T. (2001). Order-restricted score tests for homogeneity in generalized linear and nonlinear mixed models. *Biometrika* **88**, 739-751.
- Kass, R.E. and Raftery, A.B. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.
- Kleinman, K.P. and Ibrahim, J.G. (1998). A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine* **17**, 2579-2596.
- Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.
- Littel, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). *SAS System for Mixed Models*, SAS Institute, Cary, NC.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. London: Chapman & Hall.
- Morrell, C.H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics* **54**, 1560-1568.
- Mukhopadhyay, S. and Gelfand, A.E. (1997). Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association* **92**, 633-639.
- Spiegelhalter, D.J., Thomas, A, and Best, N. (2000). WinBUGS with DoodleBUGS, version 1.3. Available from MRC Biostatistics Unit, Cambridge.
- Stram, D.O. and Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171-1177.

Weiss, R.E., Wang, Y. and Ibrahim, J.G. (1997). Predictive model selection for repeated measures random effects models using Bayes factors. *Biometrics* **53**, 592-602.

Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* **86**, 79-86.

APPENDIX

Details for Implementation of MCMC Algorithm

Step 1. Letting $\mathbf{v}_{ij} = (\lambda_l \mathbf{z}_{ijl} + \sum_{m=l+1}^q \lambda_m z_{ijm} \gamma_{ml} : l = 1, \dots, q)^T$, for $j = 1, \dots, n_i$, the full conditional density of \mathbf{b}_i is

$$N_q \left(\left(\mathbf{I}_{q \times q} + \tau \sum_{j=1}^{n_i} \mathbf{v}_{ij} \mathbf{v}_{ij}^T \right)^{-1} \tau \sum_{j=1}^{n_i} \mathbf{v}_{ij} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha}), \left(\mathbf{I}_{q \times q} + \tau \sum_{j=1}^{n_i} \mathbf{v}_{ij} \mathbf{v}_{ij}^T \right)^{-1} \right),$$

for $i = 1, \dots, n$.

Step 2. Let $(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ denote the prior density, let $\mathbf{u}_{ij} = (b_{il} \lambda_m z_{ijm} : l = 1, \dots, q; m = l + 1, \dots, q)^T$, and let $\mathbf{r}_{ij} = (\mathbf{x}_{ij}^T, \mathbf{u}_{ij}^T)^T$. It follows from expression (9), after some algebra, that the full conditional of $(\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T$ is $N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ with

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \hat{\boldsymbol{\Sigma}} \left(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + \tau \sum_{i=1}^n \sum_{j=1}^{n_i} \mathbf{r}_{ij} \left(y_{ij} - \sum_{l=1}^q b_{il} \lambda_l z_{ijl} \right) \right), \\ \hat{\boldsymbol{\Sigma}} &= \left(\boldsymbol{\Sigma}_0^{-1} + \tau \sum_{i=1}^n \sum_{j=1}^{n_i} \mathbf{r}_{ij} \mathbf{r}_{ij}^T \right)^{-1}. \end{aligned}$$

Step 3. The full conditional distribution of τ is gamma(\hat{c}, \hat{d}), where $\hat{c} = \sum_{i=1}^n n_i / 2 + c_0$ and $\hat{d} = \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha} - \sum_{l=1}^q b_{il} (\lambda_l z_{ijl} + \sum_{m=l+1}^q \lambda_m z_{ijm} \gamma_{ml}))^2 / 2 + d_0$.

Step 4. For $l = 1, \dots, q$, let $t_{ijl} = z_{ijl} (b_{il} + \sum_{m=1}^{l-1} b_{im} \gamma_{ml})$. Then the linear regression model implied by (9) can be rewritten as

$$y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha} - \sum_{m \neq l} t_{ijm} \lambda_m = t_{ijl} \lambda_l + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, 1/\tau)$. This, together with the prior for λ_l in (8), results in the full conditional distribution of λ_l

$$p(\lambda_l) = \hat{\rho}_l 1(\lambda_l = 0) + (1 - \hat{\rho}_l) \frac{\phi\left(\frac{\lambda_l - \hat{m}_l}{\hat{s}_l}\right)}{\Phi\left(\frac{\hat{m}_l}{\hat{s}_l}\right)} 1(\lambda_l > 0),$$

where

$$\hat{\rho}_l = \frac{\Phi\left(-\frac{m_{l0}}{s_{l0}}\right)/\phi\left(\frac{m_{l0}}{s_{l0}}\right)}{\Phi\left(-\frac{m_{l0}}{s_{l0}}\right)/\phi\left(\frac{m_{l0}}{s_{l0}}\right) + \Phi\left(\frac{\hat{m}_l}{\hat{s}_l}\right)/\phi\left(\frac{\hat{m}_l}{\hat{s}_l}\right)},$$

with

$$\hat{m}_l = \hat{s}^2 \left(\tau \sum_{i=1}^n \sum_{j=1}^{n_i} t_{ijl} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha} - \sum_{m \neq l} t_{ijm} \lambda_m) + \frac{m_{l0}}{s_{l0}^2} \right),$$

and

$$\hat{s}_l^2 = \left(\tau \sum_{i=1}^n \sum_{j=1}^{n_i} t_{ijl}^2 + \frac{1}{s_{l0}^2} \right)^{-1}.$$

Table 1: Bayes factors for identifying regression coefficients that vary for the different study centers in the PCB study

| Covariate | $\Pr(\lambda > 0 \mathbf{y})$ | Bayes factor | Range ^a |
|--------------------------|-------------------------------|--------------|-----------------------------|
| Intercept | 0.745 | 2.922 | (1.860, 3.762) |
| PCB | 0.199 | 0.248 | (0.061, 0.339) |
| Triglyceride | 0.179 | 0.218 | (0.021, 0.434) |
| Cholesterol | 0.363 | 0.570 | (0.176, 0.908) |
| Race: White | 0.170 | 0.205 | (0.022, 0.417) |
| Race: Black | 0.182 | 0.222 | (0.024, 0.444) |
| Education: = High School | 0.003 | 0.003 | (0.00 ^b , 0.005) |
| Education: > High School | 0.003 | 0.003 | (0.00 ^b , 0.003) |
| First Born | 0.158 | 0.188 | (0.020, 0.396) |
| Breastfeeding | 0.251 | 0.335 | (0.042, 0.629) |

^a For different choices of prior

^b Values < 0.0005