MULTIVARIATE LINEAR MIXED MODELS FOR MULTIPLE OUTCOMES

MARY SAMMEL[1]*, XIHONG LIN[2] AND LOUISE RYAN[3]

[1] *Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, U.S.A.*
[2] *Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*
[3] *Department of Biostatistical Science, Dana Farber Cancer Institute, and Department of Biostatistics, Harvard University, Boston, MA, U.S.A.*

SUMMARY

We propose a multivariate linear mixed (MLMM) for the analysis of multiple outcomes, which generalizes the latent variable model of Sammel and Ryan. The proposed model assumes a flexible correlation structure among the multiple outcomes, and allows a global test of the impact of exposure across outcomes. In contrast to the Sammel–Ryan model, the MLMM separates the mean and correlation parameters so that the mean estimation will remain reasonably robust even if the correlation is misspecified. The model is applied to birth defects data, where continuous data on the size of infants who were exposed to anticonvulsant medications *in utero* are compared to controls. Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

In many medical research settings, the effect of interest cannot be characterized by a single outcome, but instead multiple outcomes need to be measured on each individual under study. A classic example of this arises in the study of birth outcomes, since teratogenic exposures often result in a *syndrome* wherein the effect is identified not through a single outcome, but by a distinctive pattern of various related defects (Holmes *et al.*[1]).

The development of statistical methods for the analysis of multiple outcomes has been an area of active research for many decades. There is a rich body of literature on statistical methods for multivariate methods for continuous outcomes (see, for example, Johnson and Wichern[2]). One of the more popular approaches, factor analysis, aims to reduce multi-dimensional data into a smaller number of latent outcomes which cannot be directly measured. This concept of latent variables was first attributed to Spearman in 1904,[3] and has become popular in a variety of applied settings (see Everitt[4]). Latent variable models have also been considered under the label 'measurement model' in the structural equations framework (Bentler and Weeks[5]). Recently Sammel and Ryan[6] generalized this model to include fixed effect covariates on both the latent and observed variables. The model of Sammel and Ryan[6] can be thought of as a formalization of the

* Correspondence to: Mary Sammel, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Blockley Hall Rm. 605, 423 Guardian Dr., Philadelphia, PA 19104–6021, U.S.A. E-mail: msammel@cceb.upenn.edu.

commonly used two-step procedure wherein one first performs a factor analysis or principal components analysis to identify a linear combination of outcomes that are most correlated with each other, then models this linear combination as a function of covariates. The Sammel–Ryan model also generalizes the class of linear mixed models (Laird and Ware,[7] Harville[8]) and can accommodate either maximum likelihood or restricted maximum likelihood (REML) estimation of unknown model parameters. While the approach has some appealing features, a disadvantage is its lack of robustness due to the fact that covariance parameters are also present in the mean.

In this paper we develop a more general approach, based on a multivariate linear mixed model (MLMM), which models the mean and covariance as a function of covariates but separates mean and correlation parameters so that the mean estimation will be more robust to misspecification of the correlation structure. We also provide a global test of the impact of exposure across outcomes. We begin by describing the data which motivated our research, based on a study of *in utero* exposure to anticonvulsant medications (Holmes *et al.*[9]). We then briefly review three existing approaches to analysing these data, including generalized estimating equations (GEEs), two-stage factor analysis (TSFA), and latent variable models (LVMs). After discussing drawbacks to these approaches, we present the new method (MLMM), then compare the performance of all four methods using the birth defect data.

2. THE BIRTH DEFECT DATA

Holmes *et al.*[9] discuss an observational cohort study of infants born at Brigham and Women's Hospital in Boston, U.S.A. The study included subjects in three different groups, including *Drug Exposed* (epileptic women who took medications during their pregnancy), *Seizure History* (epileptic women who stopped taking anticonvulsants medications during their pregnancy) and *Control* (women randomly chosen from those who gave birth at the same hospital at the same time as the *Drug Exposed* and *Seizure History* women). A variety of different outcomes were assessed on the infants, including weight, size, a variety of cranial and limb measurements, presence or absence of major and minor malformations, as well as assessments on a variety of minor physical anomalies. Although not of clinical importance themselves, these minor physical anomalies, which include features such as long nasal septum, antiverted nostril, tapered fingers etc., are important since they can serve as 'markers' for the presence of more serious and perhaps as yet unobservable effects.

There are several different scientific questions of interest. Of primary importance is the question of whether *in utero* exposure to anticonvulsant medications leads to adverse birth outcomes, or whether effects are due to the maternal epilepsy itself. This aim will not be addressed in the analysis presented here. For the illustration of our methods for multiple outcomes, we will consider only 628 subjects which comprise the exposed and control subgroups. This will allow us to study the joint effect of medications and maternal epilepsy. Also of interest is the question of which outcomes, and in particular, which minor anomalies, are useful in predicting whether a baby is 'affected'. Finally, there is interest in using the data on multiple outcomes to construct a *severity score* that can indicate how severely an individual baby has been affected. Only subjects whose exposure status was blinded from the examiner have been used.

Table I shows some summary statistics for a small subset of these continuous measurements that will be used to illustrate our results. Variables include (in order) bitemporal (side to side) head diameter, nose length, ear length and width, finger length, weight and anterior–posterior (front to back) head diameter. The table shows means and standard deviations among exposed and

Table I. Summary statistics

	$n = 176$ Exposed		$n = 452$ Control		$\hat{\beta}$	p-value
	Mean	(SE)	Mean	(SE)		
Head-bt	9·10	(0·059)	9·40	(0·024)	− 0·212	< 0·001
Nose	1·94	(0·010)	1·98	(0·015)	− 0·031	0·104
Ear length	3·60	(0·026)	3·73	(0·016)	− 0·025	0·378
Ear width	2·13	(0·017)	2·17	(0·009)	− 0·017	0·336
Finger	2·86	(0·021)	2·95	(0·017)	− 0·045	0·103
Weight	3·28	(0·005)	3·41	(0·026)	− 0·004	0·917
Head-ap	11·67	(0·055)	11·74	(0·027)	− 0·007	0·880

* Adjusted for gender and gestational age using linear regression models

control infants, along with the estimated exposure effect based on a linear regression model that also adjusts for gender and gestational age. However, only bitemporal head width shows a statistically significant ($p < 0.05$) difference between controls and exposed infants. These data convincingly illustrate the need for a multivariate approach that allows for an overall assessment of exposure effects by combining information from these related outcomes and could detect the exposure effects with better statistical power. A more standard approach based on, say, Bonferroni adjustments, would lead to the conclusion of no effect.

3. METHODS FOR ANALYSING MULTIPLE OUTCOMES

In this section, we will consider four methods for the analysis of multiple outcomes. Our objective is to characterize the effect of a binary exposure variable, z_i, on a $M \times 1$ vector of continuous outcomes, $\mathbf{y}_i = (y_{i1}, \ldots, y_{iM})^{\mathrm{T}}$, for subject $i = 1, \ldots, n$ while adjusting for additional covariates, $\mathbf{x}_{ij}$. The question of interest is how to assess an overall exposure effect using the information from the M outcomes. We will discuss advantages and disadvantages of each approach and illustrate each using the birth defects data with the first five outcomes in Table I, where interest lies in estimating an overall effect of exposure to anticonvulsants and maternal epilepsy on the birth outcomes while adjusting for the effects of gender and gestational age. More detailed comparisons of these four approaches will be given in Section 4. In that section, we will consider two models: model 1 will use only the first five outcomes from Table I, while model 2 will use all seven. The reason for considering these two models is that we want to assess the sensitivity of the different approaches to the subset of outcomes chosen for inclusion in the analysis.

3.1. Generalized estimating equations

Generalized estimating equations (GEEs) (Liang and Zeger,[10] Zeger and Liang[11]) provide one natural approach to analysing multiple outcome data. A key feature of GEE is that consistent estimates can be obtained if the means of the outcomes are correctly specified, even when the correlation between the outcomes has been misspecified. We assume that the mean model for the jth outcome is

$$E(y_{ij}) = \mathbf{x}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_j + z_i \boldsymbol{\beta}_j \qquad (1)$$

Table II. Generalized estimating equations – model 1

Outcome	σ^2	β	Robust SE	p-value
(a) Individual exposure estimates				
Bitemporal	0·272	− 0·212	− 0·058	< 0·001
Nose	0·043	− 0·030	0·019	0·102
Ear length	0·099	− 0·025	0·028	0·370
Ear width	0·037	− 0·017	0·017	0·336
Finger	0·095	− 0·045	0·028	0·104
ρ	0·080			
(b) Common exposure estimates				
Bitemporal	0·278	− 0·031	0·012	0·012
Nose	0·043			
Ear length	0·099			
Ear width	0·037			
Finger	0·095			
ρ	0·079			

where α_j and β_j, respectively, measure the effects of the covariates x and the exposure effect on the jth outcome, y_{ij}. In matrix and vector notation, we have

$$E(y_i) = x_i\alpha + z_i\beta$$

where x_i is a block diagonal matrix with x_{ij}^{T} on the diagonal and $\alpha = (\alpha_1^{\mathrm{T}}, \ldots, \alpha_M^{\mathrm{T}})^{\mathrm{T}}$ and $\beta = (\beta_1, \ldots, \beta_M)^{\mathrm{T}}$. A sensible choice for our application is an exchangeable working correlation among the M outcomes:

$$\mathrm{var}(y_i) = \Psi^{1/2}((1 - \rho)I + \rho J)\Psi^{1/2}$$

where I is the $M \times M$ identity matrix, J is the $M \times M$ matrix of ones and $\Psi = \mathrm{diag}\{\sigma_j^2\}$. Using weighted least squares, the unknown mean parameters α and β can be estimated as the solution to an estimating equation similar to equation (6) of Liang and Zeger.[10] Similarly, the 'sandwich variance' of Liang and Zeger can be used to estimate the variance of the estimated parameters. The working variance and correlation parameters $(\sigma_1^2, \ldots, \sigma_M^2)$ and ρ can be estimated using the method of moments. Results of fitting GEEs to the birth defect data are presented in Table II. Two different analyses are presented. The first allows a separate exposure effect for each outcome (that is, different β_j for each j). As expected, the results are very similar to those found in Table I for the univariate analyses. The second analysis assumes a common β. The estimated value of $\beta = -0.031$ is significantly different from zero ($p = 0.012$) based on a wald test. The estimate correlation parameter fitted under the exchangeable model was $\rho = 0.079$. Virtually identical results were obtained when a completely unstructured covariance was assumed. The correlations among the observed outcomes ranged from 0·23 (for ear length and width) to − 0·06 (between nose and finger lengths). When the exchangeable correlation was estimated separately for two groups results were also unchanged ($\beta = -0.027$, $p = 0.026$).

Although GEEs have the advantage of simplicity, a limitation is the difficulty in assessing an overall exposure effect. Specifically, testing for the exposure effect is based on a M degree-of-freedom test, that is, all $\beta_j = 0$, and could be subject to low power (see Legler et al.[12]). It is also

Table III. Two-stage factor analysis – model 1

Outcome	Factor loadings	Specific variance	% Experimental variance
Bitemporal	0·107	0·268	4%
Nose	0·033	0·041	3%
Ear length	0·245	0·040	60%
Ear width	0·056	0·034	9%
Finger	0·025	0·093	1%
β	− 0·111		

unclear how to use the estimated β_j to calculate a meaningful overall exposure effect estimate. A common practice is to use the average value of the estimated β_j as an overall metric. However, the M outcomes y_{ij} could be subject to different units and different scalings. Hence this average may not be meaningful and may be dominated by the β_j whose y_{ij}'s have relatively large scales. The same argument applies to assuming a common β in (1).

3.2. Two-stage factor analysis

Another standard approach to the problem of multiple outcomes is that of two-stage factor analysis. For our example, this method would proceed as follows. First, we remove the impact of the confounding variables gender and gestational age by fitting individual linear regression models for each outcome, and use the residuals from those fits as the outcomes to be input into the factor analysis. The first factor score from the analysis is then output, and the impact of exposure on the latent outcome is assessed using a linear regression on the exposure variable. Results from this analysis are presented in Table III. Notice that ear length then bitemporal head diameter have the two highest loadings (suggesting that these two variables will have the greatest influence on the estimated latent variable, and through it, the exposure effect). A linear regression to the factor scores estimated at the first stage yields an estimated exposure effect of − 0·111, with corresponding p-value (obtained by Generalized Likelihood Ratio) of 0·115. This method treats the factor score as if it were fixed and known, not estimated. Better parameter estimates can be obtained by using an iterative procedure, see below.

3.3. The latent variable model

Sammel and Ryan[6] proposed the use of latent variable models to characterize the effect of an exposure on multiple outcomes, while adjusting for a set of confounding variables. This model can be viewed as integrating the two-stage factor analysis into a more rigorous likelihood framework.

The latent variable model is best explained using a two-stage model. At Stage 1, we assume

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_j + \lambda_j u_i + \varepsilon_{ij} \tag{2}$$

where u_i is an unobserved (latent) variable unique to individual i which reflects how severely the ith subject is affected, and $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_j^2)$ is a random error term. At Stage 2, the u_i's are modelled as a function of exposure

$$u_i = z_i \beta + a_i \tag{3}$$

Table IV. Latent variable analysis – model 1

Outcome	Factor loadings	Specific variance	% Experimental variance
Bitemporal	0·136	0·260	7%
Nose	0·042	0·040	4%
Ear length	0·189	0·063	36%
Ear width	0·063	0·034	10%
Finger	0·038	0·093	2%
β	− 0·325		

where the $a_i \sim \mathrm{N}(0, 1)$ are i.i.d. error terms. The regression coefficient β measures the exposure effect on the latent severity score u_i. An advantage of the latent variable model is that a one degree-of-freedom likelihood ratio test can be used to test for the overall exposure effect, and could be more powerful compared to the M degree-of-freedom test used in the GEE setting. Another advantage of the latent variable model is that one can estimate individual severity score by calculating $E(u_i | y_i)$ for each subject.

It follows from equations (2) and (3) that the model for y_{ij}, given a_i is

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_j + z_i\lambda_j\beta + \lambda_j a_i + \varepsilon_{ij}. \tag{4}$$

Integrating over a_i we have the marginal distribution

$$\boldsymbol{y}_i \sim \mathrm{N}(\boldsymbol{x}_i\boldsymbol{\alpha} + z_i\boldsymbol{\lambda}\beta, \boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathrm{T}} + \boldsymbol{\Psi})$$

where $(\boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{\alpha}, \boldsymbol{\Psi})$ are defined in the same way as in Section 3.1. Model fitting can proceed in several different ways, although the EM algorithm (Dempster et al.[13]) is a natural one. Sammel and Ryan[6] use a modified EM known as the ECME (Liu and Rubin[14]) which, for subsets of the parameters, conditionally maximizes the marginal or conditional likelihood at the M-step. This algorithm (also the one used by Laird and Ware[7]) tends to be faster than the classical EM for problems of this sort. Both ML or REML estimation are easily accommodated, the latter requiring a straightforward modification of the E-step. Table IV shows the results of fitting the latent variable model to the birth defect data. Notice that consistent with the results reported in Table III, the highest loadings correspond to bitemporal head width and ear length. The overall effect of exposure ($\beta = - 0.325$) is statistically significant from 0, based on a generalized likeli-hood ratio test ($p = 0.015$), and is a threefold increase in the effect estimate obtained from the TSFA approach.

While the latent variable model is appealing in many ways, analytic considerations and simulations suggest that it may not be very robust. The approach has excellent power when the model is correct, but can behave poorly when the model is violated in certain ways. In particular, the estimate of the exposure effect can be biased when there exists a subset of uncorrelated outcomes that are associated with exposure. This is because the latent variable model assumes that the correlated outcomes are the ones associated with exposure. More precisely, since $\boldsymbol{\lambda}$ enters into both the mean and covariance specifications, misspecification of the correlation structure, for example, in the presence of a subset of uncorrelated outcomes, would bias the estimates of the exposure parameter β. Another scenario where the estimate of the exposure parameter, β, can be biased is when some of the correlated outcomes are not associated with exposure.

The second scenario is illustrated by the following example. Specifically, we re-fit the model adding the additional outcomes listed in Table I. We refer to this larger model as model 2. The two added variables (weight and anterior–posterior head diameter) are correlated, but are not significantly effected by exposure (see Table I). The overall estimate of the exposure effect under model 2 shrinks back towards zero ($\beta = -0.218$) and has a p-value of < 0.042, based on the generalized likelihood ratio test.

3.4. The multivariate linear mixed model

The robustness problems associated with the latent variable model motivate consideration of the following random effects formulation which separates mean and covariance parameters

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_j + z_i\beta_j + \lambda_j a_i + \varepsilon_{ij} \tag{5}$$

where $a_i \sim \mathrm{N}(0,1)$ and $\varepsilon_{ij} \sim \mathrm{N}(0,\sigma_j^2)$. Goldstein (reference 15, Chapter 5) discusses some similar classes of models for multivariate data. One problem with this model is that it does not lead to an easy assessment of the overall effect of exposure. While an M degree-of-freedom test could be applied, this is likely to have low power when most of the outcomes are similarly affected. In such cases, a better alternative is to assume that all the β_j's are equal to a common value, β. However, this common dose effect assumption is often too strong, since different outcomes are likely to be measured on different scales. Thus, the assumption of a common effect size may still be too strong, since the strength of the exposure effect may vary among outcomes. To allow for this possibility, one could assume different β_j's for different standardized outcomes y_{ij}/σ_j and *smooth* the β_j's, for example, by assuming $\beta_j \sim \mathrm{N}(\beta,\tau^2)$. Such an approach would be appealing since it would facilitate empirical Bayes-type shrinkage of the exposure effects on each outcome towards some central, overall effect, and the test for the exposure effect can be based on a more powerful one degree-of-freedom test. However, a problem is that the asymptotics are not right; we would need the number of outcomes to be large in order to *satisfactorily* apply such a random effects model. In the presence of only a few outcomes, this strategy does not work. A remedy for this problem is to further borrow information from different subjects in our process of smoothing the exposure effect. In other words, we smooth the exposure effect by borrowing information from different outcomes and different subjects.

To provide further motivation for this idea, suppose for a moment that the outcomes are measured in the same scale, there are no confounding variables and we can observe each person's multivariate outcome vector $\boldsymbol{y}_i$ under both treatment and control settings, say, $\boldsymbol{y}_i(1)$ and $\boldsymbol{y}_i(0)$, respectively. Each person, hence, could provide an estimate of $\beta_1, \dots \beta_M$, say, $\beta_{ij} = Y_{ij}(1) - Y_{ij}(0)$. In this setting, it would make sense to assume a first-stage model such as

$$\frac{y_{ij}}{\sigma_j} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_j + z_i\beta_{ij} + \lambda_j a_i + \varepsilon_{ij} \tag{6}$$

then smooth the β_{ij}'s using a second-stage model

$$\beta_{ij} = \beta + \delta_j b_i \tag{7}$$

with $b_i \sim \mathrm{N}(0,1)$, borrowing an idea commonly used in growth curve modelling. Under the assumption of random assignment to exposure group, conditional on other covariates, we can integrate over the distribution of the unobserved outcomes to obtain a model that can be fit in the context where we only observe each individual once. In fact, we can still fit the model

Table V. MLMM – model 1

Outcome	λ_j	δ_j	σ_j
Bitemporal	1·233	0·222	0·432
Nose	0·183	0·171	0·201
Ear length	0·103	0·918	0·233
Ear width	− 0·212	0·369	0·179
Finger	0·001	0·101	0·304
β	− 0·175		

characterized in two stages by (6) and (7) as

$$\frac{y_{ij}}{\sigma_j} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_j + z_i\beta + \lambda_j a_i + z_i\delta_j b_i + \varepsilon_{ij}. \tag{8}$$

This is the model we refer to as the multivariate linear mixed model (MLMM). Integrating over the random components yields the following multivariate linear mixed model for the marginal distribution of $\boldsymbol{y}_i$:

$$\boldsymbol{y}_i \sim \mathrm{N}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\alpha} + z_i\beta\mathbf{1}), \quad \boldsymbol{\Psi}^{1/2}(\boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathrm{T}} + z_i\boldsymbol{\delta}\boldsymbol{\delta}^{\mathrm{T}} + \mathbf{I})\boldsymbol{\Psi}^{1/2}).$$

Several features of this model are worth noting. The fact that it parameterizes the exposure effect in terms of *effect sizes* is appealing, since many clinicians are familiar with this concept. The parameter β reflects the overall or average effect size of exposure. Hence one can easily apply a one degree-of-freedom likelihood ratio test to assess global effects. Although model (8) assumes that the mean exposure effect is the same for all outcomes on the standardized scale, it allows different mean exposure effects for different outcomes on the original scales. The exposure effect on the jth outcome can be estimated using $\sigma_j\beta$. A unique feature of MLMM is that it allows one to estimate subject-specific exposure effects on the jth outcome as $\hat{\sigma}_j[\hat{\beta} + \hat{\delta}_j E(b_i|\boldsymbol{y}_i)]$.

The MLMM is related to both the GEE model and the latent variable model. Specifically, the MLMM can allow for an exchangeable correlation among outcomes by setting $\delta_j = 0$ and assuming the λ_j to be the same for all j. The latent variable model (4) can be obtained by setting $\delta_j = 0$, then replacing $\sigma_j\beta$ in the mean by $\lambda_j\beta$ and $\lambda_j\sigma_j a_i$ in the covariance specification by $\lambda_j a_i$. This suggests that, unlike the latent variable model, the mean and covariance parameters in MLMMs are loosely connected through the residual variances σ_j^2. One can easily show that the correlations among outcomes specified under the MLMM depend only on $(\boldsymbol{\lambda}, \boldsymbol{\delta})$, which are separated from the parameters $(\boldsymbol{\alpha}, \beta, \boldsymbol{\sigma})$ used in the mean specification. Hence, we expect that the MLMM will be more robust than the latent variable model when the correlation structure is misspecified, for example, when a subset of outcomes are not correlated.

Inference in MLMMs can proceed by using the maximum likelihood method via directly maximizing the log-likelihood (10) from the appendix. Alternatively, the EM algorithm or its modifications can be used (see the Appendix for details), The MLEs of the parameters in model 1 for the birth defect data are given in Table V. The results suggest highly significant exposure effect, $\beta = − 0·175$, ($p < 0·001$, based on GLR test).

Table VI. Summary of model fits

	Average effect	χ^2_1	p-value
(a) *Model 1: first five outcomes from Table I*			
GEE	-0.031	6·371	0·012
TSFA	-0.010	2·490	0·115
LV	-0.031	5·944	0·015
MLMM	-0.047	11·293	< 0.001
(b) *Model 2: Model 1 + 2 corr/unaffected*			
GEE	-0.026	4·390	0·037
TSFA	-0.025	3·580	0·109
LV	-0.038	4·130	0·042
MLMM	-0.039	7·220	0·007

4. MODEL COMPARISONS

Table VI summarizes the results of fitting GEEs, the two-stage factor analysis, a latent variable model and the MLMM model to the two different subsets of the variables described in Table I. Model 1 includes the first five variables from Table I, while model 2 includes all seven. For comparisons, an overall measure of the exposure effect is needed for each model, as well as a significance test. For the GEE model (1), an overall metric is the common exposure effect β. To maintain comparability with the GEE method, we use $\bar{\lambda}\beta$ to summarize the overall effect for the two-stage factor analysis and also for the latent variable model (3.4). For the MLMM model, we use $\bar{\sigma}\beta$, where $\bar{\lambda}$ and $\bar{\sigma}$ are average values taken over the M outcomes.

Outcomes in the first subset were selected to best fit the latent variable model, and we see comparable magnitude of the exposure test for each of the analyses. The two-stage factor analysis estimate of the exposure effect is smaller, and the test is non-significant, most likely because it does not consider the error in the estimation of the factor scores. The addition of the last two outcomes from Table I changes the estimated magnitude and strength of the association between exposure and birth outcomes. Notice that the latent variable analyses now yields only a marginally significant result, whereas the GEE and MLMM approaches maintain their significance.

In contrast to the GEE model, the TSFA, LV and MLMM analyses afford the opportunity to compute a summary score which ranks the subjects with respect to the multiple outcomes. The factor scores can be used to rank the subjects for the TSFA model (Johnson and Wichern[2]), while Sammel and Ryan[6] illustrate the use of a similar summary score for the LV model. For the MLMM model we can use the estimate of the random effect $\hat{a}_i + z_i(\beta + \bar{\delta}\hat{b}_i)$, where $\hat{a}_i = E(a_i|y_i)$ and $\tilde{b}_i = E(b_i|y_i)$. Histograms for the summary scores by exposure status are presented in Figure 1. For control subjects, the summary score is centred near zero, but the average score for those exposed to anticonvulsants is shifted to the left.

Finally, it is useful to discuss and compare the degree to which the various methods provide estimates of the exposure effects associated with individual outcomes. The GEE estimates the outcome-specific estimates by β_j. For both the TSFA and LV approaches, the exposure effect on the jth outcome can be estimated by $\lambda_j\beta$. For the MLMM model, one can estimate the marginal effect of exposure on the jth outcome by $\sigma_j\beta$.
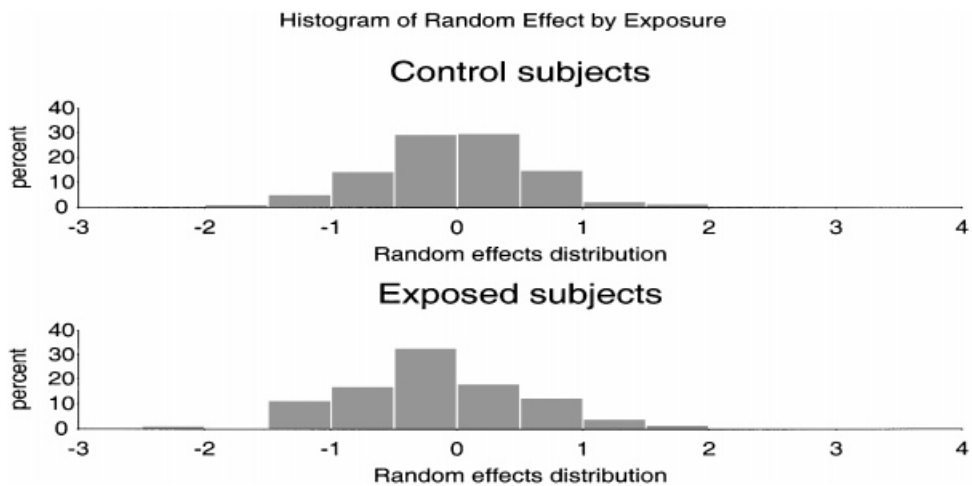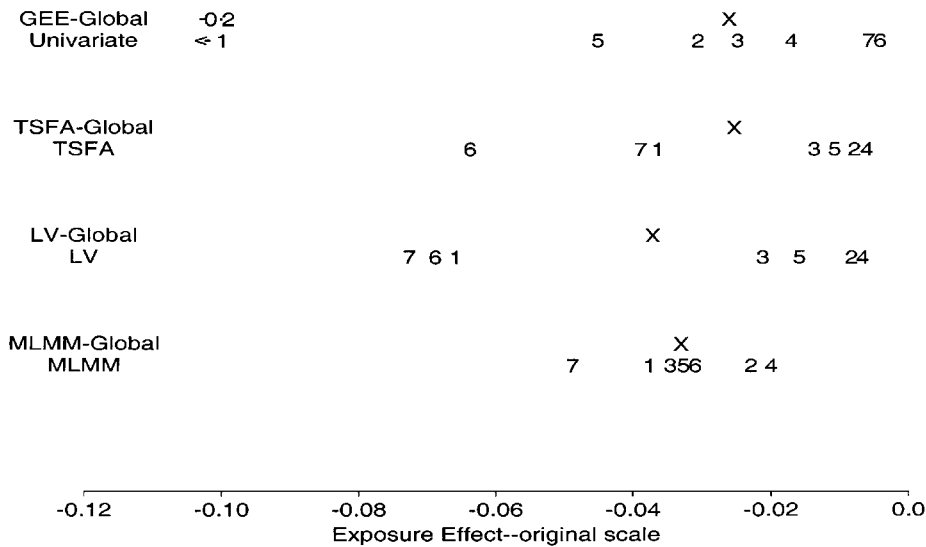
Figure 1. Subject summary scores by exposure

Figure 2. Global and individual exposure effect estimates for outcomes: 1, head-bt; 2, nose; 3, ear-l; 4, ear-w; 5, finger; 6, weight; 7, head-ap; X, global effect estimate

The two rows in each panel of Figure 2 show respectively overall exposure effect and individual effects estimated separately for each of the seven outcomes. Notice that outcome 1 (bitemporal head diameter) has a much stronger effect, -0.2, than any of the other outcomes. The two-stage factor analysis model, and to some extent the latent variable model, overestimate the exposure effect on outcomes 6 (birth weight) and 7 (anterior–posterior head diameter). This happens because the first latent factor is dominated by the variables which reflect the infant's size,

regardless of their association with exposure. The MLMM marginal estimates refer to the values $\sigma_j\beta$, and are the overall exposure effect size reflected on the original scale of each outcome. It is also useful to examine the estimated values of δ. For the seven listed in Table I, these values were: $1\cdot957$; $0\cdot166$; $0\cdot144$; $-0\cdot219$; $-0\cdot014$; $0\cdot318$; $0\cdot644$. These weights, based upon the correlation among outcomes in the exposed group, have been used to smooth the individual effect estimates to generate the overall estimate for the MLMM model.

5. DISCUSSION

Although the MLMM model is less restrictive than the LV model, it provides a flexible framework for analysing multiple outcomes. The approach retains some appealing features of the latent variable model of Sammel and Ryan.[6] For example, it allows associations between the outcomes to influence the way they are combined to create a summary score, but, in contrast to the latent variable approach, the MLMM untangles the mean and variance parameters so that the model is more robust. Through the birth defects illustration we demonstrate that the size of the exposure effect is maintained under various correlation structures.

The approach bears some similarity with Goldstein's multi-level approach to analysing multivariate data (Goldstein, reference 15, Chapters 5). However, the MLMM differs from the Goldstein approach in several important ways. One important difference is the scaling by σ prior to specifying the regression model on each outcome, so that exposure effects are characterized in terms of *effect sizes*. A second difference is the smoothing of exposure effects, as in (7).

Although the MLMM model has fewer assumptions than the LV model of Sammel and Ryan, two strong assumptions remain. The first is the assumption of a single latent variable in each of the two groups. The second is that the exposure effects are homogeneous, and therefore the smoothing of the estimates is valid. Further work is needed to test the validity of these assumptions. There are several other aspects of the MLMM worthy of further research. Questions remain regarding issues of outcome selection, extensions to include non-Gaussian outcomes and mixtures of outcomes, as well as expansions to incorporate longitudinal and/or spatial correlations. It would also be worthwhile to explore the use of robust variance estimates of the kind used in GEE estimation.

APPENDIX: ESTIMATION PROCEDURES

Likelihood equation

Define $\boldsymbol{\Psi} = \text{diag}\{\sigma_j^2\}$, x_i is a block diagonal matrix with $\{x_{ij}^{\mathrm{T}}\}$ ($j = 1, \ldots, J$) on the diagonal and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}}, \cdots, \boldsymbol{\alpha}_J^{\mathrm{T}})^{\mathrm{T}}$. Rewrite (6) and (7) for subject i, the vector of scaled outcomes as

$$y_i^{\mathrm{s}} = \boldsymbol{\Psi}^{-1/2}y_i = x_i\boldsymbol{\alpha} + z_i\mathbf{1}\beta + \boldsymbol{\lambda}a_i + z_i\boldsymbol{\delta}b_i + \boldsymbol{\varepsilon}_i$$

$$= X_i\boldsymbol{\gamma} + \boldsymbol{\lambda}a_i + z_i\boldsymbol{\delta}b_i + \boldsymbol{\varepsilon}_i \tag{9}$$

where $X_i = (x_i, z_i\mathbf{1})$ and $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \beta)^{\mathrm{T}}$. The log-likelihood contribution associated with y_i is

$$l(y_i) = -\tfrac{1}{2}\log|\boldsymbol{\Psi}| + l(y_i^{\mathrm{s}})$$

$$= -\tfrac{1}{2}\log|\boldsymbol{\Psi}| -\tfrac{1}{2}\log|V_i| - \tfrac{1}{2}(y_i^{\mathrm{s}} - X_i\boldsymbol{\gamma})^{\mathrm{T}}V_i^{-1}(y_i^{\mathrm{s}} - X_i\boldsymbol{\gamma}) \tag{10}$$

where $V_i = \boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathrm{T}} + z_i\boldsymbol{\delta}\boldsymbol{\delta}^{\mathrm{T}} + I$. The full log-likelihood is given by $l(y) = \sum_{i=1}^{n}l(y_i)$.

Given (9), the joint distribution of (y_i^s, a_i, b_i) is

$$
\begin{pmatrix} y_i^s \\ a_i \\ b_i \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} x_i\alpha + z_i\mathbf{1}\beta \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda\lambda^{\mathrm{T}} + z_i\delta\delta^{\mathrm{T}} + \mathbf{I} & \lambda & z_i\delta \\ \lambda^{\mathrm{T}} & 1 & 0 \\ z_i\delta^{\mathrm{T}} & 0 & 1 \end{pmatrix} \right].
$$

By re-expressing this in terms of the conditional distribution of y^s given a_i and b_i, it follows that the log-likelihood based on the complete data (y_i, a_i, b_i) is then

$$
\sum_{i=1}^{n} l(y_i, a_i, b_i) = \sum_{i=1}^{n} \left[-\tfrac{1}{2}\log|\Psi| + l(y_i^s|a_i, b_i) + l(a_i) + l(b_i) \right]
$$

$$
\propto -\frac{1}{2}\sum_{i=1}^{n} \left[\log|\Psi| + (y_i^s - X_i\gamma - \lambda a_i - z_i\delta b_i)^{\mathrm{T}}(y_i^s - X_i\gamma - \lambda a_i - z_i\delta b_i) + a_i^2 + b_i^2 \right]. \quad (11)
$$

EM algorithm

Beginning with suitable initial values for the parameter estimates, such as parameters from univariate regressions, iterate between the maximization and expectation steps until convergence.

Maximization step

Given the values of δ, λ and σ, we can estimate the fixed effects parameters from the marginal likelihood of y_i^s directly. Differentiation of (10) with respect to γ gives the score equation for γ as

$$
\sum_{i=1}^{n} X_i^{\mathrm{T}} V_i^{-1}(y_i^s - X_i\gamma) = 0.
$$

Hence

$$
\hat{\gamma} = \left(\sum_{i=1}^{n} X_i^{\mathrm{T}} V_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{n} X_i^{\mathrm{T}} V_i^{-1} y_i^s \right). \quad (12)
$$

The remaining parameters can be obtained by solving the expected score equations obtained from the complete data likelihood, obtained by taking expectations with respect to the posterior distribution of the missing random effects given the observed outcomes and current values of unknown parameters. From equation (11), the score equations are

$$
E\left[\frac{\partial l}{\partial \lambda}\bigg|y_i\right] = \sum_{i=1}^{n} E[a_i(y_i^s - X_i\gamma - \lambda a_i - z_i\delta b_i)|y_i] = 0
$$

$$
E\left[\frac{\partial l}{\partial \delta}\bigg|y_i\right] = \sum_{i=1}^{n} E[z_i b_i(y_i^s - X_i\gamma - \lambda a_i - z_i\delta b_i)|y_i] = 0
$$

$$
E\left[\frac{\partial l}{\partial \sigma_j}\bigg|y_i\right] = \sum_{i=1}^{n} E\left[\left\{ -\frac{1}{\sigma_j} + \frac{1}{\sigma_j}(y_{ij}^s - x_{ij}^{\mathrm{T}}\alpha_j - z_i\beta - \lambda_j a_i - z_i\delta_j b_i)y_{ij}^s \right\}\bigg|y_i\right] = 0
$$

We hence can update (λ, δ) in each M-step using

$$\delta = \frac{1}{\{\sum_{i=1}^n z_i E(b_i^2|\boldsymbol{y}_i)\}} \sum_{i=1}^n z_i [E(b_i|\boldsymbol{y}_i)(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma}) - \lambda E(a_i b_i|\boldsymbol{y}_i)]$$

$$\lambda = \frac{1}{\{\sum_{i=1}^n z_i E(b_i^2|\boldsymbol{y}_i)\}} \sum_{i=1}^n z_i [E(b_i|\boldsymbol{y}_i)(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma}) - \delta E(a_i b_i|\boldsymbol{y}_i)]$$

Since the solution to the equation for σ has no closed form, updates for this parameter can be computed using a one-step Fisher approximation for iteration $t+1$ of the form

$$\sigma_j^{(t+1)} = \sigma_j^{(t)} + \left[\sum_{i=1}^n 1/\sigma_j^{(t)2}(3V_{jj} + \mu_{ij}^2 - 1) \right]^{-1} \sum_{i=1}^n \left[\frac{1}{\sigma_j^{(t)}} y_i^s(y_i^s - \mu_{ij}) - \frac{1}{\sigma_j^{(t)}} \right]$$

where μ_{ij} and V_{jj} are the marginal mean and variance of the unscaled y_j. The calculations are very simple. Evaluation of the right hand sides of the above equations require calculating $E(a_i|\boldsymbol{y}_i)$, $E(b_i|\boldsymbol{y}_i)$, $E(a_i^2|\boldsymbol{y}_i)$, $E(b_i^2|\boldsymbol{y}_i)$, $E(a_i b_i|\boldsymbol{y}_i)$, which are given in the next section.

Expectation step

The conditional distribution of random effects given the observed data for ML estimation is

$$E(a_i, b_i|\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\sigma}) = \begin{bmatrix} \boldsymbol{\lambda}^T V_i^{-1}(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma}) \\ z_i \boldsymbol{\delta}^T V_i^{-1}(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma}) \end{bmatrix}$$

and

$$\text{cov}(a_i, b_i|\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\sigma}) = \begin{bmatrix} 1 - \boldsymbol{\lambda}^T V_i^{-1}\boldsymbol{\lambda} & -z_i \boldsymbol{\lambda}^T V_i^{-1}\boldsymbol{\delta} \\ -z_i \boldsymbol{\delta}^T V_i^{-1}\boldsymbol{\lambda} & 1 - z_i \boldsymbol{\delta}^T V_i^{-1}\boldsymbol{\delta} \end{bmatrix}. \tag{13}$$

For REML estimation the covariance becomes

$$\text{cov}(a_i, b_i|\boldsymbol{y}_i, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\sigma}) = \text{cov}(a_i, b_i|\boldsymbol{y}_i) + (\boldsymbol{\lambda} \ z_i\boldsymbol{\delta})^T V_i^{-1} X_i \left(\sum_{i=1}^n X_i V_i^{-1} X_i \right)^{-1} X_i^T V_i^{-1} (\boldsymbol{\lambda} \ z_i\boldsymbol{\delta}).$$

We then calculate (ML version)

$$E(a_i^2|\boldsymbol{y}_i) = \text{var}(a_i^2|\boldsymbol{y}_i) + \{E(a_i|\boldsymbol{y}_i)\}^2$$

$$= 1 - \boldsymbol{\lambda} V_i^{-1}\boldsymbol{\lambda} + \boldsymbol{\lambda}^T V_i^{-1}(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma})(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma})^T V_i^{-1}\boldsymbol{\lambda}$$

$$E(b_i^2|\boldsymbol{y}_i) = \text{var}(b_i^2|\boldsymbol{y}_i) + \{E(b_i|\boldsymbol{y}_i)\}^2$$

$$= 1 - z_i \boldsymbol{\delta}^T V_i^{-1}\boldsymbol{\delta} + z_i \boldsymbol{\delta}^T V_i^{-1}(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma})(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma})^T V_i^{-1}\boldsymbol{\delta}$$

$$E(a_i b_i|\boldsymbol{y}_i) = \text{cov}(a_i, b_i|\boldsymbol{y}_i) + E(a_i|\boldsymbol{y}_i)E(b_i|\boldsymbol{y}_i)$$

$$= -z_i \boldsymbol{\delta}^T V_i^{-1}\boldsymbol{\lambda} + z_i \boldsymbol{\delta}^T V_i^{-1}(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma})(\boldsymbol{y}_i^s - X_i\boldsymbol{\gamma})^T V_i^{-1}\boldsymbol{\lambda}.$$

REFERENCES

1. Holmes, L. B., Harvey, E. A., Kleiner, B. C., Leppig, K. A., Cann, C. I., Muñoz, A. and Polk, B. F. 'Predictive value of minor anomalies: II. use in cohort studies to identify teratogens', *Teratology*, **36**, 291–297 (1987).
2. Johnson, R. A. and Wichern, D. W. *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey, 1982.
3. Spearman, C. '"General intelligence" objectively determined and measured', *American Journal of Psychology*, **15**, 201–293 (1904).
4. Everitt, B. S. *An Introduction to Latent Variable Models*, Chapman and Hall, New York, 1984.
5. Bentler, P. M. and Weeks, D. G. 'Linear structural equations with latent variables', *Psychometrika*, **45**, 289–308 (1980)
6. Sammel, M. D. and Ryan, L. M. 'Latent variable models with fixed effects', *Biometrics*, **52**, 650–663 (1996).
7. Laird, N. M. and Ware, J. H. 'Random-effects models for longitudinal data', *Biometrics*, **38**, 963–974 (1982).
8. Harville, D. A. 'Maximum likelihood approaches to variance component estimation and to related problems', *Journal of the American Statistical Association*, **72**, 320–340 (1977).
9. Holmes, L. B., Harvey, E. A., Brown, K. S. and Khoshbin, S. 'Anticonvulsant teratogenesis: 1. A study design for newborn infants', *Teratology*, **49**, 202–207 (1994).
10. Liang, K. Y. and Zeger, S. L. 'Longitudinal data analysis using generalized linear models', *Biometrika*, **73**, 13–22 (1986).
11. Zeger, S. L. and Liang, K. Y. 'Longitudinal data analysis for discrete and continuous outcomes', *Biometrics*, **42**, 121–130 (1986).
12. Legler, J. M., Lefkopoulou, M. and Ryan, L. M. 'Efficiency and power of tests for multiple binary outcomes', *Journal of the American Statistical Association*, **90**, 680–693 (1995).
13. Dempster, A. P., Laird, N. M. and Rubin, D. B. 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society*, Series B, **39**, 341–353 (1977).
14. Liu, C. and Rubin, D. B. 'The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence', *Biometrika*, **81**, 633–648 (1994).
15. Goldstein, H. *Multilevel Models in Educational and Social Research*, Oxford University Press, New York, 1987.