

## Latent Variable Models for Longitudinal Data with Multiple Continuous Outcomes

Jason Roy\* and Xihong Lin

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

\* *email*: jasonroy@umich.edu

**SUMMARY.** Multiple outcomes are often used to properly characterize an effect of interest. This paper proposes a latent variable model for the situation where repeated measures over time are obtained on each outcome. These outcomes are assumed to measure an underlying quantity of main interest from different perspectives. We relate the observed outcomes using regression models to a latent variable, which is then modeled as a function of covariates by a separate regression model. Random effects are used to model the correlation due to repeated measures of the observed outcomes and the latent variable. An EM algorithm is developed to obtain maximum likelihood estimates of model parameters. Unit-specific predictions of the latent variables are also calculated. This method is illustrated using data from a national panel study on changes in methadone treatment practices.

**KEY WORDS:** EM algorithm; Factor analysis; Missing data; Multivariate response; Random effects; Repeated measures.

### 1. Introduction

Multiple outcomes are often used to properly characterize an effect of interest. Specifically, the outcome of main interest is often not observable or is difficult to measure. We instead observe several outcomes that, when taken together, characterize the endpoint of interest. For example, in teratology, measurements of multiple adverse effects are needed to identify overall severity of birth defects (Sammel and Ryan, 1996), and in phase II clinical trials, multiple endpoints are necessary to capture treatment efficacy (Pocock, Geller, and Tsiatis, 1987). Interest is often in studying the covariate effects on this unobserved endpoint of interest.

The example that motivates this paper is a national panel study examining changes in methadone treatment practices (D'Aunno, Folz-Murphy, and Lin, 1999). Methadone treatment is important in reducing illicit drug use and preventing HIV transmission and is effective when certain critical treatment practices are followed. These treatment practices include providing adequately large dose levels and long enough treatment duration. For example, Ball and Ross (1991) found that, at a methadone dose level less than 71 mg/d, up to 44% of patients continued to use heroin. Strain (1999) recommended initial dose levels of at least 50 mg/d and dose levels up to 100 mg/d for patients who do not respond to the lower dose range. It is therefore important that the maximum dose level that units are willing to provide reflect these recommendations. Further, studies have shown that better client outcomes are associated with a longer time in treatment and a small percentage of clients receiving progressively smaller doses. In other words, the effectiveness of methadone treatment practices is measured by several variables.

Despite the evidence that certain critical treatment practices must be followed in order for methadone treatment to be effective, several studies conducted in the 1980s showed that the majority of the U.S. methadone treatment units did not use effective treatment practices (Batten et al., 1992). In response to these studies, several national major initiatives were launched to improve treatment practices in the late 1980s and 1990s. For example, the Center for Substance Abuse Treatment developed a set of methadone treatment guidelines for distribution to treatment providers across the nation. In view of these activities, a national panel study was funded by the National Institute of Drug Abuse (NIDA) in 1988 and was conducted at the Institute of Social Research at the University of Michigan. The objectives of this study were (1) to investigate the extent to which overall treatment practices have improved in the last decade, (2) to study what factors affect unit treatment practices, (3) to identify units with substandard treatment practices.

This study sampled 172 methadone treatment units nationwide in 1988. These units were surveyed again in 1990 and 1995. Several measures reflecting the level of overall treatment practices were obtained at each wave. They included unit-average time in treatment, upper limit on doses, and the percentage of clients given decreasing doses. All of these outcomes are attempting to measure the same quantity, i.e., the effective treatment practices level, which is an overall measure of how well each unit is meeting desired standards at a given time. Analysis of this data set is challenging due to the fact that the outcome of major interest, the effective treatment practices level, is not observable, although several surrogates are available, and that the multiple outcomes are measured repeatedly over time.

For cross-sectional data, several authors developed global tests for common dose effects on multiple outcomes (O'Brien, 1984; Pocock et al., 1987). Legler, Lefkopoulou, and Ryan (1995) used the generalized estimating equation (GEE) methods for analyzing multiple binary outcomes. Sammel and Ryan (1996) proposed a latent variable model for multiple continuous outcomes.

For continuous longitudinal data, when only a single outcome is observed, linear mixed models are commonly used (Laird and Ware, 1982). However, limited work has been done on longitudinal data with multiple outcomes. Although separate linear mixed models can be fitted for each outcome, this approach is limited by the fact that it fails to borrow strength across the outcome variables. By exploiting the correlation structure with a multivariate model, efficiency and power could be greatly increased (O'Brien, 1984).

Shah, Laird, and Schoenfeld (1997) extended linear mixed models to allow for multiple outcomes in longitudinal data. They assumed covariates have different effects on different outcomes. Although this assumption is desirable in some applications, it does not account for the feature that the multiple outcomes are attempting to measure the same quantity in the applications we are interested in, such as the treatment practices level in the methadone data example. In their model, covariate effects are tested using many degrees of freedom. Although one could assume a common covariate effect on all outcomes to address this question, this assumption is often not appropriate and is misleading. This is because different outcomes are measured on different scales and different units. It is hence of substantial interest to develop a statistical model to account for this special feature of the data.

Additional limitations of these two existing approaches are that (1) they do not address the question of main interest—the covariate effects on the outcome of main interest, e.g., the covariate effects on the overall effective treatment practices level in the methadone example, and (2) they do not provide an estimate of the unit-specific outcome of main interest, e.g., the overall effective treatment practices level in the methadone example. These estimates are of particular interest in the methadone data example since the investigator is interested in identifying units with substandard treatment practices. Efforts could then be focused on improving the treatment practices of these units.

In this paper, we propose a latent variable model for multiple outcomes measured repeatedly over time. We consider the situation where these outcomes measure an underlying variable of main interest from different perspectives. We relate these observed outcomes using regression models to a latent variable. In particular, we view the observed outcomes as measures of the latent variable with error. A linear mixed model is then assumed to model covariate effects on the unobserved time-dependent latent variable. We allow the number and time of repeated measures to differ between units. This model provides estimates of overall covariate effects on the latent variable, which is the outcome of major interest, and borrows strength across outcomes while taking into account the correlation within each outcome over time and between outcomes on the same unit. This model also yields estimates of the unit-specific latent variables. An EM algorithm is developed to calculate maximum likelihood estimates of model

parameters. Estimates of the unit-specific latent variables are a by-product of the algorithm. We also show how to adapt this algorithm to the situation where a unit's outcomes at a given time may not be fully observed. The method is illustrated using the methadone treatment practices data.

## 2. The Latent Variable Linear Mixed Model

Suppose that, for the  $i$ th of  $n$  units, we observe data at  $K_i$  time points. At the  $k$ th time point,  $t_{ik}$  ( $k = 1, \dots, K_i$ ), we have  $J$  continuous outcomes  $Y_{ijk}$  ( $j = 1, \dots, J$ ), which attempt to characterize a latent outcome of major interest,  $U_{ik}$ , e.g., the treatment practices level in the methadone example. One way to view this problem is that each of the observed outcomes  $Y_{ijk}$  measures the latent variable  $U_{ik}$  with error. It is likely that measurement error on each outcome from the same unit is correlated over time. For example, if the  $j$ th outcome for unit  $i$  measures the latent variable with a lot of error at time 1, it is likely that it also measures the latent variable with a lot of error at the time 2. We hence assume a linear mixed model to relate  $Y_{ijk}$  and  $U_{ik}$  to account for such a correlation. Specifically, we assume

$$Y_{ijk} = \beta_{0j} + U_{ik}\beta_{1j} + b_{ij} + e_{ijk}, \quad (1)$$

where  $\beta_j = (\beta_{0j}, \beta_{1j})^T$  is a vector of regression coefficients for the  $j$ th outcome,  $e_{ijk}$  is distributed as  $N(0, \tau_j^2)$ , and each  $b_{ij}$  is distributed as  $N(0, \xi_j)$ . The random intercept  $b_{ij}$  is used to model possibly correlated measurement errors of  $Y_{ijk}$  for  $U_{ik}$  over time. In other words, it models the within-subject correlation of the repeated measures  $Y_{ijk}$  if  $U_{ik}$  were observed. We further assume the  $b_{ij}$  ( $j = 1, \dots, J$ ) are independent. This model is similar to a factor analysis model (Bartholomew, 1987) for each time point except here we allow an intercept term ( $\beta_{0j}$ ) and assume a single latent factor.

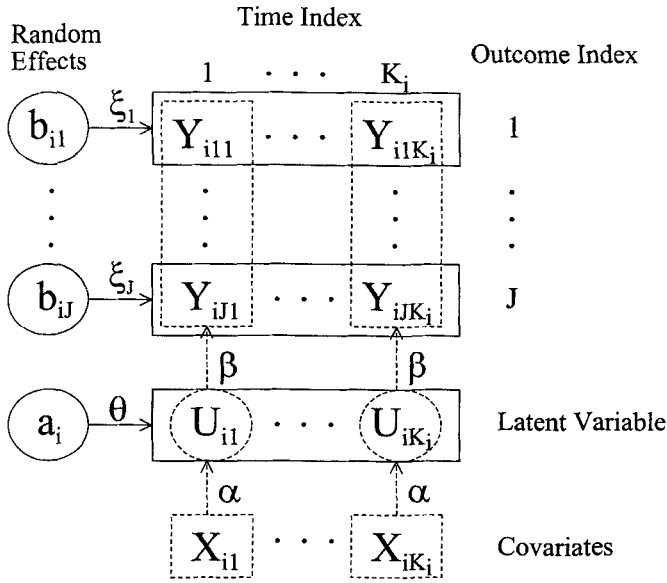
Model (1) contains several implicit assumptions. First, by assuming the  $b_{ij}$  are independent, we are assuming conditional independence of the outcomes given the latent variable; i.e., at each time point, conditional on the latent variable  $U_{ik}$ , the  $J$  outcomes  $Y_{ijk}$  ( $j = 1, \dots, J$ ) are independent. This assumption implies that cross-sectional correlation between the outcomes  $Y_{ijk}$  ( $j = 1, \dots, J$ ) is due entirely to the shared latent variable  $U_{ik}$ . In addition, model (1) assumes the relationship between the latent variable and observed outcomes is constant over time. One way to check this assumption is to compare the factor loadings obtained from fitting a factor analysis model to the outcomes at each time point.

We next assume a linear mixed model to study the effects of covariates on the latent variable  $U_{ik}$ ,

$$U_{ik} = \mathbf{X}_{ik}^T \boldsymbol{\alpha} + \mathbf{Z}_{ik}^T \mathbf{a}_i + \epsilon_{ik}, \quad (2)$$

where  $\mathbf{X}_{ik}$  and  $\boldsymbol{\alpha}$  are  $p \times 1$  vectors of covariates and regression parameters, respectively,  $\mathbf{Z}_{ik}$  is a  $q \times 1$  design vector,  $\mathbf{a}_i$  is a  $q \times 1$  vector of random effects following  $N\{0, \mathbf{D}(\boldsymbol{\theta})\}$ ,  $\mathbf{D}$  is a  $q \times q$  positive definite matrix characterized by a vector of variance components  $\boldsymbol{\theta}$ , and  $\epsilon_{ik}$  is a normally distributed error term independent of  $\mathbf{a}_i$ . The random effects  $\mathbf{a}_i$  are used to model correlation of repeated measures of the latent variable  $U_{ik}$ . For the sake of identifiability, we assume  $\epsilon_{ik}$  is  $N(0, 1)$  and  $\mathbf{X}_{ik}$  does not contain an intercept. This implies that we model a standardized latent variable.

The features of this model can be illustrated using Figure 1. Specifically, the dashed arrows indicate that, at each time



**Figure 1.** Illustration of the structure of the latent variable model. The  $j$ th random intercept  $b_{ij}$  is associated with the  $j$ th outcome vector  $\mathbf{Y}_{ij}$  with variance parameter  $\xi_j$ . The random effects  $\mathbf{a}_i$  are associated with the latent variables  $\mathbf{U}_i$  with covariance parameters  $\theta$ . The association between covariates  $\mathbf{X}_{ik}$  and latent variable  $U_{ik}$  is characterized by  $\alpha$ . The parameters  $\beta$  characterize the association between  $U_{ik}$  and outcomes  $Y_{i1k}, \dots, Y_{iJk}$ .

point, the covariates  $\mathbf{X}_{ik}$  affect the latent variable  $U_{ik}$ , which then affects the  $J$  observed outcomes  $Y_{ijk}$ , and that cross-sectional correlation of the  $J$  outcomes  $(Y_{i1k}, \dots, Y_{iJk})$  is due to the shared latent variable  $U_{ik}$ . The solid arrows show how longitudinal correlation of repeated measures of the observed outcomes and the latent variables is modeled. Specifically, conditional on the latent variable  $\mathbf{U}_i = (U_{i1}, \dots, U_{iK_i})^T$ , longitudinal correlation of repeated measures of the  $j$ th outcome  $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijK_i})^T$  due to measurement error is modeled by the random effect  $b_{ij}$ . Longitudinal correlation of repeated measures of the latent variables  $\mathbf{U}_i$  is modeled by the shared random effects  $\mathbf{a}_i$ .

Models (1) and (2) can be succinctly written using matrix notation. Let  $\mathbf{Y}_i = (\mathbf{Y}_{i1}^T, \dots, \mathbf{Y}_{iJ}^T)^T$ , with  $\mathbf{e}_i$ ,  $\mathbf{X}_i$ , and  $\mathbf{Z}_i$  defined similarly. Let  $\beta_0 = (\beta_{01}, \dots, \beta_{0J})^T$  and  $\beta_1$  be defined similarly. Models (1) and (2) become

$$\mathbf{Y}_i = \mathbf{1}_{K_i} \otimes \beta_0 + \mathbf{U}_i \otimes \beta_1 + \mathbf{1}_{K_i} \otimes \mathbf{b}_i + \mathbf{e}_i, \quad (3)$$

$$\mathbf{U}_i = \mathbf{X}_i \alpha + \mathbf{Z}_i \mathbf{a}_i + \epsilon_i, \quad (4)$$

where  $\otimes$  denotes the Kronecker product,  $\mathbf{b}_i = (b_{i1}, \dots, b_{iJ})^T$  follows  $N(\mathbf{0}, \mathbf{R}(\xi))$  with  $\mathbf{R}(\xi) = \text{diag}(\xi_1, \dots, \xi_J)$ .

Equations (3) and (4) can be thought of as the first and second stages of the model, respectively. It follows that the marginal distribution of  $\mathbf{Y}_i$  is multivariate normal with mean and covariance

$$\begin{aligned} E(\mathbf{Y}_i) &= \boldsymbol{\mu}_i = \mathbf{1}_{K_i} \otimes \beta_0 + \mathbf{X}_i \alpha \otimes \beta_1 \\ \text{cov}(\mathbf{Y}_i) &= \mathbf{V}_i = \left( \mathbf{I}_{K_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T \right) \otimes \left( \beta_1 \beta_1^T \right) \\ &\quad + \left( \mathbf{1}_{K_i} \mathbf{1}_{K_i}^T \right) \otimes \mathbf{R} \end{aligned} \quad (5)$$

$$+ \text{diag} \left( \tau_1^2 \mathbf{I}_{K_i}, \dots, \tau_J^2 \mathbf{I}_{K_i} \right). \quad (6)$$

The first term in  $\text{cov}(\mathbf{Y}_i)$  comes from the covariance of the latent variable vector  $\mathbf{U}_i$ , the second term comes from the covariance of the stage-1 random intercept vector  $\mathbf{b}_i$ , and the last term is from the covariance of the stage-1 random error vector  $\mathbf{e}_i$ . Hence, marginal correlation of the outcome vector  $\mathbf{Y}_i$  is modeled by cross-sectional correlation due to the shared latent variable  $\mathbf{U}_i$  and by longitudinal correlation due to the random effects  $\mathbf{b}_i$  and  $\mathbf{a}_i$ .

Notice that the regression coefficient vector  $\beta_1$  enters into both the marginal mean and the marginal covariance of  $\mathbf{Y}_i$ . Direct calculations of the maximum likelihood estimator are hence difficult. An EM algorithm is developed in the next section.

It is often of substantial interest to estimate the unit-specific latent variables  $\mathbf{U}_i$ . For example, in the methadone example, the latent variables  $\mathbf{U}_i$  represent the unit-specific treatment practices score at the three waves and can be used to identify which units use substandard treatment practices and need to improve. Since  $(\mathbf{Y}_i, \mathbf{U}_i)$  are jointly normally distributed, we can easily estimate the latent variables  $\mathbf{U}_i$  using the posterior means,

$$E(\mathbf{U}_i | \mathbf{Y}_i) = \mathbf{X}_i \alpha + \Lambda_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i), \quad (7)$$

where  $\Lambda_i = (\mathbf{I}_{K_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T) \otimes \beta_1$ . Its covariance is

$$\text{cov}(\mathbf{U}_i | \mathbf{Y}_i) = \left( \mathbf{I}_{K_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T \right) - \Lambda_i^T \mathbf{V}_i^{-1} \Lambda_i. \quad (8)$$

Both  $E(\mathbf{U}_i | \mathbf{Y}_i)$  and  $\text{cov}(\mathbf{U}_i | \mathbf{Y}_i)$  can be easily calculated as a by-product of the EM algorithm derived in Section 3.

### 3. Estimation Using the EM Algorithm

We discuss in this section calculations of maximum likelihood estimates of the model parameters using the EM algorithm (Dempster, Laird, and Rubin, 1977). As discussed before, maximizing the actual log likelihood directly would be difficult since  $\beta_1$  enters into both the marginal mean and covariance of  $\mathbf{Y}_i$ . An advantage of using the EM algorithm for estimation in the latent variable linear mixed models (3) and (4) is that it takes advantage of the conditional independence of  $\mathbf{Y}_i$  given  $\mathbf{U}_i$  and  $\mathbf{b}_i$ .

We treat in the EM algorithm the latent variables  $\mathbf{U}_i$  and the random effects  $(\mathbf{a}_i, \mathbf{b}_i)$  as missing data. Hence, the complete data are  $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{U}_i, \mathbf{a}_i, \mathbf{b}_i)$  and the observed data are  $(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{Z}_i)$ . It follows that the complete-data log likelihood is, apart from a constant,

$$\begin{aligned} \ell(\beta_0, \beta_1, \alpha, \tau^2, \theta, \xi) &= \sum_{i=1}^n \{ \ell(\mathbf{Y}_i | \mathbf{U}_i, \mathbf{b}_i; \beta, \tau) + \ell(\mathbf{U}_i | \mathbf{a}_i; \alpha) \\ &\quad + \ell(\mathbf{a}_i; \theta) + \ell(\mathbf{b}_i; \xi) \}, \end{aligned} \quad (9)$$

where  $\tau^2 = (\tau_1^2, \dots, \tau_J^2)$  and

$$\begin{aligned} \ell(\mathbf{Y}_i | \mathbf{U}_i, \mathbf{b}_i; \beta, \tau) &= \sum_{j=1}^J \left\{ -\frac{K_i}{2} \log \tau_j^2 \right. \\ &\quad \left. - \frac{1}{2\tau_j^2} (\mathbf{Y}_{ij} - \beta_{0j} \mathbf{1}_{K_i} \right. \end{aligned}$$

$$\begin{aligned} & - \mathbf{U}_i \beta_{1j} - b_{ij} \mathbf{1}_{K_i})^T \\ & \times (\mathbf{Y}_{ij} - \beta_{0j} \mathbf{1}_{K_i} \\ & - \mathbf{U}_i \beta_{1j} - b_{ij} \mathbf{1}_{K_i}) \} \\ \ell(\mathbf{U}_i | \mathbf{a}_i; \boldsymbol{\alpha}) &= -\frac{1}{2} (\mathbf{U}_i - \mathbf{X}_i \boldsymbol{\alpha} - \mathbf{Z}_i \mathbf{a}_i)^T \\ & \quad \times (\mathbf{U}_i - \mathbf{X}_i \boldsymbol{\alpha} - \mathbf{Z}_i \mathbf{a}_i) \\ \ell(\mathbf{a}_i; \boldsymbol{\theta}) &= -\frac{1}{2} \log | \mathbf{D}(\boldsymbol{\theta}) | - \frac{1}{2} \mathbf{a}_i^T \mathbf{D}^{-1} \mathbf{a}_i \\ \ell(\mathbf{b}_i; \boldsymbol{\xi}) &= -\frac{1}{2} \log | \mathbf{R}(\boldsymbol{\xi}) | - \frac{1}{2} \mathbf{b}_i^T \mathbf{R}^{-1} \mathbf{b}_i. \end{aligned}$$

The EM algorithm proceeds by first assuming initial estimates of the parameters, then iterating between an E-step and an M-step. The E-step takes expectations of the sufficient statistics of the complete-data log likelihood, given the observed data. The M-step maximizes the expected complete-data log likelihood conditional on the other parameters being fixed at their current values.

*E-step.* The E-step consists of the following steps:

- (1) Calculate the conditional expectations for the sufficient statistics involving the latent variable  $\mathbf{U}_i$  as  $\tilde{\mathbf{U}}_i = E(\mathbf{U}_i | \mathbf{Y}_i)$  and  $E(\mathbf{U}_i^T \mathbf{U}_i | \mathbf{Y}_i) = \tilde{\mathbf{U}}_i^T \tilde{\mathbf{U}}_i + \text{tr}\{\text{cov}(\mathbf{U}_i | \mathbf{Y}_i)\}$ , where the expressions of  $E(\mathbf{U}_i | \mathbf{Y}_i)$  and  $\text{cov}(\mathbf{U}_i | \mathbf{Y}_i)$  are given in (7) and (8).
- (2) Calculate the conditional expectations of the sufficient statistics involving the stage-2 random effects  $\mathbf{a}_i$  as  $\tilde{\mathbf{a}}_i = E(\mathbf{a}_i | \mathbf{Y}_i) = (\mathbf{Z}_i \mathbf{D} \otimes \boldsymbol{\beta}_1)^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)$  and  $E(\mathbf{a}_i \mathbf{a}_i^T | \mathbf{Y}_i) = \tilde{\mathbf{a}}_i \tilde{\mathbf{a}}_i^T + \mathbf{D} - (\mathbf{Z}_i \mathbf{D} \otimes \boldsymbol{\beta}_1)^T \mathbf{V}_i^{-1} (\mathbf{Z}_i \mathbf{D} \otimes \boldsymbol{\beta}_1)$ .
- (3) Calculate the conditional expectations of the sufficient statistics involving the stage-1 random effects  $\mathbf{b}_i$  as  $\tilde{\mathbf{b}}_i = E(\mathbf{b}_i | \mathbf{Y}_i) = (\mathbf{1}_{K_i} \otimes \mathbf{R})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)$  and  $E(\mathbf{b}_i \mathbf{b}_i^T | \mathbf{Y}_i) = \tilde{\mathbf{b}}_i \tilde{\mathbf{b}}_i^T + \mathbf{R} - (\mathbf{1}_{K_i} \otimes \mathbf{R})^T \mathbf{V}_i^{-1} (\mathbf{1}_{K_i} \otimes \mathbf{R})$ .
- (4) Calculate the conditional expectation of  $(\mathbf{U}_i \mathbf{b}_i^T | \mathbf{Y}_i)$ . Since  $(\mathbf{Y}_i, \mathbf{U}_i, \mathbf{b}_i)$  are multivariate normal, we find that  $\text{cov}(\mathbf{U}_i, \mathbf{b}_i | \mathbf{Y}_i) = -\boldsymbol{\Lambda}_i^T \mathbf{V}_i^{-1} (\mathbf{1}_{K_i} \otimes \mathbf{R})$ .

*M-step.* We maximize the expected complete-data log likelihood for the parameters  $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\tau}^2)$  at the M-step, which consists of the following steps (derived in the Appendix):

- (1) Let  $\boldsymbol{\beta}_j = (\beta_{0j}, \boldsymbol{\beta}_{1j})^T$  and  $\mathbf{U}_{i*} = (\mathbf{1}_{K_i}, \mathbf{U}_i)$ . Then the estimate of  $\boldsymbol{\beta}_j$  ( $j = 1, \dots, J$ ) is updated by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_j &= \left\{ \sum_{i=1}^n E \left( \mathbf{U}_{i*}^T \mathbf{U}_{i*} | \mathbf{Y}_i \right) \right\}^{-1} \\ & \quad \times \sum_{i=1}^n E \left\{ \mathbf{U}_{i*}^T (\mathbf{Y}_{ij} - b_{ij} \mathbf{1}_{K_i}) | \mathbf{Y}_i \right\}, \end{aligned}$$

whose detailed expression is given in Appendix A.

- (2) The estimates of the  $\tau_j$  are updated by

$$\begin{aligned} \hat{\tau}_j^2 &= \frac{1}{\sum_{i=1}^n K_i} \\ & \quad \times \sum_{i=1}^n E \left\{ (\mathbf{Y}_{ij} - \mathbf{U}_{i*} \boldsymbol{\beta}_j - b_{ij} \mathbf{1}_{K_i})^T \right. \end{aligned}$$

$$\left. \times (\mathbf{Y}_{ij} - \mathbf{U}_{i*} \boldsymbol{\beta}_j - b_{ij} \mathbf{1}_{K_i}) | \mathbf{Y}_i \right\},$$

whose detailed expression is given in Appendix A.

- (3) The estimate of  $\boldsymbol{\xi}$  is updated by  $\hat{\boldsymbol{\xi}}_j = (1/n) \sum_{i=1}^n E(b_{ij}^2 | \mathbf{Y}_i)$  for  $j = 1, \dots, J$ , where  $E(b_{ij}^2 | \mathbf{Y}_i)$  is the  $j$ th diagonal element of the matrix  $E(\mathbf{b}_i \mathbf{b}_i^T | \mathbf{Y}_i)$ .
- (4) The estimate of  $\boldsymbol{\alpha}$  is updated by  $\hat{\boldsymbol{\alpha}} = (\sum_{i=1}^n \mathbf{X}_i^T \mathbf{X}_i)^{-1} \times \sum_{i=1}^n \mathbf{X}_i^T (\tilde{\mathbf{U}}_i - \mathbf{Z}_i \tilde{\mathbf{a}}_i)$ .
- (5) If  $\mathbf{D}(\boldsymbol{\theta})$  is an arbitrary covariance matrix, then it can simply be estimated by  $\hat{\mathbf{D}} = (1/n) \sum_{i=1}^n E(\mathbf{a}_i \mathbf{a}_i^T | \mathbf{Y}_i)$ . Otherwise,  $\boldsymbol{\theta}$  can be estimated iteratively using the Fisher scoring algorithm (see Appendix A).

Maximum likelihood estimates are obtained by iterating between the outlined E- and M-steps until convergence. The covariance matrix of the parameter estimates is obtained by directly calculating the inverse of the Fisher information matrix using the marginal log likelihood of  $\mathbf{Y}_i$  at convergence (see Appendix B). The estimates  $E(\mathbf{U}_i | \mathbf{Y}_i)$  calculated from the E-step at convergence are used to estimate the unit-specific treatment practices scores in the methadone data example.

So far, we have assumed that all  $J$  outcomes are observed at each of the  $K_i$  times for every unit  $i$ . However, it is not uncommon in practice to have some outcomes missing at a given time. We will now show how the above algorithm can easily be adapted to the situation where not all of the outcomes are observed for some units at any given time. We here assume missing outcomes are missing at random (Little and Rubin, 1987).

For unit  $i$ , suppose we have  $K_{ij} \leq K_i$  repeated measures on the  $j$ th outcome. Let  $\mathbf{T}_{ij}$  be an index matrix to indicate the time points when the  $j$ th outcome is observed. Specifically,  $\mathbf{T}_{ij}$  is a  $K_{ij} \times K_i$  matrix constructed by deleting rows of  $I_{K_i}$  that correspond to missing observations on the  $j$ th outcome. For example, if unit  $i$  is only missing the second repeated measure on the  $j$ th outcome, then  $\mathbf{T}_{ij}$  is the same as  $I_{K_i}$  but with the second row removed. We then can write the observed values  $\mathbf{Y}_{ij}^o$  as  $\mathbf{Y}_{ij}^o = \mathbf{T}_{ij} \mathbf{Y}_{ij}$ . Let  $\mathbf{T}_i = \text{diag}(\mathbf{T}_{i1}, \dots, \mathbf{T}_{iJ})$ , then  $\mathbf{Y}_i^o = (\mathbf{Y}_{i1}^{oT}, \dots, \mathbf{Y}_{iJ}^{oT})^T = \mathbf{T}_i \mathbf{Y}_i$ .

The E-step is the same as before except that one replaces  $(\mathbf{Y}_i, \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i, \mathbf{V}_i, \mathbf{1}_{K_i} \otimes \mathbf{R}, \mathbf{Z}_i \mathbf{D} \otimes \boldsymbol{\beta})$  by  $\{\mathbf{Y}_i^o, \mathbf{T}_i \boldsymbol{\mu}_i, \mathbf{T}_i \boldsymbol{\Lambda}_i, \mathbf{T}_i \mathbf{V}_i \mathbf{T}_i^T, \mathbf{T}_i (\mathbf{1}_{K_i} \otimes \mathbf{R}), \mathbf{T}_i (\mathbf{Z}_i \mathbf{D} \otimes \boldsymbol{\beta})\}$ . All expectations in the M-step are taken conditional on the observed data  $\mathbf{Y}_i^o$ , and estimates of  $\boldsymbol{\beta}_j$  and  $\tau_j$  in the first two steps of the M-step need to be modified as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_j &= \left\{ \sum_{i=1}^n E \left( \mathbf{U}_{i*}^T \mathbf{T}_{ij}^T \mathbf{T}_{ij} \mathbf{U}_{i*} | \mathbf{Y}_i \right) \right\}^{-1} \\ & \quad \times \sum_{i=1}^n E \left\{ \mathbf{U}_{i*}^T \mathbf{T}_{ij}^T (\mathbf{Y}_{ij}^o - b_{ij} \mathbf{T}_{ij} \mathbf{1}_{K_i}) | \mathbf{Y}_i^o \right\} \\ \hat{\tau}_j^2 &= \frac{1}{\sum_{i=1}^n K_i} \\ & \quad \times \sum_{i=1}^n E \left[ \left\{ \mathbf{Y}_{ij}^o - \mathbf{T}_{ij} (\mathbf{U}_{i*} \boldsymbol{\beta}_j + b_{ij} \mathbf{1}_{K_i}) \right\}^T \right. \\ & \quad \left. \times \left\{ \mathbf{Y}_{ij}^o - \mathbf{T}_{ij} (\mathbf{U}_{i*} \boldsymbol{\beta}_j + b_{ij} \mathbf{1}_{K_i}) \right\} | \mathbf{Y}_i^o \right]. \end{aligned}$$

**Table 1**  
Summary statistics of the three outcome variables by year for the methadone data

Outcome	1988		1990		1995	
	Mean	SD	Mean	SD	Mean	SD
Upper limit dose	77.95	21.97	81.90	20.57	93.00	25.19
Time in treatment (months)	20.74	9.30	19.19	10.09	21.49	9.64
Percent decreasing dose	33.68	29.26	30.41	29.91	21.91	20.17

#### 4. Application to the Methadone Treatment Practices Data

We applied our method to the methadone treatment practices data described in the Introduction. The data come from a U.S. national panel study of the nation's methadone maintenance units (D'Aunno et al., 1999). The first wave data were collected in 1988 by selecting a random sample of 172 units nationwide. Unit directors and clinical supervisors of these participating units were surveyed with regard to a variety of client and unit characteristics. These units were contacted again in 1990 and 1995, and 140 units and 116 units participated, respectively. We assume in our analysis that missing data are missing at random (Little and Rubin, 1987). For more on this assumption, see the Discussion.

The interest of the study is in determining (1) how well the treatment practices of methadone maintenance units are meeting desired standards, (2) how the treatment practices have changed over time, (3) what covariates are predictive of effective practices, and (4) which units use ineffective treatment practices.

Several outcome variables, including maximum dose level ( $Y_1$ ), unit-average length of treatment ( $Y_2$ ), and percent of clients receiving decreasing doses ( $Y_3$ ), were collected at each wave. These variables have all been shown to be important in evaluating the treatment practices of methadone maintenance units. Specifically, maximum dose levels of less than 80 mg per day (mg/d), a large percentage of clients given progressively smaller doses, and treatment lengths of less than several months are all considered ineffective treatment practices. Table 1 gives the summary statistics of these treatment practices measures. It suggests that the treatment practices

improved slightly from 1988 to 1990 and improved substantially in 1995, especially when measured by maximum dose level and percent of clients receiving decreasing doses.

A log transformation was performed on the maximum dose and percent decreasing dose variables to make the normality assumption more plausible. The covariates found to be most important and therefore used in our final model were unit average age of clients (years), percentage of black clients, percentage of staff that are ex-addicts, whether unit is private-for-profit (yes/no), region (midwest, southwest, or northeast), and year (1988, 1990, 1995). Dummy variables were created for the region and year variables by using northeast and 1988 as the references, respectively.

We first conducted a preliminary analysis of the data by using the standard approach of modeling each outcome separately. We fit a random intercept model for each outcome using SAS PROC MIXED. The results are given in Table 2. Both maximum dose level and the percentage of clients receiving progressively smaller doses improved significantly in 1995 compared with 1988. Units with a larger percentage of black clients were associated with smaller upper limit dose levels, shorter treatment duration, and a larger percentage of clients receiving decreasing doses. A larger percentage of clients receiving decreasing doses was associated with a lower average age of clients and midwest units. Southwest units tended to have a smaller upper dose limit. Other variables were not significantly associated with any of the outcomes.

As discussed previously, conducting separate analyses for each outcome does not account for the fact that these outcomes are measuring the same underlying quantity, i.e., overall effective treatment practices level. In addition, it does not

**Table 2**  
Parameter estimates from fitting a random intercept model to each outcome separately<sup>a</sup>

Covariate	Log(upper limit dose)	Time in treatment	Log(percent decreasing dose)
Intercept	4.42 (0.10)	17.49 (3.67)	3.63 (0.47)
Percent black	-0.002 (0.001)	-0.05 (0.03)	0.007 (0.003)
Age	0.002 (0.003)	0.18 (0.11)	-0.03 (0.01)
Staff ex-addicts	-0.31 (0.13)	-3.81 (4.61)	0.50 (0.58)
Private-for-profit	-0.05 (0.08)	-3.48 (2.41)	0.37 (0.30)
Midwest	-0.10 (0.06)	-0.46 (1.87)	0.58 (0.24)
Southwest	-0.09 (0.04)	-0.23 (1.32)	0.18 (0.17)
1990	0.05 (0.03)	-1.75 (1.02)	-0.19 (0.13)
1995	0.16 (0.03)	1.09 (1.21)	-0.35 (0.15)
Random intercept	0.03 (0.007)	22.93 (6.83)	0.36 (0.11)

<sup>a</sup> The numbers in parentheses are standard errors.

**Table 3**  
Stage-1 parameter estimates and their estimated standard errors

Outcome	$\hat{\beta}_{0j}$	$\hat{\beta}_{1j}$	$\tau_j^2$	$\xi_j$
Log(upper limit dose)	4.31(0.08)	0.06 (0.03)	0.04 (0.01)	0.02 (0.01)
Time in treatment	19.40 (1.64)	1.13 (0.56)	73.07 (7.47)	11.60 (6.33)
Log(percent decreasing dose)	3.10 (0.31)	-0.22 (0.10)	0.88 (0.10)	0.24 (0.11)

produce estimates of the covariate effects on the overall effective treatment practices level. We therefore applied the latent variable model to the methadone data. We first assumed at stage 1 of the model that the three outcomes are measuring the latent effective treatment practices score ( $U$ ) with error. At stage 2 of the model, we were interested in how the covariates ( $\mathbf{X}$ ) (described above) were associated with the latent treatment practices score ( $U$ ). There we assumed a random intercept model, i.e.,  $\mathbf{Z}_i = \mathbf{1}$  in equation (4). A SAS program written in IML was used to implement the EM algorithm and calculate the standard errors.

Table 3 shows the stage-1 parameter estimates. Our analysis showed that the outcomes time in treatment and maximum dose level were positively associated with the latent effective practices score, while the percentage of clients receiving decreasing dose was negatively associated with this latent variable. These findings agree with previous research (D'Aunno et al., 1999).

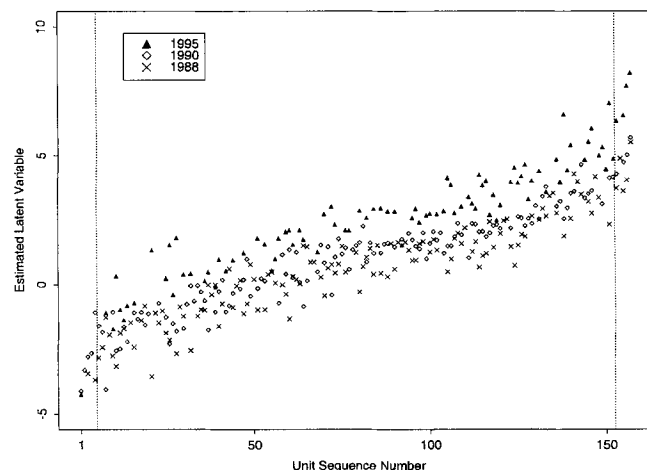
Table 4 gives the estimates of the stage-2 parameters in the linear mixed model (4) for the latent variable  $U_i$ . The results show that treatment practices did not change significantly from 1988 to 1990 ( $p$ -value = 0.22) but significantly improved in 1995 ( $p$ -value = 0.046). Units having more black clients used less effective treatment practices, while units having older clients had better practices. Northeast units had better practices compared to midwest and southwest units. Having a higher percentage of staff ex-addicts was associated with less effective treatment practices. These covariate effects were, however, either marginally significant or nonsignificant. We also examined interactions between the covariates and the time dummy variables, and none of the interaction terms were significant.

It is of substantial interest to identify units whose treatment practices effectiveness are well below those of a typical unit. Hence, efforts to improve treatment practices could fo-

cus on these units. A feature of our analysis is that estimates of the latent effective practices score via the posterior mean in equation (7) can be used to identify these units. Figure 2 shows the estimated latent variable over time for each unit. The plot is sorted by ordering the units according to average (over time) of estimated latent scores, so units with the lowest average estimated treatment practices scores over three waves are on the left side of the graph. We included vertical lines at the 2.5 and 97.5 percentiles so that units with treatment practices well below or above average can easily be identified. These units are indicated in the description of Figure 2.

**5. Discussion**

We have proposed in this paper a latent variable model for the situation where several outcomes are measured over time and are all attempting to measure the same unobservable quantity. We assume that these multiple outcomes are related to a latent variable, and we model the covariates effects on the latent variable using a random effects model. An EM algorithm was developed to estimate the model parameters. Unit-specific la-



**Figure 2.** Estimated effective practices scores over time plotted for each unit. The horizontal axis gives the unit sequence number resulting from ordering the units according to average (over time) of estimated latent scores. Units with the least effective practices are on the left-hand side of the plot. The dashed vertical lines indicate the 2.5 and 97.5 percentiles; i.e., units whose treatment practices can be viewed as substandard (units that have an average estimated latent variable less than 97% of the units in this study) appear to the left of the first vertical line (units with ID = 30020, 67, 30462, 30270, and 46). Units to the right of the second vertical line can be viewed as using more effective treatment practices (units with ID = 30440, 30218, 30386, 30094, and 30054).

**Table 4**  
Stage-2 parameter estimates and their estimated standard errors

Parameter	Estimate	$\widehat{SE}$	$P$ -value
$\alpha_1$ (percent black)	-0.03	0.02	0.065
$\alpha_2$ (age)	0.09	0.05	0.112
$\alpha_3$ (staff ex-addicts)	-4.07	2.52	0.107
$\alpha_4$ (private-for-profit)	-1.21	1.09	0.265
$\alpha_5$ (midwest)	-2.09	1.22	0.088
$\alpha_6$ (southwest)	-1.31	0.80	0.103
$\alpha_7$ (1990)	0.53	0.42	0.217
$\alpha_8$ (1995)	2.03	1.01	0.046
$\theta$	3.77	4.01	

tent variables were obtained as a byproduct of the algorithm. The proposed EM algorithm is easy to implement; however, it could converge slowly. It is worth further research to develop a more efficient algorithm to fit such models.

Our model provides a straightforward way to estimate and test for the global covariate effects since the parameters  $\alpha$  have global interpretations, e.g., they represent the effects of the covariates on the overall effective treatment practices level in the methadone data. Global testing has been shown to be more efficient when analyzing cross-sectional data with multiple outcomes (O'Brien, 1984; Legler et al., 1995). Analogous results would hold for longitudinal data with multiple outcomes. This is because global testing borrows information across multiple outcomes by modeling their correlation; it exploits the nature of the data—that multiple outcomes tend to measure the same underlying quantity. Hence, greater efficiency could be gained when the covariates are modeled for their overall effects by using many fewer parameters. A challenge in performing global testing for continuous outcomes is that the outcomes are often measured at different scales and units. Our latent variable model provides a framework for addressing this issue and performing global testing.

Our method assumes that missing data are missing at random. For the methadone data, missing data take two forms, i.e., intermittent missing outcomes and missing values due to unit dropout. There are very few intermittent missing values in this study and therefore possible nonignorability would likely not have much of an effect on inferences. However, 33% of the units dropped out of the study before study completion. The missing-at-random assumption implies that the probability the units dropped out at the current wave depends on the treatment practices outcomes at the previous wave but not on the current unobserved outcomes. An analysis examining the possibility of data being not missing at random is reported elsewhere (Roy and Lin, 1999).

We consider in this paper the situation where the multiple outcomes are continuous. In some situations, however, both continuous and discrete endpoints may be necessary to fully characterize the latent outcome. Therefore, an area in need of further research is the extension of this model to allow for several continuous and discrete outcomes.

#### ACKNOWLEDGEMENTS

This work was supported in part by U.S. National Cancer Institute grant R29 CA76404 and U.S. National Institute of Drug Abuse grant RO1 DA03272.

#### RÉSUMÉ

Des réponses multiples sont souvent utilisées pour caractériser proprement un effet d'intérêt. Ce papier propose un modèle à variable latente pour les situations où des mesures répétées dans le temps pour chaque réponse. Ces réponses sont considérées comme mesurant une quantité sous-jacente d'intérêt principal sous différentes perspectives. Nous associons les réponses observées en utilisant des modèles de régression à la variable latente, qui est elle-même modélisée comme une fonction des covariables par un modèle de régression séparé. Des effets aléatoires sont utilisés pour modéliser la corrélation entraînée par les mesures répétées sur les réponses observées et

par la variable latente. Un algorithme EM est développé pour obtenir les estimations du maximum de vraisemblance des paramètres du modèle. Les prédictions spécifiques aux unités de la variable latente sont aussi calculées. Cette méthode est illustrée en utilisant des données d'une étude sur un panel national au sujet des changements dans les pratiques de traitement par méthadone.

#### REFERENCES

- Ball, J. C. and Ross, A. (1991). *The Effectiveness of Methadone Maintenance Treatment*. New York: Springer-Verlag.
- Bartholomew, D. J. (1987). *Latent Variable Models and Factor Analysis*. Oxford: Oxford University Press.
- Batten, H. L., Prottas, J. M., Horgan, C. M., Simon, L. J., Larson, M. J., Elliott, E. A., and Marsden, M. E. (1992). *Drug services research survey*. Final Report: Phase II, Institute for Health Policy, Brandeis University, National Institute on Drug Abuse.
- D'Aunno, T., Folz-Murphy, N., and Lin, X. (1999). Variation in methadone treatment practices: Results from a national panel study, 1988–1995. *The American Journal of Drug and Alcohol Abuse* **25**, 681–699.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 341–353.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.
- Legler, J., Lefkopoulou, M., and Ryan, L. (1995). Efficiency and power of tests for multiple binary outcomes. *Journal of the American Statistical Association* **90**, 680–693.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley.
- O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* **40**, 1079–1087.
- Pocock, S. T., Geller, N. L., and Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials. *Biometrics* **43**, 487–498.
- Roy, J. (2000). Latent variable models for longitudinal data with multiple outcomes, informative dropouts, and missing covariates. Ph.D. dissertation, University of Michigan, Ann Arbor.
- Sammel, M. D. and Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics* **52**, 650–663.
- Shah, A., Laird, N., and Schoenfeld, D. (1997). A random-effects model for multiple characteristics with possibly missing data. *Journal of the American Statistical Association* **92**, 775–779.
- Strain, E. C. (1999). Methadone dose during maintenance treatment. In *Methadone Treatment for Opioid Dependence*, E. Strain and M. Stitzer (eds), 62–85. Baltimore: The Johns Hopkins University Press.

Received June 1999. Revised March 2000.

Accepted April 2000.

APPENDIX A

Detailed Calculations in the M-Step of the EM Algorithm

Differentiating the complete-data log likelihood (9) with respect to  $\beta_j$  gives

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \mathbf{U}_{i*}^T (\mathbf{Y}_{ij} - \mathbf{U}_{i*} \beta_j - b_{ij} \mathbf{1}_{K_i}).$$

Setting it to zero and taking an expectation conditional on the observed data gives

$$\begin{aligned} \hat{\beta}_j &= \left[ \sum_{i=1}^n \mathbf{E} \left( \mathbf{U}_{i*}^T \mathbf{U}_{i*} \mid \mathbf{Y}_i \right) \right]^{-1} \\ &\quad \times \sum_{i=1}^n \mathbf{E} \left[ \tilde{\mathbf{U}}_{i*}^T (\mathbf{Y}_{ij} - b_{ij} \mathbf{1}_{K_i}) \mid \mathbf{Y}_i \right] \\ &= \left[ \sum_{i=1}^n \begin{pmatrix} K_i & \mathbf{1}_{K_i}^T \tilde{\mathbf{U}}_i \\ \mathbf{1}_{K_i}^T \tilde{\mathbf{U}}_i & \mathbf{E}(\mathbf{U}_i^T \mathbf{U}_i \mid \mathbf{Y}_i) \end{pmatrix} \right]^{-1} \\ &\quad \times \sum_{i=1}^n \begin{pmatrix} \mathbf{1}_{K_i}^T (\mathbf{Y}_{ij} - \tilde{b}_{ij} \mathbf{1}_{K_i}) \\ \tilde{\mathbf{U}}_i^T (\mathbf{Y}_{ij} - \tilde{b}_{ij} \mathbf{1}_{K_i}) - \mathbf{1}_{K_i}^T \text{cov}(\mathbf{U}_i, b_{ij} \mid \mathbf{Y}_i) \end{pmatrix}. \end{aligned}$$

Maximizing (9) with respect to  $\xi$  is equivalent to maximizing

$$-\sum_{i=1}^n \sum_{j=1}^J \frac{1}{2} \log(\xi_j) - \sum_{i=1}^n \sum_{j=1}^J \frac{1}{2\xi_j} b_{ij}^2$$

with respect to  $\xi_j$  for each  $j$ . Therefore, we have, for  $j = 1, \dots, J$ ,  $\hat{\xi}_j = (1/n) \sum_{i=1}^n \mathbf{E}(b_{ij}^2 \mid \mathbf{Y}_i)$ .

The partial derivative with respect to  $\tau_j^2$  is

$$\begin{aligned} \frac{\partial \ell}{\partial \tau_j^2} &= -\sum_{i=1}^n \frac{K_i}{2\tau_j^2} + \frac{1}{2(\tau_j^2)^2} \sum_{i=1}^n (\mathbf{Y}_{ij} - \mathbf{U}_{i*} \beta_j - b_{ij} \mathbf{1}_{K_i})^T \\ &\quad \times (\mathbf{Y}_{ij} - \mathbf{U}_{i*} \beta_j - b_{ij} \mathbf{1}_{K_i}), \end{aligned}$$

which implies that

$$\begin{aligned} \hat{\tau}_j^2 &= \frac{1}{\sum_{i=1}^n K_i} \sum_{i=1}^n \mathbf{E} \left[ (\mathbf{Y}_{ij} - \mathbf{U}_{i*} \beta_j - b_{ij} \mathbf{1}_{K_i})^T \right. \\ &\quad \left. \times (\mathbf{Y}_{ij} - \mathbf{U}_{i*} \beta_j - b_{ij} \mathbf{1}_{K_i}) \mid \mathbf{Y}_i \right] \\ &= \frac{1}{\sum_{i=1}^n K_i} \left\{ \sum_{i=1}^n (\mathbf{Y}_{ij} - \beta_{0j} \mathbf{1}_{K_i} - \tilde{\mathbf{U}}_i \beta_{1j} - \tilde{b}_{ij} \mathbf{1}_{K_i})^T \right. \\ &\quad \times (\mathbf{Y}_{ij} - \beta_{0j} \mathbf{1}_{K_i} - \tilde{\mathbf{U}}_i \beta_{1j} - \tilde{b}_{ij} \mathbf{1}_{K_i}) \\ &\quad \left. + \left[ \beta_{1j}^2 \text{tr}(\text{var}(\mathbf{U}_i \mid \mathbf{Y}_i)) + K_i \text{var}(b_{ij} \mid \mathbf{Y}_i) \right. \right. \\ &\quad \left. \left. + 2\beta_{1j} \mathbf{1}_{K_i}^T \text{cov}(\mathbf{U}_i, b_{ij} \mid \mathbf{Y}_i) \right] \right\}. \end{aligned}$$

Differentiating (9) with respect to  $\alpha$  gives

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^n \mathbf{X}_i^T (\mathbf{U}_i - \mathbf{X}_i \alpha - \mathbf{Z}_i \mathbf{a}_i).$$

It follows that  $\hat{\alpha} = (\sum_{i=1}^n \mathbf{X}_i^T \mathbf{X}_i)^{-1} \sum_{i=1}^n \mathbf{X}_i^T (\tilde{\mathbf{U}}_i - \mathbf{Z}_i \tilde{\mathbf{a}}_i)$ .

Finally, if  $\mathbf{D}$  is arbitrary,  $\hat{\mathbf{D}}$  has the closed form given in the paper. Otherwise, we can use Fisher scoring to solve the following equations for  $\theta$ :

$$\begin{aligned} \frac{\partial \ell}{\partial \theta_j} &= -\frac{n}{2} \text{tr} \left( \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \theta_j} \right) \\ &\quad + \frac{1}{2} \sum_{i=1}^n \text{tr} \left[ \mathbf{E} \left( \mathbf{a}_i \mathbf{a}_i^T \mid \mathbf{Y}_i \right) \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \theta_j} \mathbf{D}^{-1} \right]. \end{aligned}$$

The Fisher scoring algorithm updates  $\theta$  using  $\theta_{\text{new}} = \theta_{\text{old}} + \hat{\mathbf{I}}_{\theta\theta}^{-1} \partial \ell / \partial \theta \mid_{\theta_{\text{old}}}$ , where  $\hat{\mathbf{I}}_{\theta\theta}$  has its  $(j, k)$ th component equal to

$$\mathbf{E} \left( -\frac{\partial^2 \ell}{\partial \theta_j \partial \theta_k} \right) = \frac{n}{2} \text{tr} \left( \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \theta_j} \mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial \theta_k} \right).$$

APPENDIX B

Calculations of the Information Matrix

The information matrix is obtained by differentiating twice the marginal multivariate normal log likelihood with mean and covariance given in equations (5) and (6) and taking an expectation of the resulting form. Let  $\mathbf{G}_i = \mathbf{1}_{K_i} \otimes \mathbf{I}_J$ ,  $\mathbf{H}_i = \mathbf{X}_i \otimes \beta_1$ , and  $\mathbf{M}_i = \mathbf{X}_i \alpha \otimes \mathbf{I}_J$ . The expected information matrix for the MLE of  $\gamma = (\beta_0, \beta_1, \alpha)$  is

$$\mathbf{I}_{\gamma\gamma} = \sum_{i=1}^n \begin{pmatrix} \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{G}_i & \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{M}_i & \mathbf{G}_i^T \mathbf{V}_i^{-1} \mathbf{H}_i \\ \mathbf{M}_i^T \mathbf{V}_i^{-1} \mathbf{G}_i & \mathbf{A} + \mathbf{M}_i^T \mathbf{V}_i^{-1} \mathbf{M}_i & \mathbf{M}_i^T \mathbf{V}_i^{-1} \mathbf{H}_i \\ \mathbf{H}_i^T \mathbf{V}_i^{-1} \mathbf{G}_i & \mathbf{H}_i^T \mathbf{V}_i^{-1} \mathbf{M}_i & \mathbf{H}_i^T \mathbf{V}_i^{-1} \mathbf{H}_i \end{pmatrix}$$

where  $\mathbf{A}$  has its  $(j, k)$ th component  $(1/2) \text{tr}[\mathbf{V}_i^{-1} (\partial \mathbf{V}_i / \partial \beta_{1j}) \times \mathbf{V}_i^{-1} (\partial \mathbf{V}_i / \partial \beta_{1k})]$ .

Let  $\delta = (\tau^2, \theta^T, \xi^T)^T$ , then

$$\begin{aligned} \mathbf{I}_{\delta_j \delta_k} &= \frac{1}{2} \sum_{i=1}^n \text{tr} \left( \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \delta_j} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \delta_k} \right) \\ \mathbf{I}_{\delta_j \beta_{1k}} &= \frac{1}{2} \sum_{i=1}^n \text{tr} \left( \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \delta_j} \mathbf{V}_i^{-1} \frac{\partial \mathbf{V}_i}{\partial \beta_{1k}} \right) \end{aligned}$$

and  $\mathbf{I}_{\delta\beta_0} = \mathbf{0}, \mathbf{I}_{\delta\alpha} = \mathbf{0}$ , where

$$\begin{aligned} \frac{\partial \mathbf{V}_i}{\partial \theta_j} &= \left( \mathbf{Z}_i \frac{\partial \mathbf{D}}{\partial \theta_j} \mathbf{Z}_i^T \right) \otimes \left( \beta_1 \beta_1^T \right), \\ \frac{\partial \mathbf{V}_i}{\partial \xi_j} &= \left( \mathbf{1}_{K_i} \mathbf{1}_{K_i}^T \right) \otimes \left( \Delta_j \Delta_j^T \right), \\ \frac{\partial \mathbf{V}_i}{\partial \beta_j} &= \left( \mathbf{I}_{K_i} + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T \right) \otimes \left( \Delta_j \beta_1^T + \beta_1 \Delta_j^T \right), \\ \frac{\partial \mathbf{V}_i}{\partial \tau_j^2} &= \mathbf{I}_{K_i} \otimes \left( \Delta_j \Delta_j^T \right). \end{aligned}$$

Here  $\Delta_j$  is a  $J \times 1$  vector with the  $j$ th element equal to one and is zero otherwise.