Bayesian Latent Variable Models for Clustered Mixed Outcomes
Author(s): David B. Dunson
Source: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, Vol. 62, No. 2 (2000), pp. 355-366
Published by: Wiley for the Royal Statistical Society
Stable URL: https://www.jstor.org/stable/3088864
Accessed: 25-03-2019 03:58 UTC

J. R. Statist. Soc. B (2000)
62, *Part* 2, *pp.* 355–366

Bayesian latent variable models for clustered mixed outcomes

David B. Dunson

National Institute of Environmental Health Sciences, Research Triangle Park, USA

[Received March 1999. Final revision October 1999]

Summary. A general framework is proposed for modelling clustered mixed outcomes. A mixture of generalized linear models is used to describe the joint distribution of a set of underlying variables, and an arbitrary function relates the underlying variables to the observed outcomes. The model accommodates multilevel data structures, general covariate effects and distinct link functions and error distributions for each underlying variable. Within the framework proposed, novel models are developed for clustered multiple binary, unordered categorical and joint discrete and continuous outcomes. A Markov chain Monte Carlo sampling algorithm is described for estimating the posterior distributions of the parameters and latent variables. Because of the flexibility of the modelling framework and estimation procedure, extensions to ordered categorical outcomes and more complex data structures are straightforward. The methods are illustrated by using data from a reproductive toxicity study.

Keywords: Gibbs sampler; Mixture model; Multihit model; Multiple outcomes; Reproductive applications

1. Introduction

Analyses of reproductive and developmental data are complicated by correlations between subunits within subjects, and between multiple outcomes on each subunit. Since outcomes are often measured on a variety of scales (continuous, binary, ordinal, unordered categorical or count), it can be difficult to model joint covariate effects. There is a need for general statistical procedures for analysing clustered mixed outcome data that are easy to implement, account for and provide information on the complex correlation structure, are applicable to a broad class of study designs and are appropriate in small samples.

Methods have been proposed for joint modelling of binary and continuous data (Catalano and Ryan, 1992; Cox and Wermuth, 1992; Chen, 1993; Fitzmaurice and Laird, 1995). These bivariate models account for dependence between one binary and one continuous outcome but do not provide information on the correlation. Recently, latent variable models have been developed for mixed outcomes (Arminger and Küsters, 1988; Moustaki, 1996; Sammel *et al.*, 1997; Shi and Lee, 1998, 2000; Moustaki and Knott, 1999). With the exception of the methods of Arminger and Küsters (1988) and Sammel *et al.* (1997), these models do not consider the effects of covariates on the latent variables or on the observed outcomes.

Arminger and Küsters (1988) assumed that each observed outcome is a manifestation of an underlying continuous variable that is linearly related to a normal latent trait. Each type of outcome is related to its own latent trait, and the set of latent traits is assumed to be multi-

Address for correspondence: David B. Dunson, Biostatistics Branch, MD A3-03, National Institute of Environmental Health Sciences, PO Box 12233, Research Triangle Park, NC 27709, USA.
E-mail: dunson1@niehs.nih.gov

 1369–7412/00/62355

variate normal with expectation potentially dependent on covariates. Sammel *et al.* (1997) developed alternative latent normal models for mixed outcomes. They assumed that each manifest variable follows a one-parameter exponential family model, and they connected the manifest variables to fixed covariates and a subject-specific latent variable through a generalized linear model.

In this paper, a new class of latent variable models is proposed for clustered mixed outcome data, and Markov chain Monte Carlo (MCMC) algorithms (Tierney, 1994) are developed for estimating the posterior distribution of the parameters and the latent variables. The modelling framework follows a similar structure to the approach of Arminger and Küsters (1988), in that the measured outcomes are assumed to be manifestations of underlying variables which are related to latent variables. However, the model generalizes the framework of Arminger and Küsters to accommodate non-normal latent variables, multilevel data, non-linear relationships between the underlying and latent variables, multiple latent variables for each outcome type and covariate-dependent modifications of the relationship between the latent and underlying variables.

The joint distribution of the underlying variables is described by a mixture of generalized linear models, where each underlying variable is assigned a distinct link function. By allowing the underlying variables to follow any distribution in the exponential family, it is trivial to modify the model and computational algorithm to accommodate a broad variety of data structures. Special cases of the general framework include the standard generalized linear mixed model (GLMM) (Zeger and Karim, 1991; Breslow and Clayton, 1993) and the model of Sammel *et al.* (1997).

With the exception of the approach of Shi and Lee (1998), methods for the analysis of mixed outcome data have been based on asymptotic maximum likelihood theory. However, the Bayesian approach has several important advantages. First, the exact posterior distributions of the parameters and latent variables can be estimated by using MCMC methods. Means and quantiles based on the estimated posteriors are appropriate regardless of the sample size. In contrast, standard errors and confidence limits for the maximum likelihood estimates are typically based on strong asymptotic normality assumptions. Second, the Bayesian approach allows for the direct incorporation of prior knowledge. This is a major advantage in structural equation modelling. Classical methods often require that a subset of the parameters is known to ensure identifiability. Although constraints on the threshold parameters and the variance of the latent variables are often reasonable, additional less justifiable constraints can be avoided by using a prior distribution to allow for prior uncertainty in the parameters. In addition, by assigning an informative prior to parameters about which there is previous information (perhaps from historical studies conducted under a similar design), more precise estimates of the parameters of interest can be obtained.

First the general modelling framework is described. Within the framework proposed, models are then developed for clustered multiple binary, unordered categorical and joint categorical and continuous outcomes. A general MCMC sampling algorithm for posterior estimation is outlined in Section 3. The methods are illustrated with an example in Section 4, and the paper concludes with a discussion.

2. Latent variable models

2.1. Modelling framework

Suppose that there are n subjects in a study ($i = 1, \ldots, n$), within each subject there are n_i subunits ($j = 1, \ldots, n_i$) and each subunit experiences L outcomes ($y_{ij1}, \ldots, y_{ijL}$). I assume

that there is a variable z_{ijk} underlying the observed outcome y_{ijk} $(k = 1, \ldots, L)$ and the vector of observed outcomes $\mathbf{y}_{ij} = (y_{ij1}, \ldots, y_{ijL})'$ is linked to the vector of underlying variables $\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ijL})'$ through the model

$$\mathbf{y}_{ij} = g(\mathbf{z}_{ij}, \boldsymbol{\xi}), \tag{1}$$

where $g(\cdot)$ is a known function and $\boldsymbol{\xi}$ is a parameter vector. The underlying variable z_{ijk} is assumed to follow an exponential family,

$$p(z_{ijk}|\theta_{ijk}, \tau_k) = \exp[\tau_k\{z_{ijk}\theta_{ijk} - a(\theta_{ijk})\} + c_k(z_{ijk}, \tau_k)], \tag{2}$$

where

$$\mu_{ijk} = E(z_{ijk}|\theta_{ijk}, \tau_k) = \frac{\mathrm{d}a(\theta_{ijk})}{\mathrm{d}\theta_{ijk}},$$

$$v_{ijk} = \mathrm{var}(z_{ijk}|\theta_{ijk}, \tau_k) = \tau_k^{-1}\frac{\mathrm{d}^2a(\theta_{ijk})}{\mathrm{d}\theta_{ijk}^2}$$

and τ_k is a scalar dispersion parameter $(k = 1, \ldots, L)$. The canonical parameter θ_{ijk} is related to covariates and latent variables through the generalized linear model

$$h_k(\theta_{ijk}) = \eta_{ijk} = \mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1} + \mathbf{u}_{ijk2}^{\mathrm{T}}\mathbf{b}_{ij2}, \tag{3}$$

where $h_k(\cdot)$ is a monotonic link function that is specific to the kth underlying variable, $\mathbf{x}_{ijk}$ is a $q \times 1$ vector of covariates, $\boldsymbol{\beta}$ is a $q \times 1$ vector of global parameters, $\mathbf{u}_{ijk1}$ is an $r_1 \times 1$ vector of covariates linking the subject level latent variables $\mathbf{b}_{i1}$ to η_{ijk} and $\mathbf{u}_{ijk2}$ is an $r_2 \times 1$ vector of covariates linking the subunit level latent variables $\mathbf{b}_{ij2}$ to η_{ijk}. The first term in model (3) describes the global effect of the covariates, the second term describes how these global effects are modified for each subject and the third term describes how each subject's effects are modified for each subunit.

The latent variables are either assigned multivariate normal densities

$$\begin{aligned} \mathbf{b}_{i1} &\sim N_{r_1}(\mathbf{0}, \boldsymbol{\Omega}_1^{-1}), \\ \mathbf{b}_{ij2} &\sim N_{r_2}(\mathbf{0}, \boldsymbol{\Omega}_2^{-1}), \end{aligned} \tag{4}$$

where $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are precision matrices, or they are linked to variables in a simple exponential family:

$$\begin{aligned} \mathbf{b}_{i1} &= \boldsymbol{\gamma}_1 d_{i1}, \\ \mathbf{b}_{ij2} &= \boldsymbol{\gamma}_2 d_{ij2}, \end{aligned} \tag{5}$$

where

$$p(d_{i1}) = \exp\{a_{01} + c_{01}(d_{i1})\},$$

$$p(d_{ij2}) = \exp\{a_{02} + c_{02}(d_{ij2})\}.$$

The second formulation clearly accommodates a broader class of distributions but is less flexible in the normal case in that it assumes that a single ranking applies to all outcomes. However, substantial gains in computing efficiency and stability are possible by using either expressions (5) or (4) with diagonal $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$.

I make the standard conditional independence assumption, i.e. z_{ijk} and $z_{ijk'}$ are independent conditionally on $\mathbf{b}_{i1}$ and $\mathbf{b}_{ij2}$. In the special case where $y_{ijk} = z_{ijk}$, $h_k(\cdot) \equiv h(\cdot)$ for all k, $\mathbf{b}_{ij2} = 0$ and $\mathbf{b}_{i1}$ is normal, models (1)–(4) describe a standard GLMM (Zeger and Karim, 1991).

Models (2) and (3) are in a general form that accommodates mixtures of underlying variables by assigning the underlying variables distinct link functions and dispersion parameters. An important special case is the normal linear mixture model

$$p(\mathbf{z}_{ij}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2}, \boldsymbol{\Psi}) = N_L(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{U}_{ij1}\mathbf{b}_{i1} + \mathbf{U}_{ij2}\mathbf{b}_{ij2}, \boldsymbol{\Psi}^{-1}), \tag{6}$$

where $\mathbf{X}_{ij}$, $\mathbf{U}_{ij1}$ and $\mathbf{U}_{ij2}$ are matrices with row vectors $\mathbf{x}_{ijk}$, $\mathbf{u}_{ijk1}$ and $\mathbf{u}_{ijk2}$ respectively ($k = 1, \ldots, L$) and $\boldsymbol{\Psi} = \text{diag}(\tau_1, \ldots, \tau_L)$ is an $L \times L$ diagonal precision matrix.

The framework of models (1)–(5) is applicable to a broad class of mixed outcome data structures, such as mixed dichotomous and count data or mixed ordinal and continuous data. Variations in the form of the observed outcome vector $\mathbf{y}_{ij}$ can be accommodated by choosing an appropriate link function in model (1) and/or by varying the distributions of the underlying and latent variables. For brevity, the discussion is limited to a few special cases. Models for clustered multiple binary and clustered unordered categorical outcomes are outlined in Sections 2.2 and 2.3 respectively. In Section 2.4 these models are extended to allow for joint modelling of multiple continuous outcomes.

2.2. Clustered multiple binary outcomes

Several models have been proposed for correlated binary data. Legler and Ryan (1997) developed a random-effects model that assumes a Poisson latent variable for the severity of the response. Their model assigns each subject a single latent variable and is therefore useful in quantifying the subject-specific severity of response. Dey and Chen (1996), Chen and Dey (1998) and Chib and Greenberg (1998) have developed Bayesian latent normal models for correlated binary data. On the basis of the framework of Section 2.1, I propose alternative models.

Suppose that the observed outcome $y_{ijk} = 1$ if an adverse event of type k occurs on subunit j within subject i, and $y_{ijk} = 0$ otherwise. I assume that $y_{ijk} = 1$ if the normal underlying variable $z_{ijk} > 0$, and $y_{ijk} = 0$ otherwise. In univariate probit models the normal variance parameter is non-identifiable by the observed data and can be set to an arbitrary constant. To ensure identifiability I set $\boldsymbol{\Psi} = \mathbf{I}_{L\times L}$ in model (6), where $\mathbf{I}_{L\times L}$ is the $L \times L$ identity matrix. It follows from model (6) that the probability of an adverse event of type k on the jth subunit of subject i is

$$\Pr(y_{ijk} = 1|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2}) = \Phi(\mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1} + \mathbf{u}_{ijk2}^{\mathrm{T}}\mathbf{b}_{ij2}), \tag{7}$$

where

$$\Phi(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$$

is the standard normal distribution function. Model (7) allows for heterogeneity among subjects and subunits within subjects, as well as general covariate effects on the probability of each type of event. A Monte Carlo EM algorithm similar to that proposed by Shi and Lee (2000) could potentially be used for maximum likelihood estimation of the model. I instead use a Gibbs sampler (Zeger and Karim, 1991; Gilks *et al.*, 1993) for estimating the posterior distribution (see Section 3). The advantages of the Bayesian approach are discussed in Section 1.

Normal underlying variable models are a special case of the general class of models outlined in Section 2.1. Alternative models can be formulated by assuming different distributions for the underlying variables. For example, a novel multihit model (Rai and Van Ryzin, 1981; Chu and Kuo, 1997) could be developed for multiple binary outcomes by assuming a Poisson distribution for the underlying variables:

$$\begin{aligned} y_{ijk} &= \mathbf{1}(z_{ijk} \geqslant \xi_k), \\ p(z_{ijk}|\mu_{ijk}) &= \text{Poisson}(\mu_{ijk}), \\ \mu_{ijk} &= \exp(\mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1} + \mathbf{u}_{ijk2}^{\mathrm{T}}\mathbf{b}_{ij2}), \end{aligned} \tag{8}$$

where ξ_k is the number of 'hits' required to induce an adverse response of type k and μ_{ijk} is the intensity of the underlying Poisson process. If $\xi_k = 1$ then model (8) describes a one-hit model, which follows a complementary log–log-form,

$$\log[-\log\{\Pr(y_{ijk} = 0)\}] = \mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1} + \mathbf{u}_{ijk2}^{\mathrm{T}}\mathbf{b}_{ij2}.$$

An extremely flexible and biologically motivated model can be developed by assuming that the number of hits ξ_k is unknown.

2.3. *Clustered multinomial outcomes*

In this section models are developed for correlated multinomial data. Suppose that y_{ijk} is a binary indicator that subunit j within subject i experiences an adverse event of type k ($k = 1, \ldots, L-1$) and $y_{ijL} = 1 - \Sigma_{k=1}^{L-1} y_{ijk}$. I assume that the outcomes are mutually exclusive so that $y_{ijk} = 1$ implies that $y_{ijk'} = 0$ for all $k' \neq k$. Following the approach of Arminger and Küsters (1988) and Albert and Chib (1993), I link the observed outcome $\mathbf{y}_{ij}$ to the latent $\mathbf{z}_{ij}$ by setting $y_{ijk} = 1$ if $z_{ijk} > z_{ijk'}$ for all $k' \neq k$, and $y_{ijk} = 0$ otherwise. I initially assume that $\mathbf{z}_{ij}$ is normal as shown in model (6). To ensure that the parameters are identifiable, $\boldsymbol{\Psi}$ is set equal to the $(L-1) \times (L-1)$ identity matrix and z_{ijL} is assumed to equal 0.

For a given subject, the number of subunits per category is multinomial whether or not the category probabilities are the same for all subunits. Since each subunit experiences one and only one of the L events, the data are not informative about the subunit level latent variables $\mathbf{b}_{ij2}$. It follows that

$$p(z_{ijk}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2} = 0, \tau_k = 1) = N(\mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1}, 1) \qquad \text{if } k \neq L \text{ and 0 otherwise,}$$

and the multinomial probabilities are given by

$$\Pr(y_{ijk} = 1|\mathbf{z}_{ij}, \boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2} = 0, \tau_k = 0) = \Pr(z_{ijk} > z_{ijk'}, \text{ for all } k' \neq k), \tag{9}$$

where $z_{ijL} = 0$. Models can be formulated that allow subjects to vary in both the tendency to experience particular adverse events and the overall probability of experiencing any adverse event. A Gibbs sampler is outlined in Section 3.

An interesting alternative model can be formulated within the framework outlined above by assigning gamma densities to the underlying variables:

$$p(z_{ijk}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2} = 0, \tau_k = 1) = \text{gamma}\{-1/(\mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1})\} \qquad \text{if } k \neq L \text{ and } \xi \text{ otherwise,}$$

where ξ is a random variable. The gamma model is more flexible than a normal model, since the relationship between the multinomial probabilities and the covariates depends not only on the relative means of the underlying densities but also on the relative skewness.

2.4. Clustered categorical and continuous outcomes

In this section, the normal underlying variable models proposed in the previous two sections are generalized to allow joint modelling of multiple continuous outcomes. Similar extensions are possible for models with non-normal underlying variables.

I first describe a model for multiple binary and continuous outcomes. Suppose that the L observed outcomes $y_{ij1}, \ldots, y_{ijL}$ consist of B dichotomous indicators of adverse events $y_{ij1}, \ldots, y_{ijB}$ and $L - B$ continuous outcomes $y_{ijB+1}, \ldots, y_{ijL}$ $(B < L)$. I assume that $y_{ijk} = z_{ijk}$ for $k > B$, and y_{ijk} is an indicator of $z_{ijk} > 0$ for $k \leqslant B$. To ensure identifiability of the parameters the precision matrix $\boldsymbol{\Psi}$ is defined as

$$\boldsymbol{\Psi} = \begin{pmatrix} \mathbf{I}_{B\times B} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_0 \end{pmatrix}_{L\times L},$$

where $\boldsymbol{\Psi}_0$ is an $(L - B) \times (L - B)$ diagonal precision matrix. It follows from model (6) that

$$E(y_{ijk}|\mu_{ijk}) = \begin{cases} \Phi(\mu_{ijk}), & \text{for } k \leqslant B, \\ \mu_{ijk}, & \text{for } k > B, \end{cases} \tag{10}$$

where $\mu_{ijk} = \mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1} + \mathbf{u}_{ijk2}^{\mathrm{T}}\mathbf{b}_{ij2}$ and $E(y_{ijk}) = \Pr(y_{ijk} = 1)$ for $k \leqslant B$.

A similar approach can be used to model multinomial and multiple continuous outcomes jointly. Suppose that the $B - 1$ observed outcomes $y_{ij1}, \ldots, y_{ijB-1}$ are mutually exclusive indicators of adverse events. Although subunit level modifiers are not identifiable on the basis of multinomial data, this is not the case when a subset of $\mathbf{z}_{ij}$ is observed directly. Following the approach outlined in Section 2.3,

$$E(y_{ijk}) = \begin{cases} \Pr(z_{ijk} > z_{ijk'}, \text{ for all } k' \neq k, k' \leqslant B), & \text{for } k \leqslant B, \\ \mu_{ijk}, & \text{for } k > B, \end{cases} \tag{11}$$

where $\mu_{ijk} = \mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1} + \mathbf{u}_{ijk2}^{\mathrm{T}}\mathbf{b}_{ij2}$, $z_{ijB} = 0$ and $E(y_{ijk}) = \Pr(y_{ijk} = 1)$ for $k \leqslant B$.

In Section 3 Gibbs sampling algorithms are presented for models (10) and (11). The Gibbs sampler can be easily adapted to allow variations on the basic form of these models. For example, it is straightforward to modify the algorithm to allow joint modelling of multinomial and multiple binary outcome data or to accommodate ordered categorical outcomes.

3. Bayesian estimation

In this section, prior distributions are chosen for the parameters, and a general MCMC algorithm is outlined for estimating the posterior distributions of the parameters and the latent variables. The prior distributions are conjugate if the underlying variables are normal.

3.1. Prior distributions

The regression parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$ are assigned a normal prior,

$$\boldsymbol{\theta} \sim N_q(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \tag{12}$$

where $\boldsymbol{\mu}_0$ is a vector of location parameters and $\boldsymbol{\Sigma}_0$ is a covariance matrix. To choose a vague prior distribution for $\boldsymbol{\theta}$, set $\mu_0 = 0$ and $\boldsymbol{\Sigma}_0 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_q^2)$ where σ_p^2 is large (e.g. $\sigma_p^2 = 100$). This ensures that the posterior is driven by the data. Wishart priors are specified for the precision matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ in expressions (4):

$$\boldsymbol{\Omega}_s \sim \text{Wishart}(\nu_s, \boldsymbol{\Lambda}_s), \tag{13}$$

with degrees of freedom ν_s and precision $\boldsymbol{\Lambda}_s$ ($s = 1, 2$). A proper prior can be assigned by choosing $\nu_s \geqslant r_s$, where r_s is the dimension of $\boldsymbol{\Omega}_s$. In the special case where $r_s = 1$, distribution (13) reduces to gamma($\nu_s/2$, $\boldsymbol{\Lambda}_s/2$). If the multiple random effects are assumed to be independent, $\boldsymbol{\Omega}_s$ is a diagonal matrix and the diagonal elements are assigned independent gamma priors. Similarly, independent gamma priors are specified for the scalar dispersion parameters τ_k:

$$\tau_k \sim \text{gamma}(\alpha_{0k}, \lambda_{0k}), \tag{14}$$

with index α_{0k} and scale λ_{0k} ($k = 1, \ldots, L$).

3.2. Posterior computations

MCMC techniques are used for posterior computations in the models proposed in Section 2. In the special case where all the underlying and latent variables are normal the MCMC algorithm is a Gibbs sampler that follows a simple form.

3.2.1. Generalized linear mixed models

In this section, an MCMC algorithm is outlined for posterior computation in the special case where $z_{ijk} = y_{ijk}$ for all i, j and k and the latent variables are multivariate normal as shown in expressions (4). It is trivial to modify the algorithm to accommodate mixture distributions of the form shown in expression (5).

Given $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, the joint posterior distribution for the parameters and latent variables is

$$\begin{aligned} p(\boldsymbol{\beta}, \mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\tau}|\mathbf{z}) &\propto p(\mathbf{z}|\boldsymbol{\beta}, \mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\tau})\, \pi(\boldsymbol{\beta}, \mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\tau}) \\ &\propto \exp\Bigg[\sum_{i,j,k} \log\{p(z_{ijk}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2}, \tau_k)\} - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0) \\ &\quad - \frac{1}{2}\sum_{i=1}^{n} \mathbf{b}_{i1}^{\mathrm{T}}\boldsymbol{\Omega}_1\mathbf{b}_{i1} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \mathbf{b}_{ij2}^{\mathrm{T}}\boldsymbol{\Omega}_2\mathbf{b}_{ij2} + \sum_{k=1}^{L}\{(\alpha_{0k} - 1)\log(\tau_k) - \lambda_{0k}\tau_k\}\Bigg], \end{aligned} \tag{15}$$

where $\Sigma_{i,j,k}$ denotes $\Sigma_{i=1}^{n}\Sigma_{j=1}^{n_i}\Sigma_{k=1}^{L}$, $\pi(\cdot)$ denotes the joint prior density, $\mathbf{b}_1 = (\mathbf{b}_{11}, \ldots, \mathbf{b}_{n1})$, $\mathbf{b}_2 = (\mathbf{b}_{112}, \ldots, \mathbf{b}_{nn_n2})$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)$ and $\mathbf{z} = (\mathbf{z}_{11}, \ldots, \mathbf{z}_{nn_n})$.

We require the full conditional distributions of each of the unknowns, which can be obtained in the standard manner (Zeger and Karim, 1991; Gilks *et al.*, 1993; Kleinman and Ibrahim, 1998). We have

$$p(\boldsymbol{\beta}|\mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\tau}, \mathbf{z}) \propto \exp\Bigg[\sum_{i,j,k} \log\{p(z_{ijk}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2}, \tau_k)\} - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\Bigg]. \tag{16}$$

If z_{ijk} is normal then this conditional distribution is normal and sampling is straightforward (see Gilks *et al.* (1993)). Otherwise, sampling can be accomplished by using either an adaptive rejection algorithm (Gilks and Wild, 1992), if the log-likelihood is concave with respect to $\boldsymbol{\beta}$, or a Metropolis–Hastings algorithm (Hastings, 1970). The full conditional distribution of $\mathbf{b}_{i1}$ is

$$p(\mathbf{b}_{i1}|\boldsymbol{\beta}, \mathbf{b}_2, \boldsymbol{\tau}, \mathbf{z}) \propto \exp\left[\sum_{i,j,k} \log\{p(z_{ijk}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2}, \tau_k)\} - \frac{1}{2}\sum_{i=1}^{n} \mathbf{b}_{i1}^{\mathrm{T}}\boldsymbol{\Omega}_1\mathbf{b}_{i1}\right], \tag{17}$$

and the full conditional distribution of $\mathbf{b}_{ij2}$ is

$$p(\mathbf{b}_{ij2}|\boldsymbol{\beta}, \mathbf{b}_1, \boldsymbol{\tau}, \mathbf{z}) \propto \exp\left[\sum_{i,j,k} \log\{p(z_{ijk}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2}, \tau_k)\} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n_i} \mathbf{b}_{ij2}^{\mathrm{T}}\boldsymbol{\Omega}_2\mathbf{b}_{ij2}\right]. \tag{18}$$

Conditional distributions (17) and (18) are normal if z_{ijk} is normal (Gilks *et al.*, 1993). Otherwise, sampling can again be accomplished by using either adaptive rejection or Metropolis steps. The full conditional distribution of τ_k is

$$p(\tau_k|\boldsymbol{\beta}, \mathbf{b}_1, \mathbf{b}_2, \mathbf{z}) \propto \tau_k^{\alpha_{0k}-1} \exp\left[\sum_{i=1}^{n}\sum_{j=1}^{n_i} \log\{p(z_{ijk}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{ij2}, \tau_k)\} - \lambda_{0k}\tau_k\right]. \tag{19}$$

This distribution is gamma if z_{ijk} is normal (Gilks *et al.*, 1993). The full conditional distributions of the precision matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are

$$p(\boldsymbol{\Omega}_1|\boldsymbol{\beta}, \mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\tau}, \mathbf{z}) = \text{Wishart}\left(\nu_1 + n, \boldsymbol{\Lambda}_1 + \sum_{i=1}^{n} \mathbf{b}_{i1}\mathbf{b}_{i1}^{\mathrm{T}}\right), \tag{20}$$

$$p(\boldsymbol{\Omega}_2|\boldsymbol{\beta}, \mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\tau}, \mathbf{z}) = \text{Wishart}\left(\nu_2 + \sum_{i=1}^{n} n_i, \boldsymbol{\Lambda}_2 + \sum_{i=1}^{n}\sum_{j=1}^{n_i} \mathbf{b}_{ij2}\mathbf{b}_{ij2}^{\mathrm{T}}\right). \tag{21}$$

The algorithm proceeds by sampling $\boldsymbol{\beta}$, $\mathbf{b}_{i1}$, $\mathbf{b}_{ij2}$, τ_k, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ from conditional distributions (16)–(21) respectively. This algorithm enables posterior computation in the case where the underlying variables are observed (i.e. $z_{ijk} = y_{ijk}$ for all i, j and k). It is straightforward to extend the algorithm to cases where the latent variables are unobservable by adding steps to sample from the full conditional distribution of z_{ijk} and the parameters $\boldsymbol{\xi}$. In the following section, such extensions will be outlined for the normal underlying variable models presented in Sections 2.2–2.4.

3.2.2. Clustered categorical and continuous outcomes

The full conditional distribution of z_{ijk} depends on the assumed link function in model (1). If $\mathbf{y}_{ij}$ consists of multiple binary outcomes then it follows from model (7) that

$$p(z_{ijk}|y_{ijk}, \boldsymbol{\beta}, \mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\tau}) = \text{truncated } N(\mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{u}_{ijk1}^{\mathrm{T}}\mathbf{b}_{i1} + \mathbf{u}_{ijk2}^{\mathrm{T}}\mathbf{b}_{ij2}, 1), \tag{22}$$

where z_{ijk} is truncated at the left by 0 if $y_{ijk} = 1$ and is truncated at the right by 0 if $y_{ijk} = 0$. If $\mathbf{y}_{ij}$ consists of multiple binary and continuous outcomes then, from model (10), $z_{ijk} = y_{ijk}$ for $l > B$ and equation (22) can be used to sample z_{ijk} for $l \leqslant B$.

If $\mathbf{y}_{ij}$ is a clustered multinomial outcome a procedure proposed by Albert and Chib (1993) can be used to sample the latent $\mathbf{z}_{ij}$. Conditional on $y_{ijk} = 1$,

$$p(\mathbf{z}_{ij}|\mathbf{y}_{ij}, \boldsymbol{\beta}, \mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\tau}) = N(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{U}_{ij1}\mathbf{b}_{i1}, \mathbf{I}) \qquad \text{subject to } z_{ijk} > z_{ijk'} \text{ for all } k' \neq k, \tag{23}$$

where $z_{ijL} = 0$ and $\mathbf{I}$ is the $(L-1) \times (L-1)$ identity matrix. To sample $\mathbf{z}_{ijk}$, sample from equation (23) and accept the draw if the ordering constraint is satisfied.

In alternative models with ordinal outcomes or non-normal underlying variables the link between the observed and latent data may involve a parameter vector $\boldsymbol{\xi}$. In this case, a step should be added to sample from the full conditional distribution of $\boldsymbol{\xi}$. This can be done as described in Albert and Chib (1993).

4. Example: reproductive toxicity study

4.1. Animal breeding data

To illustrate the methodology, I analyse animal breeding data from a reproductive toxicity study of ethylene glycol monomethyl ether (EGMME) conducted by the US National Toxicology Program (1988). EGMME is a commonly used solvent with many industrial applications, and there is evidence that it may cause neurologic symptoms in humans. In our study, breeding pairs of CD-1 albino mice were exposed through drinking water to 0.0%, 0.03%, 0.1% or 0.3% EGMME. The pairs bred continuously until five litters had been produced or the study period ended. For each litter, the number of pups was recorded and it was documented whether the time to birth was normal (20 weeks or less) or long (greater than 20 weeks).

For illustration, I analyse data from litters which were born during the study to dams in the control and 0.1% EGMME groups. For litter j from pair i, let $y_{ij1} = 1$ if the time to birth is long and $y_{ij1} = 0$ otherwise, and let y_{ij2} be the litter size. In the control group 54 out of 134 litters (40.3%) had a long time to birth and the mean litter size was 12.59. In the exposed group 117 out of 132 litters (88.6%) had a long time to birth and the mean litter size was 11.65.

4.2. The model

The data consist of a dichotomous outcome y_{ij1} and a count y_{ij2} for each litter from each breeding pair. I link the observed outcomes to underlying variables z_{ij1} and $\mathbf{z}_{ij2} = (z_{ij21}, \ldots, z_{ij2M})$ using

$$\begin{aligned} y_{ij1} &= \mathbf{1}(z_{ij1} > 0), \\ y_{ij2} &= \sum_{m=1}^{M} \mathbf{1}(z_{ij2m} = 0), \end{aligned} \tag{24}$$

where M is the maximum number of pups per litter. The underlying variables are assumed to follow Poisson distributions with intensities

$$\begin{aligned} E(z_{ij1}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{i2}) &= \exp(\beta_{1j} + \beta_2 x_i + b_{i1.1} + b_{ij2}), \\ E(z_{ij2m}|\boldsymbol{\beta}, \mathbf{b}_{i1}, \mathbf{b}_{i2}) &= \exp(\beta_{3j} + \beta_4 x_i + b_{i1.2} + b_{ij2}), \end{aligned} \tag{25}$$

where $x_i = 1$ if the ith pair is exposed and $x_i = 0$ otherwise, and the latent variables $b_{i1.1}$, $b_{i1.2}$ and b_{ij2} are assumed to follow independent normal distributions with mean 0 and precisions Ω_{11}, Ω_{12} and Ω_2 respectively. Expression (25) describes a three-factor model with the covariance between the pair-specific latent variables $b_{i1.1}$ and $b_{i1.2}$ assumed to be 0. Heterogeneity among breeding pairs in the time to birth and the number of pups per litter is accommodated through the pair-specific latent variables $b_{i1.1}$ and $b_{i1.2}$, and correlation between the outcomes is accounted for through the litter-specific latent variable b_{ij2}.

This model for y_{ij1} is based on the assumption that the time to birth is normal unless one or more adverse events occur that delay reproduction. The model for y_{ij2} is motivated by the observation that litter size distributions tend to be skewed to the left and underdispersed relative to the Poisson distribution (Williams, 1987). To accommodate these features, I assume that y_{ij2} is the number of pups out of M that do not experience an adverse event. The maximum number of pups per litter is set to $M = 20$, which is consistent with historical data on CD-1 mice.

To complete a Bayesian formulation of the model, priors are specified for β, Ω_{11}, Ω_{12} and Ω_2. Control data are available from historical studies conducted under the same design as the current study. These data are used to choose informative priors for the intercept parameters $\beta_1 = (\beta_{11}, \ldots, \beta_{15})$ and $\beta_3 = (\beta_{31}, \ldots, \beta_{35})$, and the mixture parameters Ω_{11}, Ω_{12} and Ω_2 by fitting the above model to historical control data. The prior mean for β is set to

$$\mu_0 = (0.54, -2.05, -2.34, -1.68, -1.13, 0.00, -0.70, -0.80, -1.00, -0.90, -0.96, 0.00),$$

and the prior covariance matrix is set to

$$\Sigma_0 = \text{diag}(1, 1, 1, 1, 1, 1000, 0.05, 0.05, 0.05, 0.05, 0.05, 1000).$$

Independent gamma priors are specified for the precision parameters: $\Omega_{11} \sim$ gamma(0.86, 0.30), $\Omega_{12} \sim$ gamma(8.15, 1.00) and $\Omega_2 \sim$ gamma(6.35, 0.30). The ease with which prior information can be incorporated is an attractive feature of the Bayesian approach.

4.3. The analysis

I used a Gibbs sampler within BUGS (Gilks *et al.*, 1994) to generate samples from the joint posterior distribution of the parameters and the latent variables. I ran three chains with widely varying initial values and used 10000 Gibbs iterates collected after convergence from each chain to compute posterior summaries of the parameters. Convergence of the Gibbs sampler was assessed by using several convergence diagnostic procedures recommended by Cowles and Carlin (1996). Posterior summaries of the global parameters for each outcome are shown in Table 1.

It appears that dams in the exposed group tend to have a longer time to birth and fewer pups per litter than unexposed dams. I evaluated the robustness of this result to the choice of the prior and the latent variable structure by repeating the analysis using reasonable alternative priors and latent variable distributions. For a prior with the variance inflated by a factor of 10, the posterior means were within 5% of the estimates shown in Table 1 and the standard deviation was only slightly inflated. Reasonable modifications of the prior means and of the assumed latent variable structure also had little effect on the conclusions of the analysis.

Table 1. Posterior summaries of the parameters for the reproductive toxicity example

Outcome	*Parameter*	*Mean*	*Median*	*Standard deviation*	*90% credible interval*
Time to birth	β_{11}	−0.381	−0.380	0.245	(−0.79, −0.07)
Time to birth	β_{12}	−1.364	−1.357	0.268	(−1.82, −1.03)
Time to birth	β_{13}	−0.864	−0.860	0.254	(−1.29, −0.54)
Time to birth	β_{14}	−0.725	−0.720	0.258	(−1.16, −0.40)
Time to birth	β_{15}	−0.669	−0.664	0.275	(−1.12, −0.32)
Time to birth	β_2	1.741	1.733	0.244	(1.35, 2.06)†
Litter size	β_{31}	−0.825	−0.825	0.087	(−0.97, −0.68)
Litter size	β_{32}	−0.830	−0.829	0.087	(−0.98, −0.69)
Litter size	β_{33}	−0.882	−0.882	0.088	(−1.03, −0.74)
Litter size	β_{34}	−0.935	−0.935	0.090	(−1.08, −0.79)
Litter size	β_{35}	−0.799	−0.798	0.094	(−0.95, −0.65)
Litter size	β_4	0.192	0.193	0.104	(0.02, 0.36)‡

†Exposure appears to extend the time between births.
‡Exposure appears to cause a reduction in litter size.

5. Discussion

I have proposed a flexible approach for Bayesian modelling of mixed outcome data. By using a mixture of generalized linear models to describe the joint distribution of a set of underlying variables, the model accommodates a broad class of data structures. Within the framework proposed, novel models were developed for joint binary, unordered categorical and continuous outcomes. These models can be easily adapted to accommodate ordered categorical outcomes by including threshold parameters to link each ordinal outcome to an underlying variable.

The MCMC algorithm is straightforward to implement within freely available software for Bayesian inference using Gibbs sampling (BUGS; Gilks *et al.* (1994); available from `www.mrc-bsu.cam.ac.uk/bugs`) when the likelihood is log-concave (Gilks and Wild, 1992). This approach was used in the example, and the computing time for 11000 Gibbs iterates was approximately 10 min on a Dec Alpha workstation. In models with correlated latent variables, the computing time is substantially longer. Special software is needed to implement the MCMC algorithm when the likelihood is not log-concave.

Standard approaches can be used for assessing convergence (Cowles and Carlin, 1996) and for model checking (Gelman *et al.*, 1996) and comparisons (Chib, 1995; DiCiccio *et al.*, 1997). In my experience, the Gibbs sampler tends to converge quickly for models with a relatively small number of independent latent variables (i.e. $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ diagonal and of low dimension). However, for models with multiple correlated latent variables convergence may be slow, as is the case in standard multiple random-effects models (Vines *et al.*, 1996). Convergence can be improved by constraining the latent variables to sum to 0, or potentially by reparameterizing the model. This is an area for future research.

For clarity, I have assumed that an equal number of outcomes are measured for each subunit. It is straightforward to extend the model to accommodate missing outcomes and more complex data structures with varying numbers and types of measurements. A particularly interesting case occurs when measurements are recorded on multiple levels (e.g. for parents and for each child). In such a case the outcomes on one level can easily be linked to the outcomes on another level through shared and/or correlated latent variables.

Because of the Bayesian formulation it is easy to include prior information, which is often available from historical studies with similar designs to the current study. The mixture model proposed in this paper can easily be extended to account for heterogeneity among studies. Such an extension would allow historical control data from several studies to be included in a single analysis.

Acknowledgements

I would like to thank Dr H. Zhou for his critical reading of the manuscript.

References

Albert, J. H. and Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Ass.*, **88**, 669–679.

Arminger, G. and Küsters, U. (1988) Latent trait models with indicators of mixed measurement level. In *Latent Trait and Latent Class Models* (eds R. Langeheine and J. Rost), pp. 51–73. New York: Plenum.

Breslow, N. E. and Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

Catalano, P. J. and Ryan, L. M. (1992) Bivariate latent variable models for clustered discrete and continuous outcomes. *J. Am. Statist. Ass.*, **87**, 651–658.

Chen, J. (1993) A malformation incidence dose-response model incorporating fetal weight and/or litter size as covariates. *Risk Anal.*, **13**, 559–564.
Chen, M. H. and Dey, D. K. (1998) Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions. *Technical Report 98*. Department of Statistics, University of Connecticut, Storrs.
Chib, S. (1995) Marginal likelihood from the Gibbs output. *J. Am. Statist. Ass.*, **90**, 1313–1321.
Chib, S. and Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.
Chu, H. M. and Kuo, L. (1997) Sampling based approach for one-hit and multi-hit models in quantal bioassay. *Statist. Comput.*, **7**, 183–192.
Cowles, M. K. and Carlin, B. P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Statist. Ass.*, **91**, 883–904.
Cox, D. R. and Wermuth, N. (1992) Response models for mixed binary and quantitative variables. *Biometrika*, **79**, 441–461.
Dey, D. K. and Chen, M. H. (1996) Bayesian analysis of correlated binary data models. *Technical Report 96-02*. Department of Statistics, University of Connecticut, Storrs.
DiCiccio, T. J., Kass, R. E., Raftery, A. and Wasserman, L. (1997) Computing Bayes factors by combining simulation and asymptotic approximations. *J. Am. Statist. Ass.*, **92**, 903–915.
Fitzmaurice, G. M. and Laird, N. M. (1995) Regression models for a bivariate discrete and continuous outcome with clustering. *J. Am. Statist. Ass.*, **90**, 845–852.
Gelman, A., Meng, X. L. and Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sin.*, **6**, 733–760.
Gilks, W. R., Thomas, A. and Spiegelhalter, D. J. (1994) A language and program for complex Bayesian modelling. *Statistician*, **43**, 169–177.
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. *Appl. Statist.*, **41**, 337–348.
Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.
Kleinman, K. P. and Ibrahim, J. G. (1998) A semi-parametric Bayesian approach to generalized linear mixed models. *Statist. Med.*, **17**, 2579–2596.
Legler, J. M. and Ryan, L. M. (1997) Latent variable models for teratogenesis using multiple binary outcomes. *J. Am. Statist. Ass.*, **92**, 13–20.
Moustaki, I. (1996) A latent trait and a latent class model for mixed observed variables. *Br. J. Math. Statist. Psychol.*, **49**, 313–334.
Moustaki, I. and Knott, M. (1999) Generalized latent trait models. *Psychometrika*, to be published.
National Toxicology Program (1988) Ethylene glycol monomethyl ether reproduction and fertility assessment in CD-1 mice when administered in drinking water. *NIH Publ. 88211446*. US Department of Health and Human Services, Public Health Service, National Institutes of Health, Research Triangle Park.
Rai, K. and Van Ryzin, J. (1981) A generalized multihit dose-response model for low-dose extrapolation. *Biometrics*, **37**, 341–352.
Sammel, M. D., Ryan, L. M. and Legler, J. M. (1997) Latent variable models for mixed discrete and continuous outcomes. *J. R. Statist. Soc.* B, **59**, 667–678.
Shi, J. Q. and Lee, S. Y. (1998) Bayesian sampling-based approach for factor analysis models with continuous and polytomous data. *Br. J. Math. Statist. Psychol.*, **51**, 233–252.
——— (2000) Latent variable models with mixed continuous and polytomous data. *J. R. Statist. Soc.* B, **62**, 77–87.
Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.*, **22**, 1701–1762.
Vines, S. K., Gilks, W. R. and Wild, P. (1996) Bayesian multiple random effects models. *Statist. Comput.*, **6**, 337–346.
Williams, D. A. (1987) Reader reaction: dose-response models for teratological experiments. *Biometrics*, **43**, 1013–1016.
Zeger, S. L. and Karim, M. R. (1991) Generalized linear models with random effects: a Gibbs sampling approach. *J. Am. Statist. Ass.*, **86**, 79–86.