Web-based Supplementary Materials for
Joint Models for a Primary Endpoint and Multiple Longitudinal Covariate Processes
by Erning Li, Naisyin Wang, and Nae-Yuh Wang

Appendix A
Technical Details of Bias Calculations

There are general techniques for bias calculation (e.g., Wang *et al.*, 1998, *JASA* **93**, 249–261), one of which involves comparing the correspondence between the observed data model and the misspecified model when both models share the same structure. In this section, we concentrate on the bias calculation for the SS and on a simple but representative joint model that consists of a logistic primary model and a random intercept covariate process. The bias analysis strategy is applicable to general joint models and bias calculations for the CS; more complicated joint models require no additional procedures, but are potentially harder to compute and may lead to more complicated results.

For simplicity let $g = 1$, $m_i = m$, and X_i, α and β be scalars. Assume longitudinal data $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{im})^{\mathrm{T}}, i = 1, \ldots, n$, follow a random intercept model $\boldsymbol{W}_i = \mathbf{1}X_i + \boldsymbol{U}_i$, where $\mathbf{1}$ is an $m \times 1$ vector of 1's, $\boldsymbol{U}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{\Sigma}_i = \sigma^2\{(1-\rho)\boldsymbol{I} + \rho\boldsymbol{J}\}$ for $m \times m$ identity matrix $\boldsymbol{I}$ and $\boldsymbol{J} = \mathbf{1}\mathbf{1}^{\mathrm{T}}$. That is, $\boldsymbol{\Sigma}_i$ has a compound symmetry structure. Let the primary outcome $Y_i, i = 1, \ldots, n$, be binary and follow a logistic model $\mathrm{pr}(Y_i = 1|X_i; \boldsymbol{\theta}) = H(\alpha + \beta X_i)$.

For the assumed model, the estimating equation based on the GSS for the logistic primary model is

$$\sum_{i=1}^{n} \begin{pmatrix} \psi_{\alpha i}(\boldsymbol{\theta}) \\ \psi_{\beta_i}(\boldsymbol{\theta}) \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} Y_i - \mu_i \\ (Y_i - \mu_i)(\mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{1})^{-1}(\boldsymbol{S}_i - \boldsymbol{\beta}) \end{pmatrix} = \mathbf{0},$$

where $\mu_i = H\{\alpha + (S_i - \beta/2)(\mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{1})^{-1}\beta\}$, and $S_i = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_i + Y_i\beta$. Notice the fact that $\boldsymbol{\Sigma}_i^{-1} = \sigma^{-2}\{(1+\rho)\boldsymbol{I} + \rho\boldsymbol{J}\}^{-1} = \sigma^{-2}(1-\rho)^{-1}\{\boldsymbol{I} - \rho(1-\rho+m\rho)^{-1}\boldsymbol{J}\}$, then $(\mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{1})^{-1} =$

$\sigma^2(1-\rho)\{\mathbf{1}^{\mathrm{T}}\mathbf{1}-\rho(1-\rho+m\rho)^{-1}\mathbf{1}^{\mathrm{T}}\boldsymbol{J}\mathbf{1}\}^{-1} = m^{-1}\sigma^2(1-\rho+m\rho)$, and $S_i = \sigma^{-2}(1+\rho)^{-1}\{\mathbf{1}^{\mathrm{T}} - \rho(1+\rho-m\rho)^{-1}\mathbf{1}^{\mathrm{T}}\boldsymbol{J}\}\boldsymbol{W}_i + Y_i\beta = m\sigma^{-2}(1-\rho+m\rho)^{-1}\overline{W}_i + Y_i\beta$, where $\overline{W}_i = m^{-1}\sum_{j=1}^{m} W_{ij}$. Further, because the GSS is derived based on the true $\boldsymbol{\Sigma}_i$, the correct estimating equation based on the GSS for the logistic primary model is

$$\sum_{i=1}^{n}\begin{pmatrix}\psi_{\alpha_i}(\boldsymbol{\theta})\\ \psi_{\beta_i}(\boldsymbol{\theta})\end{pmatrix} = \sum_{i=1}^{n}\begin{pmatrix}Y_i-\mu_i\\ (Y_i-\mu_i)m^{-1}\sigma^2(1-\rho+m\rho)(S_i-\beta)\end{pmatrix} = \mathbf{0}, \tag{A.1}$$

where $\mu_i = H\{\alpha + m^{-1}\sigma^2(1-\rho+m\rho)S_i\beta - m^{-1}\sigma^2(1-\rho+m\rho)\beta^2/2\}$, and $S_i = m\sigma^{-2}(1-\rho+m\rho)^{-1}\overline{W}_i + Y_i\beta$.

Suppose the true $\boldsymbol{\Sigma}_i = \sigma^2\{(1-\rho)\boldsymbol{I}+\rho\boldsymbol{J}\}$ is incorrectly specified as $\boldsymbol{\Sigma}_{i_A} = \sigma^2\boldsymbol{I}$ (equivalently, correlation is ignored by setting $\rho = 0$). Let $\boldsymbol{\theta}_* = (\alpha_*, \beta_*)^{\mathrm{T}}$ denote the asymptotic limit of $\widehat{\boldsymbol{\theta}}_* = (\widehat{\alpha}_*, \widehat{\beta}_*)^{\mathrm{T}}$ which solves (A.1) when $\boldsymbol{\Sigma}_i$ is replaced by $\boldsymbol{\Sigma}_{i_A}$ (equivalently, $\widehat{\boldsymbol{\theta}}_*$ solves the estimating equation based on the SS). Because $(\mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_{i_A}^{-1}\mathbf{1})^{-1} = \sigma^2(\mathbf{1}^{\mathrm{T}}\mathbf{1})^{-1} = m^{-1}\sigma^2$, $S_{i_A} = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_{i_A}^{-1}\boldsymbol{W}_i + Y_i\beta_* = m\sigma^{-2}\overline{W}_i + Y_i\beta_*$, and also $S_{i_A} = S_i + m\sigma^{-2}\rho(m-1)(1-\rho+m\rho)^{-1}\overline{W}_i + Y_i(\beta_*-\beta)$, we obtain the misspecified estimating equation that $\widehat{\boldsymbol{\theta}}_*$ solves is

$$\sum_{i=1}^{n}\begin{pmatrix}\psi_{\alpha_i}(\boldsymbol{\theta}_*)\\ \psi_{\beta_i}(\boldsymbol{\theta}_*)\end{pmatrix} = \sum_{i=1}^{n}\begin{pmatrix}Y_i-\mu_{i_A}\\ (Y_i-\mu_{i_A})m^{-1}\sigma^2(S_{i_A}-\beta_*)\end{pmatrix} = \mathbf{0}, \tag{A.2}$$

where $\mu_{i_A} = H[\alpha_* + m^{-1}\sigma^2 S_i\beta_* + \{-m^{-1}\sigma^2 Y_i\beta + \rho(m-1)(1-\rho+m\rho)^{-1}\overline{W}_i\}\beta_* + m^{-1}\sigma^2(Y_i - 1/2)\beta_*^2]$, $S_{i_A} = m\sigma^{-2}\overline{W}_i + Y_i\beta_*$, and $\overline{W}_i = m^{-1}\sum_{j=1}^{m} W_{ij}$. Comparing $\psi_{\alpha_i}(\boldsymbol{\theta}_*)$ (or μ_{i_A}) in (A.2) with the true $\psi_{\alpha_i}(\boldsymbol{\theta})$ (or μ_i) in (A.1), we see that coefficients for both β and β^2 are incorrectly specified in $\psi_{\alpha_i}(\boldsymbol{\theta}_*)$ and $\psi_{\alpha_i}(\boldsymbol{\theta}_*)$ contains extra terms involving Y_i and $\overline{W}_i$. Similar phenomena hold in $\psi_{\beta_i}(\boldsymbol{\theta}_*)$.

Although no closed form expression corresponds $\boldsymbol{\theta}_*$ to $\boldsymbol{\theta}$ based on (A.1) and (A.2), we can numerically calculate $\boldsymbol{\theta}_*$ and the asymptotic bias of $\widehat{\boldsymbol{\theta}}_*$ when $n \to \infty$. Note that $n^{-1}\sum_{i=1}^{n}\psi_{\alpha_i}(\widehat{\boldsymbol{\theta}}_*) \to E\{\psi_{\alpha_i}(\boldsymbol{\theta}_*)\}$ and $n^{-1}\sum_{i=1}^{n}\psi_{\beta_i}(\widehat{\boldsymbol{\theta}}_*) \to E\{\psi_{\beta_i}(\boldsymbol{\theta}_*)\}$ in probability as $n \to \infty$. The random variables involved in $\psi_{\alpha_i}(\widehat{\boldsymbol{\theta}}_*)$ and $\psi_{\beta_i}(\widehat{\boldsymbol{\theta}}_*)$ are Y_i and $\overline{W}_i = X_i + \overline{U}_i$.

To compute $E\{\psi_{\alpha_i}(\boldsymbol{\theta}_*)\}$ and $E\{\psi_{\beta_i}(\boldsymbol{\theta}_*)\}$, suppressing the subscripts i we first express all terms in them as $E\{E(\cdot|X,\overline{U})\}$ where $\overline{U}=\overline{W}-X$ and then approximate them in a Monte Carlo manner by $B^{-1}\sum_{l=1}^{B}E(\cdot|X_l,\overline{U}_l)$ for some reasonably large number B, where X_l and $\overline{U}_l$, $l=1,\dots,B$, are samples drawn from their true distributions. From (A.2), we can show that

$$\begin{aligned} E\{\psi_{\alpha_i}(\boldsymbol{\theta}_*)\} &= E\left[E\{A_1(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*)-A_2(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*)\}|X,\overline{U}\right], \\ E\{\psi_{\beta_i}(\boldsymbol{\theta}_*)\} &= E\left[E\{A_3(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*)-A_4(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*)\}|X,\overline{U}\right], \end{aligned}$$

where

$$\begin{aligned} A_1(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*) &= Y, \\ A_2(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*) &= H\{\alpha_*+\overline{W}\beta_*+m^{-1}\sigma^2(Y-1/2)\beta_*^2\}, \\ A_3(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*) &= Y\overline{W}, \\ A_4(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*) &= A_2(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*)\{\overline{W}+m^{-1}\sigma^2(Y-1)\beta_*\}. \end{aligned}$$

To remove the randomness of Y, the binary outcome that follows the logistic model, for any function $A(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*)$, we have $E\{A(Y,\overline{W},\boldsymbol{\theta},\boldsymbol{\theta}_*)|X,\overline{U}\}=A(Y=1,X+\overline{U},\boldsymbol{\theta},\boldsymbol{\theta}_*)H(\alpha+\beta X)+A(Y=0,X+\overline{U},\boldsymbol{\theta},\boldsymbol{\theta}_*)\{1-H(\alpha+\beta X)\}$. The above strategy for computing $\boldsymbol{\theta}_*$ ensures a very stable convergence, even with a moderately large number of replications M.

An interesting special case is when the number of longitudinal observations m is large. A common conception is that the effect of measurement errors is eliminated when $m\to\infty$. We note that this is not the case here. When $m\to\infty$, $E\{\psi_{\alpha_i}(\boldsymbol{\theta}_*)|X,\overline{U}\}\to H(\alpha+\beta X)-H(\alpha_*+\overline{W}\beta_*)$ and $E\{\psi_{\beta_i}(\boldsymbol{\theta}_*)|X,\overline{U}\}\to\overline{W}\{H(\alpha+\beta X)-H(\alpha_*+\overline{W}\beta_*)\}$. However, from the correct equation (A.1), we can deduce that when both $n\to\infty$ and $m\to\infty$, $E\{\psi_{\alpha_i}(\boldsymbol{\theta})|X,\overline{U}\}\to H(\alpha+\beta X)\{1-H(\alpha+\overline{W}\beta+\rho\sigma^2\beta^2/2)+H(\alpha+\overline{W}\beta-\rho\sigma^2\beta^2/2)\}-H(\alpha+\overline{W}\beta-\rho\sigma^2\beta^2/2)$ and $E\{\psi_{\beta_i}(\boldsymbol{\theta})|X,\overline{U}\}\to\overline{W}[H(\alpha+\beta X)\{1-H(\alpha+\overline{W}\beta+\rho\sigma^2\beta^2/2)+H(\alpha+\overline{W}\beta-\rho\sigma^2\beta^2/2)\}-H(\alpha+\overline{W}\beta-\rho\sigma^2\beta^2/2)]+\rho\sigma^2\beta H(\alpha+\overline{W}\beta-\rho\sigma^2\beta^2/2)\{1-H(\alpha+\beta X)\}$. This indicates that even with a large cluster size m, $\boldsymbol{\theta}_*$ differs from $\boldsymbol{\theta}$ unless the true correlation $\rho=0$.

We performed a numerical illustration for this proposed computation of $\boldsymbol{\theta}_*$ for ρ varying between 0 and 0.8 and for m ranging from 2 to ∞. The parameter configurations are

$\boldsymbol{\theta} = (\alpha, \beta)^{\mathrm{T}} = (-2.5, 3)^{\mathrm{T}}$ and $\sigma^2 = 1$. Letting random intercepts $X_i \sim \mathcal{N}(0.5, 1)$, we generated $X_l, l = 1, \ldots, B$, from this normal distribution. Notice that $\mathrm{var}(\overline{U}) = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{1}/m^2 = \sigma^2\{\rho + m^{-1}(1-\rho)\}$, thus $\overline{U}_l, l = 1, \ldots, B$, are generated from $\mathcal{N}(0, \sigma^2\{\rho + m^{-1}(1-\rho)\})$ when m is finite and from $\mathcal{N}(0, \sigma^2\rho)$ when $m \to \infty$. We chose a fairly large $B = 200,000$.

Appendix B

Details Regarding Inference and Implementation

First Order Derivatives of Score Functions GSS and GCS for the Logistic Primary Model

Denote the GSS for the logistic primary model as $\boldsymbol{\psi}_S(Y_i, \boldsymbol{W}_i, \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\psi}_{S,\boldsymbol{\alpha}} \\ \boldsymbol{\psi}_{S,\boldsymbol{\beta}} \end{pmatrix}$. With $\boldsymbol{S}_i = \boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_i + Y_i\boldsymbol{\beta}$ fixed, the first order partial derivatives of the GSS used in the Newton-Raphson updating scheme are

$$
\begin{aligned}
\frac{\partial \boldsymbol{\psi}_{S,\boldsymbol{\alpha}}}{\partial \boldsymbol{\alpha}^{\mathrm{T}}}|\boldsymbol{S}_i &= \mu_i(\mu_i - 1)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}} \\
\frac{\partial \boldsymbol{\psi}_{S,\boldsymbol{\alpha}}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}|\boldsymbol{S}_i &= \mu_i(\mu_i - 1)\boldsymbol{Z}_i(\boldsymbol{S}_i - \boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1} = \left(\frac{\partial \boldsymbol{\psi}_{S,\boldsymbol{\beta}}}{\partial \boldsymbol{\alpha}^{\mathrm{T}}}|\boldsymbol{S}_i\right)^{\mathrm{T}} \\
\frac{\partial \boldsymbol{\psi}_{S,\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}|\boldsymbol{S}_i &= \mu_i(\mu_i - 1)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}(\boldsymbol{S}_i - \boldsymbol{\beta})(\boldsymbol{S}_i - \boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1} - (Y_i - \mu_i)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}
\end{aligned}
$$

With $\boldsymbol{S}_i$ replaced by $\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_i + Y_i\boldsymbol{\beta}$, the first order partial derivatives of the GSS used in the empirical sandwich estimator are

$$
\begin{aligned}
\frac{\partial \boldsymbol{\psi}_{S,\boldsymbol{\alpha}}}{\partial \boldsymbol{\alpha}^{\mathrm{T}}} &= \mu_i(\mu_i - 1)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}} \\
\frac{\partial \boldsymbol{\psi}_{S,\boldsymbol{\alpha}}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} &= \mu_i(\mu_i - 1)\boldsymbol{Z}_i\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{W}_i + (2Y_i - 1)\boldsymbol{\beta}\}^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1} \\
\frac{\partial \boldsymbol{\psi}_{S,\boldsymbol{\beta}}}{\partial \boldsymbol{\alpha}^{\mathrm{T}}} &= \mu_i(\mu_i - 1)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{W}_i + (Y_i - 1)\boldsymbol{\beta}\}\boldsymbol{Z}_i^{\mathrm{T}} \\
\frac{\partial \boldsymbol{\psi}_{S,\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} &= \mu_i(\mu_i - 1)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{W}_i + (Y_i - 1)\boldsymbol{\beta}\}\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{W}_i \\
&\quad + (2Y_i - 1)\boldsymbol{\beta}\}^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1} + (Y_i - \mu_i)(Y_i - 1)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}
\end{aligned}
$$

Similarly, denote the GCS for the logistic primary model as $\boldsymbol{\psi}_C(Y_i, \boldsymbol{W}_i, \boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\psi}_{C,\boldsymbol{\alpha}} \\ \boldsymbol{\psi}_{C,\boldsymbol{\beta}} \end{pmatrix}$.

With $\boldsymbol{S}_i$ fixed, the first order partial derivatives of the GCS used in the Newton-Raphson updating scheme are

$$\begin{aligned}
\frac{\partial \boldsymbol{\psi}_{C,\boldsymbol{\alpha}}}{\partial \boldsymbol{\alpha}^{\mathrm{T}}}|\boldsymbol{S}_i &= \mu_i(\mu_i-1)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}} \\
\frac{\partial \boldsymbol{\psi}_{C,\boldsymbol{\alpha}}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}|\boldsymbol{S}_i &= \mu_i(\mu_i-1)\boldsymbol{Z}_i(\boldsymbol{S}_i-\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1} \\
\frac{\partial \boldsymbol{\psi}_{C,\boldsymbol{\beta}}}{\partial \boldsymbol{\alpha}^{\mathrm{T}}}|\boldsymbol{S}_i &= \mu_i(\mu_i-1)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}\{\boldsymbol{S}_i+(Y_i-2\mu_i)\boldsymbol{\beta}\}\boldsymbol{Z}_i^{\mathrm{T}} \\
\frac{\partial \boldsymbol{\psi}_{C,\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}|\boldsymbol{S}_i &= \mu_i(\mu_i-1)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}\{\boldsymbol{S}_i+(Y_i-2\mu_i)\boldsymbol{\beta}\}(\boldsymbol{S}_i-\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1} \\
&\quad -\mu_i(Y_i-\mu_i)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}
\end{aligned}$$

With $\boldsymbol{S}_i$ replaced by $\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_i+Y_i\boldsymbol{\beta}$, the first order partial derivatives of the GCS used in the empirical sandwich estimator are

$$\begin{aligned}
\frac{\partial \boldsymbol{\psi}_{C,\boldsymbol{\alpha}}}{\partial \boldsymbol{\alpha}^{\mathrm{T}}} &= \mu_i(\mu_i-1)\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}} \\
\frac{\partial \boldsymbol{\psi}_{C,\boldsymbol{\alpha}}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} &= \mu_i(\mu_i-1)\boldsymbol{Z}_i\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{W}_i+(2Y_i-1)\boldsymbol{\beta}\}^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1} \\
\frac{\partial \boldsymbol{\psi}_{C,\boldsymbol{\beta}}}{\partial \boldsymbol{\alpha}^{\mathrm{T}}} &= \mu_i(\mu_i-1)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{W}_i+2(Y_i-\mu_i)\boldsymbol{\beta}\}\boldsymbol{Z}_i^{\mathrm{T}} \\
\frac{\partial \boldsymbol{\psi}_{C,\boldsymbol{\beta}}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} &= \mu_i(\mu_i-1)(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{W}_i+2(Y_i-\mu_i)\boldsymbol{\beta}\}\{\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{W}_i \\
&\quad +(2Y_i-1)\boldsymbol{\beta}\}^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}+(Y_i-\mu_i)^2(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{D}_i)^{-1}
\end{aligned}$$

A Summary of Implementation Procedure

In sum, the estimation of $\boldsymbol{\theta}$ can be carried out in the following steps. First, combine the individual longitudinal covariate processes (1) into the multivariate random effects formulation (2). Second, obtain $\widehat{\boldsymbol{\Sigma}}_i$ as in Section 5,1. Last, with $\boldsymbol{\Sigma}_i$ replaced by $\widehat{\boldsymbol{\Sigma}}_i$, either directly perform the GSS and the GCS on (7) to (10) using the Newton-Raphson algorithm, or adopt the re-parameterization procedure described in Section 5.2; both ways provide virtually identical results. Alternatively, we can add a set of additional unbiased estimating equations corresponding to the parameters in $\boldsymbol{\Sigma}_i$ to the GSS and GCS estimating equations at (7) to (10) and then use the estimated parameters obtained in Section 5,1 as initial values and solve the whole set of estimating equations using Newton-Raphson algorithm. We have found this latter procedure more complicated with mostly incremental efficiency gains.

Appendix C

Numerical Evaluation of the Effects of Violating the Independent Between-Process Measurement Error Assumption

To numerically investigate the impact of the conditional independence assumption between measurement errors from different longitudinal covariate processes on the inferences, we considered a simulation setup with two longitudinal processes, each of which had a compound symmetry correlation structure for within-subject measurement errors. We also let the measurement errors from the two longitudinal processes to be correlated when the measurements were taken at the same time.

In particular, we generated one set of longitudinal data from the random intercept-slope model $W_{ij}^{(1)} = X_{1i} + X_{2i}t_{ij} + U_{ij}^{(1)}$, and another set of longitudinal data from the random intercept model $W_{ij}^{(2)} = X_{3i} + U_{ij}^{(2)}$, where $t_{ij} = j-1$, $j = 1,\ldots,m, (m = 10)$; $i = 1,\ldots,n, (n = 500)$. The covariance of the measurement errors from the first longitudinal process, $\boldsymbol{U}_i^{(1)}$, is $\boldsymbol{\Omega}_i^{(1)} = \sigma^2\{(1-\rho_1)\boldsymbol{I}_m + \rho_1\boldsymbol{J}_m\}$ and the covariance of the measurement errors from the second longitudinal process, $\boldsymbol{U}_i^{(2)}$, is $\boldsymbol{\Omega}_i^{(2)} = \sigma^2\{(1-\rho_2)\boldsymbol{I}_m + \rho_2\boldsymbol{J}_m\}$, where $\sigma = 0.5$ and $\rho_1 = \rho_2 = 0.25$. That is, the true $\boldsymbol{\Omega}_i^{(1)}$ and $\boldsymbol{\Omega}_i^{(2)}$ both have compound symmetry structures. Further, the correlation between measurement errors from the two longitudinal processes taken at the same time is equal to $\rho_{12} = \text{corr}(U_{ij}^{(1)}, U_{ij}^{(2)}) = 0.5$. This implies that the covariance for the overall measurement errors $\boldsymbol{U}_i = \left(\boldsymbol{U}_i^{(1)\mathrm{T}}, \boldsymbol{U}_i^{(2)\mathrm{T}}\right)^{\mathrm{T}}$ is $\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Omega}_i^{(1)} & \boldsymbol{\Omega}_i^{(1,2)} \\ \boldsymbol{\Omega}_i^{(1,2)} & \boldsymbol{\Omega}_i^{(2)} \end{pmatrix}$, where $\boldsymbol{\Omega}_i^{(1,2)} = \sigma^2\rho_{12}\boldsymbol{I}_m$. Similar to the simulations in Sections 2 and 6, we designed four scenarios of the true $(X_{1i}, X_{2i})^{\mathrm{T}}$ distribution: (1) a bivariate normal; (2) a bimodal mixture of normals with mixing proportion 30-70; (3) a bivariate skew-normal with coefficients of skewness -0.10 and 0.85 for X_{1i} and X_{2i}, respectively; and (4) a bivariate t_5 distribution. For all $(X_{1i}, X_{2i})^{\mathrm{T}}$ distribution scenarios,

$E(X_{1i}) = E(X_{2i}) = 0.5$, $\mathrm{var}(X_{1i}) = 1.0$, $\mathrm{var}(X_{2i}) = 0.64$, and $\mathrm{cov}(X_{1i}, X_{2i}) = -0.2$. X_{3i} was generated from $\mathcal{N}(0.5, 0.6)$. The primary binary observations Y_i were generated from logistic model $\mathrm{pr}(Y_i = 1|\boldsymbol{X}_i) = [1 + \exp\{-(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})\}]^{-1}$ with $\alpha = -3$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}} = (3, 2, 1)^{\mathrm{T}}$. Five hundred data sets were generated for each $(X_{1i}, X_{2i})^{\mathrm{T}}$ distribution.

In the estimation procedures, we assume that measurement errors from two different processes are independent, i.e., ignore the existence of ρ_{12}. Like Tables 1 and 2 of Section 6, we report values of RB (%, estimated relative bias in percentage), SD (Monte Carlo standard deviation), SE (average of estimated standard errors) and CP (Monte Carlo coverage probability of 95% Wald confidence interval) under the four random effects distributional scenarios, for CS, SS, GSS and GCS in Table S.3. The CS and the SS of Li *et al.* (2004) require IID assumption on measurement errors and thus they ignore all correlations (ρ_1, ρ_2, and ρ_{12}). Here, the proposed GSS and GCS under the conditional independence assumption only ignore the correlation across processes (ρ_{12}). They accounted for the within process correlations (ρ_1 and ρ_2) and were obtained via reparameterization of data using $\boldsymbol{\Omega}_i^{(1)}$ and $\boldsymbol{\Omega}_i^{(2)}$.

The results in Table S.3 show that in all cases, the GSS and the GCS with the conditional independence assumption provide promising inferences in terms of negligible bias and satisfactory coverage probabilities, although the measurement errors from different longitudinal processes are in fact correlated. Similar to the results in Tables 1 and 2 of Section 6 and Tables S.1 and S.2, the performances of the SS and the CS are poor when the IID assumption on measurement errors is violated. In general, they substantially underestimate the parameters and their coverage probabilities fall below nominal level. We have also adopted the re-parameterization procedure described in Section 5.2 using the true $\boldsymbol{\Sigma}_i$ and obtained similar but no better numerical performances than those when ρ_{12} was assumed to be 0. However, this could be due to the fact that the parameters in $\boldsymbol{\Sigma}_i$ were not estimated. In conclusion, the numerical study seems suggest that the proposed method is insensitive toward the independent between-process measurement error assumption.

Table S.1

Simulation results for the joint model with $\rho = -0.25$ under four underlying $\boldsymbol{X}_i$ distributions. In the logistic model, true $\alpha = -2.5$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}} = (3.0, 2.0)^{\mathrm{T}}$. Reported values are RB, estimated relative bias (%); SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, Monte Carlo coverage probability of 95% Wald confidence interval.

$\rho = -0.25$	Method	RB (%)	SD	SE	CP	RB (%)	SD	SE	CP
		$\boldsymbol{X}_i$ Normal				$\boldsymbol{X}_i$ Bimodal mixture			
$\widehat{\alpha}$	RRC	19.2	0.46	0.46	0.96	48.9	0.83	0.78	0.84
	GRRC	1.7	0.32	0.32	0.96	14.9	0.45	0.44	0.94
	SS	19.8	0.48	0.48	0.97	15.9	0.53	0.50	0.97
	GSS	2.1	0.32	0.32	0.96	3.0	0.40	0.38	0.96
	CS	19.7	0.48	0.49	0.96	17.1	0.56	0.52	0.97
	GCS	2.0	0.32	0.32	0.96	3.0	0.40	0.39	0.96
$\widehat{\beta}_1$	RRC	21.8	0.54	0.54	0.94	52.9	0.94	0.86	0.61
	GRRC	1.5	0.34	0.35	0.96	19.1	0.47	0.45	0.86
	SS	22.6	0.58	0.58	0.97	10.8	0.43	0.39	0.96
	GSS	2.0	0.34	0.35	0.96	2.5	0.33	0.31	0.95
	CS	22.4	0.58	0.58	0.96	11.7	0.45	0.41	0.98
	GCS	2.0	0.35	0.35	0.96	2.5	0.33	0.31	0.95
$\widehat{\beta}_2$	RRC	15.1	0.36	0.36	0.96	34.4	0.57	0.54	0.93
	GRRC	1.7	0.27	0.27	0.96	11.2	0.35	0.34	0.94
	SS	15.4	0.37	0.37	0.97	10.2	0.39	0.36	0.96
	GSS	2.0	0.27	0.27	0.96	2.7	0.32	0.30	0.95
	CS	15.2	0.36	0.38	0.96	11.1	0.40	0.37	0.96
	GCS	2.0	0.27	0.27	0.96	2.7	0.32	0.30	0.95
		$\boldsymbol{X}_i$ Skew-normal				$\boldsymbol{X}_i$ Bivariate t_5			
$\widehat{\alpha}$	RRC	18.9	0.46	0.46	0.96	11.1	0.41	0.41	0.97
	GRRC	1.3	0.32	0.32	0.95	−3.7	0.30	0.30	0.92
	SS	19.1	0.48	0.49	0.97	25.6	0.57	0.58	0.97
	GSS	1.5	0.32	0.32	0.95	1.6	0.32	0.33	0.96
	CS	18.2	0.50	0.47	0.96	25.0	0.54	0.57	0.94
	GCS	1.5	0.32	0.32	0.95	1.6	0.32	0.33	0.96
$\widehat{\beta}_1$	RRC	21.4	0.54	0.54	0.94	12.8	0.47	0.48	0.98
	GRRC	1.1	0.34	0.35	0.95	−4.5	0.31	0.33	0.91
	SS	21.6	0.56	0.58	0.97	29.4	0.70	0.71	0.97
	GSS	1.3	0.34	0.35	0.96	1.5	0.35	0.37	0.96
	CS	20.6	0.62	0.55	0.96	28.8	0.67	0.71	0.94
	GCS	1.2	0.33	0.35	0.96	1.5	0.35	0.37	0.96
$\widehat{\beta}_2$	RRC	14.9	0.36	0.36	0.96	8.4	0.34	0.34	0.97
	GRRC	1.4	0.27	0.27	0.96	−2.7	0.27	0.26	0.93
	SS	15.0	0.37	0.38	0.97	19.8	0.45	0.44	0.96
	GSS	1.6	0.27	0.27	0.96	1.5	0.28	0.28	0.96
	CS	14.4	0.37	0.36	0.96	19.3	0.43	0.44	0.95
	GCS	1.6	0.27	0.27	0.96	1.5	0.28	0.28	0.96

Table S.2

Simulation results for the joint model with $\rho = -0.50$ under four underlying $\boldsymbol{X}_i$ distributions. The rest of the setup is identical to that of Table S.1.

$\rho = -0.50$	Method	RB (%)	SD	SE	CP	RB (%)	SD	SE	CP
		$\boldsymbol{X}_i$ Normal				$\boldsymbol{X}_i$ Bimodal mixture			
$\widehat{\alpha}$	RRC	37.6	0.62	0.62	0.88	74.2	1.14	1.11	0.78
	GRRC	0.8	0.29	0.30	0.97	8.5	0.38	0.38	0.96
	SS	35.9	0.63	0.66	0.91	25.0	0.58	0.57	0.95
	GSS	0.9	0.29	0.30	0.96	1.9	0.35	0.36	0.96
	CS	33.4	0.53	0.62	0.87	27.7	0.62	0.63	0.96
	GCS	0.9	0.29	0.30	0.96	1.9	0.35	0.36	0.96
$\widehat{\beta}_1$	RRC	42.9	0.77	0.76	0.80	76.0	1.30	1.23	0.42
	GRRC	0.6	0.31	0.32	0.96	11.2	0.36	0.36	0.92
	SS	41.4	0.80	0.82	0.89	16.1	0.44	0.43	0.96
	GSS	0.7	0.31	0.32	0.96	1.7	0.29	0.29	0.96
	CS	38.6	0.64	0.76	0.82	18.1	0.48	0.48	0.96
	GCS	0.7	0.30	0.32	0.96	1.7	0.29	0.29	0.96
$\widehat{\beta}_2$	RRC	29.4	0.47	0.47	0.93	51.1	0.76	0.74	0.93
	GRRC	0.9	0.25	0.25	0.96	6.5	0.31	0.30	0.95
	SS	27.7	0.48	0.49	0.94	15.2	0.40	0.40	0.95
	GSS	1.0	0.25	0.25	0.96	1.9	0.29	0.29	0.95
	CS	25.6	0.41	0.46	0.92	17.1	0.42	0.43	0.95
	GCS	1.0	0.25	0.25	0.96	1.9	0.29	0.29	0.95
		$\boldsymbol{X}_i$ Skew-normal				$\boldsymbol{X}_i$ Bivariate t_5			
$\widehat{\alpha}$	RRC	37.5	0.65	0.62	0.86	28.2	0.54	0.55	0.93
	GRRC	0.5	0.30	0.30	0.95	−3.1	0.28	0.29	0.93
	SS	35.4	0.64	0.65	0.91	48.0	0.76	0.84	0.89
	GSS	0.5	0.29	0.30	0.95	−0.1	0.29	0.30	0.96
	CS	32.7	0.53	0.59	0.87	41.8	0.63	0.75	0.79
	GCS	0.5	0.29	0.30	0.95	−0.1	0.29	0.30	0.95
$\widehat{\beta}_1$	RRC	42.9	0.80	0.76	0.79	32.5	0.64	0.67	0.91
	GRRC	0.3	0.30	0.32	0.95	−3.8	0.29	0.31	0.92
	SS	40.6	0.78	0.80	0.90	55.7	0.98	1.10	0.83
	GSS	0.2	0.30	0.32	0.96	−0.3	0.30	0.33	0.96
	CS	37.7	0.62	0.72	0.83	48.5	0.82	0.95	0.71
	GCS	0.2	0.29	0.31	0.96	−0.4	0.30	0.33	0.96
$\widehat{\beta}_2$	RRC	29.2	0.48	0.47	0.93	21.6	0.42	0.42	0.95
	GRRC	0.7	0.25	0.25	0.96	−2.2	0.25	0.26	0.94
	SS	27.0	0.46	0.48	0.95	36.4	0.56	0.61	0.93
	GSS	0.7	0.25	0.25	0.96	0.1	0.26	0.26	0.95
	CS	25.0	0.40	0.44	0.91	31.6	0.47	0.55	0.88
	GCS	0.7	0.25	0.25	0.96	0.1	0.26	0.26	0.95

Table S.3

Simulation results for the joint model with the correlation between measurement errors from two longitudinal processes taken at the same time to be $\rho_{12} = 0.5$ under four underlying $(X_{1i}, X_{2i})^{\mathrm{T}}$ distributions. Each longitudinal process has compound symmetry structure for within-subject measurement errors. In the logistic model, true $\alpha = -3$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}} = (3, 2, 1)^{\mathrm{T}}$. Reported values are RB, estimated relative bias (%); SD, Monte Carlo standard deviation; SE, average of estimated standard errors; CP, Monte Carlo coverage probability of 95% Wald confidence interval. The GSS and the GCS are obtained under the conditional independence assumption.

	Method	RB (%)	SD	SE	CP	RB (%)	SD	SE	CP
		$(X_{1i}, X_{2i})^{\mathrm{T}}$ Normal				$(X_{1i}, X_{2i})^{\mathrm{T}}$ Bimodal mixture			
$\widehat{\alpha}$	SS	−14.9	0.31	0.29	0.60	−13.5	0.36	0.36	0.74
	GSS	1.8	0.47	0.43	0.95	2.4	0.51	0.49	0.96
	CS	−14.8	0.31	0.29	0.60	−13.6	0.36	0.35	0.74
	GCS	1.9	0.47	0.42	0.95	2.5	0.52	0.49	0.96
$\widehat{\beta}_1$	SS	−14.6	0.28	0.26	0.55	−7.3	0.27	0.27	0.84
	GSS	2.9	0.44	0.40	0.95	3.4	0.37	0.36	0.97
	CS	−14.6	0.28	0.26	0.55	−7.3	0.27	0.27	0.84
	GCS	3.0	0.44	0.40	0.96	3.5	0.38	0.36	0.97
$\widehat{\beta}_2$	SS	−11.2	0.24	0.23	0.78	−7.2	0.28	0.26	0.88
	GSS	2.9	0.33	0.30	0.95	4.1	0.36	0.33	0.96
	CS	−11.2	0.24	0.23	0.78	−7.3	0.28	0.26	0.88
	GCS	3.0	0.32	0.30	0.94	4.2	0.36	0.33	0.95
$\widehat{\beta}_3$	SS	−22.3	0.19	0.18	0.71	−23.3	0.22	0.21	0.77
	GSS	−3.0	0.27	0.25	0.93	−2.9	0.30	0.28	0.94
	CS	−22.3	0.19	0.18	0.71	−23.4	0.22	0.21	0.77
	GCS	−3.0	0.26	0.25	0.94	−2.7	0.30	0.28	0.93
		$(X_{1i}, X_{2i})^{\mathrm{T}}$ Skew-normal				$(X_{1i}, X_{2i})^{\mathrm{T}}$ Bivariate t_5			
$\widehat{\alpha}$	SS	−15.1	0.28	0.29	0.63	−17.0	0.29	0.29	0.55
	GSS	1.4	0.41	0.42	0.96	2.3	0.45	0.44	0.96
	CS	−15.0	0.28	0.29	0.63	−17.0	0.29	0.29	0.55
	GCS	1.5	0.41	0.42	0.96	2.3	0.45	0.44	0.96
$\widehat{\beta}_1$	SS	−15.0	0.27	0.26	0.55	−17.8	0.26	0.26	0.43
	GSS	2.3	0.41	0.40	0.95	3.1	0.45	0.43	0.95
	CS	−15.0	0.27	0.26	0.54	−17.7	0.26	0.26	0.44
	GCS	2.3	0.41	0.40	0.95	3.1	0.44	0.43	0.95
$\widehat{\beta}_2$	SS	−11.0	0.22	0.23	0.77	−13.6	0.23	0.23	0.73
	GSS	2.8	0.29	0.30	0.98	3.2	0.33	0.32	0.96
	CS	−11.0	0.22	0.23	0.77	−13.5	0.23	0.23	0.73
	GCS	2.9	0.30	0.30	0.98	3.2	0.33	0.32	0.96
$\widehat{\beta}_3$	SS	−21.6	0.17	0.18	0.77	−21.8	0.18	0.17	0.71
	GSS	−2.2	0.23	0.24	0.95	−2.7	0.24	0.23	0.95
	CS	−21.6	0.17	0.18	0.77	−21.7	0.18	0.17	0.72
	GCS	−2.1	0.23	0.24	0.95	−2.7	0.24	0.23	0.94

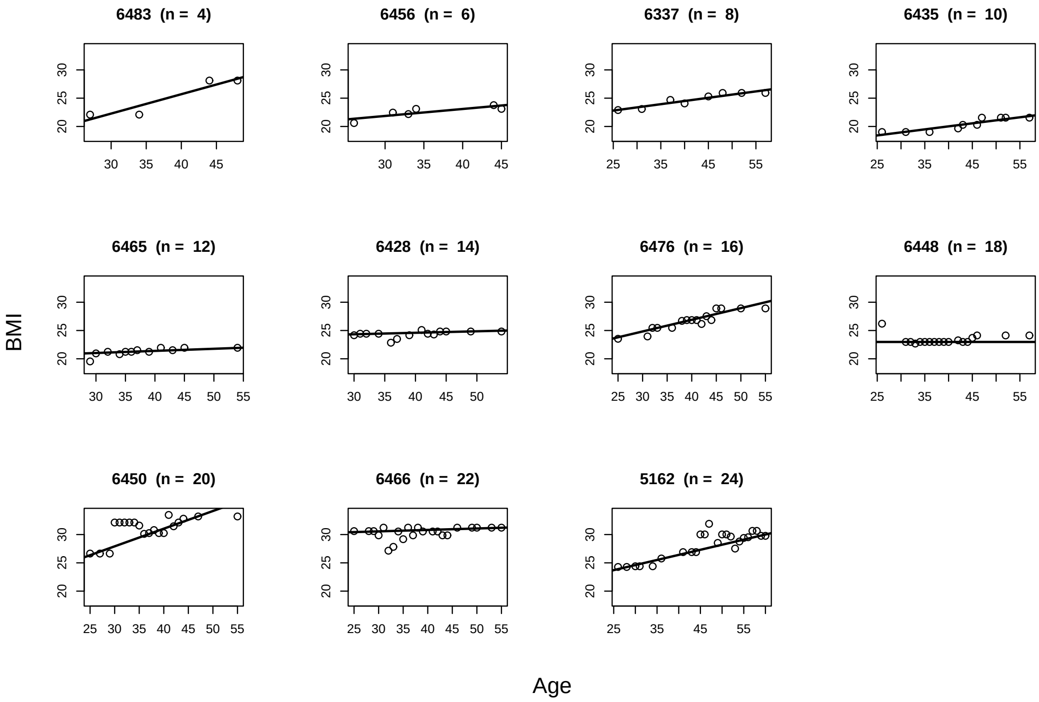

Figure S.1. Scatter Plots of Individual Longitudinal BMI versus Time. The main text of each plot contains the subject's ID and the number of observations from that subject. The fitted regression line is superimposed. The last individual (by order of ID number) of each group in which the subjects share the same number of BMI observations were selected. The number of observations per group in this figure ranges from 4 to 24.