Longitudinal Latent Variable Models Given Incompletely Observed Biomarkers and Covariates

Chunfeng Ren* Yongyun Shin
Department of Biostatistics, Virginia Commonwealth University, Richmond, Virginia 23219
**email*: renc@vcu.edu

Statistics in Medicine, v35 n26 p4729-4745 2016

SUMMARY. In this paper, we analyze a two-level latent variable model for longitudinal data from the National Growth of Health Study where surrogate outcomes or biomarkers and covariates are subject to missingness at any of the levels. A conventional method for efficient handling of missing data is to reexpress the desired model as a joint distribution of variables, including the biomarkers, that are subject to missingness conditional on all of the covariates that are completely observed, and estimate the joint model by maximum likelihood, which is then transformed to the desired model. The joint model, however, identifies more parameters than desired, in general. We show that the over-identified joint model produces biased estimation of the latent variable model, and describe how to impose constraints on the joint model so that it has a one-to-one correspondence with the desired model for unbiased estimation. The constrained joint model handles missing data efficiently under the assumption of ignorable missing data and is estimated by a modified application of the expectation-maximization (EM) algorithm.

KEYWORDS. Longitudinal data analysis; Multivariate outcomes; Random effects; Missing data; Latent variable; the EM algorithm

1. Introduction

The National Heart, Lung, and Blood Institute initiated the Growth and Health Study (NGHS) to investigate ethnic disparities in dietary, family, psychosocial and physical activity factors of obesity about 2,379 girls in 1985. It collected data on development of obesity and factors associated with the development from 1,213 African-American and 1,166 white girls. The study followed the subjects from 1987-1988 when they were 9 to 10 years old until 1996-1997 when they were 18 to 19 years old. The subjects were assessed on development of obesity and related factors annually [1].

We consider multiple biomarkers of obesity: body mass index (BMI), sum of skinfolds at triceps, subscapular, and suprailiac sites (Skinfold), maximum below-waist circumference (Waist), and percent body fat by bioelectrical impedance analysis (PercentFat). Many investigators have identified the risk factors of child obesity using one of these biomarkers as an outcome variable [2-6]. Although useful, some of these biomarkers do not differentiate the fat mass from body mass while others are measured with error. For example, BMI, the ratio of body weight in kilograms to height in meters squared, is widely used to define obesity (BMI$\geq$ 30) for men and women. Consequently, it is a broadly analyzed outcome variable as a surrogate body fat. However, it cannot distinguish muscle mass from body adiposity, in particular, for children and adolescents [7-9].

Our analysis aims to quantify child obesity via multiple biomarkers and study its risk factors simultaneously. Specifically, we want to control for ethnic and social disparities in the growth of obesity, and ask how environmental factors such as TV watching and mother's BMI influence the development of child obesity. Because obesity is not directly observable, NGHS collected the four biomarkers of obesity. We formulate a latent-variable model (LVM) of simultaneous equations where biomarkers, given the latent obesity, are independent in a measurement model, and the obesity is regressed on covariates in a structural model [10-19]. Given completely observed covariates and biomarkers having ignorable missing data [20], the LVM may be estimated by maximum likelihood (ML) via standard LVM software such as Amos [21], EQS [22], and Mplus

[23].

This paper focuses on a longitudinal multilevel model where occasions at level 1 are nested within individuals at level 2 and where missing data are present at both levels under the assumption of ignorable missing data [20, 24]. Roy and Lin [25] estimated a longitudinal LVM given nonignorable dropouts and level-1 covariates missing not at random by ML. Das et al. [26] estimated a structural equation model by a Markov Chain Monte Carlo method where continuous responses and covariates at level 1 may be missing at random in the measurement model. Both approaches handle level-1 outcomes and covariates subject to missingness.

Recent advances enable efficient handling of missing data in a hierarchical linear model (HLM) by ML [27-30] or by Bayesian approaches [27, 31-33]. Shin and Raudenbush [28] formulated a univariate HLM as a joint normal distribution of variables, including the outcome, subject to missingness conditional on completely observed covariates. The authors estimated the joint model by ML via the EM algorithm [34], and then transformed the estimated joint model to the HLM. They showed that the unconstrained joint model, in general, over-identifies the HLM and that the over-identified HLM leads to biased inferences. Therefore, the authors estimated a constrained joint model to just identify the HLM for unbiased estimation. The method, however, cannot be used for the complicated LVM of simultaneous equations. In this paper, we extend the method to the LVM where multiple biomarkers and covariates are subject to missingness at any of the levels.

We analyze the LVM given biomarkers and covariates that are subject to missingness with a general missing pattern at any of the levels. A conventional method for efficient handling of the missing data is to reexpress the LVM as a joint distribution of the variables, including the biomarkers, that are subject to missingness conditional on all of the covariates that are completely observed, and estimate the joint model which is then transformed to the LVM. The unconstrained joint model, however, identifies more parameters than desired in the LVM. Furthermore, the LVM is not nested within the joint model, in general. The consequence is that the over-identified joint model leads to biased estimation of the LVM. This paper explains how to characterize the joint model so that it is a one-to-one transformation of the LVM for unbiased estimation. To yield un-

biased estimation of the LVM while handling missing data efficiently, we estimate the constrained joint model according to the LVM within each iteration of the EM algorithm.

The next section introduces an LVM of our interest given incomplete data. Section 3 explains a joint model for efficient handling of missing data and shows how to impose proper constraints on the joint model for unbiased estimation of the LVM. Section 4 describes the EM algorithm for efficient handling of the constrained joint model. Section 5 simulates an LVM to show that the conventional method produces biased estimation of the LVM and that our approach corrects the bias. Section 6 illustrates unbiased and efficient analysis of the desired LVM given the NGHS data. Section 7 discusses the limitations and future extensions of our method.

2. Latent Variable Model

This section introduces the LVM of interest [15]. The structural model is

$$U_{ik} = A_{ik}^T\alpha + B_{ik}^T b_i + \epsilon_{ik}, \;\; b_i \overset{iid}{\sim} N(0, D), \; \epsilon_{ik} \overset{iid}{\sim} N(0,1), \tag{1}$$

where U_{ik} is a univariate latent obesity score, A_{ik} is a vector of covariates having fixed effects α, B_{ik} is a vector of known covariates having level-2 unit-specific random effects b_i independent of a level-1 unit-specific random error ϵ_{ik}, and level-1 unit or occasion k is nested within level-2 unit or subject i for $k = 1, \cdots, k_i$ and $i = 1, \cdots, n$, and D is a positive definite matrix. This model cannot be directly estimated due to unobservable U_{ik}. However, U_{ik} is related to biomarkers by a measurement model

$$R_{ik} = \gamma_0 + \gamma_1 U_{ik} + a_i + e_{ik}, \tag{2}$$

where R_{ik} is a vector of J biomarkers, $\gamma_0 = [\gamma_{01} \;\; \gamma_{02} \cdots \gamma_{0J}]^T$ is a vector of J intercepts, $\gamma_1 = [\gamma_{11} \;\; \gamma_{12} \cdots \gamma_{1J}]^T$ is a vector of the J effects or factor loadings of U_{ik}, and subject-specific random effects $a_i \overset{iid}{\sim} N(0, \oplus_{j=1}^J \xi_j)$ are independent of level-1 random errors $e_{ik} \overset{iid}{\sim} N(0, \oplus_{j=1}^J \tau_j)$ for a diagonal matrix $\oplus_{\ell=1}^J \psi_\ell = \text{diag}(\psi_1, \psi_2, \cdots, \psi_J)$ with diagonal elements or submatrices $(\psi_1, \psi_2, \cdots, \psi_J)$ and all other elements equal to zero. To make parameters identifiable in the

model (1), we assume that var(ϵ_{ik})=1 and that A_{ik} does not contain an intercept. Note that the jth and j'th biomarkers of subject i at occasion k are correlated and their covariance is equal to $\gamma_{1j}\gamma_{1j'}$var(U_{ik}).

Our goal is to identify the obesity factors A_{ik} and B_{ik} and explain their associations with the obesity U_{ik} by efficient analysis of the LVM, that is, by analyzing all available sample data without dropping any observations. The challenge is to efficiently handle missing data in (R_{ik}, A_{ik}), which is explained in the next section. In this paper, we refer to the associations as the "effects" of the factors, but do not mean causality. Such use of the term "effects" is pervasive in the literature.

3. Missing Data

To handle missing data in R_{ik} and A_{ik} efficiently, we reparameterize the LVM in terms of a joint distribution of the response variables R_{ik} and all covariates subject to missingness in A_{ik} conditional on all covariates completely observed. Because A_{ik} may have covariates subject to missingness as well as covariates completely observed, we decompose $A_{ik} = [S_{ik}^T\ Y_{2i}^T\ W_{1ik}^T\ W_{2i}^T]^T$ where p_1-vector S_{ik} and p_2-vector Y_{2i} are level-1 and -2 covariates subject to missingness, respectively, and p_3-vector W_{1ik} and p_4-vector W_{2i} are level-1 and -2 covariates completely observed, respectively. Then, the joint model is a multivariate distribution of level-1 $Y_{1ik} = [R_{ik}^T\ S_{ik}^T]^T$ and level-2 Y_{2i} that are subject to missingness conditional on W_{1ik}, W_{2i} and B_{ik} that are completely observed. In this section, we explain that this joint model over-identifies the LVM, in general. The consequence is biased estimation of the LVM as will be illustrated in Section 5. For a positive integer m, let I_m and 1_m denote an m-by-m identity matrix and a vector of m unities, respectively.

3.1. Over-identification Problem

If we were able to observe U_{ik}, we would directly analyze the structural model (1) without involving the measurement model (2). To analyze all observed data in the model (1), we would estimate the multivariate distribution of (U_{ik}, S_{ik}, Y_{2i}) given completely observed $(W_{1ik}, W_{2k}, B_{ik})$. In this simple case, we are able to not only reveal the over-identification problem explicitly, but also ex-

plain how to correct the problem clearly. In the following subsection, we extend the multivariate distribution to efficient handling of missing data in (Y_{1ik}, Y_{2k}) conditional on $(W_{1ik}, W_{2k}, B_{ik})$ for the general LVM.

If U_{ik} were observed, efficient handling of the missing data in the desired model (1) might be achieved, without the measurement model (2), by

$$\begin{bmatrix} U_{ik} \\ S_{ik} \\ Y_{2i} \end{bmatrix} = \begin{bmatrix} \beta_{u1}^T & \beta_{u2}^T \\ \beta_{s1} & \beta_{s2} \\ 0 & \beta_{22} \end{bmatrix} \begin{bmatrix} W_{1ik} \\ W_{2i} \end{bmatrix} + \begin{bmatrix} B_{ik}^T & 0 & 0 \\ 0 & I_{p_1} & 0 \\ 0 & 0 & I_{p_2} \end{bmatrix} \begin{bmatrix} b_{ui} \\ b_{si} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{uik} \\ \epsilon_{sik} \\ 0 \end{bmatrix}, \tag{3}$$

where β_{u1}^T and β_{s1} are 1-by-p_3 and p_1-by-p_3 matrices of the fixed effects of W_{1ik} on U_{ik} and S_{ik}, respectively, β_{u2}^T, β_{s2}, and β_{22} are 1-by-p_4, p_1-by-p_4 and p_2-by-p_4 matrices of the fixed effects of W_{2i} on U_{ik}, S_{ik} and Y_{2i}, respectively, and $\begin{bmatrix} b_{ui} \\ b_{si} \\ b_{2i} \end{bmatrix} \overset{iid}{\sim} N\left(0, \begin{bmatrix} T_{uu} & T_{us} & T_{u2} \\ T_{su} & T_{ss} & T_{s2} \\ T_{2u} & T_{2s} & T_{22} \end{bmatrix}\right)$ is independent of $\begin{bmatrix} \epsilon_{uik} \\ \epsilon_{sik} \end{bmatrix} \overset{iid}{\sim} N\left(0, \begin{bmatrix} \Sigma_{uu} & \Sigma_{us} \\ \Sigma_{su} & \Sigma_{ss} \end{bmatrix}\right)$. We center level-1 S_{ik} and W_{1ik} around respective sample means and level-2 Y_{2i} and W_{2i} around respective weighted sample means $\frac{\sum_i k_i Y_{2i}}{\sum_i k_i}$ and $\frac{\sum_i k_i W_{2i}}{\sum_i k_i}$ in Equation (3), except for B_{ik} that is centered around its group mean for precise estimation of the variance matrix [35]. The centering ensures that we identify the model (1) with no intercept and model (2). Shin and Raudenbush [28] expressed $[\beta_{u1}^T \ \ \beta_{u2}^T] \begin{bmatrix} W_{1ik} \\ W_{2i} \end{bmatrix} = \beta_u^T W_{uik}$, $\beta_{s1} W_{1ik} + \beta_{s2} W_{2i} = (I_{p_1} \otimes W_{uik}^T)\beta_s$ and $\beta_{22} W_{2i} = (I_{p_2} \otimes W_{2i}^T)\beta_2$, and efficiently estimated the model (3) by ML via the EM algorithm where U_{ik} was observable.

Although the conditional model (1) expresses a single effect of each covariate in S_{ik} on U_{ik}, the multivariate model (3) expresses a distinct covariance at each level between the covariate and obesity to identify p_1 extraneous parameters than desired in the model (1). The two distinct covariances identify the within-child association between the time-varying covariate and outcome that may be different from the between-child association, the association between the child-mean

covariate and outcome. The associations identify a contextual effect of the covariate that is defined as the difference between the between- and within-child associations [35, 36]. Controlling for the within-child association, the contextual effect explains the expected difference in obesity between two children who have the same value of the covariate at an occasion, but who differ by one unit in their child-mean covariates. Consequently, the multivariate model identifies a contextual effects model where each covariate in S_{ik} has a contextual effect, controlling for the within-child effect of the covariate [29]. Because the model (1) expresses no contextual effect of the covariate, implying identical between- and within-child associations between the covariate and outcome [36], the multivariate model (3) over-identifies the model (1) and expresses the single effect of each covariate in S_{ik} as a weighted average of the two associations [30, 35]. The weighted average is different from the single effect when model (1) is directly estimated [35, 36]. The consequence is that the desired model (1) is not nested within or congenial to the multivariate model (3) [37]. The over-identified model (3) yields biased estimation of the desired model (1) unless constrains are imposed on the model (3) [28]. We illustrate the over-identification problem causing biased estimation by a simulation study in Section 5.

In order to correct the bias, we impose p_1 constraints on the model (3) so that it represents a one-to-one transformation of the LVM. For clarity, we describe the constraints for a random-intercept model (1) having $B_{ik} = 1$. Appendix A explains the constraints for a random- coefficient model (1). To simplify the notation, let $\text{cov}(b_{ui}, b_{s_i}|b_{2i}) = \begin{bmatrix} T_{uu|2} & T_{us|2} \\ T_{su|2} & T_{ss|2} \end{bmatrix}$. Given Y_{2i}, we constrain the covariances between U_{ik} and each covariate in S_{ik} to equal, i.e.

$$\alpha_1^T = T_{us|2}T_{ss|2}^{-1} = \Sigma_{us}\Sigma_{ss}^{-1}, \tag{4}$$

which says that the association between U_{ik} and each of the level-1 covariates is the same at each level given Y_{2i}. The constraints imply $\text{cov}(U_{ik}, S_{ik}|Y_{2i})[var(S_{ik}|Y_{2i})]^{-1} = (T_{us|2} + \Sigma_{us})(T_{ss|2} + \Sigma_{ss})^{-1} = \alpha_1^T$ for $T_{us|2} = \alpha_1^T T_{ss|2}$ and $\Sigma_{us} = \alpha_1^T \Sigma_{ss}$, and the one-to-one transformations between

the LVM and the multivariate model (3)

$$\begin{aligned}
&\alpha_1 = \Sigma_{ss}^{-1}\Sigma_{su},\ \alpha_2 = T_{22}^{-1}(T_{2u} - T_{2s}\alpha_1),\\
&\alpha_3 = \beta_{u1} - \beta_{s_1}^T\alpha_1,\\
&\alpha_4 = \beta_{u2} - \beta_{s_2}^T\alpha_1 - \beta_{22}^T\alpha_2,\ 1 = \Sigma_{uu} - \alpha_1^T\Sigma_{ss}\alpha_1,\\
&D = T_{uu} - \alpha_2^T T_{22}\alpha_2 - 2\alpha_1^T T_{s2}\alpha_2 - \alpha_1^T T_{ss}\alpha_1.
\end{aligned} \tag{5}$$

3.2. Efficient Handling of Missing Data

Because U_{ik} is unobservable, we need to estimate the measurement model (2) in addition to the desired model (1). Because observed biomarkers are also subject to missingness, the multivariate model (3) cannot handle the missing data in both A_{ik} and R_{ik}. Instead, we formulate the joint distribution of (R_i, S_i, Y_{2i}) subject to missingness given completely observed covariates for $R_i = [R_{i1}^T\ R_{i2}^T \cdots R_{ik_i}^T]^T$ and $S_i = [S_{i1}^T\ S_{i2}^T \cdots S_{ik_i}^T]^T$ based on the aggregated models (2) and (3)

$$\begin{bmatrix} R_i \\ S_i \\ Y_{2i} \end{bmatrix} = \begin{bmatrix} 1_{k_i} \otimes \gamma_0 + (W_{ui}\beta_u + B_i b_{ui} + \epsilon_{ui}) \otimes \gamma_1 \\ W_{si}\beta_s + (1_{k_i} \otimes I_{p_1})b_{s_i} + \epsilon_{si} \\ X_{2i}\beta_2 + b_{2i} \end{bmatrix} + \begin{bmatrix} 1_{k_i} \otimes a_i \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} e_i \\ 0 \\ 0 \end{bmatrix}, \tag{6}$$

for $W_{ui} = [W_{ui1}\ W_{ui2} \cdots W_{uik_i}]^T$, $B_i = [B_{i1}\ B_{i2} \cdots B_{ik_i}]^T$ $\epsilon_{ui} = [\epsilon_{ui1}\ \epsilon_{ui2} \cdots \epsilon_{uik_i}]^T$, $e_i = [e_{i1}^T\ e_{i2}^T \cdots e_{ik_i}^T]^T$, $W_{si} = [I_{p_1} \otimes W_{ui1}\ I_{p_1} \otimes W_{ui2}\ \cdots I_{p_1} \otimes W_{uik_i}]^T$, $\epsilon_{si} = [\epsilon_{si1}^T\ \epsilon_{si2}^T \cdots \epsilon_{sik_i}^T]^T$, and $X_{2i} = I_{p_2} \otimes W_{2i}^T$. To derive estimators, we reexpress model (6) parsimoniously as

$$\begin{bmatrix} Y_{1i} \\ Y_{2i} \end{bmatrix} = \begin{bmatrix} X_{1i} & 0 \\ 0 & X_{2i} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_{1i} & 0 \\ 0 & I_{p_2} \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} \epsilon_{1i} \\ 0 \end{bmatrix} + \begin{bmatrix} a_{1i} + c_{1i} \\ 0 \end{bmatrix}, \tag{7}$$

for $Y_{1i} = \begin{bmatrix} R_i \\ S_i \end{bmatrix}$, $X_{1i} = \begin{bmatrix} I_{J\times k_i} & W_{ui} \otimes I_J & 0 \\ 0 & 0 & W_{si} \end{bmatrix}$, $\beta_1 = \begin{bmatrix} 1_{k_i} \otimes \gamma_0 \\ \beta_u \otimes \gamma_1 \\ \beta_s \end{bmatrix}$, $Z_{1i} = \begin{bmatrix} B_i \otimes I_J & 0 \\ 0 & 1_{k_i} \otimes I_{p_1} \end{bmatrix}$, $b_{1i} = \begin{bmatrix} b_{ui} \otimes \gamma_1 \\ b_{si} \end{bmatrix}$, $\epsilon_{1i} = \begin{bmatrix} \epsilon_{ui} \otimes \gamma_1 \\ \epsilon_{si} \end{bmatrix}$, $a_{1i} = \begin{bmatrix} 1_{k_i} \otimes a_i \\ 0 \end{bmatrix}$, and $c_{1i} = \begin{bmatrix} e_i \\ 0 \end{bmatrix}$, where var$(b_{1i}, b_{2i}) =$

$$\begin{bmatrix} \tau_{11} & \tau_{12} \\ \tau_{12}^T & \tau_{22} \end{bmatrix}, \text{var}(\epsilon_{1i}) = \begin{bmatrix} I_{k_i} \otimes (\Sigma_{uu}\gamma_1\gamma_1^T) & I_{k_i} \otimes (\gamma_1\Sigma_{us}) \\ I_{k_i} \otimes (\Sigma_{su}\gamma_1^T) & I_{k_i} \otimes \Sigma_{ss} \end{bmatrix}, \text{var}(a_{1i}) = \begin{bmatrix} (1_{k_i}1_{k_i}^T) \otimes (\oplus_{j=1}^J \xi_j) & 0 \\ 0 & 0 \end{bmatrix},$$

and $\text{var}(c_{1i}) = \begin{bmatrix} I_{k_i} \otimes (\oplus_{j=1}^J \tau_j) & 0 \\ 0 & 0 \end{bmatrix}$ for $\tau_{11} = \begin{bmatrix} T_{uu} \otimes (\gamma_1\gamma_1^T) & T_{us} \otimes \gamma_1 \\ T_{su} \otimes \gamma_1^T & T_{ss} \end{bmatrix}$, $\tau_{12} = \begin{bmatrix} T_{u2} \otimes \gamma_1 \\ T_{s2} \end{bmatrix}$, and $\tau_{22} = T_{22}$. Note that the joint model (7) enables us to analyze a subject who has at least a single value observed in (Y_{1i}, Y_{2i}) for efficient analysis of the LVM.

To efficiently handle missing data, let O_{1i} and O_{2i} be matrices of the observed value indicators (1 if observed, 0 otherwise) in Y_{1i} and Y_{2i}, respectively, such that they extract all observed data $Y_{1i}^o = O_{1i}Y_{1i}$ and $Y_{2i}^o = O_{2i}Y_{2i}$ from Y_{1i} and Y_{2i}, respectively [28]. The model (7) for the observed data is

$$\begin{bmatrix} Y_{1i}^\circ \\ Y_{2i}^\circ \end{bmatrix} = \begin{bmatrix} X_{1i}^\circ & 0 \\ 0 & X_{2i}^\circ \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_{1i}^\circ & 0 \\ 0 & O_{2i} \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} + \begin{bmatrix} a_{1i}^\circ + \epsilon_{1i}^\circ + e_{1i}^\circ \\ 0 \end{bmatrix}, \tag{8}$$

for $X_{1i}^\circ = O_{1i}X_{1i}$, $X_{2i}^\circ = O_{2i}X_{2i}$, $Z_{1i}^\circ = O_{1i}Z_{1i}$, $a_{1i}^\circ = O_{1i}a_{1i}$, $\epsilon_{1i}^\circ = O_{1i}\epsilon_{1i}$, and $e_{1i}^\circ = O_{1i}e_{1i}$. We reexpress the model (8) parsimoniously as $Y_i^\circ \sim N(\mu_i^\circ, V_i^\circ)$ for $Y_i^\circ = [Y_{1i}^{\circ T}\ Y_{2i}^{\circ T}]^T$,

$$\mu_i^\circ = \begin{bmatrix} X_{1i}^\circ\beta_1 \\ X_{2i}^\circ\beta_2 \end{bmatrix},\ V_i^\circ = \begin{bmatrix} Z_{1i}^\circ\tau_{11}Z_{1i}^{\circ T} + O_{1i}[\text{var}(\epsilon_{1i}) + \text{var}(a_{1i}) + \text{var}(c_{1i})]O_{1i}^T & Z_{1i}^\circ\tau_{12}O_{2i}^T \\ O_{2i}\tau_{21}Z_{1i}^{\circ T} & O_{2i}\tau_{22}O_{2i}^T \end{bmatrix}. \tag{9}$$

4. Estimation via the EM Algorithm

This section sketches efficient estimation of the joint model (7) by a modified application of the EM algorithm [34]. See Appendices B, C and D for details. The modification is due to the fact that we efficiently estimate the LVM to find the constraints (4) that will be imposed on the joint model (7) within each iteration of the EM algorithm. We view $(Y_{1i}, Y_{2i}, U_i, b_{ui}, b_{si}, a_i)$ as complete data and Y_i° observed within unit i for $U_i = [U_{i1}\ U_{i2} \cdots U_{ik_i}]$. The constraints (4) require that the parameters α of the LVM be estimated. Within each iteration of the EM algorithm, we estimate the parameters α and translate them into the parameters of the joint model (7) according to the transformations (5). To estimate α, let $A_i = [A_{i1}\ A_{i2} \cdots A_{ik_i}]^T$, $\epsilon_i = [\epsilon_{i1}\ \epsilon_{i2} \cdots \epsilon_{ik_i}]^T$, $\gamma_j = [\gamma_{0j}\ \gamma_{1j}]^T$, $U_{ik}^* = [1\ U_{ik}]^T$,

$\epsilon_{1ik} = [\epsilon_{uik} \;\; \epsilon_{sik}]^T$, $\epsilon_{1i}^* = [\epsilon_{ui} \;\; \epsilon_{si}]$ for the LVM, $b_{1i}^* = [b_{ui} \;\; b_{si}]^T$, $b_i^* = [b_{1i}^* \;\; b_{2i}]^T$, $\beta_1^* = [\beta_u \;\; \beta_s]^T$, $T_{11} = \begin{bmatrix} T_{uu} & T_{us} \\ T_{su} & T_{ss} \end{bmatrix}$, $T_{12} = \begin{bmatrix} T_{u2} \\ T_{s2} \end{bmatrix}$, $T = \begin{bmatrix} T_{11} & T_{12} \\ T_{12}^T & T_{22} \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Sigma_{uu} & \Sigma_{us} \\ \Sigma_{su} & \Sigma_{ss} \end{bmatrix}$, $W_{usi} = \begin{bmatrix} W_{ui} & 0 \\ 0 & W_{si} \end{bmatrix}$, and $T_{2|1} = T_{22} - T_{21}T_{11}^{-1}T_{12}$ for the joint model. The complete data ML estimators in iteration k are $\hat{\alpha}^{(k)} = \hat{\alpha}^{(k-1)} + \left(\sum_{i=1}^n \sum_{k=1}^{k_i} A_{ik}A_{ik}^T\right)^{-1} \sum_{i=1}^n \sum_{k=1}^{k_i} A_{ik}\epsilon_{ik}$ and $\hat{D} = \sum_i b_i b_i^T/n$ for the structural model (1) and

$$
\begin{aligned}
\hat{\gamma}_j^{(k)} &= \hat{\gamma}_j^{(k-1)} + \left(\sum_{i=1}^n \sum_{k=1}^{k_i} U_{ik}^* U_{ik}^{*T}\right)^{-1} \sum_{i=1}^n \sum_{k=1}^{k_i} U_{ik}^* e_{ikj}, \\
\hat{\xi}_j &= \frac{1}{n}\sum_{i=1}^n a_{ij}^2, \\
\hat{\tau}_j &= \frac{1}{\sum_{i=1}^n k_i} \sum_{i=1}^n \sum_{k=1}^{k_i} e_{ikj}^2, \\
\hat{\Sigma} &= \frac{1}{\sum_{i=1}^n k_i} \sum_{i=1}^n \sum_{k=1}^{k_i} \epsilon_{1ik}\epsilon_{1ik}^T, \\
\hat{T} &= \frac{1}{n}\sum_{i=1}^n b_i^* b_i^{*T}, \\
\hat{\beta}_1^{*(k)} &= \hat{\beta}_1^{*(k-1)} + \left(\sum_{i=1}^n \Sigma^{-1} \otimes (W_{usi}^T W_{usi})\right)^{-1} \sum_{i=1}^n \Sigma^{-1} \otimes (W_{usi}^T \epsilon_{1i}^*), \\
\hat{\beta}_2^{(k)} &= \hat{\beta}_2^{(k-1)} + \left(\sum_{i=1}^n T_{2|1}^{-1} \otimes (W_{2i} W_{2i}^T)\right)^{-1} \sum_{i=1}^n T_{2|1}^{-1} \otimes W_{2i} \left(b_{2i} - T_{21}T_{11}^{-1} b_{1i}^*\right)
\end{aligned}
\tag{10}
$$

for the joint model (7). At the E step, we obtain conditional expectations, $E(A_{ik}A_{ik}^T|Y_i^\circ)$, $E(A_{ik}\epsilon_{ik}|Y_i^\circ)$, $E(b_i b_i^T|Y_i^\circ)$, $E(U_{ik}|Y_i^\circ)$, $E(U_{ik}^2|Y_i^\circ)$, $E(U_{ik}e_{ikj}|Y_i^\circ)$, $E(e_{ikj}|Y_i^\circ)$, $E(e_{ikj}^2|Y_i^\circ)$, $E(a_{ij}^2|Y_i^\circ)$, $E(\epsilon_{1ik}\,\epsilon_{1ik}^T|Y_i^\circ)$, $E(b_i^*|Y_i^\circ)$, $E(b_i^* b_i^{*T}|Y_i^\circ)$, and $E(\epsilon_{1i}^*|Y_i^\circ)$ from the distribution of $Y_{1i}, Y_{2i}, U_i, e_i, \epsilon_{1i}^*$, $b_i^*, a_i|Y_i^\circ$. Let $V(A)$ denote a vector of distinct elements in a variance-covariance matrix A. At convergence, the Fisher information matrix is obtained from the observed log-likelihood of parameters $(\gamma_0, \gamma_1, \beta_1^*, \beta_2, \tau, T_{uu}, V(T_{ss}), V(T_{2s}),\; V(T_{22}), \xi,\; V(\Sigma_{ss}), \alpha_1, \alpha_2)$. The variance matrix associated with the parameter estimates in the constrained joint model (7) is produced by inverting the Fisher information matrix. We obtain the point estimates and the standard errors associated with the parameters of the LVM by the invariance property of MLEs and multivariate delta method,

respectively.

The next two sections illustrate the method by analyses of simulated and NGHS data. The convergence is taken to be the difference in the observed log-likelihoods between two consecutive iterations less than 10^{-6}.

5. Simulation

In this section, we simulate a simple LVM involving two biomarkers ($J = 2$), a level-1 covariate S_{ik}, and a level-2 covariate W_{2i}. The goal is to show that given W_{2i}, the over-identified joint model (7) of (R_{ik}, S_{ik}) leads to biased estimation of the LVM and that the constrained joint model (7), according to equations (4), corrects the bias. Next, we simulate ignorable missing data to show that our method via the constrained joint model estimates the desired LVM well given incomplete data.

5.1. Over-identification Problem

Five occasions $(k_i = 5)$ are nested within each of 1000 subjects $(n = 1000)$ in the simulated LVM

$$\begin{aligned} U_{ik} &= S_{ik} + W_{2i} + b_i + \epsilon_{ik}, b_i \sim N(0,1),\ \epsilon_i \sim N(0,1), \\ R_{ik} &= 1_2 + 1_2 U_{ik} + a_i + e_{ik},\ a_i \sim N(0, 0.25 I_2),\ e_{ik} \sim N(0, 0.25 I_2), \end{aligned} \tag{11}$$

where $\alpha_2 = \alpha_3 = 0$, $\alpha_1 = \alpha_4 = D = \gamma_{01} = \gamma_{02} = \gamma_{11} = \gamma_{12} = 1$, $\tau_1 = \tau_2 = \xi_1 = \xi_2 = 0.25$, $S_{ik} \sim N(0,1)$, and $W_{2i} \sim$ Bernoulli (0.5). We simulate the model with no missing data because the corresponding unconstrained joint model (7) identifies more parameters than desired to yield biased estimation of the LVM regardless of whether there are missing data or not. Given the simulated data, we estimate the LVM (11) by three different ML methods via the EM algorithm: direct estimation of the LVM given complete data; estimation of the corresponding constrained joint model (7), according to Equations (4), which is then transformed to the LVM; and estimation of the unconstrained joint model that is transformed to the LVM. We call the three approaches benchmark, just-identified and over-identified estimation methods. An estimation method works well if it produces all point estimates close to the benchmark counterparts. Note that we do not

simulate missing data because the complete data analysis illustrates the over-identification problem and the consequential biased estimation.

Table 1 displays the results. The benchmark estimates are shown under column heading "Benchmark". All point estimates are close to their true values. The standard errors are very small. The just-identified LVM estimates and their standard errors in the next column under heading "Just-identified" are identical to the benchmark counterparts. The last column under "Over-identified" shows over-identified LVM estimates. It is apparent that all point estimates of the model (1) and their standard errors are comparatively underestimated while the effects of U_{ik} and their standard errors in the model (2) appear overestimated relative to the benchmark counterparts.

5.2. Missing Data

To compare the performance of the just-identified and over-identified estimations given incomplete data, we simulate ignorable missing values (R_{ik}, S_{ik}) in the simulated data set of Table 1. Let M_{Rik} be 1 if R_{ik} is missing, and 0 otherwise. We define M_{Sik} for S_{ik} likewise, and draw missing values according to

$$\text{logit}(p_i) = 1 + W_{2i} + \delta_i,\ \delta_i \sim N(0,1)$$

for the W_2 simulated completely observed so that

$$M_{Rik} \sim \text{binomial}(k_i, p_i),\ \text{if logit}(p_i) > t_1$$
$$M_{Sik} \sim \text{bnomial}(k_i, 1-p_i),\ \text{if logit}(p_i) < -t_2$$

We set thresholds $t_1 = 2.09$ and $t_2 = 0.91$ which are equal to the 70^{th} and 30^{th} percentiles of logit(p_i), respectively. Consequently, we drop 28.14% and 13.14% of R_{ik} and S_{ik}, respectively. Note that the parameters of LVM (11) are distinct from those of the missing data mechanism above. Then, the missing values are missing at random or ignorable because the missing data mechanism depends on completely observed covariate W_{2i} [20].

The estimated LVMs appear in Table 2 under the same column headings as those of Table 1. Both just-identified and over-identified points estimates are close to their complete-data counter-

parts in Table 1. Due to the missing values, however, the standard errors are inflated relative to their complete-data counterparts, in general. Therefore, the just-identified LVM estimates appear unbiased under the simulated missing rate.

6. Analysis of NGHS Data

Now, we estimate a just-identified LVM to analyze the NGHS data. Each subject in the study was scheduled to visit a clinic for measurement once a year, but a number of subjects had item-nonresponse, or missed their visits to produce unit-nonresponse. We analyze all these subjects, including those having unit-nonresponse, in the joint model (7) as they have at least person-specific characteristics observed to strengthen the inferences at level 2 [29]. Table 3 summarizes the longitudinal data for analysis where level-1 variables are time-varying while level-2 variables are individual-level or base-line characteristics. The biomarkers have high correlations ranging from 0.81 to 0.92 as shown in Table 7. We reason that the high positive correlations result because they are the biomarkers of obesity. The previous studies identified influential covariates of the biomarkers as age (Age), race ethnicity (Race), single-parent family (OneParent), maturation categorizing prepuberty, puberty, post-menarche, and $\geq$ 2 years after post-menarche (Maturation), maximum parental education categorizing high school or less, and some college or more (ParentEd), household yearly income (Income, categorizing $\leq \$19,999$, $\$20,000 - \$39,999$, and $\geq \$40,000$), the weekly number of hours of TV watching (TV), overall physical activity pattern score (PhysicalAct, the higher, the more physically active), and mother's BMI (MotherBMI). Maturation and Income are coded as 0, 1, 2, 3 and 0, 1, 2, respectively. Our preliminary analysis shows that the linear associations between the coded covariates and obesity are reasonable. Specifically, we took the first principal component of the biomarkers as the obesity outcome, explaining 91.4% of the total variability in the biomarkers. Figure 1 draws the obesity outcome against the coded covariates, revealing that the linear associations are reasonable. We analyze dummy indicator variables for white students (White), single-parent family (OneParent), and and the maximum parent education of some college or more (ParentEd). Except for Age, White, OneParent and ParentEd, nine other

variables miss up to 32% of their values.

We use all available data to efficiently analyze a random-intercept LVM and a random-coefficient LVM. The random intercept LVM has R_{ik}=[BMI Skinfold PercentFat Waist]^T, S_{ik}=[Maturation TV PhysicalAct]^T, Y_{2i}=[MotherBMI Income]^T, W_{1ik}=[Age Age^2 Age $\times$ White]^T, W_{2i} = [ParentEd White OneParent]^T, and $B_{ik} = 1$, while the random-coefficient model has every component the same as the random-intercept counterpart except for B_{ik}=[1 Age$_{ik}$]^T and $D = \begin{bmatrix} D_{00} & D_{01} \\ D_{10} & D_{11} \end{bmatrix}$.

The estimated structural and measurement models of the random-intercept LVM appear in Tables 4 and 5, respectively. From the fitted structural model under column-heading "MAR" in Table 4, TV, Maturation, MotherBMI, Age, and OneParent are positively associated while PhysicalAct, quadratic Age, Age-by-White interaction and White are negatively associated with obesity, ceteris paribus. Controlling for other covariates, Income and ParentEd are not statistically significant, unlike previous studies [38-40]. The estimated measurement model in Table 5 shows that all biomarkers are highly significant and, thus, predictive of the latent obesity.

The estimated random-coefficient LVM is also displayed in Tables 4 and 5. The last column of Table 4 under column heading "MAR" shows the estimated structural model. The statistical inferences on all fixed effects stay the same as they are in the random-intercept LVM. However, the effects of linear and quadratic Age, Age-by-White interaction and White strengthen, compared to the random-intercept counterparts. In particular, the negative gap of white girls' obesity relative to black girls' triples. Besides, the variance of the random intercept in the random-coefficient LVM doubles from that of the random-intercept LVM. The measurement model in Table 5 shows that the obesity has attenuating effect on biomarkers, comparatively with the random-intercept counterparts. The likelihood ratio test for the null hypothesis $H_0 : D_{01} = D_{11} = 0$ produces the p-value$<$ 0.01. Although the p-value is conservative [41-43], the small p-value reveals evidence that the effect of age varies randomly across individuals. To confirm the evidence, we compute the Akaike's Information Criterion (AIC) for the random-coefficient model AIC$_1$=498,993.00 and the AIC for the random-intercept model AIC$_2$=507,572.40. The ΔAIC$= 8579.4 > 10$, which is the

difference between AIC_2 and AIC_1, also indicates that age has a random effect on the child obesity [44].

Figure 2 displays the effects of age for black and white girls based on the random-coefficient LVM. Adjusting for the effects of other covariates in the model, Age is positively associated with obesity [3, 45, 46]. However, we find that the positive association weakens more rapidly for white girls than for black girls toward the later stage of adolescence, thereby widening the racial gap in obesity between the two subpopulations of girls. The gap starts widening rapidly from about age 14 where a 95% confidence interval for obesity is $(0.05, 0.59)$.

Table 4 compares the complete-case analysis under column heading "MCAR" with our missing data analysis under "MAR" of the random-intercept and -coefficient structural models. We dropped 57.22% of occasions and 37.16% of subjects for the MCAR analyses. The estimated random intercept model under MCAR reveals that the effects of Maturation and Income are comparatively over-represented while the effect of Age-by-White interaction is relatively under-estimated. Furthermore, the statistical inferences of the complete-case analysis are relatively biased. The effect of Income is statistically significant under MCAR, but insignificant in our missing data analysis while the effects of quadratic Age, Age-by-White interaction, White and OneParent are statistically insignificant in the complete-case analysis, but significant under MAR. The biased inferences result mainly because the standard errors of the complete-case analysis are up to 142.22% more inflated than the MAR counterparts. For analysis of the random-coefficient model, the complete-case analysis over-represents the effects of Maturation and Income , but under-represents those of Age-by-White interaction and White, relative to the MAR counterparts. The Age-by-White interaction effect is statistically insignificant under MCAR, but significant under MAR. The biased inference is due to the MCAR standard error that is 240% as large as the MAR counterpart.

Table 6 shows the complete-case analyses of the measurement models. The effects of obesity on biomarkers and their standard errors are comparatively overestimated. Overall, the complete-case analyses appear comparatively biased and inefficient.

7. Discussion

In this paper, we presented a maximum likelihood method for unbiased estimation of a latent variable model of simultaneous equations where biomarkers are related to latent obesity in a measurement equation and the latent obesity is regressed on covariates in a structural equation. Both covariates and biomarkers may be subject to missingness with a general missing pattern at any level of the hierarchy. The method handles missing data efficiently under an assumption of ignorable missing data. To handle missing data efficiently, we reexpressed the LVM as a joint distribution of the variables, including the biomarkers, subject to missingness conditional on completely observed covariates. The joint model, however, over-identifies the desired LVM when level-1 covariates are subject to missingness. The consequence is that the over-identified LVM may produce considerably biased inferences as was illustrated in Section 5. To overcome the problem of over-identification, we constrained the joint model to be a one-to-one transformation of the LVM, efficiently estimated the constrained joint model by ML via the EM algorithm and, then, transformed the estimated joint model to the LVM for unbiased and efficient estimation. We simulated an LVM to show that the just-identified LVM estimates are unbiased while the over-identified LVM counterparts are biased.

We wrote a SAS IML program to estimate a constrained (and unconstrained) joint model, which was then transformed to the desired LVM via the one-to-one transformation formulas (5). The convergence criterion was the difference in observed log likelihoods between two-consecutive iterations, which was taken to less than 10^{-6}.

An alternative approach to our efficient ML estimation of LVM (1) given incomplete data is via multiple imputation (MI) [47]. Given the estimated joint model (7), we may randomly draw MI of completed data for subsequent analysis of the LVM [28, 30]. The MI may include the latent obesity. Existing statistical software packages cannot impute the level-1 and -2 missing data efficiently according to the joint model (7) to the best knowledge of the authors. Therefore, researchers may be tempted to use MI of missing values using standard imputation software packages such as SAS PROC MI and NORM [48], followed by complete-data analysis given the imputation by standard LVM software [47]. When MI of single-level data is applied to multilevel data, the variance-

covariance structure of the imputed data sets will not accurately represent the multilevel process (7) that generated the data, nor will the structural relations at each level be captured correctly. The resulting inferences may be substantially biased [48]. If MI is applied correctly according to the data-generating process (7), subsequent complete-data analysis of the LVM given the MI will produce estimation of the LVM comparable to the estimated LVM by our method. Both of our ML method and the MI approach require efficient estimation of the joint model (7). Following the estimation, our method requires technical transformation of the joint model to the LVM by the multivariate Delta method while the MI approach includes the cumbersome extra step of drawing MI for subsequent complete-data analysis of the LVM [30]. However, once generation of MI is automated, the MI approach will be less technical and, thus, broadly accessible to a wide range of researchers. We would like to take on this research in near future.

A limitation of the current approach is our assumption that the covariate having a random effect is completely observed. When such a covariate has missing values, it should be modeled on the left-hand side of the joint model in order to handle missing data efficiently. At the same time, the covariate appears on the right-hand side of the joint model for estimation of the random effect. Such a joint model is non-normal so that normal factorization of the joint model that leads to the desired LVM as a conditional distribution of biomarkers given covariates does not apply. One possible solution is a Bayesian approach where parameters are assumed to have their prior distributions, and the missing data are imputed from their posterior distributions given the parameters. Although the relaxed assumption will make our method more applicable, it is beyond the scope of the current research.

Another limitation of our current approach is the multivariate normal joint model to handle missing data efficiently. We analyzed discrete covariates, household income and maturation stage, subject to missingness. Although it is not appropriate to handle such discrete missing values under the joint normality, the identified model is the desired LVM we want to analyze [14, 16, 19]. The advantage is that we analyze the covariates subject to missingness by the efficient missing data method [11, 28, 29, 49]. Robust handling of a mixture of discrete and continuous missing data is

in our future research agenda.

Finally, we assumed the independence of biomarkers given obesity in the measurement model. To see how plausible the assumption is for each LVM, we computed the correlations between biomarkers implied by each fitted LVM and compared them to the corresponding sample correlations. Table 7 reveals that the random-intercept LVM explains 54 to 89% of the sample correlations while the random-coefficient LVM does 62 to 93%. The random-coefficient LVM explains high 87 to 93% of the sample correlations in three pairs involving waist circumferences while it explains comparatively low 62 to 72% of the sample correlations between other three pairs of the biomarkers. Although the random-coefficinet LVM does a better job of explaining the sample correlations than the random-intercept LVM, it can be further improved, in particular, for the biomarker pairs that do not involve the waist circumference by relaxing the independence assumption. Another way is to consider a more elaborate structural model having autoregressive random effects of the latent child obesity as the obesity is likely to be correlated between occasions within a person [50, 51].

Acknowledgements: We thank the editor, the associate editor, and two reviewers for their constructive criticism, which helped us improve the quality of the paper. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305D130033 and by the National Institutes of Health through grants R01HL113697 and 1U01HL101064.

References

1. Morrison JA. Obesity and cardiovascular disease risk factors in black and white girls: the NHLBI Growth and Health Study. *American Journal of Public Health* 1992; 82: 1613-1620.

2. Biro FM, Lucky AW, Simbartl LA, Barton BA, Daniels SR, Striegel-Moore R et al. Pubertal maturation in girls and the relationship to anthropometric changes: pathways through puberty. *Journal of Pediatrics* 2003; 142(6): 643-646.

3. Patterson ML, Stern S, Crawford PB, McMahon RP, Similo SL, Schreiber GB et al. Sociodemographic factors and obesity in preadolescence black and white girls: NHLBI's Growth and Health Study. *Journal of the National Medical Association* 1997; 89(9): 594-600.

4. Kimm SY, Glynn NW, Obarzanek E, Kriska AM, Daniels S, Barton BA et al. Relation between the changes in physical activity and body-mass index during adolescence: a multicentre longitudinal study. *Lancet* 2005; 366: 301-307.

5. Kriemler S, Puder J, Zahner L, Roth R, Meyer U, and Bedogni G. Estimation of percentage body fat in 6- to 13-year-old children by skinfold thickness, body mass index and waist circumference. *British Journal of Nutrition* 2010; 104: 1565-1572.

6. Mahoney JL. Adolescent summer care arrangements and risk for obesity the following school year. *Journal of Adolescence* 2011; 34(4): 737-749.

7. Maynard LM, Wisemandle W, Roche AF, Chumlea WC, Guo SS, and Siervogel RM. Childhood Body Composition in Relation to Body Mass Index. *Pediatrics* 2001; 107(2): 344-350.

8. Prentice AM and Jebb SA. Beyond body mass index. *Obesity Review* 2001; 2: 141-147.

9. Krebs NF, Himes JH, Jacobson D, Nicklas TA, Guilday P, and Styne D. Assessment of child and adolescent overweight and obesity. *Pediatrics* 2007; 120(Suppl. 4): S193-S228.

10. Catalano PJ, and Ryan LM. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of American Statistical Association* 1992; 87: 651-658.

11. Cox DR, and Wermuth N. Response models for mixed binary and quantitative variables. *Biometrika* 1992; 79: 441-461.

12. Fitzmaurice GM, and Laird NM. Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of American Statistical Association* 1995; 90: 845-852.

13. Sammel MD, Lin X, and Ryan L. Multivariate linear mixed models for multiple outcomes. *Statistics in Medicine* 1999; 18: 2479-2492.

14. Sammel MD, Ryan LM, and Legler JM. Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B* 1997; 59: 667-678.

15. Roy J and Lin X. Latent Variable Models for Longitudinal Data with Multiple Continuous Outcomes. *Biometrics* 2000; 56(4): 1047-1054.

16. Moustaki I. A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology* 2003; 56: 337-357.

17. Moustaki I and Steele F. Latent variable models for mixed categorical and survival responses, with an application to fertility preferences and family planning in Bangladesh. *Statistical Modeling* 2005; 5: 327-342.

18. Zhu J, Eickhoff JC, Yan P. Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* 2005; 61: 674-683.

19. Song XY, Xia YM and Lee SY. Bayesian semi-parametric analysis of structural equation models with mixed continuous and unordered categorical variables. *Statistics in Medicine* 2009; 28: 2253-2276.

20. Little R and Rubin D. *Statistical Analysis with Missing Data*, New York: Wiley, 2002.

21. Arbuckle JL. *Amos 5.0 Update to the Amos 5.0 Users Guide* [Computer Software and Manual], Chicago: Smallwaters, 2003.

22. Bentler PM. *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, 2007.

23. Muthén L and Muthén B. *Mplus user's guide* (6th ed.), Los Angeles, CA: Muthén and Muthén, 6th edition, 2010.

24. Rubin DB. Inference and missing data. *Biometrika* 1976; 63: 581-592.

25. Roy J and Lin X. Analysis of multivariate longitudinal outcomes with nonignorable dropouts and missing covariates: changes in methadone treatment practices. *Journal of the American Statistical Association* 2002; 97: 40-52.

26. Das S, Chen M, Kim S and Warren N. A Bayesian structural equations model for multilevel data with missing responses and missing covariates. *Bayesian Analysis* 2008; 3 (1): 197-224.

27. Schafer J and Yucel R. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics* 2002; 11: 437-457.

28. Shin Y and Raudenbush SW. Just-identified versus over-identified two-level hierarchical linear models with missing data. *Biometrics* 2007; 63: 1262-1268.

29. Shin Y and Raudenbush SW. The causal effect of class size on academic achievement: Multivariate instrumental variable estimators with data missing at random, *Journal of Educational and Behavioral Statistics* 2011; 36: 154-185.

30. Shin Y and Raudenbush SW. Efficient analysis of Q-level nested hierarchical general linear models given ignorable missing data. *International Journal of Biostatistics* 2013; 9(1): 109-133.

31. Goldstein H and Browne W. Multilevel factor analysis modeling using markov chain monte carlo estimation in G. Marcoulides and Moustaki I, eds., *Latent variable and latent structure models*, London: Lawrence Erlbaum, 2002.

32. Goldstein H, Carpenter J, Kenward MG, and Levin KA. Multilevel models with multivariate mixed response types. *Statistical Modeling* 2009; 9: 173-197.

33. Goldstein H and Kounali D. Multilevel multivariate modeling of childhood growth, numbers of growth measurements and adult characteristics. *Journal of the Royal Statistical Society, Series A* 2009; 172: 599-613.

34. Dempster AP, Laird NM, and Rubin DB. Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* 1977; 76: 1-38.

35. Raudenbush SW and Bryk AS. *Hierarchical Linear Models*. Newbury Park, CA: Sage, 2002.

36. Shin Y and Raudenbush SW. A Latent Cluster Mean Approach to The Contextual Effects Model with Missing Data. *Journal of Educational and Behavioral Statistics* 2010; 35: 26-53.

37. Meng X. Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statistical Science* 1994; 9: 538-558.

38. Danielzik S, Czerwinski-Mast M, Langnse K, Dilba B, and Mller MJ. Parental overweight, socioeconomic status and high birth weight are the major determinants of overweight and obesity in 5-7 y-old children: baseline data of the Kiel Obesity Prevention Study (KOPS). *International Journal of Obesity* 2004; 28: 1494-1502.

39. Lamerz A, Kuepper-Nybelen J, Wehle C, Bruning N, Trost-Brinkhues G, Brenner H, Hebebrand J, and Herpertz-Dahlmann B. Social class, parental education, and obesity prevalence in a study of six-year-old children in Germany. *International Journal of Obesity* 2005; 29: 373-380.

40. Strauss RS, Knight J. Influence of the home environment on the development of obesity in children. *Pediatrics* 1999; 103(6): e85-92.

41. Pinheiro JC and Bates DM. *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer, 2000.

42. Self SG and Liang KY. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 1987; 82: 605-610.

43. Stram DO and Lee JW. Variance components testing in the longitudinal mixed effects model. *Biometrics* 1994; 50: 1171-1177.

44. Wagenmakers E and Farrell S. AIC model selection using Akaike weights. *Psychonomic Bulletin and Review* 2004; 11 (1): 192-196.

45. Obarzanek E, Schreiber GB, Crawford PB, Goldman SR, Barrier PM, Frederick MM et al. Energy intake and physical activity in relation to indexes of body fat: the National Heart, Lung, and Blood Institute Growth and Health Study. *American Journal of Clinical Nutrition* 1994; 60(1): 15-22.

46. Chambers JA and Swanson V. A health assessment tool for multiple risk factors for obesity: age and sex differences in the prediction of body mass index. *British Journal of Nutrition* 2010; 104(2): 298-307.

47. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, Wiley, 1987.

48. Shin Y. *Inference and applications in hierarchical linear models with missing data*, University of Michigan, 2003.

49. Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall, 1997.

50. Bollen KA and Curran PJ. Autoregressive Latent Trajectory (ALT) models: A synthesis of two tradition. *Sociological methods and research* 2004; 32: 336-383.

51. Curran PJ and Bollen KA. The best of both worlds: Combining autoregressive and latent growth curve models. In Collins L. M. and Sayar, A. G. (Eds.), *New methods for the analysis of change* (PP 105-136). Washington, D.C.: American Psychological Association, 2001.

Appendix A

Transformation Formular Deviation

It is easily to derive that the responses in models (1) and (3) are distributed as

$$U_{ik}|S_{ik}, Y_{2i} \sim N(\mu_{1ik}, V_{1ik}),\ [U_{ik}\ S_{ik}^T\ Y_{2i}^T]^T \sim N(\mu_{2ik}, V_{2ik}), \tag{a}$$

where

$$\mu_{1ik} = S_{ik}^T\alpha_1 + Y_{2i}^T\alpha_2 + W_{1ik}^T\alpha_3 + W_{2i}^T\alpha_4,\ \ V_{1ik} = B_{ik}^T D B_{ik} + 1,$$

$$\mu_{2ik} = \begin{bmatrix} \beta_{u1}^T W_{1ik} + \beta_{u2}^T W_{2i} \\ \beta_{s1} W_{1ik} + \beta_{s2} W_{2i} \\ \beta_{22} W_{2i} \end{bmatrix},\ \ V_{2ik} = \begin{bmatrix} B_{ik}^T T_{uu} B_{ik} + \Sigma_{uu} & B_{ik}^T T_{us} + \Sigma_{us} & B_{ik}^T T_{u2} \\ T_{su} B_{ik} + \Sigma_{su} & T_{ss} + \Sigma_{ss} & T_{s2} \\ T_{2u} B_{ik} & T_{2s} & T_{22} \end{bmatrix}.$$

Let us express model (3) such that it recognizes the latent random effect b_{s_i} of S_{ik} as

$$[U_{ik}\ (S_{ik} - b_{s_i})^T\ b_{s_i}^T\ Y_{2i}^T]^T \sim N(\mu_{3ik}, V_{3ik}) \tag{b}$$

with

$$\mu_{3ik} = \begin{bmatrix} \beta_{u1}^T W_{1ik} + \beta_{u2}^T W_{2i} \\ \beta_{s1} W_{1ik} + \beta_{s2} W_{2i} \\ 0 \\ \beta_{22} W_{2i} \end{bmatrix},\ \ V_{3ik} = \begin{bmatrix} B_{ik}^T T_{uu} B_{ik} + \Sigma_{uu} & \Sigma_{us} & B_{ik}^T T_{us} & B_{ik}^T T_{u2} \\ \Sigma_{su} & \Sigma_{ss} & 0 & 0 \\ T_{su} B_{ik} & 0 & T_{ss} & T_{s2} \\ T_{2u} B_{ik} & 0 & T_{2s} & T_{22} \end{bmatrix}.$$

Then, a regression of U_{ik} on the other variables leads to

$$U_{ik}|S_{ik} - b_{s_i}, b_{s_i}, Y_{2i} \sim N(\mu_{4ik}, V_{4ik}) \tag{c}$$

where

$$\mu_{4ik} = (B_{ik}^T T_{us|2} T_{ss|2}^{-1} - \Sigma_{us}\Sigma_{ss}^{-1}) b_{si} + S_{ik}^T \Sigma_{ss}^{-1}\Sigma_{su} + Y_{2i}^T T_{22}^{-1} \left(T_{u2} - T_{2s} T_{ss|2}^{-1} T_{su|2}\right) B_{ik}$$
$$+ W_{1ik}^T(\beta_{u1} - \beta_{s1}^T \Sigma_{ss}^{-1}\Sigma_{su}) + W_{2i}^T \left(\beta_{u2} - \beta_{22}^T T_{22}^{-1}(T_{2u} - T_{2s} T_{ss|2}^{-1} T_{su|2}) B_{ik} - \beta_{s2}^T \Sigma_{ss}^{-1}\Sigma_{su}\right),$$
$$V_{4ik} = \Sigma_{uu} - \Sigma_{us}\Sigma_{ss}^{-1}\Sigma_{su} + B_{ik}^T (T_{uu|2} - T_{us|2} T_{ss|2}^{-1} T_{su|2}) B_{ik}.$$

Model (c) implies model (a) if $b_{si} = 0$. Model (c) with $b_{si} = 0$, however, has too strong assumption that S_{ik} does not vary across level-2 unit. The violation of the assumption leads to substantially biased inferences. Alternatively, model (c) implies model (a) if

$$\alpha_1^T = B_{ik}^T T_{us|2} T_{ss|2}^{-1} = \Sigma_{us}\Sigma_{ss}^{-1} \text{ and } \Sigma_{uu} - \alpha_1^T \Sigma_{ss}\alpha_1 = 1. \qquad (d)$$

In the following, we discuss constraints and transformation formulas for two cases: $B_{ik} = 1$ and $B_{ik}^T = [1\ X_{dik}^T]$ with p_5 covariates X_{dik} having random coefficients in model (1). If $B_{ik} = 1$, then one-to-one transformation formulas between models (a) and (c) are

$$\begin{aligned}
&\alpha_1 = \Sigma_{ss}^{-1}\Sigma_{su},\ \alpha_2 = T_{22}^{-1}(T_{2u} - T_{2s}\alpha_1),\ \alpha_3 = \beta_{u1} - \beta_{s_1}^T\alpha_1, \\
&\alpha_4 = \beta_{u2} - \beta_{s_2}^T\alpha_1 - \beta_{22}^T\alpha_2,\ D = T_{uu} - \alpha_2^T T_{22}\alpha_2 - 2\alpha_1^T T_{s2}\alpha_2 - \alpha_1^T T_{ss}\alpha_1, \qquad (e) \\
&1 = \Sigma_{uu} - \alpha_1^T\Sigma_{ss}\alpha_1,\ T_{us} = \alpha_1^T T_{ss} + \alpha_2^T T_{2s}.
\end{aligned}$$

If $B_{ik}^T = [1\ X_{dik}^T]$, then let $b_{ui} = [b_{u_0 i}\ b_{u_1 i}^T]^T$, $T_{uu} = \begin{bmatrix} T_{u_0u_0} & T_{u_0u_1} \\ T_{u_1u_0} & T_{u_1u_1} \end{bmatrix}$, $T_{us} = \begin{bmatrix} T_{u_0s} \\ 0 \end{bmatrix}$, $T_{su} = T_{us}^T$, and $T_{u2} = [T_{u_02}^T\ 0]^T$. Note that we assume $\text{cov}(b_{u_1 i}, b_{si}) = \text{cov}(b_{u_1 i}, b_{2i}) = 0$. Non-zero covariances can be estimated, but they introduce extraneous terms and make interpretable difficulty. Let $\tilde{T} = \begin{bmatrix} \alpha_2^T T_{22}\alpha_2 + 2\alpha_1^T T_{s2}\alpha_2 + \alpha_1^T T_{ss}\alpha_1 & 0 \\ 0 & 0 \end{bmatrix}$. Then the one-to-one transformation formulas for α_2, D, and T_{u_0s} are

$$\alpha_2 = T_{22}^{-1}(T_{2u_0} - T_{2s}\alpha_1),\ D = T_{uu} - \tilde{T},\ T_{u_0s} = \alpha_1^T T_{ss} + \alpha_2^T T_{2s}, \qquad (f)$$

and the others keep same as these in (e).

Appendix B

Parameter Estimation

The maximum likelihood estimators (MLE) of the complete data derived from their likelihood

$L(\theta|R_i, U_i, S_{1i}, Y_{2i}, b_{u_i}, b_{s_i}, b_{r_i})$ are

$$
\begin{aligned}
&\hat{\gamma}_j^{(k)} = \hat{\gamma}_j^{(k-1)} + \left(\sum_{i=1}^{n}\sum_{k=1}^{k_i} U_{ik}^* U_{ik}^{*T}\right)^{-1} \sum_{i=1}^{n}\sum_{k=1}^{k_i} U_{ik}^* e_{ikj},\\
&\hat{\beta}_1^{*(k)} = \hat{\beta}_1^{*(k-1)} + \left(\sum_{i=1}^{n} \Sigma^{-1} \otimes (W_{usi}^T W_{usi})\right)^{-1} \sum_{i=1}^{n} \Sigma^{-1} \otimes (W_{usi}^T \epsilon_{1i}^*),\\
&\hat{\beta}_2^{(k)} = \hat{\beta}_2^{(k-1)} + \left(\sum_{i=1}^{n} T_{2|1}^{-1} \otimes (W_{2i} W_{2i}^T)\right)^{-1} \sum_{i=1}^{n} (T_{2|1}^{-1} \otimes W_{2i}) \left(b_{2i} - T_{21} T_{11}^{-1} b_{1i}\right), \qquad \text{(g)}\\
&\hat{\xi}_j = \frac{1}{n}\sum_{i=1}^{n} a_{ij}^2, \quad \hat{\tau}_j = \frac{1}{\sum_{i=1}^{n} k_i} \sum_{i=1}^{n}\sum_{k=1}^{k_i} e_{ikj}^2,\\
&\hat{\Sigma} = \frac{1}{\sum_{i=1}^{n} k_i} \sum_{i=1}^{n}\sum_{k=1}^{k_i} \epsilon_{1ik}\epsilon_{1ik}^T, \quad \hat{T} = \frac{1}{n}\sum_{i=1}^{n} b_i^* b_i^{*T},\\
&\hat{\alpha}^{(k)} = \hat{\alpha}^{(k-1)} + \left(\sum_{i=1}^{n}\sum_{k=1}^{k_i} A_{ik} A_{ik}^T\right)^{-1} \sum_{i=1}^{n}\sum_{k=1}^{k_i} A_{ik}\epsilon_{ik}, \quad \hat{D} = \frac{1}{n}\sum_{i=1}^{n} b_i b_i^T.
\end{aligned}
$$

Given $\hat{\alpha}$ and $\hat{D}$ for a random-intercept model (1), we update the estimators $\hat{\Sigma}_{us}$, $\hat{\Sigma}_{uu}$, $\hat{T}_{uu}$, $\hat{\beta}_{u1}$, $\hat{\beta}_{u2}$, $\hat{T}_{u2}$, and $\hat{T}_{us}$ in model (7) via formulas (e). Given $\hat{\alpha}$ and $\hat{D}$ for a random-coefficient model (1), we update the estimators, $\hat{\Sigma}_{us}$, $\hat{\Sigma}_{uu}$, $\hat{T}_{uu}$, $\hat{\beta}_{u1}$, $\hat{\beta}_{u2}$, $\hat{T}_{u_0 2}$, and $\hat{T}_{u_0 s}$ in model (7) via formulas (f) and set $T_{u_1 2} = T_{u_1 s} = 0$.

At E-step, we estimate the following conditional expectations.

$$
\begin{aligned}
&\tilde{U}_{ik} = \beta_{u1}^T W_{1ik} + \beta_{u2}^T W_{2i} + \Delta_u (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),\\
&E(U_{ik}^2|Y_i^\circ) = \tilde{U}_{ik}^2 + B_{ik}^T T_{uu} B_{ik} + \Sigma_{uu} - \Delta_u (V_i^\circ)^{-1} \Delta_u^T,\\
&\tilde{e}_{ikj} = \Delta_{er}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ), \quad E(e_{ikj}^2|Y_i^\circ) = \tilde{e}_{ikj}^2 + \tau_j - \Delta_{er}(V_i^\circ)^{-1}\Delta_{er}^T,\\
&\tilde{a}_{ij} = \Delta_a (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ), \quad E(a_{ij}^2|Y_i^\circ) = \tilde{a}_{ij}^2 + \xi_j - \Delta_a (V_i^\circ)^{-1}\Delta_a^T, \qquad \text{(h)}\\
&E(\epsilon_{1i}^*|Y_i^\circ) = \Delta_{es}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ), \quad \tilde{\epsilon}_{1ik} = \Delta_e (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ),\\
&\tilde{b}_i^* = \Delta_b (V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ), \quad E(b_i^* b_i^{*T}|Y_i^\circ) = \tilde{b}_i^* \tilde{b}_i^{*T} + T - \Delta_b (V_i^\circ)^{-1}\Delta_b^T,\\
&E(U_{ik} e_{ikj}|Y_i^\circ) = \tilde{U}_{ik}\tilde{e}_{ikj} - \Delta_u (V_i^\circ)^{-1}\Delta_{er}^T, \quad E(\epsilon_{1ik}\epsilon_{1ik}^T|Y_i^\circ) = \tilde{\epsilon}_{1ik}\tilde{\epsilon}_{1ik}^T + \Sigma - \Delta_e (V_i^\circ)^{-1}\Delta_e^T,
\end{aligned}
$$

where $\Delta_u = \left[\Delta_{u1} \;\; \Delta_{u2} \;\; B_{ik}^T T_{u2}\right] O_i^T$, $\Delta_{er} = [0_{1\times((k-1)J+j-1)} \;\; \tau_j \;\; 0_{1\times(J-j)} \;\; 0_{1\times((k_i-k)J+p_1 k_i+p_2)}] O_i^T$,

$\Delta_a = [0_{1\times(j-1)k_i}\ 1_{k_i}^T\xi_j\ 0_{1\times((J-j)k_i+p_1k_i+p_2)}]O_i^T$, $\Delta_{es} = \begin{bmatrix} I_{k_i}\otimes\Sigma_{uu}\otimes\gamma_1^T & I_{k_i}\otimes\Sigma_{us} & 0 \\ I_{k_i}\otimes\Sigma_{su}\otimes\gamma_1^T & I_{k_i}\otimes\Sigma_{ss} & 0 \end{bmatrix} O_i^T$, $\Delta_e =$

$\begin{bmatrix} \Sigma_{uu}\otimes\Delta_k^T\otimes\gamma_1^T & \Delta_k^T\otimes\Sigma_{us} & 0 \\ \Sigma_{su}\otimes\Delta_k^T\otimes\gamma_1^T & \Delta_k^T\otimes\Sigma_{ss} & 0 \end{bmatrix} O_i^T$, and $\Delta_b = \begin{bmatrix} (T_{uu}B_i^T)\otimes\gamma_1^T & 1_{k_i}^T\otimes T_{us} & T_{u2} \\ (T_{su}B_i^T)\otimes\gamma_1^T & 1_{k_i}^T\otimes T_{ss} & T_{s2} \\ (T_{2u}B_i^T)\otimes\gamma_1^T & 1_{k_i}^T\otimes T_{2s} & T_{22} \end{bmatrix} O_i^T$ for $\Delta_{u1} =$

$(B_{ik}^T T_{uu} B_{ik}^T + [0_{1\times(k-1)}\ \Sigma_{uu}\ 0_{1\times(k_i-k)}])\otimes\gamma_1^T$, $\Delta_{u2} = 1_{k_i}^T\otimes(B_{ik}^T T_{us}) + [0_{1\times(k-1)p_1}\ \Sigma_{us}\ 0_{1\times(k_i-k)p_1}]$,

and Δ_k is a vector with the k^{th} element equal to 1 and zero otherwise,

In addition, we calculate $E(A_{ik}A_{ik}^T|Y_i^\circ)$, $E(A_{ik}\epsilon_{ik}|Y_i^\circ)$, and $E(b_ib_i^T|Y_i^\circ)$ in the LVM.

$$E(A_{ik}A_{ik}^T|Y_i^\circ) = \begin{bmatrix} E(S_{ik}S_{ik}^T|Y_i^\circ) & E(S_{ik}Y_{2i}^T|Y_i^\circ) & \tilde{S}_{ik}W_{1ik}^T & \tilde{S}_{ik}W_{2i}^T \\ E(Y_{2i}S_{ik}^T|Y_i^\circ) & E(Y_{2i}Y_{2i}^T|Y_i^\circ) & \tilde{Y}_{2i}W_{1ik}^T & \tilde{Y}_{2i}W_{2i}^T \\ W_{1ik}\tilde{S}_{ik}^T & W_{1ik}Y_{2i}^T & W_{1ik}W_{1ik}^T & W_{1ik}W_{2i}^T \\ W_{2i}\tilde{S}_{ik}^T & W_{2i}\tilde{Y}_{2i}^T & W_{2i}W_{1ik}^T & W_{2i}W_{2i}^T \end{bmatrix}, \quad \text{(i)}$$

$$E(A_{ik}\epsilon_{ik}|Y_i^\circ) = \begin{bmatrix} \tilde{S}_{ik}^T\tilde{\epsilon}_{ik} - \Delta_s(V_i^\circ)^{-1}\Delta_{ec}^T \\ \tilde{Y}_{2i}^T\tilde{\epsilon}_{ik} - \Delta_y(V_i^\circ)^{-1}\Delta_{ec}^T \\ W_{1ik}\tilde{\epsilon}_{ik} \\ W_{2i}\tilde{\epsilon}_{ik} \end{bmatrix}, \quad \text{(j)}$$

$$E(b_ib_i^T|Y_i^\circ) = \tilde{b}_i\tilde{b}_i^T + D - \Delta_{ac}(V_i^\circ)^{-1}\Delta_{ac}^T \quad \text{(k)}$$

where $\tilde{S}_{ik} = \beta_{s1}W_{1ik} + \beta_{s2}W_{2i} + \Delta_s(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ)$, $E(S_{ik}S_{ik}^T|Y_i^\circ) = \tilde{S}_{ik}\tilde{S}_{ik}^T + T_{ss} + \Sigma_{ss} - \Delta_s(V_i^\circ)^{-1}\Delta_s^T$, $\tilde{Y}_{2i} = \beta_{22}W_{2i} + \Delta_y(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ)$, $E(Y_{2i}Y_{2i}^T|Y_i^\circ) = \tilde{Y}_{2i}\tilde{Y}_{2i}^T + T_{22} - \Delta_y(V_i^\circ)^{-1}\Delta_y^T$, $E(S_{ik}Y_{2i}^T|Y_i^\circ) = \tilde{S}_{ik}\tilde{Y}_{2i}^T + T_{s2} - \Delta_s(V_i^\circ)^{-1}\Delta_y^T$, and $\tilde{\epsilon}_{ik} = \Delta_{ec}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ)$ for $\Delta_s = [\Delta_{s1}\ \Delta_{s2}\ 1_{k_i}^T \otimes T_{s2}]O_i^T$, $\Delta_{s1} = \left((T_{su}B_i^T) + [0_{p_1\times(k-1)}\ \ T_{su}\ \ 0_{p_1\times(k_i-k)}]\right)\otimes\gamma_1^T$, $\Delta_{s2} = 1_{k_i}^T\otimes T_{ss} + [0_{p_1\times(k-1)p_1}\ \ \Sigma_{ss}\ \ 0_{p_1\times(k_i-k)p_1}]$, $\Delta_y = [(T_{2u}B_i^T)\otimes\gamma_1^T\ \ 1_{k_i}^T\otimes T_{2u}\ T_{22}]O_i^T$, $\Delta_{ec} = [0_{1\times(k-1)J}\ \ \gamma_1^T\ \ 0_{1\times(k_iJ-kJ+p_1k_i+p_2)}]O_i^T$, $\tilde{b}_i = \Delta_{ac}(V_i^\circ)^{-1}(Y_i^\circ - \mu_i^\circ)$ and $\Delta_{ac} = [(DB_i^T)\otimes\gamma_1^T\ 0\ 0]O_i^T$.

Appendix C

Calculation of the Information Matrix

The information matrix is obtained by differentiating twice the observed marginal multivariate normal log-likelihood with mean and covariance given in (9), but we introduce new parameters α_1 and α_2, which are defined in (5). Consequently, parameters Σ_{us}, T_{u2}, T_{us}, and Σ_{uu} are the functions of α_1, α_2 and the other elements in Σ and T as

$$\begin{aligned} &\Sigma_{us} = \alpha_1^T \Sigma_{ss}, \;\; T_{u2} = \alpha_2^T T_{22} + \alpha_1^T T_{s2}, \\ &T_{us} = \alpha_1^T T_{ss} + \alpha_2^T T_{2s}, \;\; \Sigma_{uu} = 1 + \alpha_1^T \Sigma_{ss} \alpha_1. \end{aligned} \tag{l}$$

Let $W(A)$ denote a vector by horizontally arranging the elements in the matrix A and $\gamma = (\gamma_0, \gamma_1, \beta^{**})$ in which $\beta^{**} = [\beta_u^T \; W(\beta_{s1})^T \; W(\beta_{s2})^T \; W(\beta_{22})^T]^T$. The arrangement makes us easily extract the covariances between $W(\beta_{s1})$, $W(\beta_{s2})$, $W(\beta_{22})$ and α_1, α_2 to estimate the variances of α_3, α_4 and D by multivariate Delta method. let $H_i = O_i \oplus_{j=1}^3 H_{ij}$ with $H_{i1} = [1_{k_i} \otimes I_J \;\; (W_{ui}\beta_u) \otimes I_J]$, $H_{i2} = [1_{k_i} \otimes I_{p_1} \;\; W_{1i} \otimes I_{p_1} \;\;\; W_{2i} \otimes 1_{k_i} \otimes I_{p_1}]$, and $H_{i3} = [I_{p_2} \;\;\; I_{p_2} \otimes W_{2i}]$, $F_i = O_i \oplus_{j=1}^3 F_{ij}$ with $F_{i1} = [1_{k_i} \otimes I_J \;\;\; W_{ui} \otimes \gamma_1]$, $F_{i2} = H_{i2}$, and $F_{i3} = H_{i3}$, $G_i = H_i \begin{bmatrix} I_J \\ 0_{(J+p_3p_1+p_4p_2)\times J} \end{bmatrix}$, $M_i = H_i \begin{bmatrix} 0_{J\times J} \\ I_J \\ 0_{(p_3p_1+p_4p_2)\times J} \end{bmatrix}$, and $Q_i = F_i \begin{bmatrix} 0_{J\times(p_3+p_3p_1+p_4p_2)} \\ I_{p_3+p_3p_1+p_4p_2} \end{bmatrix}$. The expected information matrix for the MLE of $\gamma = (\gamma_0, \gamma_1, \beta^{**})$ is

$$I_{\gamma\gamma} = \sum_{i=1}^{n} \begin{bmatrix} G_i^T (V_i^\circ)^{-1} G_i & G_i^T (V_i^\circ)^{-1} M_i & G_i^T (V_i^\circ)^{-1} Q_i \\ M_i^T (V_i^\circ)^{-1} G_i & A + M_i^T (V_i^\circ)^{-1} M_i & M_i^T (V_i^\circ)^{-1} Q_i \\ Q_i^T (V_i^\circ)^{-1} G_i & Q_i^T (V_i^\circ)^{-1} M_i & Q_i^T (V_i^\circ)^{-1} Q_i, \end{bmatrix} \tag{m}$$

where A has its $(j,k)th$ component $\frac{1}{2}\text{tr}\left((V_i^\circ)^{-1} \frac{\partial V_i^\circ}{\partial \beta_{1j}} (V_i^\circ)^{-1} \frac{\partial V_i^\circ}{\partial \beta_{1k}}\right)$.

Define $V(A)$ a vector by vertically arranging the distinct elements of the matrix A. Let $\delta = (\xi, \tau, T_{uu}, V(T_{ss}), V(T_{2s}), V(T_{22}), V(\Sigma_{ss}), \alpha_1, \alpha_2) = (\delta_1, \delta_2, \cdots, \delta_M)$ for $M = 35$ and $M = 36$

in a random-intercept model (1) and in a random-coefficient model (1), respectively. Then

$$
\begin{aligned}
I_{\delta_j\beta_{1k}} &= \tfrac{1}{2}\sum_{i=1}^{n} \text{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \delta_j}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \beta_{1k}}\right), \\
I_{\delta_j\delta_k} &= \tfrac{1}{2}\sum_{i=1}^{n} \text{tr}\left((V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \delta_j}(V_i^\circ)^{-1}\frac{\partial V_i^\circ}{\partial \delta_k}\right),
\end{aligned} \tag{n}
$$

and $I_{\delta\gamma_0} = 0$, $I_{\delta\beta^{**}} = 0$, where $\frac{\partial V_i^\circ}{\partial \delta_m}$ is an element-wise derivative with respect to δ_m for $m = 1, 2, \cdots, M$.

Appendix D

The variance calculation of the parameters in the LVM

The variances of the estimators α_1, α_2, β_0, β_1, ξ and τ in the LVM are estimated in Appendix C. Let $\theta_1 = [\beta_{u1}^T\ W(\beta_{s1})^T\ \alpha_1^T]^T$, $\theta_2 = [\beta_{u2}^T\ W(\beta_{s2})^T\ W(\beta_{22})^T\ \alpha_1^T\ \alpha_2^T]^T$, and $\theta_3 = [T_{uu}\ V(T_{ss})^T\ V(T_{s2})^T\ V(T_{22})^T\ \alpha_1^T\ \alpha_2^T]^T$. From the transformation formulas (5) and Delta method, the covariances of $\hat{\alpha}_3$, $\hat{\alpha}_4$, and $\hat{D}$ with $B_{ik} = 1$ are estimated as

$$
\text{cov}\hat{\alpha}_3 = \widehat{\nabla f}_1 \text{cov}\hat{\theta}_1 \widehat{\nabla f}_1^T,\ \text{cov}\hat{\alpha}_4 = \widehat{\nabla f}_2 \text{cov}\hat{\theta}_2 \widehat{\nabla f}_2^T,\ \text{cov}\hat{D} = \widehat{\nabla f}_3 \text{cov}\hat{\theta}_3 \widehat{\nabla f}_3^T \tag{o}
$$

where $\text{cov}\hat{\theta}_i$ can be extracted from the inverse of the fisher information matrix in Appendix C, $\nabla f_1 = \left[I_{p_3}\ -\alpha_1^T \otimes I_{p_3}\ -\beta_{s_1}^T\right]$, $\nabla f_2 = [I_{p_4}\ -\left(\alpha_1^T \otimes I_{p_4}\right) - \left(\alpha_2^T \otimes I_{p_4}\right)\ -\beta_{s_2}^T\ -\beta_{22}^T]$, and $\nabla f_3 = \left[1\ \left(\frac{\partial D}{\partial V(T_{ss})}\right)^T\ \left(\frac{\partial D}{\partial V(T_{s2})}\right)^T\ \left(\frac{\partial D}{\partial V(T_{22})}\right)^T\ \left(\frac{\partial D}{\partial \alpha_1}\right)^T\ \left(\frac{\partial D}{\partial \alpha_2}\right)^T\right]$ for $\frac{\partial D}{\partial V(T_{ss})_j} = -\alpha_1^T \frac{\partial T_{ss}}{\partial V(T_{ss})_j}\alpha_1$, $\frac{\partial D}{\partial V(T_{s2})_j} = -2\alpha_1^T \frac{\partial T_{s2}}{\partial V(T_{s2})_j}\alpha_2$, $\frac{\partial D}{\partial V(T_{22})_j} = -\alpha_2^T \frac{\partial T_{22}}{\partial V(T_{22})_j}\alpha_2$, $\frac{\partial D}{\partial \alpha_1} = -2T_{s2}\alpha_2 - 2T_{ss}\alpha_1$, and $\frac{\partial D}{\partial \alpha_2} = -2T_{22}\alpha_2 - 2T_{2s}\alpha_1$.

Note that $\frac{\partial T_{ss}}{\partial V(T_{ss})_j}$, $\frac{\partial T_{s2}}{\partial V(T_{s2})_j}$, and $\frac{\partial T_{22}}{\partial V(T_{22})_j}$ are unknown. We know for any p-by-p matrix ϖ_1 the first derivative of the $(l,k)th$ $(k > l)$ element is

$$
\frac{\partial \varpi_1}{\partial \varpi_{1kl}} = \begin{cases} \delta_k\delta_l^T + \delta_l\delta_k^T & k > l \\ \delta_k\delta_l^T & k = l, \end{cases} \tag{p}
$$

and for any p-by-q ($p \neq q$) matrix ϖ_2 the first derivative of the $(l,k)th$ element is

$$\frac{\partial \varpi_2}{\partial \varpi_{2kl}} = \delta_k \eta_l^T, k = 1, 2, \cdots, p,\ l = 1, 2, \cdots, q \tag{q}$$

where δ_h and η_h are p-by-1 and q-by-1 vectors with the h^{th} element equal to one and zero otherwise, respectively. After we vertically arrange the distinct elements in ϖ_1 and ϖ_2, the first derivative of the j^{th} element for $j = 1, 2, \cdots, p(p+1)/2$ or $j = 1, 2, \cdots, pq$ has a one-to-one transformation with equations (p) and (q), respectively. Similarly, the variances of distinct elements in D could be estimated for a random-coefficient model (1).

Table 1: *Estimation of the simulated LVM (11) by three different estimation methods*

Model	Para.	True value	Estimate (S.E.[a])		
			Benchmark	Just-identified	Over-identified
(1)	α	1	1.031 (0.075)	1.032 (0.075)	**0.901** (0.065)
		1	1.007 (0.024)	1.007 (0.024)	**0.882** (0.021)
	D	1	0.999 (0.069)	0.999 (0.069)	**0.751** (0.052)
(2)	γ_0	1	0.993 (0.054)	0.993 (0.054)	0.993 (0.054)
		1	1.026 (0.055)	1.026 (0.055)	1.026 (0.054)
	γ_1	1	0.987 (0.014)	0.987 (0.014)	**1.129** (0.016)
		1	0.987 (0.014)	0.987 (0.014)	**1.129** (0.016)
	ξ	0.25	0.268 (0.035)	0.268 (0.035)	0.267 (0.035)
		0.25	0.291 (0.035)	0.291 (0.036)	0.291 (0.036)
	τ	0.25	0.240 (0.018)	0.240 (0.018)	0.240 (0.018)
		0.25	0.258 (0.019)	0.258 (0.019)	0.258 (0.018)

[a] standard error

Table 2: *Estimation of the simulated LVM (11) given ignorable missing data*

Model	Para.	True value	Estimate (S.E.[a])	
			Just-identified	Over-identified
(1)	α	1	0.994 (0.033)	**0.878** (0.031)
		1	1.061 (0.081)	**0.932** (0.058)
	D	1	1.010 (0.082)	**0.763** (0.040)
(2)	γ_0	1	0.987 (0.056)	0.987 (0.044)
		1	1.013 (0.056)	1.013 (0.045)
	γ_1	1	0.984 (0.019)	**1.120** (0.033)
		1	0.981 (0.019)	**1.117** (0.033)
	ξ	0.25	0.248 (0.025)	0.248 (0.031)
		0.25	0.317 (0.041)	0.318 (0.033)
	τ	0.25	0.249 (0.025)	0.249 (0.018)
		0.25	0.260 (0.025)	0.260 (0.018)

[a] standard error

Table 3: *NGHS data for analysis*

level	variable	description	mean (S.E.)	missing (%)
	BMI	BMI(kg/m^2)	22.42 (5.81)	308 (1.5)
	Skinfold	sum of skinfolds (mm)	45.11 (24.88)	783 (3.8)
	Waist	max. below-waist circumference (cm)	93.95 (12.87)	2807 (13.5)
level 1	PercentFat	percent fat by BIA	25.29 (11.49)	1694 (8.1)
	Age	age in years at time of visit	14.36 (2.99)	0 (0.0)
	TV	TV watching (hours/week)	31.35 (21.32)	4834 (23.2)
	PhysicalAct	physical activity pattern score	17.35 (17.75)	6573 (31.5)
	Maturation	maturation stage	2.10 (1.03)	1063 (5.1)
	MotherBMI	mother's BMI	27.35 (6.91)	6772 (32.4)
	ParentEd	[a]maximum parental education	0.75 (0.43)	0 (0.0)
level 2	Income	[b]household income	1.06 (0.83)	1156 (5.5)
	White	[c]race (white/black)	0.48 (0.50)	0 (0.0)
	OneParents	[d]single-parent family	0.31 (0.46)	0 (0.0)

[a] 1 if some college or more, 0 otherwise
[b] 0 if $<$ \$20k, 2 if $\geq$ \$40k, 1 otherwise
[c] 1 if white, 0 if black
[d] 1 if single parent family, 0 if two-parent family

Table 4: *Estimated model (1)*

Para.	Covariate	Estimate(S.E.)			
		Random intercept		Random coefficient	
		MCAR	MAR	MCAR	MAR
α_1	TV	0.005 ‡ (0.001)	0.004‡ (0.001)	0.006‡ (0.001)	0.004‡ (0.001)
	PhysicalAct	-0.004‡ (0.001)	-0.003‡ (0.001)	-0.002† (0.001)	-0.002† (0.001)
	Maturation	0.504‡ (0.043)	0.347‡ (0.021)	0.606‡ (0.051)	0.387‡ (0.024)
α_2	MotherBMI	0.149‡ (0.012)	0.150‡ (0.011)	0.156‡ (0.015)	0.133‡ (0.013)
	Income	-0.285† (0.127)	-0.183 (0.096)	-0.176 (0.158)	0.078 (0.114)
α_3	AGE	0.454 ‡ (0.047)	0.502‡ (0.020)	0.647‡ (0.057)	0.713‡ (0.024)
	AGE^2	-0.025 (0.013)	-0.025‡ (0.005)	-0.039† (0.015)	-0.031‡ (0.005)
	AGE×White	-0.032 (0.063)	-0.057† (0.026)	-0.081 (0.079)	-0.124‡ (0.033)
α_4	ParentEd	0.041 (0.219)	0.012 (0.155)	0.094 (0.237)	0.144 (0.179)
	White	-0.330 (0.176)	-0.309† (0.137)	-0.694 ‡ (0.225)	-0.938‡ (0.186)
	OneParent	0.357 (0.217)	0.380† (0.159)	0.610† (0.271)	0.568‡ (0.185)
D_{00}^a		8.735 (0.575)	8.040 (0.386)	16.935 (0.802)	16.482 (0.560)
D_{01}				0.818 (0.060)	0.942 (0.043)
D_{11}				0.154 (0.009)	0.155 (0.006)

[a] $D_{00} = D$ in a random-intercept model (1)
† p-value< 0.05, ‡ p-value< 0.01

Table 5: *Estimated model (2) given incomplete data*

Model (2) with	Biomarker	$\hat{\beta}_{0j}$	$\hat{\beta}_{1j}$	$\hat{\tau}_j$	$\hat{\xi}_j$
random intercept	BMI	22.74 (0.09)	1.46 (0.01)	1.06 (0.02)	1.07 (0.09)
	Skinfold	47.30 (0.37)	5.24 (0.04)	75.05 (0.83)	73.38 (2.73)
	Waist	93.46 (0.29)	4.37 (0.03)	2.02 (0.10)	29.72 (1.16)
	PercentFat	25.88 (0.19)	2.69 (0.02)	15.29 (0.18)	21.00 (0.75)
random coefficient	BMI	22.74 (0.09)	1.08 (0.01)	0.54 (0.01)	0.86 (0.08)
	Skinfold	47.37 (0.38)	3.95 (0.03)	65.06 (0.73)	69.98 (2.56)
	Waist	93.49 (0.28)	3.04 (0.02)	6.04 (0.11)	24.79 (0.97)
	PercentFat	25.85 (0.19)	1.95 (0.02)	15.40 (0.18)	21.88 (0.76)

Table 6: *Complete-case analysis of model (2)*

Model (2) with	Biomarker	$\hat{\beta}_{0j}$	$\hat{\beta}_{1j}$	$\hat{\tau}_j$	$\hat{\xi}_j$
random intercept	BMI	22.94 (0.12)	1.74 (0.02)	0.80 (0.02)	1.12 (0.10)
	Skinfold	46.98 (0.48)	6.41 (0.07)	70.74 (1.25)	58.82 (2.88)
	Waist	95.02 (0.33)	4.52 (0.04)	3.55 (0.15)	18.22 (0.97)
	PercentFat	26.52 (0.25)	3.15 (0.04)	14.58 (0.26)	19.96 (0.93)
random coefficient	BMI	23.01 (0.12)	1.21 (0.01)	0.44 (0.02)	0.93 (0.09)
	Skinfold	47.25 (0.49)	4.46 (0.06)	66.98 (1.18)	52.75 (2.77)
	Waist	95.20 (0.32)	3.05 (0.04)	5.51 (0.15)	17.88 (0.93)
	PercentFat	26.63 (0.25)	2.13 (0.03)	15.74 (0.28)	20.54 (0.94)

Table 7: *Correlations and % sample correlations explained by LVM*

Correlation between	Sample correlation	Model correlation		% of correlation explained by	
		RIM	RCM	RIM	RCM
BMI and Skinfold	0.89	0.48	0.56	53.50	62.27
BMI and WC	0.92	0.77	0.83	83.44	89.63
BMI and PBF	0.82	0.53	0.59	64.71	72.43
Skinfold and WC	0.81	0.65	0.71	79.84	87.39
Skinfold and PBF	0.81	0.45	0.51	55.29	63.07
WC and PBF	0.81	0.72	0.76	88.73	93.39

RIM: random-intercept model; RCM: random-coefficient model; WC: waist circumference; PBF: percent body fat

Figure 1: *Scatter Plots of Obesity Score against Houshold Income and Maturation Stages*

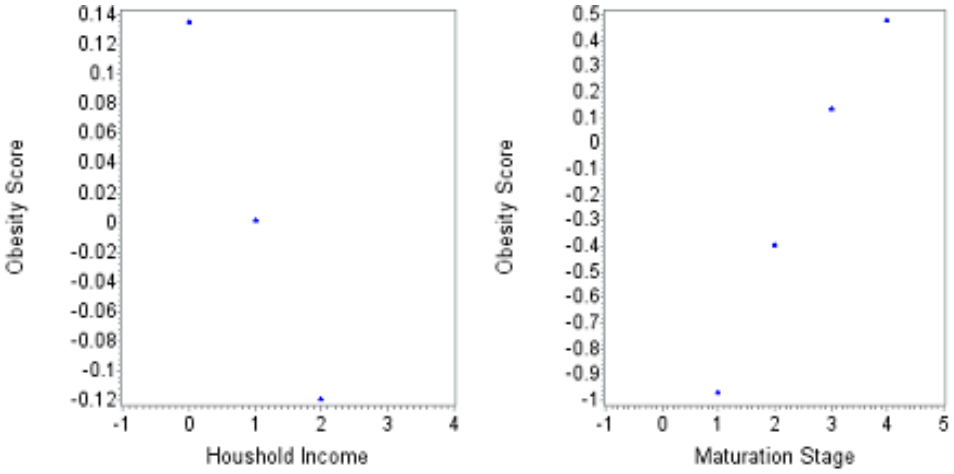

Figure 2: *Obesity growth curves for blacks and whites*