

# Bayesian Joint Models of Cluster Size and Subunit-Specific Outcomes

David B. Dunson<sup>1,\*</sup>, Zhen Chen<sup>1</sup>, and Jean Harry<sup>2</sup>

<sup>1</sup>Biostatistics Branch and <sup>2</sup> Laboratory of Molecular Toxicology,  
MD A3-03, National Institute of Environmental Health Sciences,  
P.O. Box 12233, Research Triangle Park, NC 27709

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

**SUMMARY.** In applications that involve clustered data, such as longitudinal studies and developmental toxicity experiments, the number of subunits within a cluster is often correlated with outcomes measured on the individual subunits. Analyses that ignore this dependency can produce biased inferences. This article proposes a Bayesian framework for jointly modeling cluster size and multiple categorical and continuous outcomes measured on each subunit. We use a continuation ratio probit model for the cluster size and underlying normal regression models for each of the subunit-specific outcomes. Dependency between cluster size and the different outcomes is accommodated through a latent variable structure. The form of the model facilitates posterior computation via a simple and computationally efficient Gibbs sampler. The approach is illustrated through application to developmental toxicity data, and applications to joint modeling of longitudinal and event time data are discussed.

**KEY WORDS:** Continuation ratio; Developmental toxicity; Factor analysis; Informative cluster size; Litter size; Multiple outcomes; Probit model; Random length data.

## 1. Introduction

In longitudinal follow-up studies, in developmental toxicity experiments, and in other applications involving clustered data, the number of subunits within a cluster is often correlated with outcomes measured on the individual subunits. For example, in rodent teratology studies, pregnant dams are typically randomized to dose groups for treatment with a test agent during organogenesis, the major period of organ development in the growing fetus. It is well known that fetal outcomes, such as birth weight, can be highly dependent on litter size and that this association can result in biased inferences about dose effects on fetus-specific outcomes if not accounted for (cf., Romero et al., 1992; Tanaka, 1998). Thus, considering litters as clusters and fetuses as subunits nested within clusters, developmental toxicology studies provide a clear example of informative cluster size. As opposed to earlier conditional approaches that account for this problem by incorporating cluster size as a covariate in a model for the subunit-level outcomes (cf., Catalano and Ryan, 1992; Chen, 1993; Romero et al., 1992; Ryan and Molenberghs, 1999), we propose to model cluster size and the subunit-level outcomes jointly.

Within the past decade, there has been substantial interest in the problem of joint analysis of clustered multiple outcome data, motivated largely by reproductive and developmental toxicity applications (for some recent references, refer to Dunson, 2000; Fitzmaurice and Laird, 1995; Gueorguieva and Agresti, 2001; Molenberghs and Ryan, 1999; Regan and Catalano, 1999). Almost all of the available methods have assumed that the response distribution for the individual subunits is independent of the cluster size, though cluster size can sometimes be incorporated as a covariate. To our knowledge, methods that model the cluster size distribution (cf., Chen, 1993) have conditioned on the cluster size in modeling the subunit-level outcomes. Such conditional models do not capture the covariate-dependent associations between cluster size and the subunit-level outcomes, which are common in reproductive and longitudinal studies.

To conceptualize this problem, consider an untreated female mouse (i.e., dam) who is mated with a healthy male. Given the right experimental conditions, there is a high probability that conceptions will occur, implantation sites will form, and many of the implants will grow into healthy fetuses that survive to birth. On average, if fewer fetuses are produced in a litter, then there will be more space and nutritional resources available, resulting in an inverse association between litter size and fetal weight. Now suppose that the dam is exposed to a known developmental toxicant prior to mating and throughout gestation. Depending on the mechanism of action of the compound, the dose, and on random characteristics of the dam and of the pregnancy, there may be no effect, there may be reductions in fetal weight without increases in resorptions or fetal losses, or some of the fetuses that would have survived in the absence of exposure may be spontaneously aborted. If some of the unobservable dam and fetus-specific factors contributing to exposure-induced fetal loss are also associated with fetal weight, as seems likely, then analyses that condition on litter size can produce misleading inferences about the dose effect on fetal weight. In general, it is important to model litter size and the fetal outcomes jointly since they are so closely interlinked biologically.

We propose a Bayesian modeling framework for joint analysis of cluster size and multiple categorical and continuous subunit-level outcomes. To model cluster size, we use a continuation-ratio ordinal response model (cf., Albert and Chib, 2001; Ten Have and Utal, 1994), which is a convenient form for incorporating covariate effects without restricting the baseline cluster size distribution. The multiple outcomes measured for the individual subunits are then assigned an underlying normal model that accounts for dependency between outcomes having different measurement scales by incorporating shared latent variables within outcome-specific regression models (cf., Arminger and Küsters, 1988; Dunson, 2000). We allow the same latent variables to load on cluster size and the subunit-level outcomes in order to flexibly accommodate dependency. As we demonstrate, the resulting model can be

expressed in the form of a factor analytic regression model for a vector of underlying cluster-specific variables. Due to the structure of this model, posterior computation can proceed quickly and easily using a Gibbs sampling algorithm (Gelfand and Smith, 1990; Chen, Shao and Ibrahim, 2000), which updates the regression parameters and factor loadings for the different outcomes jointly.

This methodology can also be applied to the problem of joint modeling of longitudinal and event time data, which has received substantial interest lately in the statistical literature (cf., Berzuini and Larizza, 1996; Faucett and Thomas, 1996; Hogan and Laird, 1997; Wang and Taylor, 2001; Wulfsohn and Tsiatis, 1997, among others). In many epidemiologic studies, progression to onset of disease is monitored using disease markers, and the periodically measured markers are subject to measurement error. Methods for jointly modeling the marker process and the time to onset of disease are needed to avoid bias. Most of the literature in this area has focused on simple, univariate Gaussian random effects models for the marker process, though more flexible stochastic models have been developed recently (cf., Wang and Taylor, 2001). Considering the discretized event time as a cluster size and longitudinal observations as subunits, the modeling approach proposed in this paper can be used for jointly modeling an event time and multidimensional longitudinal data having mixed measurement scales. Gray and Brookmeyer (2000) considered a related data structure but used a different approach that focused on combining information about a treatment effect.

Section 2 proposes the model and prior formulation, and discusses properties. Section 3 describes the Gibbs sampling approach. Section 4 illustrates the methods through application to data from a developmental toxicity study, and Section 5 discusses the results.

## **2. The Model**

### *2.1 General Framework*

Let  $i$  index the cluster and let  $j$  index the subunit within a cluster. For cluster  $i$  ( $i =$

$1, \dots, n$ ), let  $s_i$  denote the number of subunits and let  $\mathbf{y}_{ij} = (y_{ijk}, k = 1, \dots, p)^T$  denote the  $p$ -dimensional outcome vector for subunit  $j$  ( $j = 1, \dots, s_i$ ). We focus on the case in which the subunit-level outcomes are either binary or continuous, though the extension to categorical outcomes is straightforward. Therefore, without loss of generality, we assume  $y_{ijk}$  is continuous for  $k = 1, \dots, p_c$  and is binary for  $k = p_c + 1, \dots, p$ .

Following Muthén (1984), we link the observed outcome  $y_{ijk}$  to an underlying normal variable  $y_{ijk}^*$  through  $y_{ijk} = g_k(y_{ijk}^*)$ , where  $g_k(y^*) = I(y^* > 0)$  if the  $k$ th outcome is binary and  $g_k(\cdot)$  is a known one-to-one function (e.g., identity, log) if the  $k$ th outcome is continuous. We use a multilevel regression model for the underlying variables  $\mathbf{y}_{ij}^* = (y_{ij1}^*, \dots, y_{ijp}^*)^T$ ,

$$\mathbf{y}_{ij}^* = \boldsymbol{\mu} + \boldsymbol{\alpha}\mathbf{x}_i + \boldsymbol{\Lambda}\mathbf{W}_i\boldsymbol{\xi}_i + \boldsymbol{\Gamma}\mathbf{W}_i\boldsymbol{\eta}_{ij} + \boldsymbol{\epsilon}_{1ij}, \quad (1)$$

where  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^T$  are outcome-specific intercept parameters,  $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_p]^T$  is a  $p \times q$  matrix of regression coefficients,  $\mathbf{x}_i$  is a  $q \times 1$  vector of cluster-level covariates,  $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_p]^T$  and  $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p]$  are  $p \times r$  factor loadings matrices,  $\mathbf{W}_i = \text{diag}(w_{i1}, \dots, w_{ir})$  is a known diagonal weighting matrix,  $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{ir})^T$  and  $\boldsymbol{\eta}_{ij} = (\eta_{ij1}, \dots, \eta_{ijr})^T$  are vectors of iid  $N(0,1)$  cluster- and subunit-level latent variables, respectively,  $\boldsymbol{\epsilon}_{1ij} = (\epsilon_{1ij1}, \dots, \epsilon_{1ijp})^T \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$  are error residuals with  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ , and the elements of  $\boldsymbol{\xi}_i$ ,  $\boldsymbol{\eta}_{ij}$ , and  $\boldsymbol{\epsilon}_{1ij}$  are independent. Expression (1) accommodates within-cluster and within-subject dependency among outcomes having different measurement scales by incorporating shared latent variables in linear regression models for the underlying normal variables. When the weighting matrix  $\mathbf{W}_i$  has elements of  $\mathbf{x}_i$  along the diagonal, the model can accommodate covariate-dependent associations among the outcome variables in a similar manner to standard mixed effects models (Laird and Ware, 1982).

To model the cluster size  $s_i$  jointly with the subunit-level outcomes  $\mathbf{y}_i = (\mathbf{y}_{i1}^T, \dots, \mathbf{y}_{is_i}^T)^T$ , we let

$$\Pr(s_i = s \mid s_i \geq s, \mathbf{x}_i, \boldsymbol{\xi}_i) = F(\delta_s - \mathbf{x}_i^T \boldsymbol{\beta} - \boldsymbol{\lambda}_0^T \mathbf{W}_i \boldsymbol{\xi}_i) \quad \text{for } s = 1, \dots, T - 1, \quad (2)$$

where  $F(\cdot)$  is a known monotone 1-1 function mapping  $\mathfrak{R} \rightarrow [0, 1]$  (e.g., the standard normal distribution function,  $\Phi$ ),  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{T-1})^T \in \mathfrak{R}^{T-1}$  are parameters characterizing the baseline cluster size distribution given  $\mathbf{x}_i = \mathbf{0}$  and  $\boldsymbol{\xi}_i = \mathbf{0}$ ,  $\boldsymbol{\beta}$  are regression parameters, and  $\boldsymbol{\lambda}_0$  is a  $r \times 1$  vector of factor loadings. The cluster size is assumed to have support  $\{1, \dots, T\}$ . Expressions (1) and (2) flexibly accommodate dependency between  $s_i$  and  $\mathbf{y}_i$  through incorporation of shared latent variables. This framework allows the correlation between cluster size and the different subunit-level outcomes to vary, while reducing the dimensionality of the covariance structure. Expression (2) is in the form of a continuation ratio model for ordered response data (cf., Agresti, 1990, Chapter 9). This model does not restrict the baseline cluster size distribution, as would a typical count response model, such as a Poisson or negative binomial.

We complete a Bayesian specification of the model with prior distributions for the parameters. Letting  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$  with  $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}^T, \boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_p^T, \boldsymbol{\lambda}_1^T, \dots, \boldsymbol{\lambda}_p^T, \boldsymbol{\gamma}_1^T, \dots, \boldsymbol{\gamma}_p^T)^T$  denoting the regression parameters in expression (1) and  $\boldsymbol{\theta}_2 = (\boldsymbol{\delta}^T, \boldsymbol{\beta}^T, \boldsymbol{\lambda}_0^T)^T$  denoting the parameters in expression (2), we choose

$$\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, \boldsymbol{\Psi}_0) \quad \text{and} \quad \sigma_k^{-2} \sim \text{gamma}(c_{0k}, d_{0k}) \quad \text{for } k = 1, \dots, p_c, \quad (3)$$

where  $\boldsymbol{\theta}_0$ ,  $\boldsymbol{\Psi}_0$ ,  $c_{0k}$  and  $d_{0k}$  are investigator-specified hyperparameters, and  $\sigma_k = 1$  for  $k = p_c + 1, \dots, p$  for identifiability. These conditionally conjugate forms result in simplified posterior computation, as we describe in Section 3.

## 2.2 Univariate Normal Case

To illustrate the properties of the model, we focus first on the case where  $p = 1$ ,  $g_1(\cdot)$  is the identity link,  $r = 1$  and  $\mathbf{W}_i = 1$ . For example, pup weights may be available for litters of mice. In this case, expressions (1) and (2) imply that

$$y_{ij} = \mu + \mathbf{x}_i^T \boldsymbol{\alpha} + \lambda_1 \xi_i + \epsilon_{1ij}$$

and

$$\Pr(s_i = s | \mathbf{x}_i, \xi_i) = F(\delta_s - \mathbf{x}_i^T \boldsymbol{\beta} - \lambda_0 \xi_i) \prod_{t=1}^{s-1} \left\{ 1 - F(\delta_t - \mathbf{x}_i^T \boldsymbol{\beta} - \lambda_0 \xi_i) \right\}, \quad (4)$$

where  $\lambda_1 \geq 0$  and subunit-level effects are excluded to ensure identifiability. The correlation coefficient between  $y_{ij}$  and  $y_{ij'}$  (e.g., the weights for two different pups in the same litter) is

$$\rho(y_{ij}, y_{ij'}) = \frac{\text{cov}(y_{ij}, y_{ij'})}{\sqrt{\text{var}(y_{ij}) \text{var}(y_{ij'})}} = \frac{\lambda_1^2}{\lambda_1^2 + \sigma_1^2},$$

which equals 0 for  $\lambda_1 = 0$ .

The correlation coefficient between  $y_{ij}$  and  $s_i$  (e.g., a pup's weight and the litter size) is

$$\rho(y_{ij}, s_i) = \frac{\text{cov}(y_{ij}, s_i)}{\sqrt{\text{var}(y_{ij}) \text{var}(s_i)}} = \frac{\lambda_1 \int_{-\infty}^{\infty} \xi_i \mathbb{E}(s_i | \xi_i) d\Phi(\xi_i)}{\sqrt{(\lambda_1^2 + \sigma_1^2) \int_{-\infty}^{\infty} \text{var}(s_i | \xi_i) d\Phi(\xi_i)}}, \quad (5)$$

where the expectation and variance of the cluster size  $s_i$  conditional on  $\xi_i$  are, respectively,

$$\mathbb{E}(s_i | \mathbf{x}_i, \xi_i) = \sum_{s=1}^T s \Pr(s_i = s | \mathbf{x}_i, \xi_i) \quad (6)$$

$$\text{var}(s_i | \mathbf{x}_i, \xi_i) = \sum_{s=1}^T s^2 \Pr(s_i = s | \mathbf{x}_i, \xi_i) - [\mathbb{E}(s_i | \mathbf{x}_i, \xi_i)]^2. \quad (7)$$

It follows from (5) that  $\rho(y_{ij}, s_i) = 0$  if  $\lambda_1 = 0$  (i.e., there is no within-cluster dependency in the subunit-level outcomes). The correlation  $\rho(y_{ij}, s_i)$  also equals 0 if  $\lambda_0 = 0$ , since the expression for  $\mathbb{E}(s_i | \xi_i)$  does not involve  $\xi_i$  in this case, and hence  $\text{cov}(y_{ij}, s_i) = \lambda_1 \mathbb{E}(s_i) \int_{-\infty}^{\infty} \xi_i d\Phi(\xi_i) = \lambda_1 \mathbb{E}(s_i) \mathbb{E}(\xi_i) = 0$ . For  $\lambda_1 > 0$  and  $\lambda_0 > 0$ , the correlation is positive:  $\rho(y_{ij}, s_i) > 0$ , while for  $\lambda_1 > 0$  and  $\lambda_0 < 0$ , the correlation is negative:  $\rho(y_{ij}, s_i) < 0$ . In general, it is easy to calculate  $\rho(y_{ij}, s_i)$  from expression (5) using numeric integration.

### 2.3 Multivariate Normal Case

Now consider the more general case where  $g_k(\cdot)$  is the identity for all  $k$ , but no restrictions are placed on  $p$ ,  $r$  or  $\mathbf{W}_i$ . In this case, the covariance matrix of  $\mathbf{y}_{ij}$  is

$$\text{var}(\mathbf{y}_{ij}) = \boldsymbol{\Lambda} \mathbf{W}_i \mathbf{W}_i \boldsymbol{\Lambda}^T + \boldsymbol{\Gamma} \mathbf{W}_i \mathbf{W}_i \boldsymbol{\Gamma}^T + \boldsymbol{\Sigma}.$$

Under expression (1), observations from different clusters are uncorrelated:  $\text{cov}(\mathbf{y}_{ij}, \mathbf{y}_{i'j}) = \mathbf{0}$ , while observations for two different subunits in the same cluster have  $\text{cov}(\mathbf{y}_{ij}, \mathbf{y}_{i'j'}) = \mathbf{\Lambda} \mathbf{W}_i \mathbf{W}_i \mathbf{\Lambda}^T$ . In addition, two different outcomes on the same subunit have  $\text{cov}(y_{ijk}, y_{ijk'}) = \boldsymbol{\lambda}_k^T \mathbf{W}_i \mathbf{W}_i \boldsymbol{\lambda}_{k'} + \boldsymbol{\gamma}_k^T \mathbf{W}_i \mathbf{W}_i \boldsymbol{\gamma}_{k'}$ , where  $\boldsymbol{\lambda}_k$  and  $\boldsymbol{\gamma}_k$  are the  $k$ th row vectors of  $\mathbf{\Lambda}$  and  $\mathbf{\Gamma}$ , respectively.

The structure of expressions (1) and (2) is flexible enough to allow the correlation between the subunit-level outcomes and the cluster size to depend on outcome type. In particular,

$$\rho(y_{ijk}, s_i) = \frac{\int \boldsymbol{\lambda}_k^T \mathbf{W}_i \boldsymbol{\xi}_i \text{E}(s_i | \mathbf{x}_i, \boldsymbol{\xi}_i) d\Phi(\xi_{i1}) \cdots d\Phi(\xi_{ir})}{\sqrt{(\boldsymbol{\lambda}_k^T \mathbf{W}_i \mathbf{W}_i \boldsymbol{\lambda}_k + \boldsymbol{\gamma}_k^T \mathbf{W}_i \mathbf{W}_i \boldsymbol{\gamma}_k + \sigma_k^2) \int \text{var}(s_i | \mathbf{x}_i, \boldsymbol{\xi}_i) d\Phi(\xi_{i1}) \cdots d\Phi(\xi_{ir})}}, \quad (8)$$

where the conditional expectation and variance,  $\text{E}(s_i | \mathbf{x}_i, \boldsymbol{\xi}_i)$  and  $\text{var}(s_i | \mathbf{x}_i, \boldsymbol{\xi}_i)$ , can be calculated using a direct extension of expression (7). In general, since  $\text{E}(s_i | \mathbf{x}_i, \boldsymbol{\xi}_i)$  is non-linear in the latent variables  $\xi_{i1}, \dots, \xi_{ir}$ , the  $k$ th subunit-level outcome is uncorrelated with the cluster size (i.e.,  $\rho(y_{ijk}, s_i) = 0$ ) only if  $\lambda_{kl} \lambda_{0l} = 0$  for  $l = 1, \dots, r$ .

It is of interest to derive the bias induced by ignoring cluster size, and basing inferences about covariate effects on the subunit-level outcomes only on expression (1), without factoring expression (2) into the likelihood. Note that conditional on  $\mathbf{x}_i = \mathbf{x}$ , the marginal expectation of  $y_{ijk}$  is

$$\begin{aligned} \text{E}(y_{ijk} | \mathbf{x}_i = \mathbf{x}) &= \sum_{s=1}^T \int \text{E}(y_{ijk} | \mathbf{x}, \boldsymbol{\xi}_i) \text{Pr}(s_i = s | \mathbf{x}, \boldsymbol{\xi}_i) d\Phi(\xi_{i1}) \cdots d\Phi(\xi_{ir}) \\ &= \mu_k + \mathbf{x}^T \boldsymbol{\alpha}_k + \sum_{s=1}^T \int \boldsymbol{\lambda}_k^T \mathbf{W}_i \boldsymbol{\xi}_i \text{Pr}(s_i = s | \mathbf{x}, \boldsymbol{\xi}_i) d\Phi(\xi_{i1}) \cdots d\Phi(\xi_{ir}), \end{aligned} \quad (9)$$

integrating out cluster size and the latent variables. In contrast, the model that ignores cluster size has  $\text{E}(y_{ijk} | \mathbf{x}_i = \mathbf{x}) = \mu_k + \mathbf{x}^T \boldsymbol{\alpha}_k$ . Hence, the third term in the above expression represents the bias in the marginal expectation due to ignoring cluster size. In general, the bias will equal 0 only if  $\lambda_{kl} \lambda_{0l} = 0$  for  $l = 1, \dots, r$ , which is the same condition for  $\rho(y_{ijk}, s_i) = 0$ . If  $\boldsymbol{\lambda}_0 = \mathbf{0}$ , the joint likelihood  $L(\mathbf{y}_i, s_i; \boldsymbol{\theta}, \boldsymbol{\xi}_i, \boldsymbol{\eta}_i)$  conditional on  $\boldsymbol{\xi}_i$  and  $\boldsymbol{\eta}_i$  can be factored into the product of the subunit-level likelihood:  $L_1(\mathbf{y}_i; \boldsymbol{\alpha}, \mathbf{\Lambda}, \mathbf{\Gamma}, \boldsymbol{\xi}_i, \boldsymbol{\eta}_i)$  and

the cluster size likelihood:  $L_2(s_i; \boldsymbol{\delta}, \boldsymbol{\beta})$ . Otherwise, cluster size is non-ignorable and it is necessary to use a joint model to avoid biased inferences.

#### 2.4 General Case

In the general case where  $g_k(\cdot)$  is a link function other than the identity for some  $k$ , it is non-trivial to obtain general expressions for the covariance. However, given the structure of the model, we can simplify inferences about the covariance structure by focusing on the underlying normal outcomes  $\mathbf{y}_{ij}^*$  instead of the observed outcomes  $\mathbf{y}_{ij}$ . This strategy is standard practice in the structural equations modeling literature when outcomes are both categorical and continuous (cf., Muthén, 1984). It also allows us to utilize the simple expressions derived above.

#### 2.5 Identifiability Considerations

As in standard factor analytic models, it is necessary to place constraints on the variance components  $\boldsymbol{\Lambda}$ ,  $\boldsymbol{\Gamma}$  and  $\boldsymbol{\Sigma}$  to ensure identifiability of the model based on the data, without relying on informative priors. In the special case where  $\mathbf{W}_i = \mathbf{I}_{r \times r}$  and  $\boldsymbol{\lambda}_0 = 0$ , expression (1) is in the form of a standard multi-level factor analytic regression model, and the typical identifiability constraints apply (cf., Bartholomew, 1987). In particular, for all  $k$  such that  $y_{ijk}$  is categorical, the variance  $\sigma_k^2$  is not uniquely identified, so the standard convention sets  $\sigma_k = 1$ . In addition, appropriate restrictions need to be imposed on  $\boldsymbol{\Lambda}$  and  $\boldsymbol{\Gamma}$  so that  $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T$  and  $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$  are uniquely defined from  $\boldsymbol{\Lambda}$  and  $\boldsymbol{\Gamma}$ , respectively. For example, for  $p \geq 3$ , identifiability can be achieved by setting the upper diagonal elements of  $\boldsymbol{\Lambda}$  and  $\boldsymbol{\Gamma}$  equal to 0, restricting the diagonal elements to be positive, and choosing  $r \leq p/2$ .

In the general case, the diagonal elements of  $\mathbf{W}_i$  consist of 1s and elements of the covariate vector  $\mathbf{x}_i$ . By choosing a particular covariate as the  $l$ th diagonal element of  $\mathbf{W}_i$ , one causes the level of expression of the  $l$ th latent factor to be dependent on the level of that covariate. Incorporation of  $\mathbf{W}_i$  changes the necessary identifiability restrictions on  $\boldsymbol{\Lambda}$  and  $\boldsymbol{\Gamma}$ .

In general, in the presence of  $\mathbf{W}_i$ , one should apply the standard identifiability rules to the submatrices  $\mathbf{\Lambda}_l$  and  $\mathbf{\Gamma}_l$ ,  $l = 1, \dots, m$ , defined by choosing the rows of  $\mathbf{\Lambda}$  and  $\mathbf{\Gamma}$ , respectively, having common diagonal elements of  $\mathbf{W}_i$ . Note that  $m$  ( $m \leq q$ ) refers to the number of elements of  $\mathbf{x}_i$  included in  $\mathbf{W}_i$ .

### 3. Posterior Computation

#### 3.1 Underlying Normal Structure

As described in Section 2.1, the subunit-level outcomes  $\mathbf{y}_{ij}$  are linked to underlying normally distributed outcomes  $\mathbf{y}_{ij}^*$ . In addition, taking  $F(\cdot)$  in expression (2) to be the standard normal distribution function, we can link the cluster size  $s_i$  to a vector of normally distributed underlying variables  $\mathbf{s}_i^* = (s_{i1}^*, \dots, s_{it_i}^*)^T$ , where  $t_i = \min(s_i, T - 1)$ . In particular, note that the cluster size distribution characterized by expression (2) is stochastically equivalent to generating a sequence of random variables  $s_{ij}^* \sim N(\delta_j - \mathbf{x}_i^T \boldsymbol{\beta} - \boldsymbol{\lambda}_0^T \mathbf{W}_i \boldsymbol{\xi}_i, 1)$  for  $j = 1, \dots, T - 1$ , and then letting  $s_i = \min\{j : s_{ij}^* > 0; j = 1, \dots, T - 1\}$  if  $\max(s_{i1}^*, \dots, s_{it_i}^*) > 0$  or  $s_i = T$  otherwise. It follows that underlying  $s_i$  we have  $s_{i1}^*, \dots, s_{it_i}^*$ , where  $s_{ij}^* \sim N(\delta_j - \mathbf{x}_i^T \boldsymbol{\beta} - \boldsymbol{\lambda}_0^T \mathbf{W}_i \boldsymbol{\xi}_i, 1)$  truncated above by 0 for  $j < s_i$  and truncated below by 0 for  $j = s_i$ .

Given the definition of  $\boldsymbol{\theta}$  in Section 2.1, expression (1) is equivalent to  $\mathbf{y}_{ij}^* = \mathbf{U}_{1ij} \boldsymbol{\theta} + \boldsymbol{\epsilon}_{1ij}$ , where  $\mathbf{U}_{1ij}$  is a  $p \times \{p(1 + q + 2r) + T - 1 + q + r\}$  matrix with its  $k$ th row vector  $\mathbf{u}_{1ijk}$ :

$$(\mathbf{0}_{(k-1)}^T, 1, \mathbf{0}_{(p-k)}^T, \mathbf{0}_{(k-1)q}^T, \mathbf{x}_i^T, \mathbf{0}_{(p-k)q}^T, \mathbf{0}_{(k-1)r}^T, \boldsymbol{\xi}_i^T \mathbf{W}_i, \mathbf{0}_{(p-k)r}^T, \mathbf{0}_{(k-1)r}^T, \boldsymbol{\eta}_{ij}^T \mathbf{W}_i, \mathbf{0}_{(p-k)r}^T, \mathbf{0}_{(T-1+q+r)}^T)^T$$

for  $k = 1, \dots, p$ . Similarly, the model for the underlying variables of the cluster size is equivalent to  $s_{ij}^* = \mathbf{u}_{2ij}^T \boldsymbol{\theta} + \boldsymbol{\epsilon}_{2ij}$ , where

$$\mathbf{u}_{2ij} = (\mathbf{0}_{p(1+q+2r)}^T, \mathbf{0}_{(j-1)}^T, 1, \mathbf{0}_{(T-j-1)}^T, -\mathbf{x}_i^T, -\boldsymbol{\xi}_i^T \mathbf{W}_i)^T,$$

and  $\boldsymbol{\epsilon}_{2ij} \sim N(0, 1)$  for  $j = 1, \dots, t_i$ . Thus, letting  $\mathbf{z}_i = (\mathbf{y}_i^{*T}, \mathbf{s}_i^{*T})^T$  and

$$\mathbf{U}_i = [\mathbf{U}_{1i1}^T, \dots, \mathbf{U}_{1is_i}^T, \mathbf{u}_{2i1}^T, \dots, \mathbf{u}_{2it_i}^T]^T,$$

we have  $\mathbf{z}_i \sim \mathcal{N}(\mathbf{U}_i\boldsymbol{\theta}, \mathbf{A}_i)$ , where

$$\mathbf{A}_i = \begin{bmatrix} \boldsymbol{\Sigma} \otimes \mathbf{I}_{s_i \times s_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{t_i \times t_i} \end{bmatrix},$$

with  $\otimes$  the kronecker product operator. Due to this underlying normal linear model structure, posterior computation can be implemented easily using Gibbs sampling.

### 3.2 Gibbs Sampler

Gibbs sampling can be used to generate samples from the joint posterior density of the unknowns in the model. Bayesian inference can then be based on posterior summaries estimated from these samples. After choosing the hyperparameters and initial values for  $\boldsymbol{\theta}$  and the latent variables, our Gibbs sampler proceeds as follows:

1. Sample the underlying normal variables  $\mathbf{z}_i$  from their full conditional posterior density given the data  $\mathbf{d}_i = (\mathbf{y}_i^T, s_i)^T$ ,  $\mathbf{x}_i$  and current values for  $\boldsymbol{\theta}$  and  $\{\boldsymbol{\xi}_i, \boldsymbol{\eta}_{i1}, \dots, \boldsymbol{\eta}_{is_i}\}$ .
2. Update  $\boldsymbol{\theta}$  by sampling from its full conditional distribution.
3. Update  $\boldsymbol{\xi}_i$ , for all  $i$ , by sampling from its full conditional distributions.
4. Update  $\boldsymbol{\eta}_{ij}$ , for all  $i, j$ , by sampling from its full conditional distributions.
5. Update  $\boldsymbol{\Sigma}$  by sampling from the full conditional distributions for  $\tau_1, \dots, \tau_{p_c}$ .
6. Repeat steps (1)-(5) until apparent convergence and calculate posterior summaries based on a large number of additional iterates.

The necessary conditional distributions are provided in an Appendix. Under mild regularity conditions, samples from this Gibbs sampling algorithm will converge to a stationary distribution that is the joint posterior. Due to the block sampling structure, the algorithm is expected to sample efficiently from the joint posterior relative to MCMC algorithms that update the parameters and latent variables one at a time.

## 4. Developmental Toxicology Application

### 4.1 NTP Ethylene Glycol Study and Previous Analyses

We illustrate the methodology through application to data from a developmental toxicity study of ethylene glycol in mice conducted by the National Toxicology Program (Price, Kimmel, Tyl, and Marr 1985). During organogenesis, the major period of organ development for the growing fetuses, pregnant mice (i.e., dams) were exposed to ethylene glycol at one of four different dose levels: 0, 0.75, 1.5, and 3 mg/kg. Data for each dam consisted of the litter size, the fetal weights for the individual pups, and binary indicators of the occurrence of malformations in each pup. Table 1 presents summary statistics of the data. Dose-related trends with respect to both fetus level outcomes are evident, with mean fetal weight decreasing with increasing dose and malformation rate increasing with increasing dose. There also appears to be a dose effect on litter size, with litter size decreasing monotonically with increasing dose. Figure 1 illustrates the inverse association between litter size and fetal weight.

These data were previously analyzed by, among others, Catalano and Ryan (1992), Molenberghs and Ryan (1999), and Gueorguieva and Agresti (2001). While all of these analyses jointly modeled fetal weight and malformation, Catalano and Ryan were the only ones to consider litter size, including it as a covariate in their model. As noted in Section 1, such conditional models can fail to capture the complex dependency structure and may result in biased inferences about dose effects on fetal weight and malformation. In addition, toxicologists are typically interested in assessing dose effects on all adverse outcomes, including reductions in litter size.

The main findings from Catalano and Ryan’s analysis of the ethylene glycol data were that the dose effect was negative for fetal weight and positive for the probability of malformation. These results are consistent with the summary statistics in Table 1. They also found that litter size was negatively associated with both fetal weight and malformation,

though the correlation with malformation was small and insignificant. A limitation of the Catalano and Ryan approach is that it cannot be used to estimate certain associations among the outcomes, such as the correlation between fetal weight and malformation for the same fetus. This limitation is due to their factorization of the joint distribution of fetal weight and malformation into the marginal distribution of weight and the conditional distribution of malformation given weight.

#### 4.2 Our Analysis

We now describe our analysis of the ethylene glycol data using the proposed joint model. Let  $s_i$  denote the size of litter  $i$ , and let  $y_{ij1}$  and  $y_{ij2}$  denote the fetal weight and malformation status, respectively, for the  $j$ th pup in litter  $i$ . Furthermore, let  $y_{ij2}^*$  denote a normal variable underlying  $y_{ij2}$  such that  $y_{ij2} = I(y_{ij2}^* > 0)$ . Our model can be specified as follows:

$$\begin{aligned} y_{ij1} &= \mu_1 + \alpha_1 x_i + \lambda_1 \xi_i + \gamma_1 \eta_{ij} + \epsilon_{1ij1}, \\ y_{ij2}^* &= \mu_2 + \alpha_2 x_i + \lambda_2 \xi_i + \gamma_2 \eta_{ij} + \epsilon_{1ij2}, \\ \Pr(s_i = s \mid x_i, \xi_i) &= \Phi(\delta_s - \beta x_i - \lambda_0 \xi_i) \prod_{t=1}^{s-1} \left\{ 1 - \Phi(\delta_t - \beta x_i - \lambda_0 \xi_i) \right\}, \end{aligned}$$

$s = 1, \dots, T - 1$ , where  $\mu_1$  and  $\mu_2$  are intercepts in the weight and malformation model, respectively, with  $\mu_1$  representing the expected weight for a typical pup in a typical litter of the control group,  $\alpha_1$  and  $\alpha_2$  characterize the effect of dose ( $x_i$ ) of ethylene glycol on fetal weight and the occurrence of malformations, respectively,  $\xi_i \sim N(0, 1)$  is a latent variable for the  $i$ th litter,  $\eta_{ij} \sim N(0, 1)$  is a latent variable for the  $j$ th pup,  $\epsilon_{1ij1} \sim N(0, \sigma^2)$  is the error term in the fetal weight model,  $\epsilon_{1ij2} \sim N(0, 1)$  is the error term in the malformation model,  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{15})'$  are parameters characterizing the baseline litter size distribution among dams in the control group having  $\xi_i = 0$ ,  $\beta$  is the coefficient representing the dose effect on litter size, and  $T$  is assumed to be the maximum number of pups per litter. In this analysis, we set  $T = 16$ , which is consistent with historical data and expert opinion. The five latent variable coefficients,  $\lambda_1$ ,  $\lambda_2$ ,  $\gamma_1$ ,  $\gamma_2$ , and  $\lambda_0$ , determine the various correlations. We restrict

$\lambda_1 > 0$  and  $\gamma_1 > 0$  for identifiability.

The model was completed by specifying priors for the parameters. Summary statistics from historical studies (Catherine et al., 1984) were used to develop informative priors for  $\mu_1$ ,  $\mu_2$ , and  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{15})^T$ . In particular, point estimates from the historical data were used as the prior means and the estimated variances of these point estimates were inflated by 10 to serve as the prior variances. For the other parameters, we chose vague but proper priors. To this end, the prior mean for  $\boldsymbol{\theta}_1 = (\mu_1, \mu_2, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \gamma_1, \gamma_2)^T$  was set to  $\boldsymbol{\theta}_{10} = (.98, -2.28, 0, 0, 0, 0, 0, 0)^T$ , and the covariance matrix was set to  $\boldsymbol{\Psi}_{10} = \text{diag}(.1, 2.6, 4, 4, 4, 4, 4, 4)$ . The prior mean for  $\boldsymbol{\theta}_2 = (\delta_1, \dots, \delta_{15}, \beta, \lambda_0)^T$  was set to

$$\begin{aligned} \boldsymbol{\theta}_{20} = ( & - 5.69, -2.45, -5.69, -2.02, -2.01, -1.62, -1.31, \\ & - 1.30, -.86, -.82, -.43, -.35, -.37, .59, .25, 0, 0)^T, \end{aligned}$$

and the covariance matrix was set to

$$\boldsymbol{\Psi}_{20} = \text{diag}(2, 1.3, 2, 0.6, 0.6, 0.3, 0.2, 0.3, 0.2, 0.3, 0.3, 0.4, 0.6, 1, 3.2, 4, 4),$$

Independent gamma priors were specified for the precision parameters:  $\sigma_k^{-2} \sim (.01, .01)$ . These priors have mean 1 and variance 100, again reflecting our vague knowledge about the precision parameters. We performed sensitivity analyses where we doubled and then halved the prior variances, and found that the final results were robust.

Fortran programs were developed for generating samples from the joint posterior distribution of the parameters and the latent variables. We used simulation studies (not reported here) to validate the code. We ran a single chain and used 100,000 iterates collected after convergence to compute posterior summaries of the parameters. Convergence of the sampler was assessed by diagnostic procedures recommended by Cowles and Carlin (1996) and implemented with BOA ([www.public-health.uiowa.edu/boa](http://www.public-health.uiowa.edu/boa)).

Table 2 presents the results of fitting the model to the ethylene glycol data. Fitted dose parameters are consistent with the summary in Table 1 in that the dose coefficient is negative

for weight ( $\Pr(\alpha_1 < 0|\text{data}) > 0.99$ ), positive for malformation ( $\Pr(\alpha_2 > 0|\text{data}) > 0.99$ ), and negative for litter size (though not significant,  $\Pr(\beta < 0|\text{data}) = 0.99$ )

The parameters  $\delta_1, \dots, \delta_{15}$  characterize the baseline distribution of litter size in a typical unexposed litter. The estimates for this baseline distribution is plotted in Figure 2. This distribution is centered around 10 with standard deviation 3.

None of the 95% posterior credible intervals of the coefficients for the latent variables included 0, with  $\lambda_1$  and  $\gamma_1$  positive (reflecting the identification constraints), and  $\lambda_2$ ,  $\gamma_2$ , and  $\lambda_0$  negative. The implied structure of the associations among the three outcomes, for two litter mates  $j$  and  $j'$ , is illustrated in Figure 3. In agreement with our intuition, the intra-litter associations in both weight and malformation are positive, while the intro-fetus association between weight and malformation is negative. Litter mates tend to respond similarly and malformed fetuses tend to be smaller. Also anticipated was the negative association between litter size and weight, which has been noted repeatedly in earlier studies. An unanticipated finding was the positive association between litter size and malformation. This association could be due to the tendency for malformed fetuses to be smaller or due to survival differences between malformed fetuses in different litters.

## 5. Discussion

This article has proposed a general Bayesian framework for jointly modeling cluster size and multiple categorical and continuous subunit-level outcomes. The proposed approach is based on an underlying normal model that generalizes previous structural equations models developed in the psychometrics literature (cf., Muthén, 1984). The approach has the advantage of being both flexible in accommodating a wide variety of dependency structures, and easy to implement via a Gibbs sampling algorithm. A practical advantage of the modeling structure is that, after augmenting the data with underlying normal variables, the regression and factor loadings parameters can be updated jointly, limiting problems with slow mixing common to

Markov chain Monte Carlo algorithms for posterior computation of latent variable models.

The problem of informative cluster size is related to the problem of informative censoring sometimes encountered in longitudinal follow-up studies. In this setting, our latent variable model represents a type of multivariate extension of shared parameter models that incorporate the same random effect in models for a longitudinal outcome and for the censoring process. An earlier latent variable model proposed by Dunson and Perreault (2001) accounted for informative missingness in clustered multivariate normal data but not for categorical outcomes or for random cluster size.

Although we have focused on the case where the cluster size is the only cluster-level outcome, many applications collect additional cluster-level data. For example, in developmental toxicology studies, maternal body weight and other measures of maternal health are typically available. In the setting of longitudinal studies, time-independent outcomes may be collected along with response variables measured at each follow-up time. A straightforward generalization of our approach would accommodate multiple cluster-level outcomes having a mixture of continuous, binary and categorical measurement scales.

#### ACKNOWLEDGEMENTS

The authors would like to thank Paul Catalano for providing the data, and Beth Gladen and David Umbach for their comments.

## REFERENCES

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.
- Albert, J.H. and Chib, S. (2001). Sequential ordinal modeling with applications to survival data. *Biometrics* **57**, 829-836.
- 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.
- Bartholomew, D.J. (1987). *Latent Variable Analysis and Factor Analysis*. London: Griffen.
- Berzuini, C. and Larizza, C. (1996). A unified approach for modeling longitudinal and failure time data, with application to medical monitoring. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**, 109-123.
- Catalano, P. and Ryan, L.M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* **87**, 651-658.
- Chen, J. (1993). A malformation incidence dose-response model incorporating fetal weight and or litter size as covariates. *Risk Analysis* **13**, 559-564.
- Chen, M.-H., Shao, Q.-M. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.
- Dunson, D.B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society B* **62**, 355-366.
- Dunson, D.B. and Perreault, S.D. (2001). Factor analytic models of clustered multivariate data with informative censoring. *Biometrics* **57**, 302-308.

- Faucett, C.J. and Thomas, D.C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine* **15**, 1663-1685.
- Fitzmaurice, G.M. and Laird, N.M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association* **90**, 845-852.
- Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398-409.
- Gray, S.M. and Brookmeyer, R. (2000). Multidimensional longitudinal data: Estimating a treatment effect from continuous, discrete or time-to-event response data. *Journal of the American Statistical Association* **95**, 396-406.
- Gueorguieva, R.V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* **96**, 1102-1112.
- Hogan, J.W. and Laird, N.M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* **16**, 239-257.
- Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.
- Molenberghs, G. and Ryan, L.M. (1999). An exponential family model for clustered multivariate binary data. *Environmetrics* **10**, 279-300.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika* **49**, 115-132.

- Regan, M.M. and Catalano, P.J. (1999). Likelihood models for clustered binary and continuous outcomes: Application to developmental toxicology. *Biometrics* **55**, 760-768.
- Romero, A., Villamayor, F., Grau, M.T., Sacristan, A. and Ortiz, J.A. (1992). Relationship between fetal weight and litter size in rats - application to reproductive toxicology studies. *Reproductive Toxicology* **6**, 453-456.
- Ryan, L. and Molenberghs, G. (1999). Statistical methods for developmental toxicity - Analysis of clustered multivariate binary data. *Uncertainty in the Risk Assessment of Environmental and Occupational Hazards* **895**, 196-211.
- Tanaka, T. (1998). Effects of litter size on behavioral development in mice. *Reproductive toxicology* **12**, 613-617.
- Ten Have, T.R. and Uttal, D.H. (1994). Subject-specific and population-averaged continuation ratio logit models for multiple discrete-time survival profiles. *Applied Statistics* **43**, 371-384.
- Wang, and Taylor, J.M.G. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* **96**, 895-905.
- Wulfsohn, M.S. and Tsiatis, A.A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330-339.

## APPENDIX

### Conditional Posterior Distributions for Gibbs Sampling

*Step 1.* To sample the underlying normal variables  $\mathbf{z}_i$  from their full conditional distributions:

(i) For each continuous  $y_{ijk}$ , set the corresponding element of  $\mathbf{z}_i$  equal to  $g_k^{-1}(y_{ijk})$ ; (ii) For each binary  $y_{ijk}$ , sample the corresponding element of  $\mathbf{z}_i$  from  $N(\mathbf{u}_{1ijk}^T \boldsymbol{\theta}, 1)$  truncated above by 0 for  $y_{ijk} = 0$  and below by 0 for  $y_{ijk} = 1$ ; and (iii) For  $j = 1, \dots, s_i - 1$ , sample the  $(ps_i + j)$ th element of  $\mathbf{z}_i$  from  $N(\mathbf{u}_{2ij}^T \boldsymbol{\theta}, 1)$  truncated above by 0; and (iv) If  $s_i < T$ , sample the  $(p + 1)s_i$ th element of  $\mathbf{z}_i$  from  $N(\mathbf{u}_{2ij}^T \boldsymbol{\theta}, 1)$  truncated below by 0.

*Step 2.* Given its prior in (3), the full conditional distribution of  $\boldsymbol{\theta}$  is  $N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Psi}})$ , where

$$\hat{\boldsymbol{\Psi}} = \left( \sum_{i=1}^n \mathbf{U}_i^T \mathbf{A}_i^{-1} \mathbf{U}_i + \boldsymbol{\Psi}_0^{-1} \right)^{-1} \quad \text{and} \quad \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Psi}} \left( \sum_{i=1}^n \mathbf{U}_i^T \mathbf{A}_i^{-1} \mathbf{z}_i + \boldsymbol{\Psi}_0^{-1} \boldsymbol{\theta}_0 \right),$$

where  $\mathbf{U}_i$  and  $\mathbf{A}_i$  are defined in Section 3.1.

*Step 3.* Let  $\boldsymbol{\phi}_i = (\boldsymbol{\phi}_{1i1}^T, \dots, \boldsymbol{\phi}_{1is_i}^T, \phi_{2i1}, \dots, \phi_{2it_i})^T$  be a  $(ps_i + t_i) \times 1$  vector, with  $\boldsymbol{\phi}_{1ij} = \mathbf{y}_{ij}^* - \boldsymbol{\mu} - \boldsymbol{\alpha} \mathbf{x}_i - \boldsymbol{\Gamma} \mathbf{W}_i \boldsymbol{\eta}_{ij}$  and  $\phi_{2ij} = s_{ij}^* - \tau_j + \mathbf{x}_i^T \boldsymbol{\beta}$ . Then  $\boldsymbol{\phi}_i \sim N(\boldsymbol{\Lambda}_i \mathbf{W}_i \boldsymbol{\xi}_i, \mathbf{A}_i)$ , where  $\boldsymbol{\Lambda}_i = [\boldsymbol{\Lambda}^T \otimes \mathbf{1}_{(s_i)}^T, -\boldsymbol{\lambda}_0 \otimes \mathbf{1}_{(t_i)}^T]^T$ . Hence the conditional distribution of  $\boldsymbol{\xi}_i$  is

$$N \left( \left( \mathbf{W}_i \boldsymbol{\Lambda}_i^T \mathbf{A}_i^{-1} \boldsymbol{\Lambda}_i \mathbf{W}_i + \mathbf{I} \right)^{-1} \mathbf{W}_i \boldsymbol{\Lambda}_i^T \mathbf{A}_i^{-1} \boldsymbol{\phi}_{ij}, \left( \mathbf{W}_i \boldsymbol{\Lambda}_i^T \mathbf{A}_i^{-1} \boldsymbol{\Lambda}_i \mathbf{W}_i + \mathbf{I} \right)^{-1} \right).$$

*Step 4.* Similarly, since

$$\mathbf{y}_{ij}^* - \boldsymbol{\mu} - \boldsymbol{\alpha} \mathbf{x}_i - \boldsymbol{\Lambda}_1 \mathbf{W}_i \boldsymbol{\xi}_i = \boldsymbol{\Gamma} \mathbf{W}_i \boldsymbol{\eta}_{ij} + \boldsymbol{\epsilon}_{1ij},$$

for  $i = 1, \dots, n$ ,  $j = 1, \dots, s_i$ , the conditional distribution of  $\boldsymbol{\eta}_{ij}$  is also normal, with mean

$$\left( \mathbf{W}_i \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma} \mathbf{W}_i + \mathbf{I} \right)^{-1} \mathbf{W}_i \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}^{-1} \left( \mathbf{y}_{ij}^* - \boldsymbol{\mu} - \boldsymbol{\alpha} \mathbf{x}_i - \boldsymbol{\Lambda}_1 \mathbf{W}_i \boldsymbol{\xi}_i \right)$$

and variance

$$\left( \mathbf{W}_i \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma} \mathbf{W}_i + \mathbf{I} \right)^{-1},$$

respectively.

*Step 5.* Given their priors in (3), the full conditional distributions of the  $\tau_k$  are

$$\text{gamma} \left( \frac{1}{2} \sum_{i=1}^n s_i + c_{0k}, \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{s_i} (y_{ijk}^* - \mathbf{u}_{1ijk}^T \boldsymbol{\theta})^2 + d_{0k} \right),$$

for  $k = 1, \dots, p_c$ .

Table 1: NTP developmental toxicity study of ethylene glycol

Dose (mg/kg)	No. Dams	Mean litter size	Mean fetal weight (g)	Malformation (%)
0	25	11.9	0.97	0.3
0.75	24	11.5	0.88	9.4
1.50	22	10.4	0.76	38.9
3.00	23	9.8	0.70	57.1

Table 2: Posterior summaries of the parameters for the ethylene glycol example

Variable	Parameter	Mean	SD	95% Credible interval
<i>Weight model</i>				
Intercept	$\mu_1$	0.962	0.015	(0.933, 0.990)
Dose	$\alpha_1$	-0.092	0.009	(-0.110, -0.075)
Litter-level latent variable	$\lambda_1$	0.090	0.007	(0.077, 0.106)
Fetus-level latent variable	$\gamma_1$	0.032	0.018	(0.007, 0.069)
<i>Malformation model</i>				
Intercept	$\mu_2$	-2.435	0.310	(-3.091, -1.933)
Dose	$\alpha_2$	0.962	0.135	(0.719, 1.235)
Litter-level latent variable	$\lambda_2$	-0.642	0.114	(-0.867, -0.435)
Fetus-level latent variable	$\gamma_2$	-0.470	0.295	(-1.061, -0.042)
<i>Litter size model</i>				
Baseline	$\delta_1$	-2.512	0.429	(-3.346, -1.694)
	$\delta_2$	-2.650	0.305	(-3.232, -2.055)
	$\delta_3$	-2.462	0.431	(-3.313, -1.690)
	$\delta_4$	-2.052	0.189	(-2.430, -1.689)
	$\delta_5$	-2.220	0.203	(-2.619, -1.830)
	$\delta_6$	-1.793	0.143	(-2.066, -1.516)
	$\delta_7$	-1.310	0.117	(-1.535, -1.075)
	$\delta_8$	-1.475	0.141	(-1.767, -1.219)
	$\delta_9$	-1.105	0.115	(-1.320, -0.877)
	$\delta_{10}$	-0.838	0.124	(-1.077, -0.593)
	$\delta_{11}$	-0.620	0.127	(-0.860, -0.366)
	$\delta_{12}$	-0.286	0.140	(-0.545, -0.001)
	$\delta_{13}$	-0.257	0.174	(-0.592, 0.081)
	$\delta_{14}$	0.467	0.225	(0.037, 0.914)
	$\delta_{15}$	0.609	0.404	(-0.133, 1.422)
Dose	$\beta$	-0.018	0.047	(-0.108, 0.072)
Litter-level latent variable	$\lambda_0$	-0.157	0.079	(-0.319, -0.011)