Statistics
in Medicine

Research Article

Statistics
in Medicine

Estimators for longitudinal latent exposure models: examining measurement model assumptions

Brisa N. Sánchez[*,1], **Sehee Kim**[1], **Mary D. Sammel**[2]

Latent variable (LV) models are increasingly being used in environmental epidemiology as a way to summarize multiple environmental exposures and thus minimize statistical concerns that arise in multiple regression. LV models may be especially useful when multivariate exposures are collected repeatedly over time. LV models can accommodate a variety of assumptions, but at the same time present the user with many choices for model specification particularly in the case of exposure data collected repeatedly over time. For instance, the user could assume conditional independence of observed exposure biomarkers given the latent exposure, and, in the case of longitudinal latent exposure variables, time invariance of the measurement model. Choosing which assumptions to relax is not always straightforward. We were motivated by a study of prenatal lead exposure and mental development, where assumptions of the measurement model for the time-changing longitudinal exposure have appreciable impact on (maximum-likelihood) inferences about the health effects of lead exposure. Although we were not particularly interested in characterizing the change of the latent variable itself, imposing a longitudinal latent variable structure on the repeated multivariate exposure measures could result in high efficiency gains for the exposure-disease association. We examine the biases of maximum likelihood estimators when assumptions about the measurement model for the longitudinal latent exposure variable are violated. We adapt existing instrumental variable (IV) estimators to the case of longitudinal exposures and propose them as an alternative to estimate the health effects of a time-changing latent predictor. We show that IV estimators remain unbiased for a wide range of data generating models and have advantages in terms of mean squared error.

Keywords: estimating equations; measurement model invariance; instrumental variables

1. Introduction

Structural equation models (SEM) with latent variables (LV) [1, 2, 3] are increasingly being used in environmental epidemiology as a way to summarize multiple environmental exposure measures and reduce the dimensionality of predictor and outcome variables. For example, LV models have been employed to elucidate relationships between exposure to methyl mercury and development [4], air pollution and cardiovascular disease [5], lead exposure and physical growth [6], among others. In these applications, a measurement model is used to relate the observed exposures to a hypothesized

[1] *Department of Biostatistics, University of Michigan, Ann Arbor, MI USA 48109*
[2] *Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, USA*
[*] *Correspondence to: brisa@umich.edu, Department of Biostatistics, University of Michigan, Ann Arbor, MI USA 48109*

Prepared using simauth.cls *[Version: 2010/03/10 v3.00]*

underlying latent exposure, and then the association between the latent exposure and observed (or latent) outcomes is described in a (structural) model for the outcome. The parameters in the structural model for the health outcome are often of primary interest, since they encode the environmental health effects. A key advantage of SEMs with LV in these applications is that, compared to typical regression analysis, a much smaller set of tests is conducted due to the dimension reduction that occurs in the exposure measurement model.

There is also increased interest in epidemiological studies in examining how exposures measured repeatedly over time are related to a subsequent health outcome, e.g., [7, 8]. A feature of these studies is that while the exposure measures are longitudinal, the health outcome is often univariate and not measured at the same time as the exposures. An increasing number of studies are also now collecting multiple exposure biomarkers over time, e.g., [9, 10], rather than a multiple exposures at a single time point or a single exposure at multiple time points. Given the even larger number of exposure biomarkers available in these studies with multivariate exposure measures taken repeatedly, it is important to examine the applicability of latent variable models. For instance, latent variable models can be used not only to reduce dimensionality of exposure biomarkers taken at a single point in time and thus reduce the number of tests conducted, but also to potentially exploit the correlations among biomarkers measured over time and thus gain additional efficiency in estimating exposure-outcome associations.

We were motivated by a study of prenatal lead exposure and mental development, conducted as part of the Early Life Exposures in MExico City to Neuro-Toxicants (ELEMENT) project [11, 12]. The study collected several biomarkers of lead exposure during prenatal visits occurring at each trimester of pregnancy to examine the effects of lead exposure on child development later in life. At any one visit, a latent variable model can be developed to summarize exposure information captured by the biomarkers. Because the biomarkers are measured over time, a model that posits a latent variable whose value possibly changes with time can be proposed to summarize the exposure data. Exposure summaries can then be related to the distal health outcome, in this case child's mental development at 24 months. Figure 1 shows path diagrams [1] describing two potential models for these data. The diagrams represent two similar models for a time varying latent exposure, but differ in terms of the assumptions made regarding the measurement model, i.e., how observed exposures relate to latent exposures at each trimester.

Although the use of LVs greatly reduces statistical concerns such as multiple testing and collinearity, correctly specified covariance structures, in addition to mean structures (e.g., linearity), are needed for correct estimation and inference of

mean parameters. The covariance structure of the observed data is in part determined by the measurement model. For example, a classic assumption in the LV model literature is that the items are assumed to be conditionally independent given the latent variable [1]. In the lead example, lack of conditional independence can arise from lead concentrations being measured in the same media or by the same lab, or due to serial correlation among repeated measures of the same item over time. While these correlations can be estimated, a model where all correlations are unconstrained is not identifiable, and it is not always straightforward to choose which correlations to estimate or constrain to zero as various models sometimes may have equivalent fit [13, 14]. Nevertheless, model fit statistics and specification tests, e.g., [15], can used to help select the covariance structure or identify non-zero correlations. However, re-specifying the model multiple times can increase Type I error [16]. In the lead example, one might choose correlations to estimate based on what is known about serial and within lab correlations, but it is possible that other non-zero correlations exist.

Another important consideration is the stability, or invariance, of the measurement model over time, which also plays a key role in defining the model-implied marginal covariance structure of the observed items. Violations of measurement invariance could arise if the measurement properties of an observed exposure, also called item, change over time. For instance, in the lead example, the correlation between lead measured in plasma and other media changes as the concentration of lead in plasma (but not red cells) becomes diluted due to plasma volume expansion during pregnancy [17]. Recommendations to identify and appropriately handle violations of measurement invariance, such as starting with the most flexible model and then testing if constraints can be imposed, have been described [18, 19, 20], but the issues of model selection and inflated Type I error remain[13, 14, 16]. While the potential use of longitudinal latent exposure models is very appealing, the consequences of violating assumptions of time invariance of the longitudinal measurement model have not been systematically studied.

Several estimators have been proposed for LV models, some of which may be more robust to misspecification errors of the measurement model. When observed variables are continuous, most estimators are derived by minimizing a measure of distance between the model-implied and observed/empirical covariance matrix, regardless of whether or not distributional assumptions are made [1]. These estimators may propagate bias due to model misspecification in one part of the model to correctly specified portions of the model [13, 21]. Among this set of estimators, maximum likelihood (ML) is by far the most frequently used approach due to its wide availability (e.g., [22, 23]). In contrast, instrumental variable (IV) estimators [24], also called two stage least square estimators (2SLS), have been shown to contain bias induced due to misspecification

in one part of the model from contaminating other parts of the model. These IV estimators [24] have been adapted for the case of longitudinal latent outcomes in order to impose the assumption of time invariance of the measurement model [25, 26]. As pointed out in [15], other IV estimators [27, 28] developed prior to [24] either use ML to obtain estimates for some of the model parameters or do not admit any correlated errors. These other IV estimators will unlikely be robust in longitudinal exposure models since correlated errors will likely be present, and are thus not considered further.

A series of robustness studies in the SEM literature have been conducted (see [29] and background section in [24] for excellent reviews). In particular, Reddy evaluated ML estimators under misspecification of residual error structure in the context of a model with three latent variables, but did not include any IV estimators [21]. Sammel and Ryan [30] investigated how violations of the error covariance structure among a set of observed outcomes impacted tests of the effect of an observed predictor on those outcomes. Sánchez et al. [31] investigated violations of the impact of misspecifying the error structure of longitudinal outcomes when the predictor of interest was a latent variable. Bollen [32] described specific conditions under which his [24] IV estimator should yield robust estimates for some equations of the SEM, even when misspecifications were present in other components of the model. Bollen et al. [33] carried out simulation studies comparing IV and ML estimators when items were incorrectly assumed to not load on particular LVs. They also varied the sample size and the number of instruments used in the IV estimator. Nestler [26] examined the degree of bias introduced in ML and IV estimates of the coefficient of an observed predictor when the time pattern of a longitudinal latent outcome was misspecified. However, specific examination of the bias due to incorrectly assuming time invariance have not been conducted.

In the present article we conduct simulation studies within the setting of longitudinal latent exposures to examine the impact of violating assumptions of the exposure measurement model over time on MLE and IV estimators. We focus on how misspecification of the exposure measurement model, and hence the model-implied item covariance structure, ultimately impacts estimation and inference regarding the associations between the latent exposures and a health outcome. In Section 2 we give the algebraic representation of the longitudinal exposure models in Figure 1, and show how the model can be embedded within the broader SEM with LVs framework. Section 3 describes the estimators used, and Section 4 compares the estimators via simulation. Section 5 presents the analysis of data from the ELEMENT study, and we end in Section 6 with a summary of our findings and further discussions.

2. Modeling approach

There are several available notations for linear structural equation models (SEM) with latent variables (LV) [1, 2, 34]. Here we borrow from a common form of the LISREL model [32] to write our longitudinal latent exposure model. We then explain how our model fits within the general LISREL notation.

2.1. Longitudinal latent exposure model

As discussed in the introduction, we are interested specifically on surrogate exposure measures that are measured repeatedly over time. To devote our focus to the exposure measurement model assumptions, we make the simplifying assumption that the health outcome of interest, Y_i, measured on $i = 1, \dots N$ individuals, is univariate and continuous. The discussion points to literature for non-continuous Y_i. We use $X_{it} = (x_{it1}, ..., x_{itK})^\top$ to denote measurements on $k = 1, ..., K$ exposure markers collected repeatedly on $t = 1, \dots, \ell$ occasions. All of the observed exposure measures can be collected into a block vector $X_i = (X_{i1}^\top, ..., X_{i\ell}^\top)^\top$, where each block represents the multivariate exposure measures at each occasion. In the motivating example there are $\ell = 3$ measurement occasions (Figure 1).

We posit that exposure information X_{it} is a manifestation of a univariate latent exposure variable ξ_{it} that arises at each occasion $t = 1, ..., \ell$. The vector of latent exposures for the i^{th} individual $\xi_i = (\xi_{i1}, \dots, \xi_{i\ell})^\top$ is assumed to have mean $\alpha_{\xi_i} = (\alpha_{\xi_{i1}}, \dots \alpha_{\xi_{i\ell}})^\top$ and covariance matrix Σ_ξ. We use the subscript i in α_{ξ_i} to denote the possibility that the mean of the latent exposure changes according to person-level covariates such as the time of measurement (see Section *2.2.3*). For simplicity we assume that exposure biomarkers at occasion t, X_{it} are not associated with the latent variable at another occasion, $\xi_{t'}$, given ξ_t; this assumption can be relaxed (see Section *2.3*). Then, at any one measurement occasion, t, the *exposure measurement model* is:

$$X_{it} = \alpha_{xt} + \Lambda_{xt}\xi_{it} + \delta_{it} \quad t = 1, \dots, \ell. \tag{1}$$

We assume the observed health outcome is impacted by the latent exposures through the *outcome model*:

$$Y_i = \alpha_y + \gamma^\top \xi_i + \epsilon_i. \tag{2}$$

In (1), the vector of intercepts $\alpha_{xt} = (\alpha_{xt1}, \alpha_{xt2}, \dots \alpha_{xtK})^\top$ and the vector of factor loadings $\Lambda_{xt} = (\lambda_{t1}, \dots \lambda_{tK})^\top$ capture the assumed linear relationships between the observed biomarkers X_{it} and its corresponding latent variable ξ_{it}. The vector δ_{it} denotes exposure measurement errors, is independent of the vector of latent variables ξ_i, and has zero mean and covariance $\Sigma_{\delta t}$. In (2), α_y is the typical intercept and $\gamma = (\gamma_1, \dots, \gamma_\ell)^\top$ captures the association between the

latent exposures and the outcome, and ϵ_i is a residual error with zero mean and variance σ^2_ϵ, and is independent of ξ_i and $\delta_i = (\delta_{i1}^\top, \ldots, \delta_{i\ell}^\top)^\top$. The exposure coefficients γ are of primary interest in this study. Additional covariates or adjustment factors can be included in (1)-(2), in addition to the intercepts. For simplicity, we leave out such adjustment factors since including them is rather straightforward, albeit with more complex notation.

Identifiability constraints, e.g. [1], are needed in order to estimate model parameters. A common constraint is to set the scale of each latent variable equal to the scale of one of its surrogates by constraining the corresponding factor loading to one. The particular surrogate with the factor loading equal to one is called the *scaling item*. The location can be set either by constraining the mean of the latent variable to zero, or constraining the intercept of the scaling item to zero. Without loss of generality, we assume the scaling item at each occasion is x_{it1}, the first element of X_{it}. It is also common to assume a diagonal structure for the covariance matrix of the measurement errors, although this is not required. In our exposure model, we assume that the measurement error in the scaling item is independent of the measurement errors in the remaining items. Prior to discussing additional constraints that can be made in the longitudinal latent exposure model, we describe the exposure model for the lead example to help gain clarity of the model.

In the lead example (Figure 1), the exposure measurement model at each trimester $t = 1, 2, 3$ is

$$\begin{array}{rll} x_{it1} = & \xi_{it} + \delta_{it1} & \text{Plasma Lead} \\ x_{it2} = & \alpha_{xt2} + \lambda_{t2}\xi_{it} + \delta_{it2} & \text{Blood Lead (Laboratory 1)} \\ x_{it3} = & \alpha_{xt3} + \lambda_{t3}\xi_{it} + \delta_{it3} & \text{Blood Lead (Laboratory 2)} \\ x_{i34} = & \alpha_{x34} + \lambda_{34}\xi_{i3} + \delta_{i34} & \text{Cord Blood,} \end{array}$$

where plasma lead measures at each trimester t, x_{it1}, have been selected as the scaling items.

2.2. *Exposure measurement model assumptions*

Given the assumed longitudinal nature of the latent exposure variable, several assumptions can be made on the exposure model parameters across the measurement occasions, including assumptions about $\alpha_{xt}, \Lambda_{xt}, \Sigma_{\delta t}$, the covariances between the error vectors at different time points $Cov(\delta_{it}, \delta_{it'})$, and the mean $\alpha_{\xi_i t}$ and variance of the latent exposure. If appropriate, these additional constraints would reduce the overall number of parameters estimated, and thus have the potential to boost efficiency of model parameter estimates.

2.2.1. Time invariance of the measurement model. Time invariance of the measurement model is the assumption that at least some of the exposure model parameters $\alpha_{xt}, \Lambda_{xt}, \Sigma_{\delta t}$ are equal for all t. There are several types of time invariance that can be discussed [18, 19, 20], the strongest type being when intercepts, factor loadings and variance matrices are all equal.

Equality of the factor loadings, i.e., in our case $\Lambda_{x1} = \Lambda_{x2} = \cdots = \Lambda_{x\ell}$, with other parameters left unrestricted is the most basic type of measurement invariance, and we focus on it in the simulation study. In the lead example, differences in the factor loadings could arise, implying $\overline{TI}$, due to correlations among lead biomarkers changing over time; for example, due to plasma volume expansion during pregnancy [17] as previously described. We use TI to denote cases when the time invariance of all parameters holds, and $\overline{TI}$ when it does not hold.

2.2.2. Variance structure for measurement errors. In many applications of latent variable models, conditional independence of all items given the latent variables is assumed (i.e., diagonal structure for $var(\delta) = \Sigma_\delta$). Although this assumption can be relaxed, there are many choices about which off-diagonal terms to estimate. In the longitudinal exposure setting, there may be "between-occasion" correlations due to serial correlation among the measurement errors related to measuring the same biomarker repeatedly over time (i.e., serial correlation), or there may be "within-occasion correlations" among measurement errors in biomarkers measured at the same occasion. For instance, positive correlation among blood lead levels measured at laboratory 1 ($k = 2$) across trimesters could be present due to specific measurement techniques at the laboratory (i.e., $Cov(\delta_{it2}, \delta_{it'2}) > 0$). When such correlations are zero, we say that serial independence, denoted as as SI, among the measurement errors of a specific item measured repeatedly holds. Alternatively, when there is non-zero between-occasion correlations in the measurement errors of a given item, then serial independence does not hold, denoted as $\overline{SI}$. Within-occasion correlation among the residual errors of the items measured at the same visit could exist as well; in our example, for instance, due to lead being measured in the same media (e.g., blood vs. plasma) in our example. Although the term conditional independence is broader and includes serial correlation, here we use the symbol (CI) to denote conditional independence, given the latent variable, of biomarkers or items taken at the same time (i.e., CI implies $var(\delta_t) = \Sigma_{\delta t}$ is a diagonal matrix for each t). Otherwise, lack of conditional independence is present (denoted as $\overline{CI}$).

2.2.3. Mean and variance structure for latent exposure vector. Since the latent exposure is longitudinal, it is conceivable that the mean of the latent exposure changes over time. In that case, the mean vector α_{ξ_i} could depend on covariates such as the actual time of measurement. Similarly, although the covariance matrix among the latent exposure variables, Σ_ξ, is typically left unstructured, more parsimonious assumptions could be used, such as an autoregressive or compound symmetry structure or random effects [35]. Parsimonious representations of α_{ξ_i} and Σ_ξ may be advantageous when there

are a large number of occasions ℓ, especially in studies with small sample size. Since in the lead example $\ell = 3$, and time of measurement is coded as discrete occasions (trimesters), we leave α_ξ and Σ_ξ unstructured.

2.3. Connection to general linear structural equation modeling

In general, SEMs with LV are denoted with a larger system of equations than (1)-(2) and encompass latent outcomes in addition to latent exposures. The general SEMs with LV are composed of a structural equation which defines the relationships among m latent outcomes $\eta_i = (\eta_1, \ldots, \eta_m)^\top$, and latent exposures ξ_i, i.e., $\eta_i = \alpha_\eta + B\eta_i + \Gamma\xi_i + \epsilon_i^*$, and two measurement model equations. One measurement equation describes the relationships between latent predictors and their surrogates, i.e., $X_i = \alpha_x + \Lambda_x\xi_i + \delta_i$, similar to (1) above. The second describes the relationships between latent outcomes and a multivariate vector of p observed outcome surrogates, i.e., $Y_i = \alpha_y + \Lambda_y\eta_i + \zeta_i$. Measurement error vectors ζ_i and δ_i are independent of ϵ_i^* and ξ_i. The intercept vectors α_η, α_x and α_y, and vector α_ξ parametrize the means of the variables corresponding to their subscript, and Σ, with a corresponding subscript, is used to denote covariance matrices. The coefficient matrix B, with diagonal entries equal to zero, encodes the relationships among the endogenous latent variables η_i; each of the m rows of the $m \times \ell$ matrix Γ contains the effects of the latent exposures on the latent outcome variable corresponding to the row; Λ_y and Λ_x are factor loading matrices defining which, and how strongly, observed variables relate to the latent variables.

Our measurement model (1) can be embedded in this general notation by defining the factor loading matrix, Λ_x as a block diagonal matrix with diagonal blocks Λ_{xt}, and letting $\alpha_x = (\alpha_{x1}^\top, \ldots, \alpha_{x\ell}^\top)^\top$ and $\delta = (\delta_1^\top, \ldots, \delta_\ell^\top)^\top$. In the lead example, the parameters are $\alpha_x = (0, \alpha_{x12}, \alpha_{x13}, 0, \alpha_{x22}, \alpha_{x23}, 0, \alpha_{x32}, \alpha_{x33}, \alpha_{x34})^\top$ and

$$\Lambda_x = \begin{pmatrix} 1 & \lambda_{12} & \lambda_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda_{22} & \lambda_{23} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \lambda_{32} & \lambda_{33} & \lambda_{34} \end{pmatrix}^\top .$$

Our outcome model (2) is obtained as a special case of the structural equation and the measurement model for the outcome where $p = 1$ and $m = 1$. In this case, the only observed outcome, univariate Y_i, becomes the scaling item, i.e., $\Lambda_y = 1$. We set the location of the latent outcome by letting $\alpha_\eta = 0$. Since $m = 1$, then $B = 0$, and Γ has only one row, which we denote as $\gamma^\top$. Then, the simplified structural equation $\eta_i = \gamma^\top\xi + \epsilon_i^*$ can be plugged in to the simplified measurement model, $Y_i = \alpha_y + \eta_i + \zeta_i$, to yield our outcome model $Y_i = \alpha_y + \gamma^\top\xi + \epsilon_i$, where the residual $\epsilon_i = \epsilon_i^* + \zeta_i$ combines the residual from the structural equation and the measurement error for the observed outcome.

Since our model (1)-(2) can be embedded within the general LISREL notation, it is straightforward to relax the assumption that exposure biomarkers at occasion t are not associated with the latent variable at another occasion, $\xi_{t'}$,

given ξ_t. Specifically, this can be done by freeing some of the factor loadings that have been constrained to zero in Λ_x above.

3. Estimation

3.1. Maximum Likelihood Estimation

Let θ represent all model parameters. Assuming normality of ϵ, δ, and ξ, the observed marginal likelihood can be written as $L(\theta) = \prod_{i=1}^{N} f_{\mathrm{Y|X}}(Y_i|X_i;\theta) f_{\mathrm{X}}(X_i;\theta)$, where $f_{\mathrm{Y|X}}, f_{\mathrm{X}}$ are normal densities. Letting $\ell(\theta) = \log L(\theta)$, the ith's subject contribution to the likelihood score equation for γ is

$$\frac{\partial \ell_i}{\partial \gamma} = \widehat{\xi}_i^{\mathrm{LV}}(Y_i - \mu^i_{\mathrm{Y|X}})/\sigma^2_{\mathrm{Y|X}} + \frac{\Sigma_{\xi|\mathrm{X}}\gamma}{2(\sigma^2_{\mathrm{Y|X}})^2}\left[(Y_i - \mu^i_{\mathrm{Y|X}})^2 - \sigma^2_{\mathrm{Y|X}}\right] \tag{3}$$

where $\widehat{\xi}_i^{\mathrm{LV}} = E(\xi_i|X_i) = \alpha_{\xi_i} + \Sigma_\xi \Lambda_x^\top \Sigma_{\mathrm{X}}^{-1}(X_i - \mu^i_{\mathrm{X}})$ is the expected value of the latent variable given the items; $\mu^i_{\mathrm{Y|X}} = \alpha_y + \gamma^\top \widehat{\xi}_i^{\mathrm{LV}}$ and $\sigma^2_{\mathrm{Y|X}} = \gamma^\top \Sigma_{\xi|\mathrm{X}}\gamma + \sigma^2_\epsilon$ are the conditional mean and variance of Y_i given X_i; and $\Sigma_{\xi|\mathrm{X}} = \mathrm{var}(\xi_i|X_i) = \Sigma_\xi - \Sigma_\xi \Lambda_x^\top \Sigma_{\mathrm{X}}^{-1} \Lambda_x \Sigma_\xi$; $\mu^i_{\mathrm{X}} = E(X_i) = \alpha_x + \Lambda_x \alpha_{\xi_i}$; $\Sigma_{\mathrm{X}} = Var(X_i) = \Lambda_x \Sigma_\xi \Lambda_x^\top + \Sigma_\delta$. Parameter estimates are obtained by solving the full set of score equations (see Appendix A) using standard numerical procedures. Variances for parameter estimates can be obtained by inverting the information matrix, $I_\theta = -E(\partial^2 \ell(\theta)/\partial\theta\partial\theta^\top)$, or by computing robust variances $\widehat{var}_R(\widehat{\theta}) = B^{-1}AB^{-\top}$, where $A = 1/N \sum_{i=1}^{N} \{\partial \ell_i(\theta)/\partial\theta\}\{\partial \ell_i(\theta)/\partial\theta\}^\top$, $B = -1/N \sum_{i=1}^{N} \partial \ell_i(\theta)/\partial\theta$.

Since we are primarily interested in the bias in $\widehat{\gamma}$, we examine the expected value of (3). In contrast to the score equation for a regression on observed predictors, (3) has two components. The first term is the same as what would be used if $\widehat{\xi}_i^{\mathrm{LV}}$ was the observed predictor, and the second term is due to the residual error variance of the outcome given the observed items, $\sigma^2_{\mathrm{Y|X}}$, being dependent on γ. Letting the superscript ^{0} denote the true parameters from the data generating model and the superscript * the (asymptotic) parameter estimates that would be obtained for a given model, and assuming Y and X are conditionally independent given ξ and the relationship between Y and ξ in (2) is correct, the expected value of (3)

is

$$E\left[\frac{\partial \ell_i}{\partial \gamma}\right] = \frac{1}{\sigma^{2,*}_{\mathrm{Y}|\mathrm{X}}}\left\{C^*\Sigma^0_{\mathrm{X}}(C^0-C^*)^\top + \alpha^0_{\xi_i}(\alpha^*_{\xi_i}-\alpha^0_{\xi_i}+C^*(\mu^{i,0}_{\mathrm{X}}-\mu^{i,*}_{\mathrm{X}}))^\top\right\}\gamma \quad (4)$$

$$+\frac{\Sigma^*_{\xi|\mathrm{X}}}{2(\sigma^{2,*}_{\mathrm{Y}|\mathrm{X}})^2}(\sigma^{2,*}_{\mathrm{Y}|\mathrm{X}}-\sigma^{2,0}_{\mathrm{Y}|\mathrm{X}})\gamma \quad (5)$$

$$+\frac{\Sigma^*_{\xi|\mathrm{X}}}{2(\sigma^{2,*}_{\mathrm{Y}|\mathrm{X}})^2}\left\{\gamma^\top(C^0-C^*)\Sigma^0_{\mathrm{X}}(C^0-C^*)^\top\gamma\right\}\gamma \quad (6)$$

$$+\frac{\Sigma^*_{\xi|\mathrm{X}}}{2(\sigma^{2,*}_{\mathrm{Y}|\mathrm{X}})^2}\left\{(\alpha^*_{\xi_i}-\alpha^0_{\xi_i}+C^*(\mu^{i,0}_{\mathrm{X}}-\mu^{i,*}_{\mathrm{X}}))^\top\gamma\right\}^2\gamma \quad (7)$$

where $C=\Sigma_\xi\Lambda_x^\top\Sigma_{\mathrm{X}}^{-1}$ are the weights given to the items X_i to obtain $\widehat{\xi}^{\mathrm{LV}}_i=\alpha_{\xi_i}+\Sigma_\xi\Lambda_x^\top\Sigma_{\mathrm{X}}^{-1}(X_i-\mu^i_{\mathrm{X}})$. Hence, even when the means are correctly specified, if components of the marginal covariance of X, $\Lambda_x, \Sigma_\xi, \Sigma_\delta$, are not correctly modeled, the expected value of (3) will not be zero. The magnitude of bias will depend on how poorly the estimated weight matrix C^* approximates the truth, C^0. When the approximation is relatevely poorer, $\widehat{\xi}^{\mathrm{LV}}_i$ will be a more biased estimate of the latent exposure. Thus, regressing Y_i on $\widehat{\xi}^{\mathrm{LV}}_i$, as the first term of (3) suggests, yields a biased estimate of γ. The term (4) represents the bias in the score equation for γ associated with $\widehat{\xi}^{\mathrm{LV}}_i$ not being a consistent estimate of ξ_i when the covariance components of X are misspecified. The remaining three terms (5) – (7) correspond to the second term of (3). Importantly, if all elements of γ are zero, then there is no bias. However, if at least one of the elements of γ is not zero, then the zero elements of γ can have bias, as will be demonstrated in the simulations.

3.2. Bollen's two-stage least squares (2SLS) estimator

Here we describe Bollen's 2SLS estimator in the context of our model, whereas [24] and [32] describe the estimator for the full LISREL model. The first step to construct the estimator is to re-write the model (1)-(2) using only observed variables. This is achieved by sorting and partitioning the vector of observed exposures X_i into

$$X_i=(\widetilde{X}^\top_{i1},\widetilde{X}^\top_{i2})^\top$$

where $\widetilde{X}_{i1}=(x_{i11},x_{i21},\ldots,x_{i\ell 1})^\top$ is a vector containing the scaling indicators across all measurement occasions, and $\widetilde{X}_{i2}$ contains the remaining exposure measures for the individual. Thus,

$$\widetilde{X}_{i1}=\xi_i+\widetilde{\delta}_{i1}$$

where $\widetilde{\delta}_{i1}$ is the vector containing the subset of measurement errors corresponding only to the scaling items. This later equation can be used to solve for ξ_i:

$$\xi_i=\widetilde{X}_{i1}-\widetilde{\delta}_{i1},$$

and is substituted into (2) to get

$$Y_i = \alpha_y + \gamma^\top \widetilde{X}_{i1} + (\epsilon_i - \gamma^\top \widetilde{\delta}_{i1}). \tag{8}$$

Similarly, a reduced measurement model for $\widetilde{X}_{i2}$ can be obtained,

$$\widetilde{X}_{i2} = \alpha_{\widetilde{X}_{i2}} + \Lambda_{\widetilde{X}_2} \widetilde{X}_{i1} + (\widetilde{\delta}_{i2} - \Lambda_{\widetilde{X}_2} \widetilde{\delta}_{i1}). \tag{9}$$

Parameters in these reduced equations (8)-(9) cannot be estimated using ordinary least squares since the residual terms (e.g., $\epsilon_i - \gamma^\top \widetilde{\delta}_{i1}$) are correlated with the observed predictors $\widetilde{X}_{i1}$. Instead, instrumental variables (IV) for $\widetilde{X}_{i1}$ can be used. Although an instrumental variable approach is available to jointly estimate all parameters (8)-(9), the estimates are the same as estimating parameters in each equation of the system separately [32]. Since we are primarily interested in the outcome model parameters, we follow the equation-by-equation approach in [24] and focus on IV estimation for (8).

IV estimation requires the availability of at least ℓ variables or instruments that: (a) are correlated with $\widetilde{X}_{i1}$, and (b) are uncorrelated with the composite error term $u_i = \epsilon_i - \gamma^\top \widetilde{\delta}_{i1}$ [24]. Variables not included in the model (1)-(2) that meet conditions (a) and (b) could be used as IVs. However, Bollen's IV estimator [24] is built upon the idea that $\widetilde{X}_{i2}$ form a pool of potential *model implied* IVs. Variables from $\widetilde{X}_{i2}$ that meet conditions (a) and (b) are used as IVs. To assess condition (a), the sample correlations between $\widetilde{X}_{i1}$ and candidate IV variables in $\widetilde{X}_{i2}$ can be used. In our case study and simulations, correlations among $\widetilde{X}_{i1}$ and $\widetilde{X}_{i2}$ were significantly different from zero. To verify condition (b), model assumptions are needed. We used an automated approach [36] to verify that all variables in $\widetilde{X}_{i2}$ satisfy condition (b) under our assumed model. We thus use investigate the use of the full set of $\widetilde{X}_{i2}$ as the IV variables in our simulation studies. Nevertheless, some investigators have found that dropping items in $\widetilde{X}_{i2}$ can sometimes yield parameter estimates with better properties [33]. Thus, we also investigate the use of subsets of $\widetilde{X}_{i2}$ in our simulations.

Estimation proceeds in two stages: The first is to carry out a multivariate regression of $\widetilde{X}_{i1}$ on $\widetilde{X}_{i2}$, followed by the regression of Y_i on the predicted values from the first step. To formalize this, let $\boldsymbol{Y}$ contain the n outcome values, $\boldsymbol{u}$ the n residuals $u_i = \epsilon_i - \gamma^\top \widetilde{\delta}_{i1}$, $\boldsymbol{Z}$ be an $N \times (1 + \ell)$ matrix that contains a 1 and the ℓ values $\widetilde{X}_{i1}$ for each subject in each row, and $A = (\alpha_y, \gamma_1, \gamma_2, \gamma_3)^\top$. Thus, (8) becomes:

$$\boldsymbol{Y} = \boldsymbol{Z}A + \boldsymbol{u}.$$

Similarly, let $\boldsymbol{V}$ be a matrix with N rows, with each row $i = 1, \ldots, N$ containing a 1 and the $\widetilde{X}_{i2}^\top$ values for the i^{th} subject. Then the first stage consists of the multivariate regression $E(\boldsymbol{Z}) = \boldsymbol{V}G$, where G is a $(k-1)\ell \times (1+\ell)$ matrix of coefficients that can be estimated as $\widehat{G} = (\boldsymbol{V}^\top \boldsymbol{V})^{-1} \boldsymbol{V}^\top \boldsymbol{Z}$. Then, the predicted values can be obtained as $\widehat{\boldsymbol{Z}} = \boldsymbol{V}\widehat{G}$.

Regressing $\boldsymbol{Y}$ on $\widehat{\boldsymbol{Z}}$ at the second stage then yields

$$\widehat{A} = (\widehat{\boldsymbol{Z}}^\top \widehat{\boldsymbol{Z}})^{-1} \widehat{\boldsymbol{Z}}^\top \boldsymbol{Y}.$$

Bollen [24] proposes using $\widehat{var}(\widehat{A}) = \widehat{\sigma}_u^2 (\widehat{\boldsymbol{Z}}^\top \widehat{\boldsymbol{Z}})^{-1}$, where $\widehat{\sigma}_u^2 = (\boldsymbol{Y} - \boldsymbol{Z}\widehat{A})^\top (\boldsymbol{Y} - \boldsymbol{Z}\widehat{A})/N$ to obtain asymptotic standard errors. Notice that the proposed $\widehat{var}(\widehat{A})$ is not the naive variance where $(\boldsymbol{Y} - \widehat{\boldsymbol{Z}}\widehat{A})^\top (\boldsymbol{Y} - \widehat{\boldsymbol{Z}}\widehat{A})$ is used in lieu of $(\boldsymbol{Y} - \boldsymbol{Z}\widehat{A})^\top (\boldsymbol{Y} - \boldsymbol{Z}\widehat{A})$.

3.3. Modifications to Bollen's 2SLS estimator to improve efficiency

Although potentially more robust, Bollen's 2SLS is likely to suffer a large loss of efficiency due in part to the large number of parameters in G, i.e., up to $(\ell + 1) \cdot (k - 1)\ell$ coefficients. In addition to dropping weak instruments as a way to improve the properties of IV estimators, imposing constraints in G can also reduce the number of parameters estimated and improve efficiency. When the TI assumption is satisfied, we propose constraining the diagonal blocks of G to be equal. This is a viable constraint since the association of x_{it1} with $x_{it2}, ..., x_{itk}$ conditional on all other instruments is the same for all t when the TI assumption is satisfied. We call these estimates $\widehat{\gamma}_{\text{IV2}}$, and refer to the 2SLS estimates obtained in Section 3.2 as $\widehat{\gamma}_{\text{IV1}}$. In the simulation studies we also consider using fewer instruments instead of the full list of available variables, and call them by either $\widehat{\gamma}_{\text{IV1R}}$ $\widehat{\gamma}_{\text{IV2R}}$, with the R denoting a reduced number of instruments. Appendix C of the supplementary materials describes how to impose constraints on G. It also briefly discusses other potential constraints on G that, although intuitive in a longitudinal setting, did not perform well in simulations.

3.4. Connection between Bollen's 2SLS and Regression Calibration

A rich literature has been devoted to investigate approaches to correct bias in regression coefficients that is due to measurement error in covariates. Within this literature, classical measurement error models assume that a true exposure is associated with an outcome, but the exposure is unobserved (i.e., it is a latent exposure like ξ, and model (2) is assumed. Typically an unbiased surrogate is assumed to be observed, i.e., similar to $\widetilde{X}_{i1}$ above, in order to enable estimation of regression coefficients in the outcome model. Regression calibration (RC) is an approach used to obtain unbiased exposure coefficients in the outcome model. Although several RC estimation procedures are available, a simple algorithm [37, Section 6.3.3] consists of two steps: regressing the observed unbiased surrogates $\widetilde{X}_{i1}$, on available instruments (e.g., $\widetilde{X}_{i2}$), and then using predicted values of $\widetilde{X}_{i1}$ as regressors in the outcome model instead of the unobserved ξ_i. Obviously,

this simple RC estimator yields exactly the same estimates as Bollen's 2SLS estimator, even though the RC estimator is motivated without the assumption of a full LISREL model. One difference between the RC and Bollen's approaches, however, is that in RC, the variance of the estimator is derived using estimating equations approach (see Appendix B), and thus explicitly accounts for the uncertainty in the predicted values used as regressors in the second stage of the model.

4. Simulation Studies

We evaluate the impact of violations to exposure measurement model assumptions on inferences in the outcome model parameters, specifically the exposure-outcome associations captured by γ. We examine the impact of violations of (a) conditional independence of the observed exposures, X_{it}, given the latent variables, ξ_{it}; (b) misspecification of the serial correlation of a given item across time; and (c) time invariance of the factor loadings in the exposure measurement model. Aside from using a factorial design to consider combinations of violations of these three measurement model assumptions, we also consider varying degrees of the misspecification, different sample sizes, and use a varying number of instruments in the IV estimators.

4.1. Simulation set up

We simulated data assuming the existence of one latent variable at each of three occasions. Values of the latent variables for each individual were simulated using a multivariate normal distribution assuming exchangeable correlation and unit variance (off diagonal elements of Σ_ξ are 0.25). The outcome was simulated as $Y_i = \gamma^\top \xi_i + \epsilon_i$ with $var(\epsilon_i) = 2$, and $\gamma = (0, 0, 0)$ to evaluate Type I error probabilities, and $\gamma = (-2, -2, -2)$ or $\gamma = (-2, -2, 0)$ to examine bias and relative efficiency or mean squared error.

Data for 15 items X_i (5 at each of three occasions) were initially simulated using a factorial design with three factors representing conditional independence (CI) or lack of conditional independence ($\overline{CI}$); serial independence (SI) or serial correlation ($\overline{SI}$); and measurement invariance over time (TI) or lack of measurement invariance ($\overline{TI}$). Hence, data were generated assuming a total of 8 different "true" models. We denote a given true model by the combinations of TI or $\overline{TI}$, SI or $\overline{SI}$, and CI or $\overline{CI}$. Lack of measurement invariance over time ($\overline{TI}$) was simulated by setting the factor loadings for item X_{itk}, $k = 2, \ldots, 5$ at 1 for $t = 1$, 1.2 at $t = 2$, and 1.5 at $t = 3$. Lack of conditional independence was simulated by having items 2, 4, and 5 share a random intercept at the 2nd time point only (i.e., inducing a compound symmetry structure among the within-occasion measurement errors of these items). By simulating $\overline{CI}$ at only one time point, we

would be able to assess how covariance misspecification at only one time point induces bias in the coefficient of exposure at a different time. Similarly, $\overline{SI}$ was simulated by letting each item $k = 2, 3, 4, 5$ share a random intercept with the same item at different time points (i.e., induce a compound symmetry structure among the between-occasion measurement error of the same type of biomarker). The variance of the measurement error for each item was set at 1. The variances for the random intercepts inducing conditional or serial dependence were initially set at 1; i.e., the correlation among items, conditional on the latent variables, was $1/(1 + 1) = 0.5$. For each combination of parameters we simulated 500 data sets each with N=300 or $N = 600$. For each true model, we used MLE to fit eight "working models" also defined by whether or not TI, SI or CI were assumed. We use italics to denote the true model and non-italics to denote the working model used at the estimation stage (e.g., data model $TISICI$ vs. working model TISICI).

A second set of simulations was devoted to assess the impact of different magnitudes of any one type of misspecification on mean squared error (MSE), which serves as a measure of bias-variance trade off. To examine lack of time invariance, we simulated data under the true model $\overline{TI}SICI$, assuming the factor loadings for item $j = 2, \ldots, 4$ were all 1 at $t = 1$, 1.2 at $t = 2$, and λ^* at $t = 3$, with λ^* ranging from 1 to 1.5. Various magnitudes of violation of CI or SI were simulated by generating data from the true models $TI\overline{SI}CI$ or $TISI\overline{CI}$ using random intercepts as previously described, but letting the random intercept variance range from 0 to 1. In this second set of simulations we focused only on IV1 (given the results of the first set of simulations), and fitted working models with the 'classic' assumption of TISICI or TI$\overline{\text{SI}}$CI. The latter is a reasonable first model for a longitudinal latent exposure model where $\overline{SI}$ is likely, and is consistent with the strategy of fitting a reasonable model that can then be improved according to model modification/selection criteria.

Maximum likelihood estimates for model parameters were obtained using M$plus$ [22]. Instrumental variable estimates were obtained using R functions available from the authors. Previous studies have shown that IV estimators that use fewer than all available items may have better properties [33]. Thus, for the above scenarios we fitted IV estimators that included either all available instruments, or only two of the four possible instruments at each time point. For the latter, we either included items with $k = 2, 3$ or $k = 2, 4$; items with $k = 2, 4$ have correlated errors among themselves at time point 2, thus enabling us to investigate how choices of specific item subsets can impact IV estimators.

4.2. Simulation Results

4.2.1. Type I error. When $\gamma = (0, 0, 0)^\top$, all estimators had empirical Type I error probabilities between 0.04 and 0.06 for tests at the 0.05 significance level regardless of the underlying exposure measurement model used to simulate data (results

not shown). However, when $\gamma = (-2, -2, 0)^\top$, the rejection rates for $H_o : \gamma_3 = 0$ were up to 50% for some of the working models fitted using MLE, and some of the restricted IV estimators (IV2), largely due to bias (below). The IV1 estimator and the most flexible working model $\overline{\text{TISICI}}$ fitted with MLE had rejection probabilities between 0.04 and 0.06 for this test.

4.2.2. Bias for MLE. Figure 2 displays the bias of various estimators for γ for the 8 true models used to generate data in the first set of simulations and true $\gamma = (-2, -2, -2)$ (i.e., 8 true models shown in the legend). Each of the working models fitted using MLE (shown along the x-axis) corresponds to one of the true models. For a given true model, the correctly specified working model is unbiased, as are working models that estimate additional covariance parameters. Although the most flexible working model, $\overline{\text{TISICI}}$ is always unbiased, arriving at the correct model specification may be difficult in practice. Since sample size did not have an appreciable impact on bias, our discussion on bias below does not differentiate between the sample sizes used.

Violation of the TI assumption resulted in positive bias for $\widehat{\gamma_1}$, small bias for $\widehat{\gamma_2}$, and negative bias for $\widehat{\gamma_3}$ (e.g.,in Figure 2, see true models $+\ \overline{TI}SICI$ or $\diamondsuit\ \overline{TISI}CI$ when the working model TI$\overline{\text{SI}}$CI is fitted). This pattern of bias across the coefficients may be heuristically explained by the fact that the factor loadings increased with t in both of these true models: the working models assuming TI resulted in estimated factor loadings close to the average of the factor loadings over the measurement times, and the average was close to the factor loadings at the second time point (not shown). Because the estimated factor loadings at the second time point were approximately correct, the bias for $\widehat{\gamma_2}$ is smaller. However, the estimated factor loadings at $t = 1$ were over estimated, resulting in positive bias for $\widehat{\gamma_1}$, and under estimated for $t = 3$, resulting in negative bias for $\widehat{\gamma_3}$.

Violation of CI at $t = 2$ resulted in positive bias for $\widehat{\gamma_1}$ and $\widehat{\gamma_3}$, but negative bias for $\widehat{\gamma_2}$ (e.g., see true models $\triangledown\ TISI\overline{CI}$ or $*\ TI\overline{SICI}$ when the working models TISICI or TI$\overline{\text{SI}}$CI are fitted). Interestingly, this bias was greatly reduced when the time invariance assumption on the factor loadings was removed (e.g., see true model denoted with $\triangledown\ TISI\overline{CI}$ when the working model $\overline{\text{TI}}$SICI vs. when TISICI is fitted). Even though in the true model the factor loadings are the same across time, allowing them to differ in the working model 'absorbs' the bias that would have otherwise corrupted the outcome model coefficients γ.

Finally, violation of the serial independence assumption resulted in negative bias for all outcome model regression coefficients. Unlike violations of the CI assumption, the bias due to violation of SI persisted unless serial dependence

was estimated in the model.

Simulation results when true $\gamma = (-2, -2, 0)$ were generally similar, with the exception of violations of SI. When SI did not hold, $\widehat{\gamma_1}$ and $\widehat{\gamma_2}$ had negligible bias, but $\widehat{\gamma_3}$ had severe negative bias. That is, although $\widehat{\gamma_1}$ and $\widehat{\gamma_2}$ were correctly estimated, a spurious association between ξ_3 and the outcome appeared when incorrectly assuming SI held (see Appendix Figure S1, $\times$ symbol).

4.2.3. Bias in IV estimators. IV1 estimators had a small amount of bias that decreased with sample size, both when true $\gamma = (-2, -2, -2)$ or true $\gamma = (-2, -2, 0)$ and regardless of whether the full set (IV1) or only a reduced set of instruments (IV1R$_{23}$ or IV1R$_{24}$) were used (Figure 2 and Supplementary Figure 1). At the same sample size, the estimators with fewer items, IV1R$_{23}$ and IV1R$_{24}$, had even lower bias as had been previously noted [33]. IV2 estimators were only unbiased for two true models, $\triangle\ TISICI$ and $\times\ TI\overline{SI}CI$, as would be expected since IV2 restricts coefficients in the stage 1 regression based on the TI assumption. Again, higher sample size reduced the observed bias as did using fewer instruments for both of these true models. One may hypothesize that IV2 could also be unbiased for the $TISI\overline{CI}$ model, since TI also holds. However, $\overline{CI}$ was generated at only one time point in these simulations, which makes the associations between the scaling indicators and the instruments different over time. The fact that IV2R$_{23}$ has lower bias than IV2R$_{24}$ for the true models $\triangledown\ TISI\overline{CI}$ and $*\ TI\overline{SICI}$ illustrates that the bias in IV2 comes primarily from having $\overline{CI}$ at $t = 2$. Recall that IV2R$_{23}$ uses items with $k = 2, 3$, which are conditionally independent for all t, whereas items with $k = 2, 4$ have correlated measurement errors at $t = 2$.

4.2.4. Standard errors. Figure 3 shows the variation for $\widehat{\gamma_1}$ (left) and $\widehat{\gamma_2}$ (right); the patterns of variation for $\widehat{\gamma_3}$ were similar to those of $\widehat{\gamma_2}$ (not shown). The top two figures show the variability when data are generated with true model, $TISICI$, whereas the bottom shows the variability when the data are with true model $\overline{TISICI}$. Since the variances of the estimates are greatly reduced by doubling the sample size from 300 to 600, we normalize all the variances by the empirical variance of the MLE where from working model TISICI (top) or $\overline{\text{TISICI}}$ (bottom) to enable us to show both sample sizes in the same plot.

When data were generated with true model $TISICI$, relaxing the assumption of serial independence or conditional independence among the MLEs had little impact on efficiency ($< 5\%$ e.g., comparing working model TISI$\overline{\text{CI}}$ or TI$\overline{\text{SI}}$CI to TISICI), in contrast to relaxing the time invariance assumption (up to $\approx 45\%$ comparing working model $\overline{\text{TI}}$SICI to

TISICI). This is partly due to the relatively fewer parameters needed to relax the SI and CI assumptions compared to the number of parameters needed to relax time invariance. For all MLEs, the average of the asymptotic variances were only slightly lower than the variance of the estimates; the difference was reduced with higher sample size.

IV estimators are inevitably less efficient; the empirical variance of the IV estimates (+ symbol) was at least 75% larger than that of the most parsimonious MLE. IV2 had smaller variance than IV1, as we expected given the fewer coefficients estimated in the stage 1 regression. For IV1, the variance formula proposed by Bollen and the estimating equations variance formula were close to the empirical variance, and were closer at higher sample sizes. However, for IV2, the variance estimator proposed by Bollen was consistently larger than the empirical variance it does not account for the potentially reduced variance due to constraining G. The estimators utilizing fewer items ($IV1R_{23}$ and $IV1R_{24}$, or $IV2R_{23}$ and $IV2R_{24}$) had about 30% higher variance than IV1 or IV2, respectively, since information to predict the scaling items is lost when dropping instruments.

When data were generated under the true model, $\overline{TISICI}$, Figure 3 (bottom) shows the variation in the estimates of γ obtained from IV are closer to the correct MLE. This shows that for the cases when IV estimators would be most needed (most flexible data generating model), the loss of efficiency is relatively lower than when they are not. Since parsimonious MLEs incur large degrees of bias when they are misspecified, but IVs have more comparable variability when data arise from more general data generating models, IV estimators can potentially achieve better bias-variance tradeoff, i.e., lower mean squared error (MSE).

4.2.5. Mean squared error (MSE). MSE allows us to compare estimators with respect to their compromise between bias and efficiency. Figure 4 shows the MSEs obtained from the second set of simulations when true $\gamma = (-2, -2, -2)$ (top) and $\gamma = (-2, -2, 0)$ (bottom) for parsimonious working models estimated via MLE, and IV1 and IV2, under violations of TI, SI, or CI. The IV estimators with fewer items are not included in these figures since they have about the same bias but larger variance and thus larger MSE compared their corresponding IV1 or IV2 estimator shown.

When $\gamma = (-2, -2, -2)$ (Figure 4, top), IV1 had a better bias-variance trade-off for $\widehat{\gamma}_1$ when the degree of misspecification in the MLE was large. While the MSE of IV1 remained relatively constant across a range of violations of TI (left), SI(middle), or CI (right), the MSE of the MLE increased rapidly due to increases in bias. The degree of misspecification at which IV1 was better than MLE (crossing point of the MSE curve for MLE and IV1) depended on the sample size. MLE tended to be better for smaller sample sizes because while the bias of both estimators was relatively

constant with sample size, the MLE variance was relatively lower than that of the IV variance at the lower sample size. The pattern of the MSE comparing IV1 and MLE estimators for $\widehat{\gamma}_2$ and $\widehat{\gamma}_3$ was very similar (see Appendix Figure S2) to those of $\widehat{\gamma}_1$ in the figure. IV2 demonstrated lower MSE compared to IV1 for lower degrees of misspecification when TI and CI were violated, and consistently lower when SI was violated. This is expected since IV2 has lower variance than IV1, and is unbiased regardless of whether SI holds or not. The MSE in IV2 increased with increasing magnitude of the violations of TI, or CI at $t = 2$, as increases in bias overtook its advantage with lower variance. The improvement in MSE comparing IV2 to IV1 was more apparent at lower sample sizes.

When true $\gamma = (-2, -2, 0)$ (Figure 4, bottom), the MSE for $\widehat{\gamma}_1$ was consistently lower for MLE compared to IV when SI (middle) did not hold in the true model. This is due to the fact that the bias in the MLE for $\widehat{\gamma}_1$ was small when true γ_3 is 0 and SI did not hold, but IV had consistently larger variance compared to the MLE. When either TI or CI do not hold, we again see that IV1 is better than MLE when the degree of $\overline{TI}$ or $\overline{CI}$ increases (left and right bottom panels). The patterns of MSE for $\widehat{\gamma}_2$ across various degrees of $\overline{SI}$, $\overline{TI}$ or $\overline{CI}$ was the same as for $\widehat{\gamma}_1$ (Appendix Figure S3). However, when SI did not hold in the data generating model, the MSE for the MLE of $\widehat{\gamma}_3$ increased rapidly due to large increases in bias (Supplementary Figure 2). Interestingly, the MSE for MLE estimates of γ_3 was consistently lower than MSE for IV1 estimates when $\overline{TI}$ or $\overline{CI}$ increases (Supplementary Figure 2). Again IV2 demonstrated lower MSE compared to IV1 for lower degrees of misspecification when TI and CI were violated, and consistently lower when SI was violated, as described for the case above when $\gamma = (-2, -2, -2)$.

5. Prenatal lead exposure and mental development

We use data from the Early Life Exposures in MExico City to Neuro-Toxicants (ELEMENT) study, where prospective mothers were recruited at or before conception. One goal of the study was to quantify effects of lead exposure during each trimester of pregnancy on child development [11]. Women were followed during pregnancy to assess their exposure to lead. Various measures of exposure (lead concentrations in whole blood and plasma) were collected on the mother during each trimester of pregnancy. Measures of lead concentration in blood and plasma during pregnancy are the closest surrogate measures of fetal exposure. Other information, such as maternal age and IQ was also collected [11, 12]. Children were followed after birth to assess their cognitive development using the mental development index (MDI) of Bayley's scale of mental development [38]. We use MDI measured at 24 months of age as the outcome of interest. Figure 1 shows

potential models for this data.

We present data from 341 mother-child pairs. To be included in the analysis, the mother had to have measurements on at least one of the surrogate measurements of fetal exposure, and have completed an IQ test. The children in the sample completed Bayley's MDI, and had a concurrent blood lead measurement at the time of the MDI assessment. Descriptive statistics are given in Supplementary Table 1. Because almost all variables had some missing data, we constructed five imputed data sets using sequential regression methods [39]. While MLE could account for missing data within the estimation procedure, the IV estimates could not. We estimated model parameters using the methods described in Section 3 on all imputed data sets, and combined them using standard formulae [40].

Table 1 gives the regression coefficients for the outcome model obtained by maximum likelihood estimation (MLE) under 8 possible combinations of assumptions regarding serial or conditional independence and time invariance. The estimates shown range from a most restricted working model (TISICI, top of Figure 1), to the least restrictive model ($\overline{\text{TI}}\overline{\text{SI}}\overline{\text{CI}}$, bottom of Figure 1); they differ substantially both in terms of estimated effect and their standard errors, reflecting compromises between bias and variance. Our working models that assumed $\overline{\text{CI}}$ included correlations among blood lead concentrations within each trimester, since blood measures are replicate measures assayed in different laboratories; for $\overline{\text{SI}}$ we included serial correlations for the items; and for $\overline{\text{TI}}$ we allowed factor loadings, item means, and residual variances to vary over time. Other model specifications can arise from, for example, assuming partial time invariance (e.g., constrain the factor loadings for some factors to be constant across time, but not all) or freeing other of the 45 possible pair-wise correlations. For simplicity, the working models shown assumed that $\overline{\text{TI}}$, $\overline{\text{SI}}$, and/or $\overline{\text{CI}}$ held or not. Its noteworthy that when we used the working model TISICI, modification indices revealed a need to relax constraints of conditional or serial independence, but not a need to relax the assumption of time invariance.

Table 1 also shows estimates obtained using IV estimators, and multiple linear regression (MLR) using plasma lead levels at first trimester as the exposure measure. The MLR estimate is much smaller compared to all other estimates, reflecting attenuation due to measurement error. The IV1 estimate is within the range of estimates obtained via MLE, and the standard error is comparable to those obtained with the more flexible MLEs. The IV2 estimate is somewhat attenuated compared to IV1, and follows the patterns of the MLEs where working models that assume TI have smaller coefficients than those that don't.

Table 2 shows the parameter estimates for the exposure model using the working models TI$\overline{\text{SI}}$CI and $\overline{\text{TI}}\overline{\text{SI}}$CI. The

table shows that factor loadings and item means and variances change with time (estimates under $\overline{\text{TISI}}$CI). Lack of time invariance in the measurement model is primarily responsible for the large changes in the regression coefficients in the estimated MLEs (Table 1), as would be expected from the simulation results. Table 2 also shows that there is significant covariance among the residual errors of blood lead concentrations assayed in laboratory 1 ($\sigma_{x_{21},x_{22}}, \sigma_{x_{21},x_{23}}, \sigma_{x_{22},x_{23}}$ not zero), as suspected given the longitudinal nature of the exposure biomarkers.

6. Discussion

Models which assume longitudinal latent exposures are of increasing interest in environmental epidemiology given the new tendency to collect multivariate exposure measures over time [9, 10]. Since these models are highly parameterized, it is important to examine the impact of model assumptions on the estimates of key parameters of interest. While approaches to check and relax linearity and distributional assumptions have been studied [41, 42], including the availability of distribution-free estimators such as [43], less work has been done to examine measurement model assumptions that are particularly relevant to longitudinal settings. We presented a model with longitudinal latent exposures and examined the impact of three types of misspecification of the exposure measurement model, namely, conditional independence, serial independence of repeated items, and time invariance of the factor loadings. We found that incorrectly assuming that factor loadings are constant over time can have a major impact on outcome model regression coefficients estimated via MLE, that ignoring positive serial correlation in items measured repeatedly across time results in bias toward the null in regression coefficients, and that incorrect assumption of conditional independence resulted in bias toward the null for some coefficients and away from the null for others. We examined properties of IV estimators of the exposure effect, which were more robust although predictably less efficient because they make fewer assumptions about the covariance structure among exposure biomarkers. Differences in bias-variance trade off (mean squared error) favored IV estimators, compared to the most parsimonious ML estimators, in situations where there was a medium to large degree of misspecification.

Prior robustness studies comparing ML and IV estimators have focused on examining violations of conditional independence [21], missing paths [33] or misspecifying latent variable means [26], but, to our knowledge, none examined time invariance assumptions or combinations of types of model misspecification in the measurement model of longitudinal latent variables. We found the bias-variance trade-off between IV and MLE estimates obtained from a working model assuming TI favors IV once the difference in factor loadings in the true model exceeds about 20%. Our simulation study

demonstrated that ML estimators were always sensitive to lack of serial independence. Since lack of serial independence is likely in longitudinal studies, it is recommended to always include correlations among individual items across time. We found that including additional parameters to the model to relax the serial independence assumption incurred essentially no cost in terms of efficiency of the MLE. We also found that lack of conditional independence within items measuring the latent exposure variable at one occasion biases regression coefficient of exposure parameters at other occasions. This is likely due to residual measurement error in the plug in estimate for at least one of the latent variables, i.e., the fact that measurement error in one covariate can bias coefficients for other predictors if predictors are correlated [44]. Interestingly, we found that the magnitude of bias in the exposure coefficient estimates due to lack of conditional independence depended on whether time invariance of the measurement model is assumed. When time invariance was not assumed, the bias due to lack of conditional independence was smaller. Relaxing the assumption regarding time invariance indirectly implies that the marginal covariance at one point in time can be different than at another time point by allowing the factor loadings which define the latent variable to differ. When there is lack of conditional independence, the factor loadings will be biased but allowing the factor loadings to differ at the different time points will capture the residual correlation caused by lack of conditional independence, and contain some of the bias from contaminating outcome model parameters. Finally, we found that using a smaller set of instruments can lead to small improvements in small sample bias of the IV estimator, but pronounced increases in its variance. Previous studies that used fewer instruments than the available model-implied instruments when estimating IV1 dropped the weakest instruments and showed either negligible loss of efficiency or even slight improvements [33]. In our simulations all items had the same measurement error variance (i.e., had about the same strength as instruments), which may partially account for the differences in our results compared to previous studies. When we used smaller measurement error variance for the instruments, we also observed relatively smaller losses of efficiency when dropping instruments (not shown).

The issue of whether MLE or IV estimators are best ultimately depends on the purpose of the analysis; some issues to take into account are bias vs. efficiency, testing vs. estimation, whether the exposure model parameters are of interest, and missing data. Although MLEs will be most efficient when the specified model is correct and distributional assumptions hold, IV estimators will be unbiased in more situations, thus IV methods would be preferable for estimation. Since loss of efficiency, and thus reduced power, can be noted as a problem for IV models, MLEs could be favored for testing when distributional assumptions hold, at least approximately [1, 43]. However, because potentially many (low powered)

model selection steps may be involved before achieving a well fitting model, the potential for inflated Type I error can quickly arise in the MLE framework. One scenario where MLE is potentially advantageous is when one is interested in certain exposure measurement parameters such as reliability of one item vs. others. That is, MLE may be better suited for exploratory or explanatory purposes, but IV better suited for testing and estimation of exposure effects. One setting where MLE could be seen as having an advantage is in the presence of missing data; however, this may not be readily the case. Although MLE can easily handle cases with data missing at random (MAR), MLE relies on correct model specifications in order for inferences with MAR data to be valid [40]. On the other hand, IV estimators would require a data processing step, such as multiple imputation, before the method can be applied. Thus, IV estimators would require a correct imputation model.

The use of longitudinal latent variable models as discussed here has some limitations. In the case of the example, we posited the existence of one latent variable at each occasion. The observed biomarkers could be modeled in different configurations instead of the one we posited (e.g., separate latent variables for blood lead vs. plasma lead). The model however, can be easily extended to more than one latent variable at each occasion if so desired, and our bias analyses would still be relevant. Another issue is that of assigning meaning to the longitudinal latent exposure variable when time invariance does not hold. However, if at least the scaling item keeps the same relationship with the latent variable across time (i.e., $x_{i1t} = \xi_{it} + \delta_{i1t}$), then not all interpretation is lost. In the lead example, we assume plasma lead concentration is an unbiased measure of the underlying fetal exposure regardless of time, since lead in plasma is what would be more likely bio-available to the fetus (in contrast to lead in red blood cells). In this manner, we can assume the units of the latent variable remain constant through time (the interpretation of the latent variable is the same), despite possible changes in the relationship between the latent variable and other exposure biomarkers.

Several additional extensions of our study may be worth conducting. Given the also increased interest in longitudinal latent outcome models [45, 35, 41, 46, 47, 48] it is important to evaluate how estimates of coefficients for observed predictors are influenced by incorrectly assuming time invariance in the measurement model for a longitudinal latent *outcome*. While MLE estimates will probably have the similar degree of bias as observed here, the robustness of the IV estimator will largely depend on whether time invariance constraints [25, 26] are imposed or not. Our study focused on continuous exposure surrogates, and we assumed multivariate normality to obtain ML estimates. Several other available estimators for models with continuous outcomes do not require normality, but still require a correctly specified covariance

matrix for the observed data [1, 43]. Hence, bias for those estimators would likely be of similar magnitude as for ML. When observed variables are not continuous, ML estimators can still be computed [49], as well as IV estimators [50]. Examining these estimators in the presence of missing data may also be desirable.

In summary, we found that potentially large biases in exposure coefficients can result from violations of the time invariance assumption for the measurement model, conditional independence among the items measuring one latent variable, and from violation of the assumption of serial independence across individual items measured repeatedly over time. Although our simulation design does not cover all possible model misspecifications, this study highlights the importance of examining measurement model assumptions in longitudinal latent exposure variable models.

Acknowledgements

The authors thank ELEMENT investigators for proving data for the example, as well as the following NIEHS grants that supported data collection: P01 ES012874; P42 ES05947; R01 ES013744; R01 ES014930; R01 ES007821. The authors also acknowledge salary support from grants NIEHS R01 ES016932, R01 ES017022, and 1-P20-SE018171-01.

References

1. Bollen KA. *Structural Equations with Latent Variables*. Wiley Series in Probability and Mathematical Statisics, Wiley: New York, NY, 1989.
2. Muthén B. Beyond sem: General latent variable modeling. *Behaviormetrika* 2002; **29**:81–117.
3. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall: Boca Raton, FL, 2004.
4. Budtz-Jørgensen E, Keiding N, Grandjean P, Weihe P, White RF. Statistical methods for the evaluation of health effects of prenatal mercury exposure. *Environmetrics* 2003; **14**(2):105–120.
5. Nikolov MC, Coull BA, Catalano PJ, Diaz E, Godleski JJ. Statistical methods to evaluate health effects associated with major sources of air pollution: a case-study of breathing patterns during exposure to concentrated boston air particles. *Journal of the Royal Statistical Society Series C-Applied Statistics* 2008; **57**:357–378.
6. Tao Y, Sanchez B, B M. Latent variable models for geneenvironment interactions in longitudinal studies with multiple correlated exposures. *Statistics in Medicine* 2015; **34**(7):227–1241.
7. Chen Y, KK F, JD M, McElrath T, B M. Statistical methods for modeling repeated measures of maternal environmental exposure biomarkers during pregnancy in association with preterm birth. *Environmental Health* 2015; **14**:9.
8. Ferguson K, McElrath T, Chen YH, Mukherjee B, Meeker J. Longitudinal profiling of inflammatory cytokines and c-reactive protein during uncomplicated and preterm pregnancy. *Am J Reprod. Immunol.* 2014; **72**(3):326–336.
9. Ferguson K, Cantonwine D, Rivera-Gonzalez L, Loch-Caruso R, Mukherjee B, Anzalota Del Toro L, Jimenez-Velez B, Calafat A, Ye X, Alshawabkeh A, *et al.*. Urinary phthalate metabolite associations with biomarkers of inflammation and oxidative stress across pregnancy in puerto rico. *Environ. Sci. Technol.* 2014; **48**(12):7018–7025.
10. Stingone J, Luben J TJ Daniels, Fuentes M, Richardson D, Aylsworth A, Herring A, Anderka M, Botto L, Correa A, Gilboa S, *et al.*. Maternal exposure to criteria air pollutants and congenital heart defects in offspring: Results from the national birth defects prevention study. *Environ Health Perspect* 2014; **122**(8):863–872.
11. Hu H, Tellez-Rojo MM, Bellinger D, Smith D, Ettinger AS, Lamadrid-Figueroa H, Schwartz J, Schnaas L, Mercado-Garcia A, Hernandez-Avila M. Fetal lead exposure at each stage of pregnancy as a predictor of infant mental development. *Environ Health Perspect* 2006; **114**(11):1730–5.
12. Tellez-Rojo MM, Hernandez-Avila M, Lamadrid-Figueroa H, Smith D, Hernandez-Cadena L, Mercado A, Aro A, Schwartz J, Hu H. Impact of bone lead and bone resorption on plasma and whole blood lead levels during pregnancy. *Am J Epidemiol* 2004; **160**(7):668–78.
13. Kaplan D. The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research* 1988; **23**(1):69–86.
14. Maccallum RC, Wegener DT, Uchino BN, Fabrigar LR. The problem of equivalent models in applications of covariance structure-analysis. *Psychological Bulletin* 1993; **114**(1):185–199.
15. Kirby J, Bollen KA. Using instrumental variable (iv) tests to evaluate model specification in latent variable structural equation models. *Sociological Methodology* 2009; **39**(1):327–355.

16. Cribbie RA. Evaluating the importance of individual parameters in structural equation modeling: the need for type i error control. *Personality and Individual Differences* 2000; **29**(3):567–577.
17. Faupel-Badger JM, Hsieh CC, Troisi R, Lagiou P, Potischman N. Plasma volume expansion in pregnancy: implications for biomarkers in population studies. *Cancer Epidemiol Biomarkers Prev* 2007; **16**(9):1720–3.
18. Kaplan D. Power of the likelihood ratio test in multiple group confirmatory factor-analysis under partial measurement invariance. *Educational and Psychological Measurement* 1989; **49**(3):579–586.
19. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling* 2002; **9**(2):233–255.
20. Vandenberg RJ, Lance CE. A review and synthesis of the measurment invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods* 2000; **3**(1):4–70.
21. Reddy SK. Effects of ignoring correlated measurement error in structural equation models. *Educational and Psychological Measurement* 1992; **52**(3):549–570.
22. Muthén L, Muthén B. M*plus*: Statistical analysis with latent variables user's guide 2006.
23. Rabe-Hesketh S, Pickles A, Skrondal A. *GLLAMM Manual, technical report*. Kings College, Institute of Psychiatry, Dept. Biostatistics and Computing: London, 2001.
24. Bollen KA. An alternative two stage least squares (2sls) estimator for latent variable equations. *Psychometrika* 1996; **61**(1):109–121.
25. Nestler S. How the 2sls/iv estimator can handle equality constraints in structural equation models: A system-of-equations approach. *British Journal of Mathematical and Statistical Psychology* 2014; **67**:353–369.
26. Nestler S. Using instrumental variables to estimate the parameters in unconditional and conditional second-order latent growth models. *Structural Equation Modeling* 2015; **22**:461–473.
27. Hagglund G. Factor-analysis by instrumental variables methods. *Psychometrika* 1982; **47**(2):209–222.
28. Lance C, Cornwell J, Mulaik S. Limited information parameter estimates for latent or mixed manifest and latent variable models. *Multivariate Behavioral Research* 1988; **23**:155–167.
29. Hoogland JJ, Boomsma A. Robustness studies in covariance structure modeling - an overview and a meta-analysis. *Sociological Methods & Research* 1998; **26**(3):329–367.
30. Sammel MD, Ryan LM. Effects of covariance misspecification in a latent variable model for multiple outcomes. *Statistica Sinica* 2002; **12**(4):1207–1222.
31. Sánchez BN, Budtz-Jørgensen E, Ryan LM. An estimating equations approach to fitting latent exposure models with longitudinal health outcomes. *Annals of Applied Statistics* 2009; **3**(2):830–856.
32. Bollen KA. Two-stage least squares and latent variable models: simultaneous estimation and robustness to misspecification. *Structural Equation Modeling: Present and Future, edited by R. Cudeck , S. Du Toit , and D. Sorbom. Lincolnwood, IL: Scientific Software* 2001; :119–138.
33. Bollen KA, Kirby JB, Curran PJ, Paxton PM, Chen FN. Latent variable models under misspecification - two-stage least squares (2sls) and maximum likelihood (ml) estimators. *Sociological Methods & Research* 2007; **36**(1):48–86.
34. Jöreskog KG, Sörbom D. *Lisrel 7: A Guide to the Program and Applications*. SPSS: Chicago, 1989.
35. Roy J, Lin XH. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* 2000; **56**(4):1047–1054.
36. Bollen KA, Bauer D. Automating the selection of model-implied instrumental variables. *Sociological methods and research* 2004; **32**(4):425–452.
37. Carroll RJ, Ruppert D, Stefanski L, Crainiceanu C. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd edn., Monographs on Statistics and Applied Probability, Chapman & Hall: Boca Raton, FL, 2006.
38. Bayley N. *Bayley Scales of Infant Development*. 2nd edn., The Psychological Corporation: San Antonio, TX, 1993.
39. Raghunathan TE, Solenberger P, Van Hoewyk J. Iveware: Imputation and variance estimation software user guide 2002.
40. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd edn., John Wiley & Sons, Inc: New Jersey, 2002.
41. Proust C, Jacqmin-Gadda H, Taylor JMG, Ganiayre J, Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* 2006; **62**(4):1014–1024.
42. Sánchez BN, Houseman EA, Ryan LM. Residual-based diagnostics for structural equation models. *Biometrics* 2009; **65**(1):104–115.
43. Browne M. Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* 1984; **37**:127–141.
44. Budtz-Jørgensen E, Keiding N, Grandjean P, Weihe P, White RF. Consequences of exposure measurement error for confounder identification in environmental epidemiology. *Statistics in Medicine* 2003; **22**(19):3089–3100.
45. Farrell AD. Structural equation modeling with longitudinal data: Strategies for examining group differences and reciprocal relationships. *Journal of Consulting and Clinical Psychology* 1994; **62**(3):477–487.
46. Leiby BE, Sammel MD, Ten Have TR, Lynch KG. Identification of multivariate responders and non-responders by using bayesian growth curve latent class models. *Journal of the Royal Statistical Society Series C-Applied Statistics* 2009; **58**:505–524.
47. Chavance M, Escolano S, Romon M, Basdevant A, de Lauzon-Guillain B, Charles MA. Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology. *BMC Med Res Methodol* 2010; **10**:37.
48. Barbosa-Leiker C, Wright BR, Burns GL, Parks CD, Strand PS. Longitudinal measurement invariance of the metabolic syndrome: Is the assessment of the metabolic syndrome stable over time? *Annals of Epidemiology* 2011; **21**(2):111–117.
49. Shi JQ, Lee S. Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society* 2000; **62**(1):77–87.
50. Nestler S. A monte carlo study comparing piv, uls and dwls in the estimation of dichotomous confirmatory factor analysis. *Br J Math Stat Psychology* 2013; **66**(1):127–143.

Table 1. Estimated association between prenatal lead exposure at trimester 1 and mental development at 24 months of age.

Estimator*	Est.	S.E.**	Est./S.E.
TISICI	-2.72	0.84	-3.23
TISI$\overline{\text{CI}}$	-2.71	0.84	-3.22
TI$\overline{\text{SI}}$CI	-2.70	0.84	-3.23
TI$\overline{\text{SICI}}$	-2.67	0.84	-3.18
$\overline{\text{TI}}$SICI	-3.47	1.07	-3.24
$\overline{\text{TI}}$SI$\overline{\text{CI}}$	-3.43	1.06	-3.23
$\overline{\text{TI}}$$\overline{\text{SI}}$CI	-3.37	1.07	-3.16
$\overline{\text{TISICI}}$	-3.31	1.07	-3.09
IV1 (all)	-2.83	1.09 (1.22)	-2.60 (-2.32)
IV1R$_{Lab1}$	-2.82	1.10 (1.20)	-2.56 (-2.35)
IV1R$_{Lab2}$	-2.68	1.11 (1.29)	-2.42 (-2.08)
IV2 (all)	-2.62	1.09 (1.10)	-2.41 (-2.37)
IV2R$_{Lab1}$	-2.83	1.12 (1.14)	-2.52 (-2.48)
IV2R$_{Lab2}$	-2.49	1.05 (1.08)	-2.38 (-2.31)
MLR, x_{11}	-1.97	1.15	-1.72

* First 8 lines refer to working models fitted via MLE
** SE using Bollen (estimating equations) formula

Table 2. Estimated exposure model parameters

			$TI\overline{\text{SI}}CI^{\dagger}$			$\overline{\text{TISI}}\text{CI}$		
Variable	t	Parameter	Est	SE	Est/SE	Est	SE	Est/SE
Factor Loadings								
Blood lead (Lab 1)	1	λ_{21}	1.73	0.15	11.94	2.07	0.26	8.09
	2	λ_{22}				1.35	0.15	8.97
	3	λ_{23}				1.55	0.10	15.97
Blood lead (Lab 2)	1	λ_{31}	1.52	0.10	15.02	1.90	0.17	11.49
	2	λ_{32}				1.18	0.15	7.92
	3	λ_{33}				1.32	0.08	16.51
Cord lead	3	λ_{43}	1.15	0.11	10.19	1.01	0.11	9.26
Item Means								
Blood lead (Lab 1)	1	$\alpha_{x_{21}}$	-4.30	0.55	-7.84	-5.78	0.96	-6.05
	2	$\alpha_{x_{22}}$				-2.92	0.54	-5.41
	3	$\alpha_{x_{23}}$				-3.50	0.41	-8.63
Blood lead (Lab 2)	1	$\alpha_{x_{31}}$	-3.61	0.40	-8.97	-5.20	0.65	-7.96
	2	$\alpha_{x_{32}}$				-2.40	0.55	-4.39
	3	$\alpha_{x_{33}}$				-2.81	0.31	-9.06
Cord lead	3	$\alpha_{x_{43}}$	-2.01	0.43	-4.74	-1.42	0.40	-3.52
Residual variances of items								
Plasma lead	1	$\sigma^2_{x_{11}}$	0.46	0.02	25.28	0.62	0.07	8.57
	2	$\sigma^2_{x_{12}}$				0.35	0.03	10.21
	3	$\sigma^2_{x_{13}}$				0.30	0.03	8.74
Blood lead (Lab 1)	1	$\sigma^2_{x_{21}}$	0.12	0.03	3.90	0.27	0.10	2.81
	2	$\sigma^2_{x_{22}}$				0.06	0.02	3.05
	3	$\sigma^2_{x_{23}}$				0.06	0.03	2.20
Blood lead (Lab 2)	1	$\sigma^2_{x_{31}}$	0.12	0.02	6.72	0.08	0.05	1.58
	2	$\sigma^2_{x_{32}}$				0.13	0.02	5.82
	3	$\sigma^2_{x_{33}}$				0.12	0.03	3.75
Cord lead	3	$\sigma^2_{x_{43}}$	0.51	0.06	8.53	0.52	0.06	8.19
Between-occasion covariances among item measurement errors								
Blood lead (Lab 1)	1,2	$\sigma_{x_{21},x_{22}}$	0.037	0.017	2.18	0.056	0.024	2.33
	1,3	$\sigma_{x_{21},x_{23}}$	0.046	0.023	2.00	0.055	0.027	2.04
	2,3	$\sigma_{x_{22},x_{23}}$	0.097	0.034	2.85	0.048	0.017	2.82
Blood lead (Lab2)	1,2	$\sigma_{x_{31},x_{32}}$	-0.005	0.014	-0.36	-0.009	0.020	-0.45
	1,3	$\sigma_{x_{31},x_{33}}$	0.017	0.017	1.00	0.013	0.017	0.76
	2,3	$\sigma_{x_{32},x_{33}}$	-0.020	0.024	-0.83	-0.009	0.019	-0.47

† Factor loadings, item means and measurement error variances are constrained to be equal for a given item for all t

Figure 1. Path diagrams showing relationships between exposure biomarkers, latent prenatal lead exposure measured at three trimesters, and mental development. Both models assume the same structural model. Measurement model in top panel assumes time invariance (TI), serial independence, SI and conditional independence CI of items given the latent variables; in bottom panel, TI, SI, CI assumptions are relaxed.

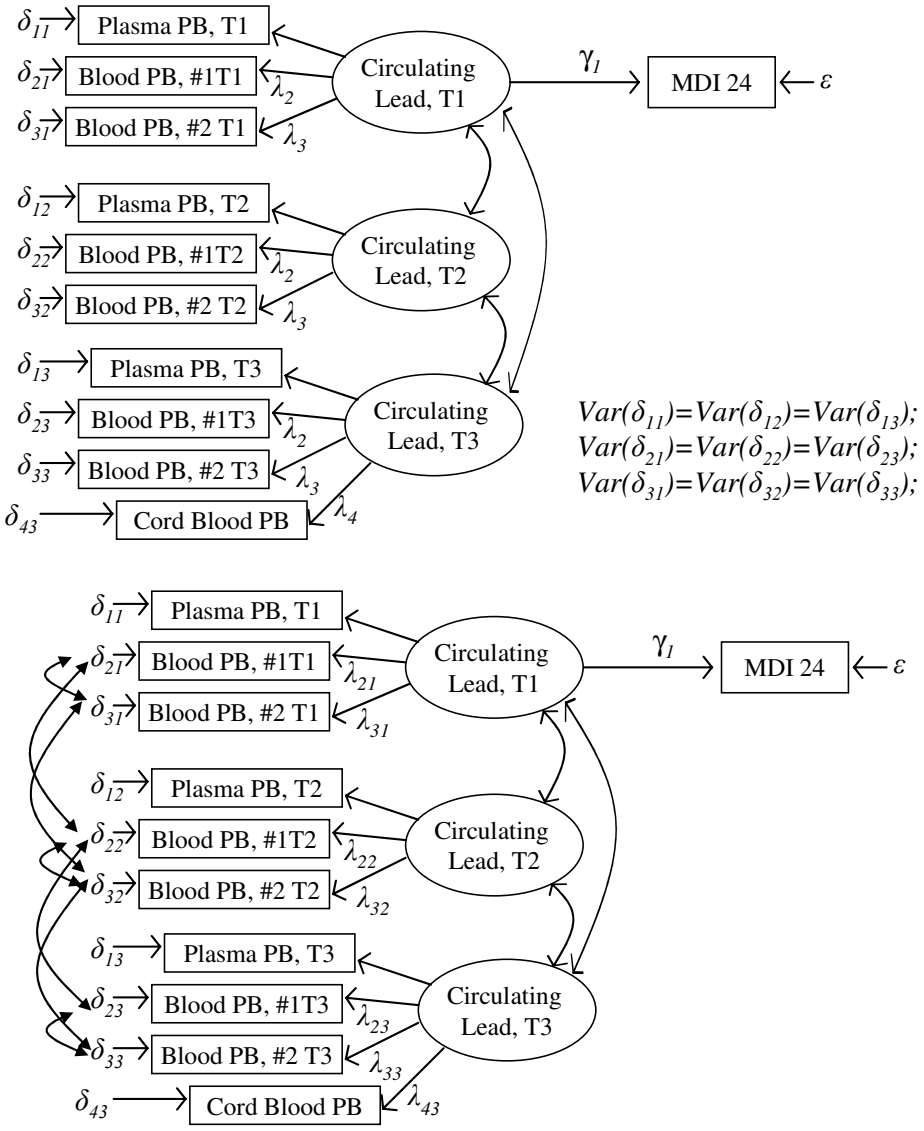


Figure 2. Percent bias in $\widehat{\gamma}_1$ (top), $\widehat{\gamma}_2$ (middle), $\widehat{\gamma}_3$ (bottom) for instrumental variable and maximum likelihood estimators with various working models (x-axis) when data are generated under a variety of true models (denoted by different symbols in the legend) and true $\gamma = (-2, -2, -2)$.

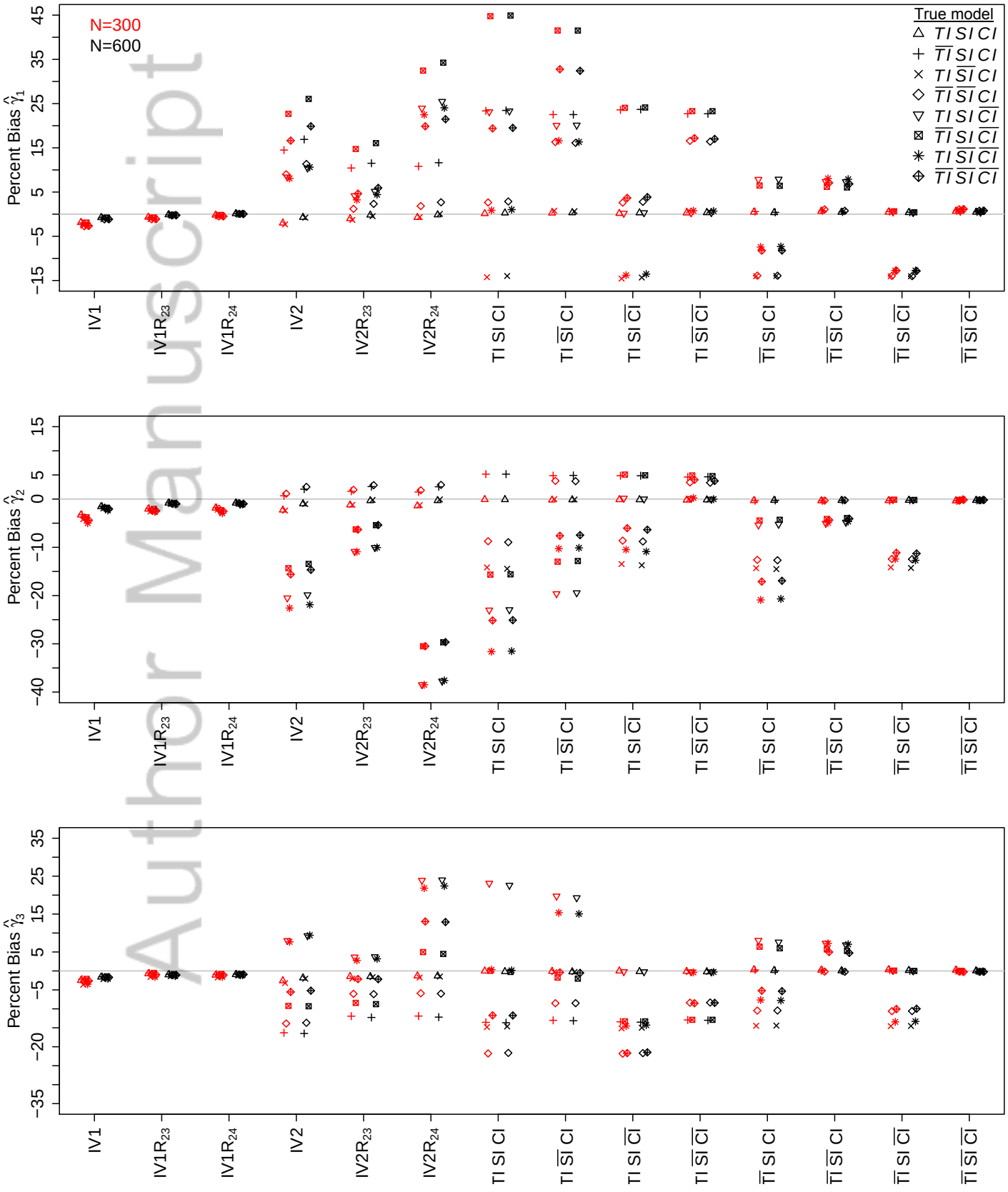
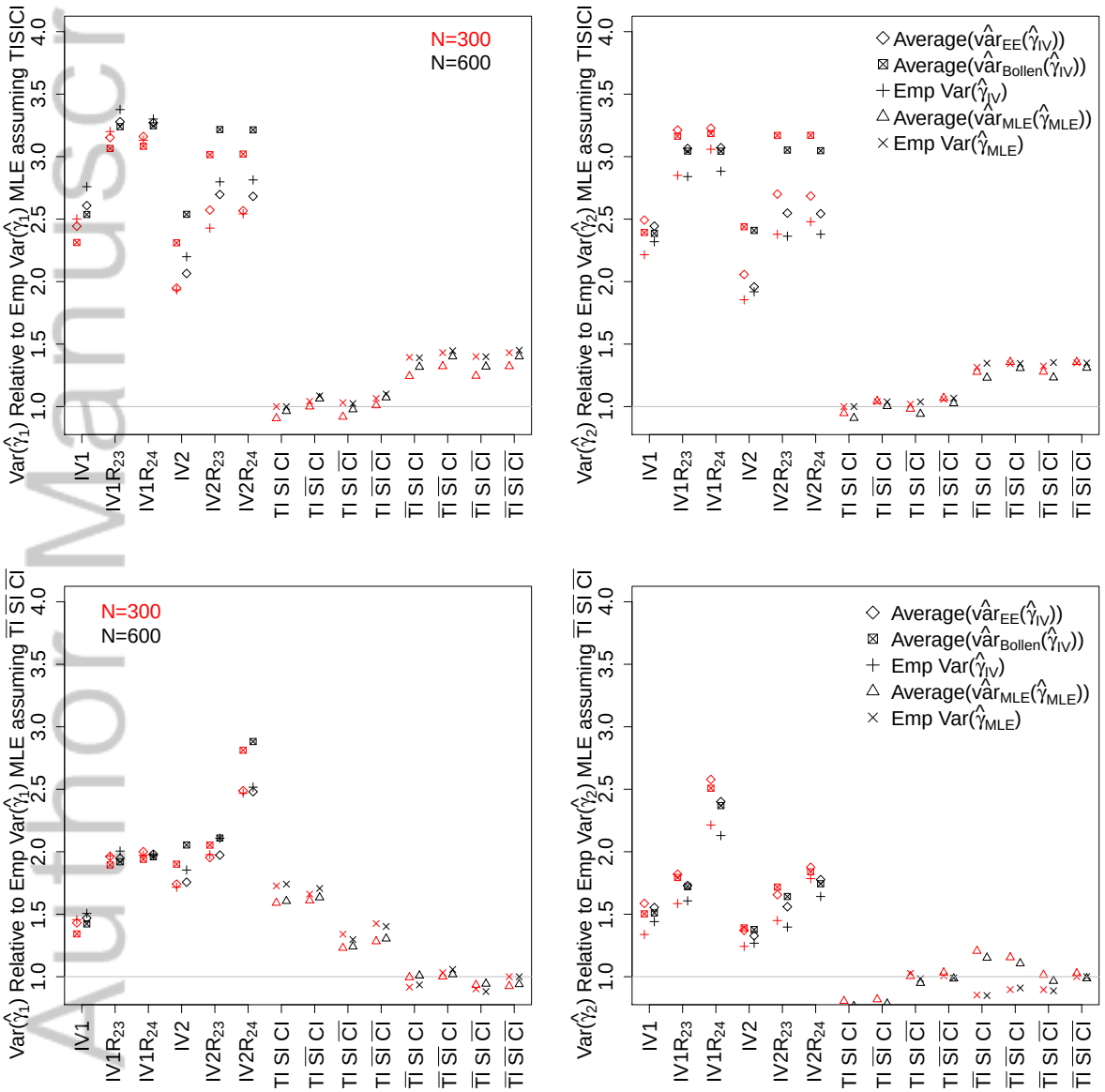


Figure 3. Variability in $\widehat{\gamma}_1$ (left) and $\widehat{\gamma}_2$ (right) for instrumental variable and maximum likelihood estimators with various working models (x-axis) when data are generated from the true model $TISICI$ (top) or true model $\overline{TI}SICI$ (bottom), and true $\gamma = (-2, -2, -2)$. Variance patterns for $\widehat{\gamma}_3$ were similar to those of $\widehat{\gamma}_2$.



Prepared using simauth.cls

Figure 4. MSE of $\widehat{\gamma}_1$ for instrumental variable and maximum likelihood estimators when data are generated from true $\gamma = (-2, -2, -2)$ (top) or true $\gamma = (-2, -2, 0)$ (bottom) and various degrees of lack of time invariance (left); various degrees of serial dependence (middle); various degrees of lack of conditional independence (right).