Structural Equation Modeling: A Multidisciplinary Journal, 24: 283–300, 2017

ISSN: 1070-5511 print / 1532-8007 online
DOI: 10.1080/10705511.2016.1250635

Pattern Mixture Models for Quantifying Missing Data Uncertainty in Longitudinal Invariance Testing

Sonya K. Sterba
Vanderbilt University

Many psychology applications assess measurement invariance of a construct (e.g., depression) over time. These applications are often characterized by few time points (e.g., 3), but high rates of dropout. Although such applications routinely assume that the dropout mechanism is ignorable, this assumption may not always be reasonable. In the presence of nonignorable dropout, fitting a conventional longitudinal factor model (LFM) to assess longitudinal measurement invariance can yield misleading inferences about the level of invariance, along with biased parameter estimates. In this article we develop pattern mixture longitudinal factor models (PM-LFMs) for quantifying uncertainty in longitudinal invariance testing due to an unknown, but potentially nonignorable, dropout mechanism. PM-LFMs are a kind of multiple group model wherein observed missingness patterns define groups, LFM parameters can differ across these pattern-groups subject to identification constraints, and marginal inference about longitudinal invariance is obtained by pooling across pattern-groups. When dropout is nonignorable, we demonstrate via simulation that conventional LFMs can indicate longitudinal noninvariance, even when invariance holds in the overall population; certain PM-LFMs are shown to ameliorate this problem. On the other hand, when dropout is ignorable, PM-LFMs are shown to provide results comparable to conventional LFMs. Additionally, we contrast PM-LFMs to a latent mixture approach for accommodating nonignorable dropout—wherein missingness patterns can differ across latent groups. In an empirical example assessing longitudinal invariance of a harsh parenting construct, we employ PM-LFMs to assess sensitivity of results to assumptions about nonignorable missingness. Software implementation and recommendations for practice are discussed.

Keywords: longitudinal factor model, longitudinal invariance, nonignorable missing data, pattern mixture model

Psychologists often are interested in assessing whether repeatedly measured manifest indicators (e.g., biting and threatening) represent an underlying construct (e.g., aggression) equally well across time (e.g., Horn & McArdle, 1992; Liu et al., 2016; Pitts, West, & Tein, 1996; Widaman, Ferrer, & Conger, 2010). This kind of assessment is important because unrecognized qualitative change in the relations between measure(s) and their underlying construct can interfere with the interpretation and assessment of quantitative change and stability in the construct (e.g., Edwards & Wirth, 2009; Leite, 2007; Widaman, 1991; Wirth, 2008). A common way to assess whether measures are indeed tapping the same underlying (i.e., latent) construct across time, to the same degree, is through assessing longitudinal measurement invariance (MI) using a longitudinal factor model (e.g., Tisak & Meredith, 1991).

A review of recent psychology applications involving invariance testing using longitudinal factor models indicates that these applications often use few time points (e.g., three) that are chronologically spaced quite far apart (e.g., several years) and accompanied by high rates of dropout. For instance, approximate dropout percentages based on reported information were: Brydges, Fox, Reid, and Anderson (2014) had 41%; Fagerstrom, Lindwall, Ingeborg, and Rennemark (2012) 53%; King (2011) 27%;

Correspondence should be addressed to Sonya K. Sterba, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203. E-mail: sonya.sterba@vanderbilt.edu

Mason, Lauterbach, McKibben, Lawrence, and Fauerbach (2013) 43%; Mogos et al. (2015) 40%; Richerson, Watkins, and Beaujean (2014) 23%; Small, Hertzog, Hultsch, and Dixon (2003) 54%; Varni, Limbers, Newman, and Seid (2008) 52%; Wang, Elhai, Daic, and Yao (2012) 17%; and Wang and Su (2013) 56%. Indeed, dropout (attrition) was the most common form of missingness in these longitudinal factor modeling applications. Furthermore, these applications all assumed that the missingness mechanism was ignorable. Under an *ignorable missingness mechanism*, the process that generates the outcomes (e.g., the aggression symptoms) and the process that generates the missingness are independent or conditionally independent—conditional on observed variables in the model (Rubin, 1976).

Considered in the particular context of longitudinal invariance testing, the ignorability assumption implies, among other things, that the measurement parameters relating the outcomes to their underlying construct are the same across missingness pattern (e.g., patterns of no dropout, dropout at Time 2, or dropout at Time 3, etc.). This assumption might be plausible in some substantive contexts. However, when studying the stability of cognitive ability in older adults, for instance, it could be that those who drop out or die before the end of the study have more unstable item–construct relations (factor loadings) because of undetected, unfolding disease processes. As another example, when studying the stability of psychopathological syndromes across adolescence, it could be that more severely ill adolescents with higher symptom levels (item intercepts) are also more likely to drop out because their symptoms interfere with their ability to make appointments. Or, when studying language comprehension throughout high school, it could be that students for whom comprehension is more poorly measured (e.g., nonnative English speakers and students with inattention problems) are also more prone to drop out of high school. These scenarios are examples of nonignorable missingness mechanisms. Under a *nonignorable missingness mechanism*, the process that generated the outcomes and the process that generated the missingness are conditionally dependent and must be jointly modeled (Rubin, 1976). In the presence of a nonignorable missingness mechanism, simply fitting a conventional longitudinal factor model to the repeatedly measured outcomes to assess invariance could yield biased parameter estimates and misleading inferences about the level of invariance supported.

Unfortunately, there exists no general test of whether the missingness mechanism is ignorable versus nonignorable (e.g., Little & Rubin, 2002). Hence, methodologists have widely and increasingly suggested employing *sensitivity analyses* to quantify the impact of uncertainty about the missingness mechanism on key results (e.g., Akl et al., 2015; Jones, Mishra, & Dobson, 2015; Little et al., 2012; Mallinckrodt, Lin, & Molenberghs, 2013; Minini & Chavance, 2004; Molenberghs & Kenward, 2007; Nguyen, Lee, & Carlin, 2015; O'Kelly & Ratitch, 2014; Rombach, Rivero-Arias, Gray, Jenkinson, & Burke, 2016; Rubin, 1977; Schafer & Graham, 2002; Souza et al., 2016; Thomas, Harel, & Little, 2016; Verbeke, Molenbergs, & Beunckens, 2008). To date, such sensitivity analysis methods have mainly concerned models for *univariate* repeated outcomes, such as univariate growth models (e.g., Enders, 2011; Feldman & Rabe-Hesketh, 2012; Gottfredson, Bauer, & Baldwin, 2014; Hedeker & Gibbons, 1997; Molenberghs & Verbeke, 2001; Muthén, Asparouhov, Hunter, & Leuchter, 2011; Sterba & Gottfredson, 2015; Xu & Blozis, 2011; Yang & Maxwell, 2014). For instance, these methods have often been geared toward assessing how departures from the assumption of ignorable missingness affect estimates of the mean intercept and slope of time in univariate growth models. However, *multivariate*, rather than univariate, repeated outcomes data are used in testing longitudinal MI. Such data pose unique challenges (described later) that interfere with the extension of some of these methods from the context of univariate to multivariate repeated measures. Of the few sensitivity analysis approaches developed to quantify the impact of missingness mechanism uncertainty on multivariate longitudinal data analyses (e.g., Hafez, Moustaki, & Kuha, 2015; Sterba, 2016), none have been geared toward assessing sensitivity to results of invariance testing—a main objective of applying longitudinal factor models.

The purpose of this article is to develop and demonstrate pattern mixture (PM) models for quantifying uncertainty in longitudinal invariance testing results that is due to an unknown but potentially-nonignorable dropout mechanism. Following a brief review of the conventional longitudinal factor model and its use for longitudinal invariance testing, we describe alternative modeling approaches that could be used to allow for nonignorable dropout in this context. We introduce and motivate a family of PM models for this purpose and describe their advantages over alternative modeling approaches. Next, we employ PM longitudinal factor models (PM-LFMs) in a simulation to demonstrate that when nonignorable missingness is not investigated, it is troublingly possible to, for instance, find evidence of measurement noninvariance when in the overall population invariance holds. Subsequently, we use an example to empirically illustrate the application of PM-LFMs to quantify missing-data uncertainty in longitudinal invariance testing involving a harsh parenting construct. We also use this empirical example to explain when a researcher could consider extending PM models (which have *observed groups*, as in multiple-group models; Jöreskog, 1971) to latent missingness class models (which have *unobserved groups*, as in the conventional mixture models [McLachlan & Peel, 2000] that are discussed throughout this special issue of *Structural Equation Modeling*). Finally, we empirically contrast the performance of PM and latent missingness class approaches and draw practical implications from this contrast. In the discussion we describe recommendations, software implementation, and future directions. Note that, although both PM and latent missingness class approaches allow pattern-specific inferences or marginal (i.e., pooled across pattern) inferences about longitudinal invariance, here we focus on marginal inferences. This focus is conventional for PM applications (see Little, 1993, 1995) and is akin to an *indirect application* of mixtures (McLachlan & Peel, 2000).

REVIEW OF THE LONGITUDINAL FACTOR MODEL

We begin with a review of the conventional longitudinal factor model (LFM; e.g., Tisak & Meredith, 1991) and its use for testing MI across time points $t = 1 \ldots T$. We initially do not introduce missing data. At time point t there are J manifest indicators ($j = 1 \ldots J$) measured on N persons. The response vector for person i ($i = 1 \ldots N$) at time point t specifically is $\mathbf{y}_{it} = (y_{i1t} \ldots y_{iJt})'$. The full response vector for person i is then $\mathbf{y}_i = (\mathbf{y}'_{i1} \ldots \mathbf{y}'_{iT})'$. At time point t, the manifest indicators in $\mathbf{y}_{it}$ are represented as a linear combination of indicator-specific intercepts in $\boldsymbol{\tau}_t$ ($J \times 1$) plus p underlying latent variables (factors) in $\boldsymbol{\eta}_{it}$ ($p \times 1$) that are weighted by indicator-specific slopes (factor loadings) in $\boldsymbol{\Lambda}_t$ ($J \times p$), plus indicator-specific residuals in $\boldsymbol{\varepsilon}_{it}$ ($J \times 1$). For all time points taken together, this yields:

$$\mathbf{y}_i = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$$

where

$$\begin{bmatrix} \mathbf{y}_{i1} \\ \vdots \\ \mathbf{y}_{iT} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\tau}_1 \\ \vdots \\ \boldsymbol{\tau}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_T \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_{i1} \\ \vdots \\ \boldsymbol{\eta}_{iT} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{i1} \\ \vdots \\ \boldsymbol{\varepsilon}_{iT} \end{bmatrix} \tag{1}$$

The factor(s) at time t, $\boldsymbol{\eta}_{it}$, have means in $\boldsymbol{\mu}_t$ ($p \times 1$) and individual-specific deviations from the mean in $\boldsymbol{\zeta}_{it}$ ($p \times 1$) . For all time points taken together this yields:

$$\boldsymbol{\eta}_i = \boldsymbol{\mu} + \boldsymbol{\zeta}_i$$

where

$$\begin{bmatrix} \boldsymbol{\eta}_{i1} \\ \vdots \\ \boldsymbol{\eta}_{iT} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_T \end{bmatrix} + \begin{bmatrix} \zeta_{i1} \\ \vdots \\ \zeta_{iT} \end{bmatrix} \tag{2}$$

Individual-specific deviations are typically assumed normally distributed and are allowed to covary within and across time (such that $\boldsymbol{\Psi}$ is unstructured), as follows:

$$\zeta_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$$

where

$$\begin{bmatrix} \zeta_{i1} \\ \vdots \\ \zeta_{iT} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Psi}_{1,1} & \cdots & \boldsymbol{\Psi}_{1,T} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{T,1} & \cdots & \boldsymbol{\Psi}_{T,T} \end{bmatrix} \right) \tag{3}$$

Indicator-specific residuals at time point t, $\boldsymbol{\varepsilon}_{it}$, are typically assumed normally distributed and are conventionally not allowed to covary *within time* ($\boldsymbol{\Theta}_{t,t}$ is diagonal). If additionally indicator-specific residuals are not allowed to covary *across time*, then $\boldsymbol{\Theta}$ is also diagonal. The latter assumption can be relaxed if, for instance, the residuals for the jth indicator at consecutive time points are theoretically anticipated to covary due to shared method variance (i.e., $\text{cov}(\varepsilon_{ijt}, \varepsilon_{ij(t+1)}) \neq 0$; e.g., Pitts et al., 1996). For all time points taken together, this yields:

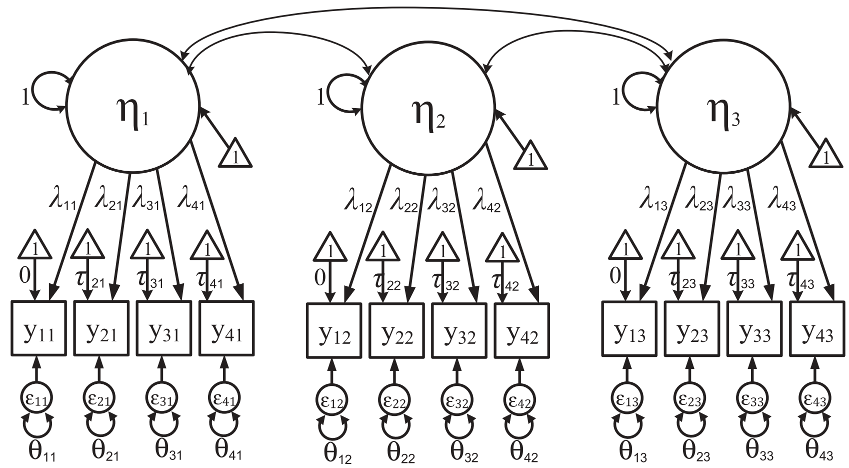
$$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Theta})$$

where

$$\begin{bmatrix} \boldsymbol{\varepsilon}_{i1} \\ \vdots \\ \boldsymbol{\varepsilon}_{iT} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Theta}_{1,1} & \cdots & \boldsymbol{\Theta}_{1,T} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Theta}_{T,1} & \cdots & \boldsymbol{\Theta}_{T,T} \end{bmatrix} \right) \tag{4}$$

In the foregoing, for simplicity we consider a typical scenario where there is only $p = 1$ latent factor at each time point (thus replacing $\boldsymbol{\eta}_{it}$, $\boldsymbol{\mu}_t$, $\boldsymbol{\zeta}_{it}$, and $\boldsymbol{\Psi}_{t,t}$ with η_{it}, μ_t, ζ_{it}, and $\psi_{t,t}$). This LFM is diagrammed in Figure 1. In general, the number of latent factors at each time point would be determined by substantive theory in a fully confirmatory approach or could be informed by model fit comparisons involving alternate p. The extension to multiple latent factors at each time point is not integral to the present developments.

For model identification, we must set the location and scale of each latent factor. Here this is accomplished by constraining the intercept of one referent indicator (here, the first indicator) at each time point to 0 (i.e.,τ_{1t}= 0) and constraining the variance of the latent factor at each time point to 1 ($\psi_{t,t}$= 1). See Vandenberg and Lance (2000) for implications of referent indicator choice. Whereas these constraints alone are sufficient to allow fitting the model and obtaining a unique solution for parameter estimates, they do not guarantee that we are measuring the same construct across time.



Where:

Weak longitudinal invariance assuming ignorable dropout requires: $\lambda_{jt} = \lambda_j$, for all j

Strong longitudinal invariance assuming ignorable dropout also requires: $\tau_{jt} = \tau_j$, for all j

Strict longitudinal invariance assuming ignorable dropout also requires: $\theta_{jt} = \theta_j$, for all j

FIGURE 1 Conventional longitudinal factor model (LFM), assuming ignorable dropout. *Note*. Shown with $p = 1$ and $J = 4$ and $T = 3$ for illustration. Indicator-specific residual covariances (if present) are not shown.

REVIEW OF ASSESSING LONGITUDINAL MEASUREMENT INVARIANCE

To investigate the extent to which we are indeed measuring the same construct over time we can assess longitudinal MI, which typically proceeds in the following steps (e.g., Meredith, 1993; Meredith & Horn, 2001; Widaman et al., 2010). First we can assess whether the LFM with the same J manifest items serving as indicators of the same latent factor(s) at each time point has adequate fit (configurally invariant LFM). If so, we can then assess whether the indicator–construct relationships are constant across time. We can do so by testing whether the factor loading for each indicator is equal across time (i.e., $\boldsymbol{\Lambda}_1 = \ldots = \boldsymbol{\Lambda}_T$; called testing *weak invariance*). Specifically, we compare the more invariant model (imposing this equality constraint) to a less invariant model (freeing this equality constraint) using model selection indexes such as information criteria (e.g., Bayesian information criterion) or a χ^2 difference test. If model selection indexes prefer the more constrained model (i.e., fit does not meaningfully deteriorate relative to the number of parameter constraints imposed), we can proceed to successively test across-time equality constraints on other types of measurement parameters: indicator-specific intercepts (i.e., $\boldsymbol{\tau}_1 = \ldots = \boldsymbol{\tau}_T$, called testing *strong invariance*), and then indicator-specific residual variances ($Diag(\boldsymbol{\Theta}_{1,1}) = \ldots = Diag(\boldsymbol{\Theta}_{T,T})$, called testing *strict invariance*). See Figure 1 for a summary of these consecutive steps.

If at least strong MI holds, change or stability in structural parameters (e.g., factor means and covariances) can be unambiguously interpreted as capturing quantitative change or stability in the construct rather than qualitative change in the meaning of the construct over time. If model selection indexes indicate that a given set of across-time constraints is not warranted, one convention is to stop MI testing and report that invariance at the previous level of testing was achieved. Another approach is to examine whether there is support for imposing further levels of MI, but only for a subset of manifest indicators (called partial invariance; see Byrne, Shavelson, & Muthén, 1989; Vandenberg & Lance, 2000; Edwards & Wirth, 2009).

APPROACHES FOR JOINTLY MODELING OUTCOME AND MISSINGNESS PROCESSES IN THE CONTEXT OF LONGITUDINAL FACTOR ANALYSIS

Presently we focus on missingness due to dropout, not only because it was by far the most common kind of missingness found in our review of LFM applications, but also because it is often considered to pose the greatest risk of nonignorable missingness (e.g., Hedeker & Gibbons, 1997; Muthén et al., 2011). Intermittent missingness, if present, would still be assumed ignorable so long as conventional full information maximum likelihood (FIML) estimation methods are used for model fitting. Also, in subsequent examples we consider the scenario where there is no dropout at Time 1, as that scenario was typical of applications in our literature review.

Designate a $(T-1) \times 1$ vector $\mathbf{m}_i$ of dropout indicators where element $m_{it} = 1$ if person i dropped out at time t, $m_{it} = 0$ if person i has not yet dropped out, and $m_{it} = .$ if person i dropped out at a previous time point. As mentioned earlier, nonignorable dropout implies conditional dependence between $\mathbf{y}_i$ and $\mathbf{m}_i$ that would need to be modeled to prevent bias in outcome-model parameter estimates and inferences. Three general approaches for jointly modeling outcome and dropout processes are selection models, random coefficient selection models, and PM models. To date, these approaches have been used mainly in the context of univariate repeated measures, and there are obstacles to extending some to multivariate repeated measures. *Selection models* accommodate dependence between $\mathbf{y}_i$ and $\mathbf{m}_i$ by allowing dropout indicators to directly depend on values of the repeated outcomes (e.g., regressing m_{it} on $\mathbf{y}_{it}$ and on $\mathbf{y}_{i(t-1)}$). However, selection models would entail impractical computational burden in the context of multivariate repeated measures with sizable J, T, or both. Specifically, this approach would require $J \times (T-1)$ dimensions of integration (because obtaining the joint likelihood of outcomes and missingness indicators for such a model requires integration over missing $\mathbf{y}_i$ variables). For instance, eight dimensions of integration would be required for the situation in Figure 1. In contrast, this approach can be practical for univariate repeated measure settings, where $J = 1$ (e.g., Diggle & Kenward, 1994). A variation on a selection model called a *random coefficient selection model* (e.g., Wu & Carroll, 1988) is more computationally tractable for multivariate repeated measures because it instead accommodates conditional dependence between $\mathbf{y}_i$ and $\mathbf{m}_i$ by allowing dropout indicators to depend on the latent factors (e.g., regressing m_{it} on η_{it} and on $\eta_{i(t-1)}$; e.g., Hafez et al., 2015). This approach would instead require $2 \times (T-1)$ dimensions of integration when $p = 1$.[1]

An alternative approach, the PM[2] model (e.g., Little, 1993, 1995; Thijs, Molenberghs, Michiels, Verbeke, & Curran, 2002), is a kind of manifest multiple group model that accommodates conditional dependence between $\mathbf{y}_i$ and $\mathbf{m}_i$ by allowing parameters governing the distribution of $\mathbf{y}_i$ to vary across observed missingness patterns (here, dropout patterns) implied by $\mathbf{m}_i$. Let d_i denote a manifest grouping variable that can take on values $g = 1 \ldots G$, for G dropout patterns. For instance, $G = 3$ dropout patterns might be: complete case, dropout at Time 3, dropout at Time 2. The marginal distribution of $\mathbf{y}_i$, denoted $f(\mathbf{y}_i)$, is a weighted average of pattern-specific distributions:

$$f(\mathbf{y}_i) = \sum_{g=1}^{G} f(\mathbf{y}_i | d_i = g) p(d_i = g) \tag{5}$$

Here $p(d_i = g)$ is the observed probability of membership in the gth dropout pattern. Previously PMs had been applied mainly in the context where $f(\mathbf{y}_i | d_i = g)$ is a multivariate normal density implied by a growth model specific to group g (e.g., Enders, 2010, 2011; Hedeker & Gibbons, 1997; Muthén et al., 2011; Yang & Maxwell, 2014). Here, $f(\mathbf{y}_i | d_i = g)$ is a multivariate normal density implied by an LFM specific to group g (potentially with group-specific parameters; e.g., $\mathbf{\Lambda}^{(g)}$, $\boldsymbol{\tau}^{(g)}$, $\mathbf{\Theta}^{(g)}$, $\boldsymbol{\mu}^{(g)}$, $\mathbf{\Psi}^{(g)}$, as discussed later). Hence we refer to this as a PM-LFM. An illustrative PM-LFM is diagrammed in Figure 2.

PM-LFMs afford pragmatic advantages for investigating nonignorable dropout in the context of longitudinal MI testing. Unlike selection and random coefficient selection models, PM-LFMs are not computationally burdensome, as they do not require numerical integration. PM-LFMs also readily allow investigation of the sensitivity of longitudinal invariance results to manipulation of missingness assumptions about each kind of measurement parameter separately (procedures discussed next). In contrast, random coefficient selection models assume that the dependency between missingness and outcomes arises solely from the structural (latent variable) level, rather than in the measurement submodel.

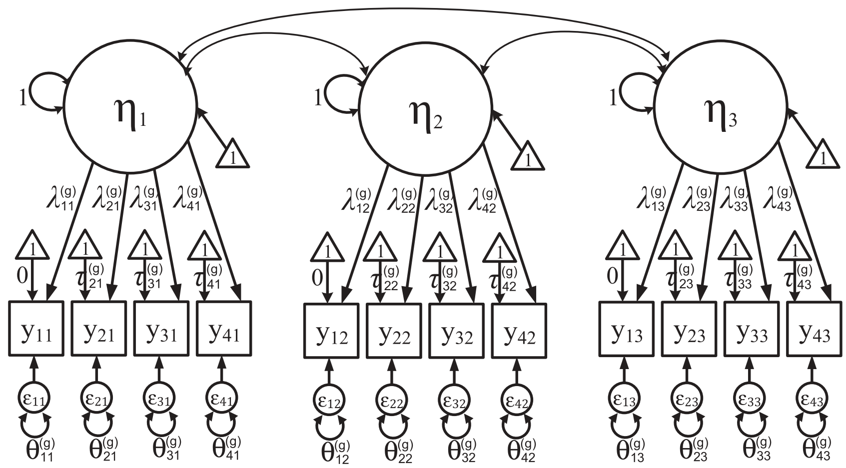
PATTERN MIXTURE MODELS FOR INVESTIGATING THE IMPACT OF MISSINGNESS MECHANISM UNCERTAINTY IN LONGITUDINAL MI TESTING

Nonignorable missingness corresponds with any parameter (measurement or structural) differing across dropout pattern in the PM-LFM. Hence, ignorable missingness corresponds with all parameters being held equal across dropout pattern.[3] PM-LFMs can be used to investigate and pinpoint sensitivity to departures from ignorable missingness with respect to each kind of parameter. Of particular interest in this context is investigating sensitivity to departures from ignorable missingness with respect to loadings (Is $\mathbf{\Lambda}^{(1)} = \ldots = \mathbf{\Lambda}^{(G)}$?), item intercepts (Is $\boldsymbol{\tau}^{(1)} = \ldots = \boldsymbol{\tau}^{(G)}$?), or residual variances (Is $\mathbf{\Theta}^{(1)} = \ldots = \mathbf{\Theta}^{(G)}$?), but similar investigations could be done for structural parameters as well. Most important, with respect to our goal of longitudinal MI testing, we can examine whether marginal (aggregated across dropout pattern) results of each

[1] Note that Hafez et al.'s (2015) model regresses m_{it} only on $\eta_{i(t-1)}$ and involves $p = 1$, thus requiring $T-1$ dimensions of integration.

[2] Although we use the term *pattern mixture model* here to be consistent with prior literature, note that the groups in this model are observed, rather than latent, unlike the conventional mixture models discussed later in this article and elsewhere in this special issue.

[3] Imposing this equality constraint parallels the approach of Allison (1987), which has been employed previously in other modeling contexts.



Where, marginally (pooling across missingness-pattern):

Weak longitudinal invariance allowing for nonignorable dropout requires:

$$\sum_{g=1}^{G} \lambda_{jt}^{(g)} p(d_i = g) = \sum_{g=1}^{G} \lambda_{j}^{(g)} p(d_i = g), \text{ for all } j$$

Strong longitudinal invariance allowing for nonignorable dropout also requires:

$$\sum_{g=1}^{G} \tau_{jt}^{(g)} p(d_i = g) = \sum_{g=1}^{G} \tau_{j}^{(g)} p(d_i = g), \text{ for all } j$$

Strict longitudinal invariance allowing for nonignorable dropout also requires:

$$\sum_{g=1}^{G} \theta_{jt}^{(g)} p(d_i = g) = \sum_{g=1}^{G} \theta_{j}^{(g)} p(d_i = g), \text{ for all } j$$

FIGURE 2 Pattern-mixture longitudinal factor model (PM-LFM), allowing for nonignorable dropout. *Note*. Shown with $p = 1$, $J = 4$, and $T = 3$ for illustration. g = dropout pattern ($g = 1 \ldots G$). Indicator-specific residual covariances (if present) are not shown. Structural parameters could be allowed to differ across dropout pattern, although this is not shown.

level of testing (e.g., weak, strong, strict) are sensitive to departures from ignorable missingness. In other words, we can see whether testing results for weak, strong, and strict MI differ under the assumption of ignorable versus nonignorable missingness. Allowing for nonignorable missingness, we test each level of longitudinal MI of marginal parameters (aggregated across dropout pattern) using PM-LFM by testing the following equalities consecutively.

Weak longitudinal invariance allowing for nonignorable dropout requires

$$\sum_{g=1}^{G} \mathbf{\Lambda}_1^{(g)} p(d_i = g) = \ldots = \sum_{g=1}^{G} \mathbf{\Lambda}_T^{(g)} p(d_i = g).$$

Strong longitudinal invariance allowing for nonignorable dropout also requires

$$\sum_{g=1}^{G} \boldsymbol{\tau}_1^{(g)} p(d_i = g) = \ldots = \sum_{g=1}^{G} \boldsymbol{\tau}_T^{(g)} p(d_i = g).$$

Strict longitudinal invariance allowing for nonignorable dropout also requires:

$$\sum_{g=1}^{G} Diag(\mathbf{\Theta}_{1,1}^{(g)}) p(d_i = g) = \ldots$$
$$= \sum_{g=1}^{G} Diag(\mathbf{\Theta}_{T,T}^{(g)}) p(d_i = g).$$

Standard errors for the marginal estimates (aggregated across pattern) can be obtained via the delta method (e.g., Little, 1993, 1995; Raykov & Marcoulides, 2004). As illustrated in the online Appendix[4] syntax, this can be accomplished using the *model constraint* command option in M*plus* (Muthén & Muthén, 1998–2016). Although our focus in this article is on marginal parameters, PM-LFMs also allow researchers the opportunity to inspect pattern-specific parameter estimate results. Pattern-specific results may be useful for particular substantive purposes to understand how different risk factors and processes might manifest across pattern groups.[5]

Number of Patterns, G, in the PM-LFM

PM models require a small enough number of dropout patterns G to support estimation of pattern-specific parameters. Thus, PM models are well suited to LFM applications because in our review these applications had few time points T (typically three), and thus few dropout patterns. In other modeling contexts where T might be larger (e.g., univariate growth modeling applications) or where intermittent missingness is considered nonignorable, methodologists have suggested creating summary pattern indicators, which entails the assumption that persons within summary pattern are interchangeable (see Enders, 2011; Gottfredson et al., 2014; Hedeker & Gibbons, 1997; Rose, von Davier, & Xu, 2010; Roy, 2007).

Identification Considerations for the PM-LFM: Choosing Identifying Constraints

Unlike selection models and random coefficient selection models, PM-LFMs do not require the researcher to provide a parametric model specification relating the outcome-generating process and missingness-generating process. For selection models, these parametric specifications typically include posited logistic regressions relating $\mathbf{m}_i$ and $\mathbf{y}_i$, and for random coefficient selection models typically include posited logistic regressions relating $\mathbf{m}_i$ and $\boldsymbol{\eta}_i$. Such specifications are known to lack robustness to small departures from their parametric assumptions (for reviews, see Beunckens, Molenberghs, Thijs, & Verbeke, 2007; Sterba & Gottfredson, 2015; Verbeke et al., 2008) and it is difficult to systematically manipulate these assumptions. Fundamental departures from these parametric assumptions might be impossible to detect because they rest on data and processes that are unobserved. Without such parametric assumptions, however, these models would be inestimable.

Although PM-LFMs do not require explicit specification of parametric relationships between $\mathbf{m}_i$ and $\mathbf{y}_i$, they do require the explicit imposition of constraints (so-called *identifying restrictions*; Little, 1993, 1995) to enable estimating particular pattern-group-specific parameters. For instance, without imposing some constraints on the model it would not be possible to estimate measurement parameters at Time 3 for the pattern-group that dropped out at Time 2. There are different options for identifying constraints, and implementing these different options could be conceptualized as implementing slightly different assumptions about the missingness process. Hence, a compelling strength of the PM modeling framework is that it is straightforward to define a range of different missingness assumptions, translate them into model constraints, and manipulate them in the context of a sensitivity analysis. This makes the missingness assumptions investigated in a PM model sensitivity analysis more transparent and explicitly operationalized than for selection models and random-coefficient selection models (Enders, 2010, 2011; Little, 1993, 1995).

There are many possible identifying constraints that could be employed in our PM-LFM. Because these identifying constraints impose different assumptions about the missingness process, none can be ruled out using fit to the observed data; in fact, they will have identical χ^2-based fit indexes. Parameter estimates will be most accurate for the identifying constraint that is most consistent with the population outcome-generating process and missingness-generating process (e.g., Demirtas & Schafer, 2003; Yang & Maxwell, 2014). An often-recommended approach is to consider several alternative identifying constraints. Constraints might be chosen because they, for instance, (a) reflect substantive theory about the missingness process, and/or (b) have been conventionally used in PM versions of other fitted models. We describe five illustrative identifying constraints that allow for nonignorable missingness. We contrast these with a sixth constraint, which assumes ignorable missingness. Here $\boldsymbol{\varphi}_t$ generically represents a vector of all estimated parameters at time t.

1. *Nearest neighbor (NN) identifying constraint*. Constrain inestimable parameters at time t in pattern g to their values in the most similar (i.e., nearest neighbor) pattern h where those parameters are estimable; that is, $\boldsymbol{\varphi}_t^{(g)} = \boldsymbol{\varphi}_t^{(h)}$. (This is an across-group within-time restriction.)
2. *Complete case (CC) identifying constraint*. Constrain inestimable parameters at time t in pattern g to their values in the pattern with complete case data, here designated as pattern 1: $\boldsymbol{\varphi}_t^{(g)} = \boldsymbol{\varphi}_t^{(1)}$. (This is an across-group within-time restriction.)
3. *Available case (AC) identifying constraint*. Constrain inestimable parameters at time t in pattern g to the weighted average of their values in all other patterns

[4] The online Appendix is available at http://www.vanderbilt.edu/peabody/sterba/.

[5] Many recent applications of invariance testing using LFM continue to analyze only complete-case data despite the fact that the FIML estimation methods used did not require this (e.g., Brydges et al., 2014; King, 2011; Mason et al., 2013; Mogos et al., 2015; Richerson et al., 2014; Varni et al., 2008; Wang et al., 2012; Wang & Su, 2013). This is akin to looking at results for only one specific missingness pattern-group and, as such, these authors would be unaware of whether results differ across missingness pattern-group.

where they are estimable at time t. The AC constraint considered here corresponds with the use of the term in Demirtas and Schafer (2003) and Enders (2011). (This is an across-group within-time restriction.)

4. *Last observation carried forward (LOCF) identifying constraint*. Constrain inestimable parameters at time t in pattern g to their values in the same pattern g at the previous time point, where estimable: $\boldsymbol{\varphi}_t^{(g)} = \boldsymbol{\varphi}_{t-1}^{(g)}$. (This is a within-group across-time restriction.) The LOCF constraint considered here corresponds with the use of the term in Little (2008).
5. *Nearest neighbor difference (NND) identifying constraint*. Constrain inestimable parameters at time t in pattern g to $\boldsymbol{\varphi}_t^{(g)} = \boldsymbol{\varphi}_{t-1}^{(g)} - (\boldsymbol{\varphi}_{t-1}^{(h)} - \boldsymbol{\varphi}_t^{(h)})$ where h is the nearest neighbor pattern. Note that if the parameter estimates are the same at the previous two time points in pattern h this simplifies to LOCF. Note that this is a variation on an identification constraint in Little (2008), tailored to the current context.
6. *Ignorable missingness constraint*. Constrain all parameters at time t equal across G patterns $\boldsymbol{\varphi}_t^{(g)} = \boldsymbol{\varphi}_t$. Note that this constraint is more restrictive than what is minimally needed for identification and thus will yield different model fit than the first five constraints, which all allow nonignorable missingness. (This is an across-group within-time restriction.)

By inspecting this illustrative set of constraint options, we can see that it is not possible to simultaneously allow for full measurement noninvariance across missingness pattern-group and allow measurement noninvariance across time. For identification purposes, we either have to impose some degree of parameter invariance within time across missingness pattern-group (NN, CC, AC), which in exchange affords us greater ability to test longitudinal invariance across time, or we have to impose some degree of parameter invariance across time within missingness pattern (LOCF, NND), which in exchange affords us greater ability to test invariance across missingness pattern-group.

Sensitivity Analysis Strategies for the PM-LFM

One sensitivity analysis strategy is to (a) specify multiple PM-LFMs under alternative identifying constraints, (b) compute marginal (aggregated across pattern) parameter estimates for each PM-LFM, and (c) test longitudinal MI of marginal parameters using each PM-LFM (see Figure 2). Then it can be reported whether parameter estimates and MI testing results differ among nonignorable PM-LFMs and whether any and all results differ from fitting a conventional LFM (that assumes ignorable missingness; see Figure 1).

Currently, researchers fitting a conventional LFM using FIML estimation assume ignorable missingness as a default approach, often without contemplating theoretically why it would be plausible. As noted earlier, the ignorability assumption in a conventional LFM is equivalent to a special case constraint on the PM-LFM (i.e., the last constraint earlier) that is more restrictive than the constraints needed to identify a PM-LFM allowing for nonignorable missingness. Hence, this ignorability constraint does imply a particular missingness generation theory, which might or might not be valid, just as the first five constraints imply different missingness generation theories. Thus, the ignorability constraint can be considered just one of many possible constraints that could be imposed. Our perspective is that researchers should either justify theoretically *why* only this particular ignorability constraint is plausible, rather than automatically adopting it as a default, or be compelled to also consider alternative constraints.

SIMULATION DEMONSTRATION

In this section we use a simulation demonstration to illustrate the benefits of sensitivity analysis using PM-LFM to quantify missingness-mechanism uncertainty. Specifically, we illustrate four main points.

1. If a conventional LFM is fit in the presence of *nonignorable* missingness, it is possible that non-MI across missingness patterns can masquerade, marginally, as non-MI across time.[6]
2. If PM-LFMs are fit unnecessarily in the presence of *ignorable* missingness, non-MI across time can still be correctly detected.
3. If PM-LFMs are fit unnecessarily in the presence of *ignorable* missingness, MI across time can still be correctly supported.
4. The ability of PM-LFMs to improve on the performance of the conventional LFM in the presence of *nonignorable* missingness depends on the realism of their identifying constraints.

In our simulation demonstration there are four different generating conditions, described later. Points 1 and 4 will be illustrated using generating Conditions 1 and 4 whereas points 2 and 3 will be illustrated using generating Conditions 2 and 3.

Simulation Design

We focus our invariance testing illustration on testing weak MI (i.e., of factor loadings); however, our same points could be illustrated by testing a different level of MI, or

[6] Although not demonstrated here, when fitting a conventional LFM in the presence of nonignorable missingness, it would also be possible for non-MI across missingness group to masquerade marginally as MI across time.

testing multiple levels of MI, or even testing structural invariance. The test of weak invariance is the most commonly conducted test in empirical applications investigating MI, according to Vandenberg and Lance (2000). In this simulation we do not include observed covariates because longitudinal MI testing applications in our review almost always used unconditional LFMs. We use a T and J typical of applications in our literature review (T = 3, J = 4). We use a baseline N near the median with respect to our literature review (N = 750). Although, in SEM applications more generally, small sample sizes are common (e.g., <N = 150; MacCallum & Austin, 2000), we found that LFM applications typically involved secondary data analysis of large data sets (e.g., N = 1,402 [Fagerstrom et al., 2012], N = 730 [Wu, 2015], N = 5,991 [Varni et al., 2008], N = 560 [Wang et al., 2012], N = 299 [Mason et al., 2013], N = 2,001 [Mogos et al., 2015], N = 484 [Small et al., 2003], N = 24,599 [Wang & Su, 2013], N = 228 [Brydges et al., 2014], N = 457 [Richerson et al., 2014], N = 2,022 [King, 2011]). Furthermore, we used dropout patterns of 64.5% complete case, 17.7% dropout at Time 3, and 17.7% dropout at Time 2; this degree of dropout was conservative with respect to our literature review (as described earlier).

Four alternative generating LFMs were considered that differed in their generating factor loading values (Conditions 1–4 next). In each generating model, item residual variances were chosen in concert with their respective loading values to ensure that each item had a total variance of 1.0. For simplicity, the models do not include residual covariances within item across time; however, the empirical example presented later does include such residual covariances for illustration purposes. In each generating model, there was longitudinal MI of item intercepts $\tau_{1t} = 0, \tau_{2t}=\tau_{3t}=\tau_{4t}=1$ and there were increasing factor means over time $\mu_1=.5, \mu_2=1, \mu_3=1.5$. Additionally, in each generating model $\psi_{1,1} = \psi_{2,2} = \psi_{3,3} = 1$ and $\psi_{1,2}=.35, \psi_{2,3}=.50, \psi_{1,3}=.25$. Allowing more parameters to differ across missingness pattern in the generating model would simply amount to greater departures from ignorable missingness and thus greater potential for misleading test results from a fitted conventional LFM. The four generating conditions differed in their factor loadings in the following manner.

- Condition 1: Loadings were MI across time marginally but non-MI across missingness pattern.
- Condition 2: Loadings were non-MI across time marginally but MI across missingness pattern.
- Condition 3: Loadings were MI across time marginally and MI across missingness pattern.
- Condition 4: Loadings were MI across time marginally but non-MI across missingness pattern. Additionally, loadings were non-MI across time within missingness pattern.

The generating loading values from these four conditions are given in Table 1. Conditions 1 and 4 entail nonignorable missingness mechanisms. Conditions 2 and 3 entail ignorable missingness mechanisms.

Five hundred data sets were generated conforming to each condition; data generation was done in M*plus* 7.4 and SAS 9.4. Each data set was analyzed with a pair of alternative specifications of each of six fitted models. The pair of alternative specifications either relaxed weak MI (i.e., marginal loadings freely estimated across time) or imposed weak MI (i.e., marginal loadings constrained equal across time). The six fitted models were: a conventional LFM (i.e., constraint (6) from earlier) and five PM-LFMs (imposing each of the constraints (1)-(5) from earlier). For each fitted model, the fit of the MI versus non-MI specifications were compared using a likelihood ratio test (LRT). Data analysis used M*plus* 7.4.

Simulation Results

The proportion of samples where weak MI was rejected is given in Table 2. Columns 1 through 4 of Table 2 provide results from generating Conditions 1 through 4, respectively. Rows of Table 2 provide results for each fitted model type. Due to the different generating conditions, results in Columns 1, 3, and 4 are Type I error rates, whereas results in Column 2 are power. The first point illustrated by Table 2 is that when missingness is *nonignorable*, fitting a conventional LFM can lead to detecting non-MI across time when in fact MI holds across time; this happened 20% of the time in Condition 1 (bottom of Column 1) and 71% of the time in Condition 4 (bottom of Column 4). The second point illustrated by Table 2 is that when missingness is *ignorable*, unnecessarily fitting a PM-LFM can correctly detect non-MI as often as does the conventional LFM; this happened 100% of the time in Condition 2 (see Column 2). The third point illustrated by Table 2 is that when missingness is *ignorable*, unnecessarily fitting a PM-LFM can correctly detect MI approximately as often as does the conventional LFM; this happened 93% to 96% of the time in Condition 3 (see Column 3). The last point illustrated by Table 2 is that—consistent with Demirtas and Schafer (2003), Enders (2011), and Yang and Maxwell (2014)—when missingness is *nonignorable*, fitting a PM-LFM can substantially improve on the performance of an LFM for invariance testing if its identifying constraints are more reflective of the generating process; however, a PM-LFM can perform similarly or slightly worse than a conventional LFM if its identifying constraints are not reflective of the generating process. For instance, in Condition 1, compared to the 20% Type I error rate for LFM, there was a Type I error rate of 5% for PM-LFM with NND constraints but a Type I error rate of 29% for PM-LFM with AC constraints. Likewise, in Condition 4, compared to the 71% Type I error rate for

TABLE 1
Simulation Design

	Condition 1: (Nonignorable Missingness) MI Across Time Marginally Non-MI Across Missingness Pattern			*Condition 2: (Ignorable Missingness) Non-MI Across Time Marginally MI Across Missingness Pattern*			*Condition 3: (Ignorable Missingness) MI Across Time Marginally MI Across Missingness Pattern*			*Condition 4: (Nonignorable Missingness) MI Across Time Marginally Non-MI Across Missingness Pattern Non-MI Across Time Within Pattern*		
	Time 1 $\lambda_{11}-\lambda_{41}$	*Time 2* $\lambda_{12}-\lambda_{42}$	*Time 3* $\lambda_{13}-\lambda_{43}$	*Time 1* $\lambda_{11}-\lambda_{41}$	*Time 2* $\lambda_{12}-\lambda_{42}$	*Time 3* $\lambda_{13}-\lambda_{43}$	*Time 1* $\lambda_{11}-\lambda_{41}$	*Time 2* $\lambda_{12}-\lambda_{42}$	*Time 3* $\lambda_{13}-\lambda_{43}$	*Time 1* $\lambda_{11}-\lambda_{41}$	*Time 2* $\lambda_{12}-\lambda_{42}$	*Time 3* $\lambda_{13}-\lambda_{43}$
Whole-population marginal loadings (aggregating across missingness pattern)	.72	.72	.72	.80	.65	.50	.70	.70	.70	.41	.41	.41
Loadings for Pattern 1 (no dropout)	.80	.80	.80	.80	.65	.50	.70	.70	.70	.20	.30	.41
Loadings for Pattern 2 (dropout at Time 3)	.65	.65	.65	.80	.65	.50	.70	.70	.70	.90	.62	.41
Loadings for Pattern 3 (dropout at Time 2)	.50	.50	.50	.80	.65	.50	.70	.70	.70	.70	.62	.41

Note. MI = measurement invariance (here, weak); non-MI = nonmeasurement invariance (here, weak).

TABLE 2
Simulation Results: Proportion of Samples out of 500 Where MI Rejected

	Condition 1: Nonignorable Missingness, Marginal MI (Type I Error Rate Shown)	*Condition 2: Ignorable Missingness, Marginal Non-MI (Power Shown)*	*Condition 3: Ignorable Missingness, Marginal MI (Type I Error Rate Shown)*	*Condition 4: Nonignorable Missingness, Marginal MI (Type I Error Rate Shown)*
Pattern mixture-LFM nearest neighbor constraint	.284	1.00	.046	.078
Pattern mixture-LFM complete case constraint	.314	1.00	.050	.167
Pattern mixture-LFM available case constraint	.292	1.00	.074	.173
Pattern mixture-LFM last observation carried forward constraint	.058	1.00	.054	.562
Pattern mixture-LFM Nearest neighbor difference constraint	.046	1.00	.042	.430
Conventional LFM	.202	1.00	.042	.714

Note. MI = measurement invariant (here, weak); LFM = longitudinal factor model. Conventional LFM assumes ignorable missingness. Pattern mixture-LFM allows nonignorable missingness.

LFM, there was a Type I error rate of 8% for PM-LFM with NN constraints but a Type I error rate of 56% for PM-LFM with LOCF constraints.

Additionally, note that results for parameter estimate bias mirror these testing results. That is, when missingness was ignorable (Conditions 2 and 3), percent absolute relative bias of parameter estimates was largely trivial (e.g., 1% or less) regardless of whether a conventional LFM or any PM-LFM were fit. When missingness was nonignorable (Conditions 1 and 4), percent absolute relative bias was trivial when fitting a PM-LFM with identifying constraints that were more realistic, but could be similarly nontrivial (e.g., > 10%) when fitting either a conventional LFM or a PM-LFM with identifying constraints that were not reflective of the missingness process. (Percent absolute relative bias for parameter estimates when fitting the non-MI model are provided in the Appendix for each of the six fitted models, in each of the four generating conditions.)

Implications of the Simulation

In practice, researchers would not know definitively if their missingness was ignorable or nonignorable. It is reassuring that, in the presence of ignorable missingness, unnecessarily fitting PM-LFMs did not provide misleading conclusions with respect to longitudinal invariance testing. It is also important to know that, in the presence of nonignorable missingness, fitting PM-LFMs can substantially improve on a conventional LFM when the identifying constraints imposed by the PM-LFM are more reflective of the generating process, but can also do slightly worse than the conventional LFM if these identifying constraints are least realistic. This simulation illustration underscores the utility of considering several PM identification constraints; if results are similar regardless of choice of identifying constraint (as in our later empirical example), we can be confident of the insensitivity of results to departures from the ignorability assumption. This simulation demonstration also underscores the utility of substantively informing the choice of identification constraints. Identifying constraints need to be imposed mainly at later time points because that is where more patterns have missing data. Hence, differences in parameter estimates across pattern groups at earlier time points can suggest and inform substantive decisions about what constraints to impose at later time points.

EMPIRICAL EXAMPLE

In this section we illustrate an empirical investigation of the sensitivity of MI testing to departures from the ignorable missingness assumption. Our substantive interest lies in assessing MI of a harsh or poor parenting construct across early childhood. This illustration uses observationally coded parenting quality data from the 14-, 24-, and 36-month home visits of the Early Head Start Research and Evaluation Study (EHSRES; Administration for Children and Families, 2002). At each visit the focal parent (here, mother) and child were observed in a semi-structured videotaped play session and aspects of parenting quality were rated using 7-point scales by teams of coders trained to a criterion level of 85% agreement. Here we consider three manifest indicators of harsh parenting:

unsupportive behavior, detached behavior, and negative regard. Unsupportive behavior includes lack of responsivity to the child's signals or needs; negative regard refers to expressing disapproval or rejection of the child; and detachment refers to being perfunctory or indifferent toward the child.

Data from 2,382 families[7] were present for at least one of the three time points. The most common type of missingness was dropout; 26% of families dropped out—12% dropped out at Time 2 and 14% dropped out at Time 3. Although there were small amounts of intermittent missingness, presently we assume these to be ignorable; however, this assumption is revisited later. Prior studies have expressed concern over the implications of dropout in the EHSRES (e.g., Carlson, 2009) and suggested that parents who were less engaged and more distracted during study visits were more likely to drop out (Roggman, Cook, Peterson, & Raikes, 2008). It is also possible that these parents who drop out are more self-conscious and distracted by being videotaped interacting with their child and so exhibit less naturalistic behavior during this task. This could imply that they exhibit weaker loadings relating parenting quality indicators to the underlying construct, compared to parents who do not drop out. Hence, here we investigate potential sensitivity of MI testing results across dropout pattern using an unconditional LFM with three indicators at each time point. Residuals were allowed to covary within item across consecutive time points.

First a configurally invariant conventional LFM was fit assuming ignorable missingness. This model had adequate to fair fit (comparative fit index [CFI] = .959, standardized root mean square residual [SRMR] = .039, root mean square error of approximation [RMSEA] = .064, CI [.056, .072]). Weak longitudinal invariance was then imposed on this conventional LFM; an LRT rejected the null hypothesis of loading invariance, $\chi^2(6) = 196.7$, $p < .05$, and the CFI (.914), SRMR (.119), and RMSEA (.080) no longer indicated adequate fit. Hence, further invariance testing was not warranted in this case. Consistent with this finding, Table 3 shows a decrease in some item loadings across time in this conventional LFM. Hence, when interpreting the factor mean change and factor correlation change over time we cannot unambiguously determine whether apparent structural changes are actually due to instability in the measurement of the harsh parenting construct across 14 to 36 months. Such instability might arise, for instance, if detachment and negative regard are less indicative of poor parenting in some situations for older children, as compared to infants. For older children, some parental detachment might be constructive, in providing flexibility for the child's active and independent exploring, and some parental negative regard might be needed in a dangerous situation to prevent an accident.

To examine the sensitivity of these findings to the ignorable dropout assumption, five PM-LFMs were then fit with the alternative identifying constraints discussed previously. Item loadings, intercepts, residual variances, and residual covariances were allowed to differ across dropout-pattern aside from the constraints imposed for identification. Note that CC identifying constraints on residual covariances were used to facilitate model convergence. These five PM-LFMs all had RMSEA = .076 and CFI = .962.

As in the conventional LFM, for each PM-LFM, weak invariance could be rejected (LRT statistics ranged from $\chi^2(6) = 142.7–197.7$, $p < .05$. When imposing weak invariance, RMSEA increased from .076 to between .095 and .103 (depending on the PM-LFM identification method) and CFI decreased from .962 to between .916 and .929 (again depending on the PM-LFM identification method). Consistent with this finding, Table 3 shows a decrease in some marginal loadings (pooled across missingness pattern) over time. This means that our substantive conclusion about the level of MI met by the harsh parenting construct is not sensitive to this perturbation of the ignorable dropout assumption. It is the case, however, that some loading estimates are lower for all PM-LFMs (regardless of identifying constraint), as compared to the conventional LFM. Specifically, Table 3 shows that the PM-LFMs have lower loadings than the conventional LFM for negative regard at all time points and for detachment and unsupportiveness at later time points. The conventional LFM weights individuals with complete data more heavily than in the PM-LFM (e.g., Yang & Maxwell, 2014), perhaps leading to this discrepancy. If MI testing or other results had been particularly sensitive to the relaxation of the ignorable dropout assumption, we could then try to explain and examine whether membership in dropout pattern groups is distinguished by particular observed covariates, such as whether families were assigned to receive services fostering child development.

OVERALL DISCUSSION

Across the psychology, medical, public health, and statistics literatures there have been increasingly widespread calls for researchers to conduct sensitivity analyses investigating the robustness of results to departures from ignorable missingness. In some cases these recommendations have been made by oversight committees such as the National Research Council (2010), who stated explicitly, "Examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting" (p. 111). Heeding this recommendation requires the development of sensitivity

[7] Note that families came from 17 sites, but site codes were not made available in the data file "due to confidentiality concerns." See http://doi.org/10.3886/ICPSR03804.v5.

TABLE 3
Parameter Estimates From Empirical Example Sensitivity Analyses

Parm	*Conventional LFM*	*Pattern Mixture LFMs*				
		NN	*CC*	*AC*	*LOCF*	*NND*
λ_{11}	.92	.96	.96	.96	.96	.96
λ_{21}	.64	.64	.64	.64	.64	.64
λ_{31}	.42	.33	.33	.33	.33	.33
λ_{12}	.96	.98	.98	.97	.98	.98
λ_{22}	.59	.51	.50	.50	.53	.53
λ_{32}	.41	.34	.34	.33	.34	.34
λ_{13}	.89	.83	.83	.83	.87	.83
λ_{23}	.51	.30	.30	.30	.39	.34
λ_{33}	.39	.24	.24	.24	.27	.24
θ_{11}	.16	.19	.19	.19	.19	.19
θ_{21}	.59	.59	.59	.59	.59	.59
θ_{31}	.83	.52	.52	.52	.52	.52
θ_{12}	.08	.09	.08	.08	.09	.08
θ_{22}	.65	.50	.48	.48	.50	.50
θ_{32}	.83	.57	.56	.56	.56	.59
θ_{13}	.21	.18	.18	.18	.17	.18
θ_{23}	.74	.26	.26	.26	.35	.30
θ_{33}	.85	.32	.32	.32	.39	.36
τ_{21}	−1.06	−1.08	−1.08	−1.08	−1.08	−1.08
τ_{31}	.08	.07	.07	.07	.07	.07
τ_{22}	−.74	−.65	−.62	−.62	−.70	−.70
τ_{32}	.06	.05	.05	.05	.04	.06
τ_{23}	−.45	−.26	−.26	−.26	−.46	−.37
τ_{33}	.16	.10	.10	.10	.07	.10
κ_1	4.21	4.23	4.23	4.23	4.23	4.23
κ_2	4.11	4.11	4.11	4.11	4.11	4.11
κ_3	4.95	4.95	4.95	4.95	4.95	4.95
$\psi_{1,2}$	.61	.61	.61	.61	.61	.61
$\psi_{1,3}$	.62	.61	.61	.61	.61	.61
$\psi_{2,3}$	.53	.52	.52	.52	.52	.52
$\theta_{11,12}$	.05	.01	.01	.01	.01	.01
$\theta_{21,22}$	.20	.11	.11	.10	.11	.11
$\theta_{31,32}$	.16	.08	.08	.08	.08	.08
$\theta_{12,13}$	−.01	.00	.00	.00	.00	.00
$\theta_{22,23}$	.19	.07	.07	.07	.07	.07
$\theta_{32,33}$	.30	.13	.13	.13	.13	.13

Note. y_{1t} = unsupportiveness; y_{2t} = detachment; y_{3t} = negative regard; LFM = longitudinal factor model; Parm = parameter; NN = nearest neighbor identifying constraint; CC = complete case identifying constraint; AC = available case identifying constraint; LOCF = last observation carried forward identifying constraint; NND = nearest neighbor difference identifying constraint.

analysis methods applicable to common longitudinal analytic goals and also requires the dissemination of strengths and limitations of these sensitivity analysis methods. A common analytic goal is to test longitudinal invariance of relations between symptom items and an underlying construct to determine whether a measure can be used to represent the same construct across age. Although psychology studies employing longitudinal factor analysis for testing invariance often are characterized by high rates of dropout, psychologists lacked methods for investigating the sensitivity of their results to potentially nonignorable dropout.

To fill this gap, this article first presented a family of PM-LFMs that allow for nonignorable dropout. We suggested an approach to sensitivity analysis that entailed comparing MI testing results while imposing varying assumptions about the dropout process—including ignorable dropout (a conventional LFM) as well as different kinds of nonignorable dropout operationalized as different “identifying constraints” on the PM-LFM. Additionally, we used a simulation to demonstrate that, reassuringly, applying PM-LFM can provide MI testing results similar to a conventional LFM when missingness is ignorable. This simulation also showed that PM-LFM might provide more accurate MI testing results than a conventional LFM when missingness is nonignorable; however, the amount of improvement greatly depended on the realism of the identifying constraints imposed on the PM-LFM. Under unrealistic identifying constraints, PM-LFM can perform as poorly as a conventional LFM (relatedly, see Enders, 2011; Yang & Maxwell, 2014). Syntax (in M*plus*) for fitting each PM-LFM discussed in this article is provided in the online Appendix.

Previously, PM models had been rarely applied outside the context of univariate growth modeling. However, we explained why PM models are particularly well suited to the context of LFM analyses because typically T is small and dropout is a predominant source of missingness. We explained why alternative methods for accommodating nonignorable missingness pose some practical drawbacks in this setting. In the remaining subsections we discuss limitations, extensions, and overall take-home points.

Limitations

Several limitations of this study can be noted, which serve as directions for future research. First, structural parameters also could be subject to bias from nonignorable missingness, for instance, if they differ depending on missingness pattern, but a conventional LFM were fit. This was not demonstrated here but could also be addressed by application of the PM-LFM. Second, a researcher could use substantive theory to formulate PM-LFM identifying constraints other than the five we listed and illustrated here (see also Little, 2008). Third, researchers could incorporate covariates into the PM-LFM, for instance, by regressing the factors on covariates. Researchers could then decide whether to allow these slopes to differ across missingness pattern in the PM-LFM. Fourth, obtaining marginal parameter estimates from the PM-LFM requires computing weighted averages of pattern-specific parameter estimates. As mentioned previously, this is straightforward to do using, for instance, the *model constraint* command in M*plus*. However, it also would be possible to obtain these marginal parameter estimates without having to explicitly do these computations if one first multiply imputes missing data from a PM-LFM and then fits a conventional LFM to the imputed data. Such a multiple imputation approach has been used occasionally when implementing PM models in the growth modeling context (e.g., Demirtas & Schafer, 2003).

Fifth, the simulation and empirical example relied on only one approach—a traditional null hypothesis significance testing (NHST) approach—for assessing configural, weak, strong, and strict invariance. Nonetheless, it is important to note that the modeling developments in this article can still be employed if one chooses instead to assess longitudinal MI using a newer equivalence testing approach (Yuan & Chan, 2016) rather than an NHST approach. Specifically, when assessing longitudinal MI using either approach, researchers can employ PM-LFMs with the five alternative identification constraints from Table 2 and can use the definitions of marginal weak, strong, and strict invariance from Figure 2, that allow for nonignorable dropout. Equivalence testing (Yuan & Chan, 2016; Yuan, Chan, Marcoulides, & Bentler, 2016) involves testing whether the misfit in the configurally invariant model exceeds a user-specified tolerance amount, and then testing whether the difference in fit between subsequent models (weak, strong, and strict) exceeds another user-specified tolerance. Yuan and Chan (2016) showed that equivalence testing provides better control of Type I error rates than NHST under many conditions and they suggested potential tolerance amounts for practice; however, because conclusions about the level of MI supported would differ depending on the tolerance amount chosen, note that the choice of tolerance could itself be a worthwhile focus of sensitivity analysis.

Extension: Pattern Mixture Versus Latent Mixture Approaches for Investigating Missingness Mechanism Uncertainty in Longitudinal MI Testing

Sometimes researchers concerned about nonignorable dropout might have sparse dropout patterns. Other times researchers also might be concerned about nonignorable intermittent (i.e., nonmonotonic) missingness for the set of J items. In such circumstances, researchers might consider methods that effectively replace a larger number of manifest missingness pattern-groups with a typically small number of latent (unobserved) classes that each draw cases from different missingness patterns—as has been done in the context of growth modeling (e.g., Haviland, Jones, & Nagin, 2011; Lin, McCulloch, & Rosenheck, 2004; Tsonaka, Verbeke, & Lesaffre, 2009). For instance, to specify a latent missingness class LFM we can let c_i represent a latent classification variable with latent classes $k = 1 \ldots K$ and let the $\mathbf{m}_i$ missingness indicators serve as effect indicators of c_i.[8] Endorsement probabilities for the $\mathbf{m}_i$ missingness indicators can vary across latent class. Measurement parameters (e.g., loadings) governing the outcome process can also vary across latent class. The outcome and missingness processes are assumed conditionally independent—conditional on latent missingness class:

$$f(\mathbf{y}_i, \mathbf{m}_i) = \sum_{k=1}^{K} f(\mathbf{y}_i | c_i = k) f(\mathbf{m}_i | c_i = k) p(c_i = k) \tag{6}$$

The class membership probabilities $p(c_i = k)$ can be obtained using a multinomial logistic specification (see, e.g., Sterba, 2013). Marginal parameters of $f(\mathbf{y}_i)$ can be obtained as weighted sums across latent classes. Whereas this extension from manifest to latent missingness classes has some compelling motivation, it also poses extra issues and complexities vis-à-vis the PM-LFM. Three issues—involving the number, interpretation, and identification of classes—are summarized next.

Regarding the *number* of classes and patterns, in the PM approach G is explicitly determined by the known number of dropout patterns or dropout pattern summaries. In the latent missingness class approach, K can be selected using model selection indexes (such as the Bayesian information criterion). However, there is ambiguity about what level of longitudinal MI to impose on the within-class LFM when selecting K. The decision about selecting K could affect the decision about what level of longitudinal MI is supported for marginal LFM parameters. These decisions are interdependent. Specifically, simulations on related latent mixtures (e.g., Lubke & Neale, 2006) suggest that using an overly constrained within-class model when selecting K could lead to overextracting K, whereas using an under-constrained within-class model when selecting K could lead to underextracting K. Researchers might elect to select K using a less constrained (configurally invariant) LFM within-class if their greatest concern was avoiding over-extracting K. This was done in the extended empirical illustration described later.

Regarding the *interpretation* of classes and patterns, just as having $G > 1$ in the PM approach itself does not imply a departure from ignorable missingness, merely selecting $K > 1$ in the latent missingness class approach also does not imply a departure from ignorable missingness. In the latent missingness class approach, multiple classes could be distinguished by class separation exclusively on outcome process parameter estimates (for a variety of reasons, including simply violating distributional assumptions, e.g., Bauer, 2007) even when missingness probabilities are similar across classes. Indeed, when the latent missingness class approach was applied to the empirical example data set,

[8] Here we continue to define $\mathbf{m}_i$ as consisting of T – 1 dropout indicators, for consistency with the earlier presentation. If we were considering intermittent missingness for the set of J items, we could redefine our vector $\mathbf{m}_i$ of binary missingness indicators to be of dimension $T \times 1$ where element $m_{it} = 1$ when $\mathbf{y}_{it}$ is missing and $m_{it} = 0$ when $\mathbf{y}_{it}$ is observed. (Similar results for the latent missingness class approach were obtained when $\mathbf{m}_i$ did vs. did not include intermittent missingness.) Finally, note that we are not discussing *item-specific* intermittent missingness here (e.g., missingness on Item 3 but not on Items 1, 2, and 4 at time t). This kind of missingness, if present, would still be assumed ignorable.

$K > 1$ was preferred by the BIC despite the fact that dropout probabilities were quite similar across the K classes. Specifically, $K = 3$ had the lowest BIC without encountering convergence problems, and probabilities of dropout at Times 2 and 3 were similar across classes (.14 and .19 for $K = 1$, .12 and .13 for $K = 2$, and .19 and .16 for $K = 3$). Using these $K = 3$ latent classes, marginal LFM parameters were computed after either imposing or relaxing weak MI. Consistent with our previous empirical results using the manifest PM-LFM, when applying the latent missingness class LFM, weak longitudinal invariance of marginal parameters could again be rejected. Thus, MI testing results were again found to be insensitive to allowing for nonignorable missingness.

Regarding *identification*, beyond fixing the scale and location of each factor in the within-class LFM, identification considerations differ in the PM-LFM versus the latent missingness class LFM. For the PM-LFM, identifying constraints (AC, CC, LOCF, NN, NND, etc.) can be decided on in advance and the implications of each are clear based on their definitions. For the latent missingness class approach, a variety of additional constraints might be needed to ensure empirical identification (see McLachlan & Peel, 2000), ranging from reducing K to placing across-class equality constraints on particular parameters. For instance, our empirical implementation of the latent missingness class LFM encountered estimation problems if $K > 3$ because then there was no variability in one or more missingness indicators within class; also, we needed to constrain all parameters other than loadings to be equal across latent class to achieve convergence. Taken together, this entailed imposing more across-class constraints than in the PM-LFM. Additionally, in the PM-LFM approach it is conventional to investigate variation in results across different identifying constraints. A similar sensitivity analysis using the latent missingness class LFM could entail investigating variation in results across K under different minimally sufficient sets of constraints.

In summary, from practical and interpretational standpoints, the PM-LFM and latent missingness class LFM approaches have different strengths and weaknesses for investigating sensitivity of MI testing to departures from nonignorable missingness. The definition of a pattern and the interpretation of across-pattern variation in parameter estimates are more straightforward under the PM-LFM. The operationalization of alternative identifying constraints is more explicit in the PM-LFM, rendering it arguably straightforward to inspect how results vary across alternative choices. When researchers are interested in investigating the possibility of intermittent nonignorable missingness, rather than nonignorable dropout, however, the latent missingness class approach may be preferable.

Overall Take-Home Points

When fitting conventional LFMs with FIML to test MI, ignorable missingness is assumed. We have discussed that this assumption amounts to requiring that parameter estimates be constrained equal across missingness pattern. This requirement can be conceptualized as a special case constraint on the PM-LFM. We encourage researchers to also consider the substantive plausibility of alternative kinds of PM-LFM constraints that allow for nonignorable missingness (e.g., the first five identifying constraints listed earlier). When—as in our empirical example on harsh parenting—the same conclusions about MI are reached despite allowing for departures from ignorable missingness, we gain confidence in the robustness of our MI testing results. It might instead be the case—as in some cells of our simulation design—that the same conclusions about MI are not reached when allowing for departures from ignorable missingness. If so, we can report this source and extent of uncertainty in results and also can choose to focus interpretation on those PM-LFM specifications that are most substantively reasonable.

ACKNOWLEDGMENTS

The author would like to thank Kristopher J. Preacher for helpful comments on a previous version of this article.

REFERENCES

Administration for Children and Families. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of early head start* (Vol. I–III). Washington, DC: U.S. Department of Health and Human Services.

Akl, E. A., Shawwa, K., Kahale, L. A., Agoritsas, T., Brignardello-Petersen, R., Busse, J. W., …Guyatt, G. H. (2015). Reporting missing participant data in randomised trials: Systematic survey of the methodological literature and a proposed guide. *British Medical Journal Open*, *5*, 1–7. doi:10.1136/bmjopen-2015-008431

Allison, P. (1987). Estimation of linear models with incomplete data. *Sociological Methodology*, *17*, 71–103. doi:10.2307/271029

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*, 757–786. doi:10.1080/00273170701710338

Beunckens, C., Molenberghs, G., Thijs, H., & Verbeke, G. (2007). Incomplete hierarchical data. *Statistical Methods in Medical Research*, *16*, 457–492. doi:10.1177/0962280206075310

Brydges, C., Fox, A., Reid, C., & Anderson, M. (2014). The differentiation of executive functions in middle and late childhood: A longitudinal latent-variable analysis. *Intelligence*, *47*, 34–43. doi:10.1016/j.intell.2014.08.010

Byrne, B., Shavelson, R., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466. doi:10.1037/0033-2909.105.3.456

Carlson, B. (2009, August). *Analysis of nonresponse bias in the early head start research and evaluation project*. Paper presented at the Joint Statistical Meetings' Section on Survey Research Methods, Washington, D.C.

Demirtas, H., & Schafer, J. L. (2003). On the performance of random-coefficient pattern mixture models for non-ignorable drop-out. *Statistics in Medicine*, *22*, 2553–2575. doi:10.1002/(ISSN)1097-0258

Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data-analysis. *Applied Statistics*, *43*, 49–93. doi:10.2307/2986113

Edwards, M. C., & Wirth, R. J. (2009). Measurement and the study of change. *Research in Human Development*, *6*, 74–96. doi:10.1080/15427600902911163

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, *16*, 1–16. doi:10.1037/a0022640

Fagerstrom, C., Lindwall, M., Ingeborg, A., & Rennemark, M. (2012). Factorial validity and invariance of the Life Satisfaction Index in older people across groups and time: Addressing the heterogeneity of age, functional ability, and depression. *Archives of Gerontology and Geriatrics*, *55*, 349–356. doi:10.1016/j.archger.2011.10.007

Feldman, B., & Rabe-Hesketh, S. (2012). Modeling achievement trajectories when attrition is informative. *Journal of Educational and Behavioral Statistics*, *37*, 703–736. doi:10.3102/1076998612458701

Gottfredson, N., Bauer, D., & Baldwin, S. (2014). Modeling change in the presence of nonrandomly missing data: Evaluating a shared parameter mixture model. *Structural Equation Modeling*, *21*, 196–209. doi:10.1080/10705511.2014.882666

Hafez, M. S., Moustaki, I., & Kuha, J. (2015). Analysis of multivariate longitudinal data subject to nonrandom dropout. *Structural Equation Modeling*, *22*, 193–201. doi:10.1080/10705511.2014.936086

Haviland, A. Jones, B. & Nagin, D. (2011). Group-based trajectory modeling extended to account for nonrandom participation attention. *Sociological Methods and Research*, 40, 367–390.

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, *2*, 64–78. doi:10.1037/1082-989X.2.1.64

Horn, J., & McArdle, J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117–144. doi:10.1080/03610739208253916

Jones, M., Mishra, B., & Dobson, A. (2015). Analytical results in longitudinal studies depended on target of inference and assumed mechanism of attrition. *Journal of Clinical Epidemiology*, *68*, 1165–1175. doi:10.1016/j.jclinepi.2015.03.011

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426. doi:10.1007/BF02291366

King, B. (2011). *Unbiased measurement of health-related quality-of-life* (Unpublished doctoral dissertation). University of Amsterdam, Amsterdam, The Netherlands.

Leite, W. (2007). A comparison of latent growth models for constructs measured by multiple items. *Structural Equation Modeling*, *14*, 581–610. doi:10.1080/10705510701575438

Lin, H., McCulloch, C. E., & Rosenheck, R. A. (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics*, *60*, 295–305. doi:10.1111/j.0006-341X.2004.00173.x

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, *88*, 125–134.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112–1121. doi:10.1080/01621459.1995.10476615

Little, R. J. A. (2008). Selection and pattern-mixture models. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 409–431). London, UK: CRC.

Little, R. J. A., D'Agostino, R., Cohen, M., Dickerson, K., Emerson, S. S., Farrar, J., … Stern, H. (2012). The prevention and treatment of missing data in clinical trials: Special report. *The New England Journal of Medicine*, *367*, 1355–1360. doi:10.1056/NEJMsr1203730

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.

Liu, Y., Millsap, R. E., West, S. G., Tein, J.-Y., Tanaka, R., & Grimm, K. J. (2016). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*. doi:10.1037/met0000075

Lubke, G., & Neale, M. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, *41*, 499–532. doi:10.1207/s15327906mbr4104_4

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Reviews of Psychology*, *51*, 201–226. doi:10.1146/annurev.psych.51.1.201

Mallinckrodt, C., Lin, Q., & Molenberghs, M. (2013). A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. *Pharmaceutical Statistics*, *12*, 1–6. doi:10.1002/pst.1547

Mason, S., Lauterbach, D., McKibben, J., Lawrence, J., & Fauerbach, J. (2013). Confirmatory factor analysis and invariance of the Davidson Trauma Scale (DTS) in a longitudinal sample of burn patients. *Psychological Trauma: Theory, Research, Practice, and Policy*, *5*, 10–17. doi:10.1037/a0028002

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543. doi:10.1007/BF02294825

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.

Minini, P., & Chavance, M. (2004). Sensitivity analysis of longitudinal normal data with drop-outs. *Statistics in Medicine*, *23*, 1039–1054. doi:10.1002/sim.1702

Mogos, M., Beckstead, J., Kip, K., Evans, M., Boothroyd, R., Aiyer, A., & Reis, S. (2015). Assessing longitudinal invariance of the Center for Epidemiologic Studies–Depression Scale among middle-aged and older adults. *Journal of Nursing Measurement*, *23*, 302–314. doi:10.1891/1061-3749.23.2.302

Molenberghs, G., & Kenward, M. G. (2007). *Missing data in clinical studies*. Chichester, UK: Wiley.

Molenberghs, G., & Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modeling*, *1*, 235–269. doi:10.1191/147108201128195

Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*, *16*, 17–33. doi:10.1037/a0022634

Muthén, L. K., & Muthén, B. (1998–2016). *M*plus (Version 7.4) [Computer software]. Los Angeles, CA: Muthén & Muthén.

National Research Council. (2010). *The prevention and treatment of missing data in clinical trials*. Panel on handling missing data in clinical trials. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

Nguyen, C., Lee, K., & Carlin, J. (2015). Posterior predictive checking of multiple imputation models. *Biometrical Journal*, *57*, 676–694. doi:10.1002/bimj.201400034

O'Kelly, M., & Ratitch, B. (2014). *Clinical trials with missing data: A guide for practitioners*. New York, NY: Wiley.

Pitts, S., West, S., & Tein, J.-Y. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning*, *19*, 333–350. doi:10.1016/S0149-7189(96)00027-4

Raykov, T., & Marcoulides, G. A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling*, *11*, 621–637. doi:10.1207/s15328007sem1104_7

Richerson, K., Watkins, M., & Beaujean, A. (2014). Longitudinal invariance of the Wechsler Intelligence Scale for Children–Fourth Edition in a referral sample. *Journal of Psychoeducational Assessment*, *32*, 597–609. doi:10.1177/0734282914538802

Roggman, L., Cook, G., Peterson, C., & Raikes, H. (2008). Who drops out of early Head Start home visiting programs? *Early Education and Development*, *19*, 574–599. doi:10.1080/10409280701681870

Rombach, I., Rivero-Arias, O., Gray, A., Jenkinson, C., & Burke, O. (2016). The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: A review of the current literature. *Quality of Life Research*, *25*, 1613–1623. doi:10.1007/s11136-015-1206-1

Rose, N., von Davier, M., & Xu, X. (2010). *Modeling non-ignorable missing data with item response theory.* Princeton, NJ: Educational Testing Company. Retrieved from https://www.ets.org/Media/Research/pdf/RR-10-11.pdf

Roy, J. (2007). Latent class models and their application to missing-data patterns in longitudinal studies. *Statistical Methods in Medical Research*, *16*, 441–456. doi:10.1177/0962280206075311

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592. doi:10.1093/biomet/63.3.581

Rubin, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association*, *72*, 538–543. doi:10.1080/01621459.1977.10480610

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037/1082-989X.7.2.147

Small, B., Hertzog, C., Hultsch, D., & Dixon, R. (2003). Stability and change in adult personality over 6 years: Findings from the Victoria Longitudinal Study. *Journal of Gerontology*, *58*, 166–176. doi:10.1093/geronb/58.3.P166

Souza, R., Eisen, R., Perera, S., Bantoto, B., Bawor, M., Dennis, B., … Thabane, L. (2016). Best (but oft-forgotten) practices: Sensitivity analyses in randomized controlled trials. *American Journal of Clinical Nutrition*, *103*, 5–17. doi:10.3945/ajcn.115.121848

Sterba, S. K. (2013). Understanding linkages among mixture models. *Multivariate Behavioral Research*, *48*, 775–815. doi:10.1080/00273171.2013.827564

Sterba, S. K. (2016). A latent transition analysis model for latent-state-dependent nonignorable missingness. *Psychometrika, 82*, 506–534. doi:10.1007/s11336-015-9442-4

Sterba, S. K., & Gottfredson, N. (2015). Diagnosing global case influence on MAR versus MNAR model comparisons. *Structural Equation Modeling*, *22*, 294–307. doi:10.1080/10705511.2014.936082

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., & Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics*, *3*, 245–265. doi:10.1093/biostatistics/3.2.245

Thomas, N., Harel, O., & Little, R. J. A. (2016). Analyzing clinical trial outcomes based on incomplete daily diary reports. *Statistics in Medicine, 35*, 2894–2906. doi:10.1002/sim.6890

Tisak, J., & Meredith, W. (1991). Longitudinal factor analysis. In A. von Eye (Ed.), *Statistical methods in longitudinal research: Vol. 1. Principles and structuring change* (pp. 125–150). San Diego, CA: Academic.

Tsonaka, R., Verbeke, G., & Lesaffre, E. (2009). A semi-parametric shared-parameter model to handle nonmonotone nonignorable missingness. *Biometrics*, *65*, 81–87. doi:10.1111/j.1541-0420.2008.01021.x

Vandenberg, R., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70. doi:10.1177/109442810031002

Varni, J., Limbers, C., Newman, D., & Seid, M. (2008). Longitudinal factorial invariance of the PedsQL 4.0 Generic Core Scales Child Self-Report Version: One year prospective evidence from the California State Children's Health Insurance Program (SCHIP). *Quality of Life Research*, *17*, 1153–1162. doi:10.1007/s11136-008-9389-3

Verbeke, G., Molenbergs, G., & Beunckens, C. (2008). Formal and informal model selection with incomplete data. *Statistical Science*, *23*, 201–218. doi:10.1214/07-STS253

Wang, M., Elhai, J., Daic, X., & Yao, S. (2012). Longitudinal invariance of posttraumatic stress disorder symptoms in adolescent earthquake survivors. *Journal of Anxiety Disorders*, *26*, 263–270. doi:10.1016/j.janxdis.2011.12.009

Wang, Z., & Su, I. (2013). Longitudinal factor structure of general self-concept and locus of control among high school students. *Journal of Psychoeducational Assessment*, *31,* 554–565. doi:10.1177/0734282913481651

Widaman, K. (1991). Qualitative transitions amid quantitative development: A challenge for measuring and representing change. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 204–217). Washington, DC: American Psychological Association.

Widaman, K., Ferrer, E., & Conger, R. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*, 10–18. doi:10.1111/j.1750-8606.2009.00110.x

Wirth, R. J. (2008). *The effects of measurement non-invariance on parameter estimation in latent growth models* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill, NC.

Wu, M., & Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, *44*, 175–188. doi:10.2307/2531905

Wu, P.-C. (2015). Longitudinal measurement invariance of Beck Depression Inventory–II in early adolescents. *Assessment*. Advance online publication. doi:10.1177/1073191115608941

Xu, S., & Blozis, S. (2011). Sensitivity analyses of mixed models for incomplete longitudinal data. *Journal of Educational and Behavioral Statistics*, *36*, 237–256. doi:10.3102/1076998610375836

Yang, M., & Maxwell, S. (2014). Estimation of treatment effects in randomized longitudinal designs with different types of non-ignorable dropout. *Psychological Methods*, *19*, 188–210. doi:10.1037/a0033804

Yuan, K.-H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests. *Psychological Methods*, *21*, 405–426. doi:10.1037/met0000080

Yuan, K.-H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indices. *Structural Equation Modeling*, *23*, 319–330. doi:10.1080/10705511.2015.1065414

APPENDIX

Percent Absolute Relative Bias for Parameter Estimates From Simulation

	Condition 1						*Condition 2*						*Condition 3*						*Condition 4*					
		PM-LFM						*PM-LFM*						*PM-LFM*						*PM-LFM*				
Parm.	*LFM*	*NN*	*CC*	*AC*	*LOCF*	*NND*	*LFM*	*NN*	*CC*	*AC*	*LOCF*	*NND*	*LFM*	*NN*	*CC*	*AC*	*LOCF*	*NND*	*LFM*	*NN*	*CC*	*AC*	*LOCF*	*NND*
λ_{11}	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	1	1	1	1	1
λ_{21}	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	1	1	1	1	1
λ_{31}	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0
λ_{41}	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	1	1	1	1	1
λ_{12}	7	4	7	3	0	0	0	0	0	3	4	0	0	0	0	3	0	0	1	0	13	13	4	8
λ_{22}	7	4	7	3	0	0	0	0	0	3	4	0	0	0	0	3	0	0	1	0	13	13	4	8
λ_{32}	7	4	7	3	0	0	0	0	0	3	4	0	0	0	0	3	0	0	1	0	14	14	3	8
λ_{42}	7	4	7	3	0	0	0	0	0	3	4	0	0	0	0	3	0	0	0	0	13	13	4	8
λ_{13}	10	11	11	11	0	0	1	1	1	1	16	0	0	0	0	0	0	0	2	1	1	1	22	19
λ_{23}	11	11	11	11	0	0	1	1	1	1	17	1	0	0	0	0	0	0	3	1	1	1	22	20
λ_{33}	10	11	11	11	0	0	0	0	0	0	16	0	0	0	0	0	0	0	1	1	1	1	21	19
λ_{43}	11	12	12	12	0	1	1	1	1	1	17	1	1	1	1	1	0	1	3	1	1	1	22	20
θ_{11}	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
θ_{21}	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	3	3	3
θ_{31}	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	3	3	3
θ_{41}	1	0	0	0	0	0	1	1	1	1	1	1	1	2	2	2	2	2	1	3	3	3	3	3
θ_{12}	12	6	14	15	0	0	1	1	1	4	8	1	1	1	1	4	1	1	4	1	5	0	4	6
θ_{22}	12	7	15	16	0	0	1	1	1	4	8	1	1	1	1	4	1	1	4	2	5	0	4	5
θ_{32}	12	6	14	15	0	0	0	1	1	4	7	1	0	1	1	4	1	1	5	1	5	1	3	6
θ_{42}	12	6	14	15	0	0	0	1	1	4	7	1	0	1	1	4	1	1	5	1	5	1	4	6
θ_{13}	22	22	22	22	1	1	1	1	1	1	14	1	0	0	0	0	1	1	1	1	1	1	12	7
θ_{23}	22	22	22	22	0	1	0	0	0	0	14	1	0	0	0	0	1	1	1	1	1	1	12	7
θ_{33}	22	22	22	22	1	1	0	0	0	0	14	1	0	0	0	0	1	1	0	0	0	0	12	6
θ_{43}	22	22	22	22	0	0	1	1	1	1	14	1	1	1	1	1	1	1	1	1	1	1	13	7
τ_{21}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
τ_{31}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
τ_{41}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
τ_{22}	1	1	1	4	1	1	1	1	1	4	1	1	1	1	1	4	1	1	2	1	1	5	1	2
τ_{32}	0	0	0	4	0	0	0	1	0	4	0	0	0	0	0	4	0	0	1	1	1	4	0	1
τ_{42}	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0	3	0	0	1	1	1	4	1	1
τ_{23}	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	2	2	2	2	2	2
τ_{33}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
τ_{43}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	3
κ_1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	1	1	1	1	1
κ_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	2	2	2	2	2
κ_3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	2	2	2	2	2
$\psi_{1,2}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
$\psi_{2,3}$	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
$\psi_{1,3}$	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	1	1	1	2	1

Note. LFM = longitudinal factor model; PM-LFM = pattern-mixture longitudinal factor model; NN = nearest neighbor identifying constraint; CC = complete case identifying constraint; AC = available case identifying constraint; LOCF = last observation carried forward identifying constraint; NND = nearest neighbor difference identifying constraint.