PSYCHOMETRIKA—VOL. 81, NO. 2, 506–534
JUNE 2016
DOI: 10.1007/s11336-015-9442-4

A LATENT TRANSITION ANALYSIS MODEL FOR LATENT-STATE-DEPENDENT NONIGNORABLE MISSINGNESS

SONYA K. STERBA

VANDERBILT UNIVERSITY

Psychologists often use latent transition analysis (LTA) to investigate state-to-state change in discrete latent constructs involving delinquent or risky behaviors. In this setting, latent-state-dependent nonignorable missingness is a potential concern. For some longitudinal models (e.g., growth models), a large literature has addressed extensions to accommodate nonignorable missingness. In contrast, little research has addressed how to extend the LTA to accommodate nonignorable missingness. Here we present a shared parameter LTA that can reduce bias due to latent-state-dependent nonignorable missingness: a parallel-process missing-not-at-random (MNAR-PP) LTA. The MNAR-PP LTA allows outcome process parameters to be interpreted as in the conventional LTA, which facilitates sensitivity analyses assessing changes in estimates between LTA and MNAR-PP LTA. In a sensitivity analysis for our empirical example, previous and current membership in high-delinquency states predicted adolescents' membership in missingness states that had high nonresponse probabilities for some or all items. A conventional LTA overestimated the proportion of adolescents ending up in a low-delinquency state, compared to an MNAR-PP LTA.

Key words: nonignorable missing data, latent transition analysis, missing not at random, shared parameter model, mixture model.

Psychologists are often interested in modeling stage-sequential change in a discrete latent construct over time. For instance, Velicer, Martin, and Collins, (1996) were interested in modeling individuals' patterns of change across latent states of addiction recovery: precontemplation, contemplation, preparation, action, and maintenance. When the discrete latent construct (e.g., addiction recovery) is measured by multiple observed indicators at each timepoint, a latent transition analysis[1] (LTA) model can be used to address research questions about latent state-to-state change (for reviews, see Collins & Lanza, 2010; Collins & Wugalter, 1992, Langeheine & Van de Pol, 1990, Reboussin, Reboussin, Liang, & Anthony, 1998). Specific research questions that LTA can address involve the number of latent states per timepoint, the proportion of persons in each latent state per timepoint, and the nature of change across states (forward-only, backward-only, no-change, etc.). Recent applications of LTA have investigated stage-sequential change in latent constructs such as peer victimization (Nylund, Asparouhov, & Muthén, 2006; Williford, Boulton, & Jenson, 2014), sexual risk behavior (Lanza & Collins, 2008; Mackesy-Amiti et al., 2014), underage or high-risk drinking (Cleveland, Lanza, Ray, Turrisi, & Mallet, 2012; Cochran, Field, & Caetano, 2013), pre-adolescent drug use (Hopfer, Hecht, Lanza, Tan, & Xu, 2013), and intimate partner violence (Bair-Merritt, Ghazarian, Burrell, & Duggan, 2012).

Psychology researchers considering the application of LTA must contend with missing responses in their longitudinal, multivariate data. Fortunately, some kinds of missing data are accommodated by full information maximum likelihood estimation methods conventionally employed to fit LTA models, typically using the Expectation-Maximization (EM) algorithm (see, e.g., Lee & Song, 2003; McLachlan & Peel, 2000). For instance, such methods accommodate

Electronic supplementary material The online version of this article (doi:10.1007/s11336-015-9442-4) contains supplementary material, which is available to authorized users.

Correspondence should be made to Sonya K. Sterba, Quantitative Methods Program, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203, USA. Email: sonya.sterba@vanderbilt.edu

[1] The LTA model is also known as a multiple-indicator latent Markov model (e.g., Langeheine & Van de Pol, 1990).

outcome missingness that is at-random (MAR, Rubin, 1976). Under MAR, the probability of missingness for a given outcome at time t can depend on observed variables (such as observed outcomes at time $t-1$), but *not* on unobservables in the outcome-generating model (such as current or previous 'addiction recovery' latent state). These methods also can accommodate outcome missingness that is completely-at-random (MCAR), wherein the probability of missingness depends neither on such observables nor unobservables. When missing outcomes are MAR or MCAR, all cases are retained when fitting a LTA model using the EM algorithm, and no parameter bias is incurred. On the other hand, conventional estimation methods for LTA do not accommodate outcome missingness that is not-at-random (MNAR), such as if the probability of missingness for a given outcome at time t depended on 'addiction recovery' latent state at time t or $t-1$, or directly depended on the score of that or another missing outcome at time t. When missing outcomes are MNAR—also called nonignorably missing (e.g., Molenberghs & Kenward, 2007)—avoiding parameter bias requires constructing a *joint* model for the outcomes of substantive interest and the missingness indicators (e.g., observed binary indicators of missingness for all repeated outcomes), as they are interdependent. In other words, a joint MNAR model must be constructed for the outcome-generating mechanism of substantive interest and the missingness-generating mechanism (Little & Rubin, 2002). A sensitivity analysis may then be employed to assess changes in parameter estimates when fitting a joint MNAR model versus a model only for the outcome-generating mechanism under MAR assumptions (e.g., Committee on National Statistics, 2010; Little & Rubin, 1999; Molenberghs & Verbeke, 2005).

Some popular longitudinal outcome models—predominately, growth models—have already been extended to form joint models for accommodating nonignorable missingness (e.g., multilevel growth models, groups-based trajectory models, and growth mixture models; Diggle & Kenward, 1994; Haviland, Jones, & Nagin, 2011; Roy, 2003; Wu & Carroll, 1988). Application of joint MNAR specifications for growth models has been increasing in the psychology literature (e.g., Enders, 2010, 2011; Feldman & Rabe-Hesketh, 2012; Hedeker & Gibbons, 1997; Lu & Zhang, 2014; Lu, Zhang, & Lubke, 2011; Muthén, Asparouhov, Hunter, & Leuchter, 2011; Sterba & Gottfredson, 2014; Xu & Blozis, 2011; Yang & Maxwell, 2014). In contrast, there has been little attention to the handling of nonignorable missingness in the context of LTA (see Sect. 3). The lack of attention to joint MNAR specifications for LTA is an important gap because substantive contexts in which LTA is applied in psychology can present an elevated risk of nonignorable missingness (Groves, Dillman, Eltinge, & Little, 2002). For instance, LTA is often applied in substance abuse research (e.g., Cleveland et al., 2012; Cochran et al., 2013) where there is the risk of missingness due to current membership in a 'high-substance-use' latent state and/or due to a latent pattern of improvement, even after conditioning on observables (Albert & Follmann, 2003; Little et al., 2012). Also, psychologists often apply LTA to outcomes that are illicit or socially undesirable, such as intimate partner violence, sexual risk behavior, and peer victimization (e.g., Mackesy-Amiti et al., 2014; Bair-Merritt et al., 2012). This raises the possibility that participants currently in a 'high-risk' latent outcome state have a greater propensity for intermittent missingness just on sensitive outcomes.

In this *Case Study* we develop a joint MNAR LTA model to accommodate intermittent missingness that may occur in combination with dropout. This model was motivated by our empirical application, in which the aim was to model state-to-state change in a discrete latent conduct problem construct in adolescents from the foster care system. This empirical application involves a combination of intermittent missingness and dropout. For substantive reasons, missingness was suspected to be latent-state dependent (described subsequently).

The remainder of this manuscript is organized as follows. First, we review alternative frameworks for joint MNAR models that have been described *outside* the context of LTA—selection, pattern-mixture, and shared parameter varieties. Second, we review the conventional LTA model. Third, we develop a particular kind of shared parameter joint MNAR model for LTA that accommodates latent-state-dependent missingness—here termed a parallel-process MNAR LTA (MNAR-PP LTA). We relate this model to existing joint MNAR models used outside the context of LTA. Fourth, we convey the necessity for the MNAR-PP LTA through a simulation demonstration. The proposed MNAR-PP LTA is shown to be able to avert bias in LTA parameters under a plausible MNAR mechanism where the latent missingness state at time t depends on the latent outcome state at time t and/or $t-1$ (and on the latent missingness state at time $t-1$). Importantly, even if the MNAR-PP LTA is mistakenly/unnecessarily applied under an MAR mechanism, the MNAR-PP LTA can recover parameters as well as does a conventional LTA, on average. Fifth, we implement the MNAR-PP LTA in an empirical example on conduct problems in adolescents leaving foster care, and we interpret MNAR-PP LTA results in the context of a sensitivity analysis. We conclude with a discussion of extensions and future directions.

1. Alternative Frameworks for Joint MNAR Modeling

Let $\mathbf{y}_{it}$ be person i's ($i = 1 \ldots N$) vector of responses to J outcomes ($j = 1 \ldots J$) measured at time t, where $\mathbf{y}_{it} = (y_{i1t} \ldots y_{iJt})'$. For univariate repeated measures, $J = 1$; for multivariate repeated measures $J > 1$. Let $\mathbf{y}_i$ correspond with person i's vector of J outcome scores repeated at timepoints $t = 1 \ldots T$, that is $\mathbf{y}_i = \{\mathbf{y}'_{i1} \ldots \mathbf{y}'_{iT}\}'$. The vector $\mathbf{y}_i$ may be partially unobserved for person i. We use a period as a missing data code such that $y_{ijt} = .$ indicates that the jth outcome at time t for person i is unobserved. The missingness pattern of $\mathbf{y}_i$ can be represented by a vector of missingness indicators $\mathbf{m}_i$, where $\mathbf{m}_i = \{\mathbf{m}'_{i1} \ldots \mathbf{m}'_{iT}\}'$ and $\mathbf{m}_{it} = (m_{i1t} \ldots m_{iJt})'$. Let element $m_{ijt} = 1$ if $y_{ijt} = .$ and let $m_{ijt} = 0$ if y_{ijt} is observed. Denote the joint distribution of the outcomes and missingness indicators by $f(\mathbf{y}_i, \mathbf{m}_i|\boldsymbol{\theta})$. Let $\boldsymbol{\theta} = \{\boldsymbol{\theta}_y, \boldsymbol{\theta}_m\}$ where $\boldsymbol{\theta}_y$ is a vector of parameters of substantive interest describing the outcome-generating mechanism and $\boldsymbol{\theta}_m$ is a vector of (typically nuisance) parameters describing the missingness mechanism. Several general MNAR modeling frameworks have been proposed which entail different factorizations of this joint distribution. These different factorizations correspond to different conceptualizations of the missingness-generating process—as follows.

One framework, employed in outcome-dependent selection models (e.g., Diggle & Kenward, 1994), allows missingness indicators to depend directly on the values of the repeated measures: $f(\mathbf{y}_i, \mathbf{m}_i|\boldsymbol{\theta}) = f(\mathbf{y}_i|\boldsymbol{\theta}_y) f(\mathbf{m}_i|\mathbf{y}_i, \boldsymbol{\theta}_m)$. This approach is often employed to model dropout (where $m_{ijt} = .$ if dropout already occurred) rather than intermittent missingness or their combination (e.g., Little, 1995). Another framework, employed in pattern mixture models (e.g., Little, 2008; Thijs, Molenberghs, Michiels, Verbeke, & Curran, 2002), uses a different factorization, $f(\mathbf{y}_i, \mathbf{m}_i|\boldsymbol{\theta}) = f(\mathbf{y}_i|\mathbf{m}_i, \boldsymbol{\theta}_y) f(\mathbf{m}_i|\boldsymbol{\theta}_m)$, which specifies the distribution of repeated outcomes to vary across missingness patterns. The marginal distribution of the repeated measures is a weighted average of pattern-specific distributions. This approach is also often used to model dropout. A third framework, shared parameter models, is often used to model intermittent missingness and/or dropout. In this framework, $\mathbf{y}_i$ and $\mathbf{m}_i$ are conditionally independent, given one or more shared latent variables. The shared latent variables, here generically denoted $\mathbf{b}_i$, could be continuously distributed random effect(s) (e.g., Molenberghs & Kenward, 2007; Ten Have et al., 1998) or discretely distributed latent classification variable(s) (e.g., Haviland et al., 2011) leading to a factorization such as $f(\mathbf{y}_i, \mathbf{m}_i, \mathbf{b}_i|\boldsymbol{\theta}) = f(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\theta}_y) f(\mathbf{m}_i|\mathbf{b}_i, \boldsymbol{\theta}_m) f(\mathbf{b}_i|\boldsymbol{\theta}_b)$. It can be integrated or summed (respectively) across $\mathbf{b}_i$ to obtain $f(\mathbf{y}_i, \mathbf{m}_i|\boldsymbol{\theta})$. Furthermore, instead of assuming $\mathbf{y}_i$ and $\mathbf{m}_i$ depend on the *same* set of latent variables, a less restrictive option is to assume $\mathbf{y}_i$ and $\mathbf{m}_i$ are conditionally independent given a set of associated latent variables (e.g., Lin, Liu, & Zhou, 2010; Muthén et al., 2011; Rizopoulos, Verbeke, Lessafre, & Vanrenterghem, 2008). We develop the latter kind of shared parameter MNAR model for LTA in Sect. 3.

2. Conventional Latent Transition Analysis (LTA) Model

Having now introduced several joint MNAR modeling frameworks in general terms, we turn next to a review of the conventional LTA (assuming MAR) before presenting the proposed joint MNAR model for LTA. In this section, we assume complete case data. Furthermore, we focus on LTAs with binary outcomes because the vast majority of LTA applications in psychology employ binary outcomes (see Collins & Lanza, 2010). Similarly, because of its common use in practice, we focus on LTAs with a first-order Markov structure, wherein discrete latent states at time t are regressed only on latent states at time $t-1$ (and not at times $t-2$, $t-3$, etc.).

For the jth binary outcome measured on person i at time t, let $y_{ijt} = 1$ if endorsed and $y_{ijt} = 0$ if not endorsed. At time t, the J outcomes for person i are indicators of a categorical latent variable, c^y_{it}, at time t. Denote the response pattern for person i as $\mathbf{y}_i = (\mathbf{y}'_{i1} \ldots \mathbf{y}'_{iT})'$ where $\mathbf{y}_{it} = (y_{i1t} \ldots y_{iJt})'$. At time t there are K_t discrete latent states for the categorical latent variable c^y_{it}, where $k_t = 1 \ldots K_t$. Typically, in empirical applications $K_t = K$, which corresponds to configural invariance of the discrete latent construct across time. K can be chosen using a model selection approach, where the fit of LTA models positing alternate K is compared. For instance, the Bayesian Information Criterion (BIC; Schwarz, 1978) is one model selection index often used in selecting the number of latent states or classes in mixture models (e.g., Nagin, 2005; Nylund, Asparouhov, & Muthén, 2007; Tofighi & Enders, 2007).

The LTA model implies that the probability of obtaining response pattern $\mathbf{y}_i$ for person i is

$$p(\mathbf{y}_i|\boldsymbol{\theta}_y) = \sum_{k_1=1}^{K_1} \cdots \sum_{k_T=1}^{K_T} \pi_{k_1} \left(\prod_{t=2}^{T} \tau_{k_t|k_{t-1}} \right) \prod_{t=1}^{T} \prod_{j=1}^{J} \left(\rho_{yjt|k_t}^{y_{ijt}} (1-\rho_{yjt|k_t})^{1-y_{ijt}} \right). \tag{1}$$

In Eq. (1), π_{k_1} is the marginal probability of membership in initial latent state k_1 of c^y_{i1}, which can be obtained via the following multinomial logistic specification:

$$\pi_{k_1} = \exp(\omega_{k_1}) \Big/ \sum_{\dot{k}_1=1}^{K_1} \exp(\omega_{\dot{k}_1}), \tag{2}$$

where ω_{k_1} is a multinomial intercept. For identification purposes, $\omega_{K_1} = 0$ for reference state K_1. In Eq. (1), $\tau_{k_t|k_{t-1}}$ is a transition probability from time $t-1$ to time t latent state. Psychologists applying LTA are often most substantively interested in these transition probabilities, which describe latent stage-sequential change. For $t \geq 2$, the probability of transitioning from latent state k_{t-1} of c^y_{it-1} to latent state k_t of c^y_{it} can be obtained from the following multinomial logistic specification (e.g., Reboussin et al., 1998):

$$\tau_{k_t|k_{t-1}} = \frac{\exp\left(\alpha_{k_t} + \boldsymbol{\beta}'_{k_t|c^y_{t-1}} \mathbf{d}_{ik_{t-1}}\right)}{\sum_{\dot{k}_t=1}^{K_t} \exp\left(\alpha_{\dot{k}_t} + \boldsymbol{\beta}'_{\dot{k}_t|c^y_{t-1}} \mathbf{d}_{ik_{t-1}}\right)}, \qquad \text{where} \quad t \geq 2 \tag{3}$$

For identification purposes, $\alpha_{K_t} = 0$ and $\boldsymbol{\beta}_{K_t|c^y_{t-1}} = \mathbf{0}$ for the outcome reference state K_t. α_{k_t} is a multinomial intercept and $\boldsymbol{\beta}_{k_t|c^y_{t-1}}$ is a $(K_{t-1}-1) \times 1$ vector of multinomial slopes for $K_{t-1}-1$ dummy codes in $\mathbf{d}_{ik_{t-1}}$; these dummy codes represent the latent states of c^y_{it-1} at time $t-1$. The reference category for the dummy codes is the K_{t-1} latent state. For example, to compute the probability of transitioning from $c^y_{it-1} = 1$ to $c^y_{it} = 2$ where $K_{t-1} = K_t = 3$, we would have $\boldsymbol{\beta}'_{k_t|c^y_{t-1}} = \left[\beta_{2|c^y_{t-1}=1} \; \beta_{2|c^y_{t-1}=2} \right]$ in Eq. (3). The interpretation of $\exp(\beta_{2|c^y_{t-1}=1})$ is as an odds ratio

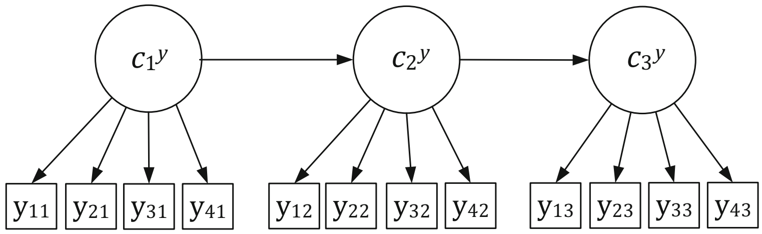

FIGURE 1.
Conventional latent transition analysis (LTA) model that requires assuming MAR (Eq. 1). Shown for $T = 3$ and $J = 4$. *Notes* Circles represent latent classification variables and they are connected by directed arrows representing regression paths. Squares represent measured variable indicators of latent classification variables. *MAR* missingness at random.

(OR). Specifically, it is the multiplicative change in the odds of being in state 2 (vs. K_t) at time t due to being in state 1 (rather than state K_{t-1}) at time $t - 1$. Marginal latent state probabilities for $t > 1$ can be computed using probabilities from previous timepoints. For example,

$$\pi_{k_t} = \sum_{k_{t-1}=1}^{K_{t-1}} \pi_{k_{t-1}} \tau_{k_t|k_{t-1}} \quad \text{where} \quad t > 1. \tag{4}$$

Also in Eq. (1), $\rho_{yjt|k_t}$ is the endorsement probability for the jth outcome at time t for persons in state k_t. $\rho_{yjt|k_t}$ is obtained from the following specification:

$$\rho_{yjt|k_t} = 1/(1 + \exp(\upsilon_{yjt|k_t})). \tag{5}$$

In Eq. (5), $\upsilon_{yjt|k_t}$ is an estimated threshold parameter for the jth outcome at time t for persons in state k_t. Because $\rho_{yjt|k_t}$ describes the relationship between a manifest outcome and a categorical latent variable, it is a measurement parameter; Eq. (5) may be considered a measurement submodel. In contrast, π_{k_1} and $\tau_{k_t|k_{t-1}}$ are structural parameters describing relationships among categorical latent variables; Eqs. (2) and (3) may be considered the structural submodel. Often measurement invariance is assumed, wherein thresholds are constrained equal across time within state. In LTA, the J outcomes at time t are assumed locally independent conditional on latent state k_t. This assumption is reflected in Eq. (1) in that, for members of state k_t, the joint probability of response pattern $\mathbf{y}_{it}$ is the product of response probabilities for outcomes $j = 1 \ldots J$.

Finally, $\boldsymbol{\theta}_y$ is a vector of all estimated model parameters discussed above. This vector includes $\omega_{k_1=1} \ldots \omega_{k_1=(K_1-1)}$. For all $t \geq 2$, this vector also includes $\{\alpha_{k_t=1} \ldots \alpha_{k_t=(K_t-1)}\}$ and $\{\boldsymbol{\beta}_{k_t=1|c^y_{t-1}} \ldots \boldsymbol{\beta}_{k_t=(K_t-1)|c^y_{t-1}}\}$. For all j and t, this vector also includes $\{\upsilon_{yjt|k_t=1} \ldots \upsilon_{yjt|k_t=K_t}\}$. A heuristic path diagram of the Eq. (1) conventional LTA is provided in Fig. 1.

3. Joint MNAR Model for LTA

3.1. Background

To date, there has been scant attention to expanding the LTA into a joint model suitable for accommodating nonignorable missingness. An exception is White and Erosheva (2013) who, in the context of modeling rolling enrollment, employed a model conceptually related to a pattern-mixture model with a LTA specification for the outcome-generating process. However, their model applied only to monotone missingness, required time homogeneity of item endorsement probabilities and transition probabilities, and could not be fit with standard mixture modeling software. None of those restrictions will be necessary in the model developed here.

Although there has been some attention to expanding *single-indicator,* manifest Markov models into joint MNAR models, we do not directly extend any of these specifications in constructing our (multiple-indicator) joint MNAR LTA because interpretational and/or practical drawbacks arise. For instance, one existing MNAR model involves a single-indicator Markov submodel for outcomes and a selection submodel for outcome-dependent missingness (e.g., Albert, 2000; Chen, Yi & Cook, 2011; Cole, Bonetti, Zaslavsky, & Gelber, 2005; Kurland & Heagerty, 2004; Liu, Waternaux & Petkova, 1999). For such an MNAR specification, computational burden is a function of the number of timepoints with missing data. However, if extended from univariate repeated measures (a single-indicator Markov submodel for outcomes) to multivariate repeated measures (a LTA submodel for outcomes), computational burden can quickly become impractical—increasing multiplicatively according to $J\times$ the number of timepoints with missing data. Specifically, obtaining the joint likelihood of outcomes and missingness indicators for such a model requires integration over missing y variables. For the empirical example discussed later (with 8 repeated outcomes each measured at 3 timepoints, where missingness starts at time 2), such a model would require 16 dimensions of integration. In contrast, the joint MNAR LTA proposed here does not require integration.

Another existing MNAR model involves a single-indicator Markov submodel for outcomes along with continuously distributed random coefficient(s) (i.e., latent factor(s)) as shared parameter(s) in the outcome and missingness submodels (e.g., Albert & Follmann, 2003; Yang, Shoptaw, Liu & Belin, 2007). We chose not to extend this approach because the conventional LTA does not involve latent factors within-state in the outcome-generating model. Adding latent factor(s) within-state at time t relaxes the LTA's local independence assumption, which would change and complicate substantive interpretation of latent outcome states (see Sterba, 2013). Outcome process parameters would not retain the same interpretation as in LTA. In contrast, the joint MNAR LTA proposed here retains the LTA's local independence assumption for the outcome process.

3.2. Overview of the Parallel-Process MNAR LTA (MNAR-PP LTA)

First we provide an overview of major features of the MNAR-PP LTA. Later, its specification is described in detail. When responses in $\mathbf{y}_i$ are missing, corresponding elements of $\mathbf{m}_i$ are 1 (else, 0). Intermittent missingness or dropout may occur *only* for one or a subset of the J outcomes at time t (e.g., a sensitive or socially undesirable outcome) and can also occur for all J outcomes at time t. The MNAR-PP LTA accommodates the combination of both, which is most typical of practice.

The MNAR-PP LTA employs a measurement submodel for the missingness indicators wherein elements of $\mathbf{m}_{it}$ are indicators of a categorical latent variable at time t, c_{it}^m. At time t, c_{it}^m can have Q_t latent missingness states, where $q_t = 1 \ldots Q_t$. Associations among missingness indicators at time t are explained by between-state differences in propensities for missingness at time t. Researchers may typically choose to constrain $Q_t = Q$, which affords configural invariance of the latent states across time. Model-building strategies involving the selection of Q will be discussed later; it is also possible for Q to be specified a priori based on theory.

The MNAR-PP LTA extends the conventional LTA by relating the outcome-generating mechanism from the LTA to a missingness mechanism in the following manner. The MNAR-PP LTA allows membership in latent *missingness* states of c_{it}^m at time t to depend on membership in latent *outcome* states at times t and $t-1$ (i.e., c_{it}^y and c_{it-1}^y), as well as latent missingness states at time $t-1$, c_{it-1}^m. There are no direct dependencies between $\mathbf{m}_{it}$ and $\mathbf{y}_{it}$ in the measurement submodels; rather, all dependencies between $\mathbf{m}_{it}$ and $\mathbf{y}_{it}$ arise indirectly through structural relations among latent missingness states and latent outcome states. In the MNAR-PP LTA, there is conditional independence of $\mathbf{m}_{it}$ and $\mathbf{y}_{it}$ given current and prior latent outcome states. For an outcome-generating process involving stage-sequential change among latent states of, for instance, antisocial or victimization behavior, the MNAR-PP LTA posits a highly substantively plausible missingness mechanism. Later, in Sect. 4, it is demonstrated that the application of MNAR-PP LTA can, on average, prevent the increase in bias in LTA parameters when missingness depends on time t and $t-1$ latent outcome states. Importantly, the MNAR-PP LTA and the conventional LTA allow the same interpretation of parameters pertaining to the outcome-generating mechanism.

The MNAR-PP LTA can be related to the taxonomy of joint MNAR modeling frameworks from Sect. 1. Specifically, the MNAR-PP LTA can be considered a kind of *shared parameter* MNAR model because $\mathbf{m}_{it}$ and $\mathbf{y}_{it}$ are conditionally independent given (current and prior) latent outcome states. Additionally, this MNAR-PP LTA can be viewed as an extension of parallel-process growth trajectory MNAR models (e.g., Lin et al., 2009) to the context of discrete, stage-sequential latent change. From another vantage point, the MNAR-PP LTA can be viewed as an extension of a constrained parallel-process LTA (e.g., Flaherty, 2008; Sterba, 2013) to the MNAR context, where one process represents missingness. Next, the MNAR-PP LTA model specification is presented in detail.

3.3. Specification of the MNAR-PP LTA

First we present a specification of the MNAR-PP LTA, Eq. (6), suited for the common situation where there is no missingness at time 1; this situation is encountered in our empirical example. (An alternative more general specification, presented later in Eq. (11), accommodates missingness starting at time 1.)

$$p(\mathbf{y}_i, \mathbf{m}_i|\boldsymbol{\theta}) = \sum_{k_1=1}^{K_1}\cdots\sum_{k_T=1}^{K_T}\sum_{q_2=1}^{Q_2}\cdots\sum_{q_T=1}^{Q_T}\pi_{k_1}\tau_{q_2|k_1,k_2}\left(\prod_{t=2}^{T}\tau_{k_t|k_{t-1}}\right)\left(\prod_{t=3}^{T}\tau_{q_t|k_{t-1},k_t,q_{t-1}}\right)$$
$$\times\prod_{j=1}^{J}\left(\left(\prod_{t=1}^{T}\rho_{yjt|k_t}^{y_{ijt}}(1-\rho_{yjt|k_t})^{1-y_{ijt}}\right)\left(\prod_{t=2}^{T}\rho_{mjt|q_t}^{m_{ijt}}(1-\rho_{mjt|q_t})^{1-m_{ijt}}\right)\right) \quad (6)$$

The sample full data log likelihood can be obtained by taking the log of Eq. (6) then summing across $i = 1\ldots N$, assuming independence across individuals. A heuristic path diagram of the Eq. (6) MNAR-PP LTA is presented in Fig. 2 Panel A, where $J = 4$ and $T = 3$. In Fig. 2, circles represent latent classification variables and they are connected by directed arrows representing regression paths. Squares represent measured variable indicators of latent classification variables.

As in the conventional LTA, in Eq. (6) at time t there are K_t latent outcome states, where $k_t = 1\ldots K_t$. The quantities of substantive interest for interpretation in the outcome-generating submodel— π_{k_t}, $\tau_{k_t|k_{t-1}}$, and $\rho_{jt|k_t}$—are computed and interpreted as in the conventional LTA (see Eqs. 2–5). The new quantities $\tau_{q2|k_1,k_2}$, $\tau_{q_t|k_{t-1},k_t,q_{t-1}}$, and $\rho_{mjt|q_t}$ are defined next.

In Eq. (6), for each timepoint $t > 1$ there are Q_t latent missingness states of c_{it}^{m} where $q_t = 1\ldots Q_t$. Variation in missing data patterns $\mathbf{m}_{it}$ across persons is a prerequisite for fitting the MNAR-PP LTA. In Eq. (6), $\tau_{q2|k_1,k_2}$ is the probability of transitioning into latent missingness state q_2 at time 2 given membership in latent outcome state k_1 at time 1 and k_2 at time 2. $\tau_{q2|k_1,k_2}$ is obtained from the following multinomial logistic regression specification:

$$\tau_{q2|k_1,k_2} = \frac{\exp(\alpha_{q_2} + \boldsymbol{\beta}'_{q2|c_1^y}\mathbf{d}_{ik_1} + \boldsymbol{\beta}'_{q2|c_2^y}\mathbf{d}_{ik_2})}{\sum_{\dot{q}_2=1}^{Q_2}\exp(\alpha_{\dot{q}_2} + \boldsymbol{\beta}'_{\dot{q}2|c_1^y}\mathbf{d}_{ik_1} + \boldsymbol{\beta}'_{\dot{q}2|c_2^y}\mathbf{d}_{ik_2})}. \quad (7)$$

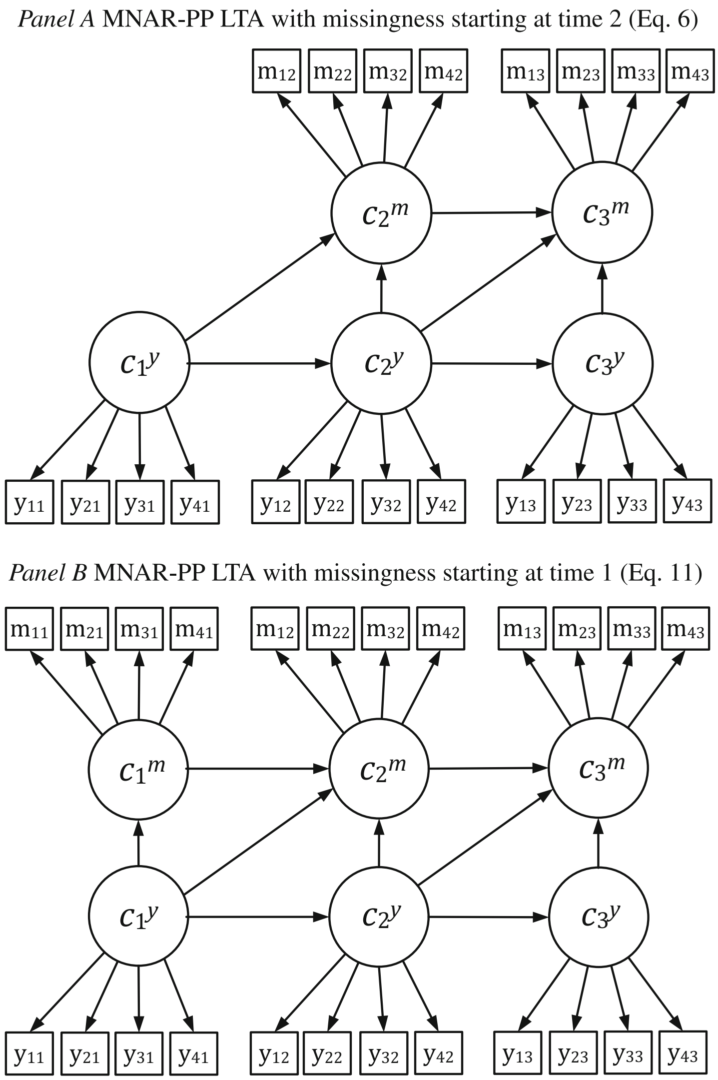

FIGURE 2.
Missing-not-at-random parallel-process latent transition analysis (MNAR-PP LTA) model. Shown for $T = 3$ and $J = 4$. Notes See Fig. 1 notes.

Here, α_{q_2} is a multinomial intercept, $\boldsymbol{\beta}_{q_2|c_1^y}$ is a vector of multinomial slopes for the $K_1 - 1$ dummy codes in $\mathbf{d}_{ik_1}$ (reference category $= K_1$) representing time 1 latent outcome states of c_{i1}^y, and $\boldsymbol{\beta}_{q_2|c_2^y}$ is a vector of multinomial slopes for the $K_2 - 1$ dummy codes in $\mathbf{d}_{ik_2}$(reference category $= K_2$) representing time 2 latent outcome states of c_{i2}^y. For identification, $\alpha_{Q_2} = 0$, $\boldsymbol{\beta}_{Q_2|c_1^y} = \mathbf{0}$, and $\boldsymbol{\beta}_{Q_2|c_2^y} = \mathbf{0}$.

In Eq. (6), when $t \geq 3$, $\tau_{q_t|k_{t-1},k_t,q_{t-1}}$ is the probability of transitioning into the time t latent missingness state q_t conditional on membership in latent outcome states k_{t-1} and k_t at times $t - 1$ and t, and conditional on membership in latent missingness state q_{t-1} at time $t - 1$. Again, $\tau_{q_t|k_{t-1},k_t,q_{t-1}}$ is obtained from a multinomial logistic regression specification:

$$\tau_{q_t|k_{t-1},k_t,q_{t-1}} = \frac{\exp\left(\alpha_{q_t} + \boldsymbol{\beta}'_{q_t|c^y_{t-1}}\mathbf{d}_{ik_{t-1}} + \boldsymbol{\beta}'_{q_t|c^y_t}\mathbf{d}_{ik_t} + \boldsymbol{\beta}'_{q_t|c^m_{t-1}}\mathbf{d}_{iq_{t-1}}\right)}{\sum_{\dot{q}_t=1}^{Q_t}\exp\left(\alpha_{\dot{q}_t} + \boldsymbol{\beta}'_{\dot{q}_t|c^y_{t-1}}\mathbf{d}_{ik_{t-1}} + \boldsymbol{\beta}'_{\dot{q}_t|c^y_t}\mathbf{d}_{ik_t} + \boldsymbol{\beta}'_{\dot{q}_t|c^m_{t-1}}\mathbf{d}_{iq_{t-1}}\right)}, \quad \text{where } t \geq 3. \tag{8}$$

Here, α_{q_t} is a multinomial intercept and $\boldsymbol{\beta}_{q_t|c^y_{t-1}}$, $\boldsymbol{\beta}_{q_t|c^y_t}$, and $\boldsymbol{\beta}_{q_t|c^m_{t-1}}$ are vectors of multinomial slopes for dummy codes in $\mathbf{d}_{ik_{t-1}}$ (reference category $=K_{t-1}$), $\mathbf{d}_{ik_t}$ (reference category $= K_t$), and $\mathbf{d}_{iq_{t-1}}$ (reference category $=Q_{t-1}$). These, respectively, represent outcome states of c^y_{it-1}, outcome states of c^y_{it}, and missingness states of c^m_{it-1}. For identification, $\alpha_{Q_t} = 0$, $\boldsymbol{\beta}_{Q_t|c^y_{t-1}} = \mathbf{0}$, $\boldsymbol{\beta}_{Q_t|c^y_t} = \mathbf{0}$, and $\boldsymbol{\beta}_{Q_t|c^m_{t-1}} = \mathbf{0}$.

In Eq. (6), $\rho_{mjt|q_t}$ is the probability that $m_{ijt} = 1$(i.e., the probability that $y_{ijt} = .$) given membership in latent missingness state q_t at time t. It is obtained from

$$\rho_{mjt|q_t} = 1/(1 + \exp(\upsilon_{mjt|q_t})), \tag{9}$$

$\upsilon_{mjt|q_t}$ is an estimated threshold parameter for the jth missingness indicator at time t in missingness state q_t. Measurement invariance of these thresholds across time within-state, $\upsilon_{mjt|q_t} = \upsilon_{mj|q_t}$, can be imposed for parsimony; this is done later to allow latent missingness states to have the same interpretation across time. At time t, elements of $\mathbf{m}_{it}$ are assumed locally independent conditional on latent state q_t in Eq. (6).

Finally, $\boldsymbol{\theta}$ on the left-hand side of Eq. (6) is a vector containing all estimated model parameters, where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_y, \boldsymbol{\theta}_m\}$. Parameters in $\boldsymbol{\theta}_y$ are familiar from the conventional LTA and now contain parameters of the outcome-generating submodel in the MNAR-PP LTA. All remaining MNAR-PP LTA parameters are contained in $\boldsymbol{\theta}_m$, and they constitute parameters of the missingness-generating submodel. The $\boldsymbol{\theta}_y$ are the focus of substantive interpretation.

3.4. *Other Probabilities of Interest in the Missingness Submodel of the MNAR-PP LTA*

Certain latent state probabilities in the missingness-generating submodel of the MNAR-PP LTA are not directly represented in Eq. (6) but may nevertheless be of interpretive interest. Here we discuss how several key examples—π_{q_t}, $\tau_{q_t|q_{t-1}}$, and $\tau_{q_t|k_t}$—are computed from already-presented probabilities.

Some latent missingness states likely will be more prevalent than others at time t. π_{q_t} is the marginal probability of membership in latent missingness state q_t. Table 1, row 1, illustrates the computation of π_{q_2}, π_{q_3}, and π_{q_z}, where $z \geq 4$.

Another kind of probability of potential interpretive interest is $\tau_{q_t|q_{t-1}}$, the conditional probability of transitioning into latent missingness state q_t from latent missingness state q_{t-1}. Examining $\tau_{q_t|q_{t-1}}$ allows assessment of stability in the prevalence of missingness states across time. Sequential change in latent missingness states can also be assessed—for instance, from a state with elevated missingness probabilities only for illicit drug use outcomes to a state with high missingness probability for all outcomes. Table 1, row 2, provides formulas for computing transition probabilities $\tau_{q_3|q_2}$, $\tau_{q_4|q_3}$, and $\tau_{q_z|q_{z-1}}$ (where $z \geq 5$).

Finally, although the MNAR-PP LTA expression in Eq. (6) already contains latent missingness state probabilities conditional on *both* current and prior latent outcome states, the probability of latent missingness state q_t given *only* current latent outcome state k_t, $\tau_{q_z|k_z}$, may also be of interest. For example, we may wish to learn which missingness states at time t are implied by the model to be highly associated with a high-risk outcome state at time t. Table 1, row 3, shows how transition probabilities $\tau_{q_2|k_2}$, $\tau_{q_3|k_3}$, and $\tau_{q_z|k_z}$(where $z \geq 4$) are computed.

TABLE 1.
Calculating probabilities of interest in the missingness submodel of the MNAR-PP LTA that are not directly represented in Eq. (6).

π_{q_t}, probability of membership in latent missingness state q_t	
At time 2:	$\pi_{q_2} = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \tau_{k_2 \mid k_1}$
At time 3:	$\pi_{q_3} = \sum_{k_1=1}^{K_1} \ldots \sum_{k_3=1}^{K_3} \sum_{q_2=1}^{Q_2} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \left(\prod_{t=2}^{3} \tau_{k_t \mid k_{t-1}} \right) \tau_{q_3 \mid k_2,k_3,q_2}$
At time z : (where $z \geq 4$)	$\pi_{q_z} = \sum_{k_1=1}^{K_1} \ldots \sum_{k_z=1}^{K_z} \sum_{q_2=1}^{Q_2} \ldots \sum_{q_{z-1}=1}^{Q_{z-1}} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \left(\prod_{t=2}^{z} \tau_{k_t \mid k_{t-1}} \right) \left(\prod_{t=3}^{z} \tau_{q_t \mid k_{t-1},k_t,q_{t-1}} \right)$
$\tau_{q_t \mid q_{t-1}}$, probability of transitioning to latent missingness state q_t from latent missingness state q_{t-1}	
From time 2 to 3:	$\tau_{q_3 \mid q_2} = \left(\sum_{k_1=1}^{K_1} \ldots \sum_{k_3=1}^{K_3} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \left(\prod_{t=2}^{3} \tau_{k_t \mid k_{t-1}} \right) \tau_{q_3 \mid k_2,k_3,q_2} \right) \Big/ \pi_{q_2}$
From time 3 to 4:	$\tau_{q_4 \mid q_3} = \left(\sum_{k_1=1}^{K_1} \ldots \sum_{k_4=1}^{K_4} \sum_{q_2=1}^{Q_2} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \left(\prod_{t=2}^{4} \tau_{k_t \mid k_{t-1}} \right) \left(\prod_{t=3}^{4} \tau_{q_t \mid k_{t-1},k_t,q_{t-1}} \right) \right) \Big/ \pi_{q_3}$
From time $z-1$ to z: (where $z \geq 5$)	$\tau_{q_z \mid q_{z-1}} = \left(\sum_{k_1=1}^{K_1} \ldots \sum_{k_z=1}^{K_z} \sum_{q_2=1}^{Q_2} \ldots \sum_{q_{z-2}=1}^{Q_{z-2}} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \left(\prod_{t=2}^{z} \tau_{k_t \mid k_{t-1}} \right) \left(\prod_{t=3}^{z} \tau_{q_t \mid k_{t-1},k_t,q_{t-1}} \right) \right) \Big/ \pi_{q_{z-1}}$
$\tau_{q_z \mid k_z}$, probability of membership in latent missing state q_t given only current latent outcome state k_t	
At time 2:	$\tau_{q_2 \mid k_2} = \left(\sum_{k_1=1}^{K_1} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \tau_{k_2 \mid k_1} \right) \Big/ \pi_{k_2}$
At time 3:	$\tau_{q_3 \mid k_3} = \left(\sum_{k_1=1}^{K_1} \ldots \sum_{k_2=1}^{K_2} \sum_{q_2=1}^{Q_2} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \left(\prod_{t=2}^{3} \tau_{k_t \mid k_{t-1}} \right) \tau_{q_3 \mid k_2,k_3,q_2} \right) \Big/ \pi_{k_3}$
At time z: (where $z \geq 4$)	$\tau_{q_z \mid k_z} = \left(\sum_{k_1=1}^{K_1} \ldots \sum_{k_{z-1}=1}^{K_{z-1}} \sum_{q_2=1}^{Q_2} \ldots \sum_{q_{z-1}=1}^{Q_{z-1}} \pi_{k_1} \tau_{q_2 \mid k_1,k_2} \left(\prod_{t=2}^{z} \tau_{k_t \mid k_{t-1}} \right) \left(\prod_{t=3}^{z} \tau_{q_t \mid k_{t-1},k_t,q_{t-1}} \right) \right) \Big/ \pi_{k_{z-1}}$

MNAR-PP LTA Missing-not-at-random parallel-process LTA.

3.5. A MAR LTA as a Constrained Special Case of the MNAR-PP LTA

Constraints can be placed on the MNAR-PP LTA in order to yield a LTA that requires assuming MAR when fit with, for instance, the EM algorithm. If Eq. (6) is modified such that $\tau_{q_2 \mid k_1,k_2} = \pi_{q_2}$ and modified such that, for $t > 2$, $\tau_{q_t \mid k_{t-1},k_t,q_{t-1}} = \tau_{q_t \mid q_{t-1}}$, then submodels for the missingness mechanism and outcome-generating mechanism would no longer be associated. This is shown in Eq. (10).

$$p(\mathbf{y}_i, \mathbf{m}_i \mid \boldsymbol{\theta}) = \sum_{k_1=1}^{K_1} \cdots \sum_{k_T=1}^{K_T} \sum_{q_2=1}^{Q_2} \cdots \sum_{q_T=1}^{Q_T} \pi_{k_1} \pi_{q_2} \left(\prod_{t=2}^{T} \tau_{k_t \mid k_{t-1}} \right) \left(\prod_{t=3}^{T} \tau_{q_t \mid q_{t-1}} \right) \times \prod_{j=1}^{J} \left(\left(\prod_{t=1}^{T} \rho_{yjt \mid k_t}^{y_{ijt}} (1 - \rho_{yjt \mid k_t})^{1-y_{ijt}} \right) \left(\prod_{t=2}^{T} \rho_{mjt \mid q_t}^{m_{ijt}} (1 - \rho_{mjt \mid q_t})^{1-m_{ijt}} \right) \right) \quad (10)$$

That is, current latent outcome state membership (k_t) would depend *only* on prior latent outcome state membership (k_{t-1}), and current latent missingness state membership (q_t) would depend *only* on prior latent missingness state membership (q_{t-1}). Here $\rho_{mjt \mid q_t}$ depends neither directly nor indirectly on unobservables in the outcome-generating process; this is consistent with MAR.

Since the MAR LTA in Eq. (10) and the MNAR-PP LTA in Eq. (6) have the same set of dependent variables, it would be possible to compare their fit as part of a sensitivity analysis.

(In contrast, the conventional LTA in Eq. (1) does not have the same set of dependent variables as Eq. (6) so the fit of these models could not be compared.) Of course a MAR versus MNAR-PP LTA model fit comparison should not be considered a *general* test of MNAR because the MNAR-PP LTA posits only one substantively plausible MNAR mechanism out of many possible MNAR mechanisms. Indeed, for this reason no MNAR model specification allows a general test of MNAR (Little & Rubin, 2002). Additionally, there is a debate over the usefulness of MAR versus MNAR model fit comparisons in general because fit is being compared only with respect to the observed data (see Ibrahim, Chen, Lipsitz, & Herring, 2005; Jansen, Hens, & Molenberghs, 2006; Muthén et al., 2011; Sterba & Gottfredson, 2014). Here, in later examples, we conservatively elect to only compare parameter estimates across MAR versus MNAR-PP LTA specifications, as this is always a main feature of sensitivity analyses pertaining to missingness mechanisms.

3.6. *Additional Specifications of the MNAR-PP LTA*

The MNAR-PP-LTA specification in Eq. (6) accommodated missingness beginning at time 2; this is an extremely common situation in practice. In Eq. (11) an additional specification of the MNAR-PP LTA is presented where missingness begins at time 1.

$$p(\mathbf{y}_i, \mathbf{m}_i|\boldsymbol{\theta}) = \sum_{k_1=1}^{K_1}\sum_{q_1=1}^{Q_1}\cdots\sum_{k_T=1}^{K_T}\sum_{q_T=1}^{Q_T}\pi_{k_1}\tau_{q_1|k_1}\left(\prod_{t=2}^{T}\tau_{k_t|k_{t-1}}\tau_{q_t|k_{t-1},k_t,q_{t-1}}\right) \times\prod_{t=1}^{T}\prod_{j=1}^{J}\left(\rho_{yjt|k_t}^{y_{ijt}}(1-\rho_{yjt|k_t})^{1-y_{ijt}}\rho_{mjt|q_t}^{m_{ijt}}(1-\rho_{mjt|q_t})^{1-m_{ijt}}\right) \quad (11)$$

Even if the baseline occasion in a study (for instance, grade 4) contains complete data, this additional MNAR-PP LTA specification can be relevant if the timepoints used in an analysis (for instance, grades 7, 8, and 9) do not include the baseline occasion (e.g., Rodgers et al., 2013). Figure 2 Panel B gives a heuristic path diagram of Eq. (11) MNAR-PP LTA for $J = 4$ and $T = 3$.

4. Simulation Demonstration

Now that the MNAR-PP LTA has been described, we use a small simulation as a proof-of-concept demonstration of the following two points. (1) As compared to fitting a conventional LTA assuming MAR, the MNAR-PP LTA can reduce bias in LTA parameters when there is latent-state-dependent missingness (e.g., latent missingness state at time t that depends on latent outcome states at t and/or $t-1$). (2) When a researcher suspects MNAR missingness, but in fact missingness is MAR, unnecessarily fitting a MNAR-PP LTA can still allow parameters to be recovered as well as if a conventional LTA were fit, on average.

First we review which parameters are at risk of bias when the missingness mechanism is latent-state-dependent MNAR, but a MAR LTA is fit instead. Outcome process structural coefficients are at little risk of bias due to MNAR when they involve timepoints at which there is no missingness. For instance, if MNAR missingness begins at $t = 2$, multinomial coefficients used in computing $\tau_{k_t|k_{t-1}}$ and π_{k_t} for $t \geq 2$ are at greatest risk of bias, whereas those used in computing π_{k_1} are at little risk of bias because there is no missingness at t =1. (If missingness begins at $t = 1$—not considered here—multinomial intercepts used to compute π_{k_1} can also be meaningfully biased). Measurement parameters $\rho_{yjt|k_t}$ should show little bias on average since the dependency between outcome-generating and missingness mechanisms occurs in the structural model.

TABLE 2.
Simulation results: percent absolute relative bias (%ARB) for multinomial coefficient structural parameters in the outcome process.

Parameter	Pop. value	MNAR missingness mechanism				MAR missingness mechanism			
		Fit MNAR-PP LTA		Fit conventional LTA		Fit MNAR-PP LTA		Fit conventional LTA	
		Avg. Est.	%ARB	Avg. Est.	%ARB	Avg. Est.	%ARB	Avg. Est.	%ARB
$\omega_{k_1=1}$	−1.2	−1.198	0.14	−1.248	3.99	−1.203	0.23	−1.204	0.29
$\omega_{k_1=2}$	−0.9	−0.884	1.76	−0.902	0.27	−0.890	1.16	−0.890	1.11
$\alpha_{k_2=1}$	−1.2	−1.203	0.24	−1.316	9.63	−1.207	0.57	−1.206	0.52
$\alpha_{k_2=2}$	−0.9	−0.901	0.16	−0.985	9.40	−0.879	2.32	−0.878	2.49
$\alpha_{k_3=1}$	−1.2	−1.204	0.35	−1.382	15.13	−1.211	0.94	−1.213	1.09
$\alpha_{k_3=2}$	−0.9	−0.896	0.49	−0.965	7.19	−0.902	0.20	−0.899	0.13
$\beta_{k_2=1\mid c_1^y=1}$	2.0	2.017	0.87	1.816	9.20	2.018	0.92	2.019	0.97
$\beta_{k_2=1\mid c_1^y=2}$	1.0	1.005	0.49	0.768	23.20	1.003	0.28	1.002	0.23
$\beta_{k_2=2\mid c_1^y=1}$	1.0	0.985	1.52	0.716	28.44	0.990	1.02	0.991	0.89
$\beta_{k_2=2\mid c_1^y=2}$	1.5	1.517	1.13	1.301	13.25	1.501	0.08	1.504	0.27
$\beta_{k_3=1\mid c_2^y=1}$	1.75	1.768	1.03	1.656	5.35	1.780	1.69	1.780	1.70
$\beta_{k_3=1\mid c_2^y=2}$	1.0	1.001	0.11	0.760	23.98	1.010	0.98	1.011	1.12
$\beta_{k_3=2\mid c_2^y=1}$	1.0	1.002	0.23	0.765	23.52	1.011	1.13	1.012	1.20
$\beta_{k_3=2\mid c_2^y=2}$	1.5	1.509	0.61	1.292	13.89	1.509	0.61	1.510	0.67

LTA latent transition analysis, *MNAR-PP LTA* missing-not-at-random parallel-process LTA, *Pop. value* population parameter value, *MAR* missing-at-random, *Avg. Est.* average estimate.

4.1. Methods

In this simulation demonstration, 500 samples with $T = 3$ and $J = 5$ were generated; these T and J are typical of LTA applications in psychology (e.g., Cain, Epler, Steinley, & Sher 2012; Catts, Compton, Tomblin, & Bridges, 2012; Collins & Lanza, 2010; Rodgers et al., 2013; Witkiewitz, 2008). Repeated measures and missingness were generated from the MNAR-PP LTA in Eq. (6) and the MAR LTA in Eq. (10). In both population models, outcome indicators and missingness indicators were measurement invariant over time and latent outcome states and latent missingness states were configurally invariant over time ($K = 3$, $Q = 2$). Complete case data consisted of $N = 5{,}000$ cases per sample. Similar $N's$ are common in LTA applications (e.g., recently Baggio et al., 2014, Cook, Pflieger, Connell, & Connell, 2014, and Kroesen, 2014).

In both population models, parameters for the outcome submodel were the same. For the outcome submodel, structural parameters are given in Table 2 (column 1). Measurement parameters, in Table 3 (column 1), were chosen to be similar to Nylund et al. (2006).

In both population models, parameters for the missingness submodel implied no missingness at $t = 1$, on average 25.5 % missingness at $t = 2$, and on average 34.5 % missingness at $t = 3$. These missingness proportions are similar to those in the empirical example described later. In *both* population models, measurement parameters for the missingness submodel were the same:

$$[\upsilon_{mjt|q_t=1} \ldots \upsilon_{mJt|q_t=1}] = [-.76, -.46, -.92, -.60, -.65];$$
$$[\upsilon_{mjt|q_t=2} \ldots \upsilon_{mJt|q_t=2}] = [1.62, 2.25, 1.77, 2.50, 2.00].$$

TABLE 3.
Simulation results: percent absolute relative bias (%ARB) for threshold measurement parameters in the outcome process.

Parameter	Pop. value	MNAR missingness mechanism				MAR missingness mechanism			
		Fit MNAR-PP LTA		Fit conventional LTA		Fit MNAR-PP LTA		Fit conventional LTA	
		Avg. Est.	%ARB	Avg. Est.	%ARB	Avg. Est.	%ARB	Avg. Est.	%ARB
$\upsilon_{y1t\|k_t=1}$	−0.90	−0.905	0.58	−0.944	4.91	−0.904	0.46	−0.904	0.48
$\upsilon_{y2t\|k_t=1}$	−1.69	−1.709	1.13	−1.785	5.63	−1.714	1.43	−1.716	1.51
$\upsilon_{y3t\|k_t=1}$	−2.20	−2.217	0.79	−2.276	3.45	−2.209	0.42	−2.210	0.45
$\upsilon_{y4t\|k_t=1}$	−1.25	−1.244	0.49	−1.296	3.66	−1.251	0.10	−1.252	0.14
$\upsilon_{y5t\|k_t=1}$	−1.48	−1.487	0.44	−1.543	4.26	−1.492	0.81	−1.493	0.84
$\upsilon_{y1t\|k_t=2}$	0.32	0.332	3.63	0.293	8.47	0.327	2.09	0.326	1.81
$\upsilon_{y2t\|k_t=2}$	0.95	0.962	1.22	0.896	5.67	0.960	1.09	0.959	0.97
$\upsilon_{y3t\|k_t=2}$	−0.20	−0.188	5.90	−0.225	12.45	−0.201	0.30	−0.202	0.80
$\upsilon_{y4t\|k_t=2}$	0.15	0.152	1.20	0.106	29.33	0.151	0.87	0.151	0.40
$\upsilon_{y5t\|k_t=2}$	0.55	0.560	1.89	0.512	6.93	0.553	0.62	0.553	0.45
$\upsilon_{y1t\|k_t=3}$	2.10	2.107	0.35	2.083	0.80	2.107	0.31	2.106	0.30
$\upsilon_{y2t\|k_t=3}$	2.67	2.679	0.32	2.647	0.86	2.671	0.05	2.671	0.04
$\upsilon_{y3t\|k_t=3}$	1.80	1.807	0.37	1.782	0.99	1.810	0.53	1.809	0.52
$\upsilon_{y4t\|k_t=3}$	2.55	2.566	0.61	2.538	0.48	2.571	0.82	2.570	0.78
$\upsilon_{y5t\|k_t=3}$	2.30	2.302	0.08	2.274	1.12	2.311	0.49	2.311	0.47

LTA latent transition analysis, *MNAR-PP LTA* missing-not-at-random parallel-process LTA, *Pop. value* population parameter value, *MAR* missing-at-random, *Avg. Est.* average estimate.

They imply that in state 1 at time t missingness across the J items averaged 66 % and in state 2 at time t missingness averaged 12 %. In the generating MNAR-PP LTA from Eq. (6), structural parameters for the missingness submodel were $\alpha_{q2=1} = -4$, $\boldsymbol{\beta}'_{q2=1|c_1^y} = [2.75, 2.00]$, $\boldsymbol{\beta}'_{q2=1|c_2^y} = [2.50, 2.00]$, $\alpha_{q3=1} = -3.25$, $\boldsymbol{\beta}'_{q3=1|c_2^y} = [2.75, 1.75]$, $\boldsymbol{\beta}'_{q3=1|c_3^y} = [2.75, 1.75]$ and $\boldsymbol{\beta}'_{q3=1|c_2^m} = [0.5]$. For the generating MAR LTA in Eq. (10), structural parameters for the missingness submodel were $\alpha_{q2=1} = -1.08$, $\alpha_{q3=1} = -0.47$, $\boldsymbol{\beta}'_{q3=1|c_2^m} = [0.5]$. M*plus* 7.11 (Muthén & Muthén, 1998-2014) and SAS 9.3 were used for generation.

Each sample (generated from either the MNAR-PP LTA or the [MAR] LTA) was then fit with *both* the MNAR-PP LTA in Eq. (6) and the conventional (MAR) LTA in Eq. (1). Hence, for each sample, one fitted model is true and the other fitted model has a misspecified missingness mechanism. (Here, a misspecified missingness mechanism arises when a MNAR mechanism is specified but missingness is actually MAR, or when a MAR mechanism is specified but missingness is actually MNAR.) Model fitting employed the EM algorithm (Dempster, Laird, & Rubin, 1977) in M*plus* 7.11. See Song and Lee 2003 for a description of the EM for mixture models with missing outcomes. For a generic parameter ϑ, percent absolute relative bias (%ARB) was computed as $\left|\left((\bar{\hat{\vartheta}} - \vartheta)/\vartheta\right) \times 100\right|$. Though not a focus of the simulation demonstration, results for standard error bias $\left|\left(\overline{SE}_\vartheta - SD_\vartheta\right)/SD_\vartheta\right) \times 100\right|$ will also be briefly discussed (with full results in the Online Appendix). Here, $\overline{SE}_\vartheta$ is the average analytic standard error for the generic parameter ϑ, and SD_ϑ is its empirical repeated sampling standard deviation.

TABLE 4.
Simulation results: percent absolute relative bias (%ARB) for multinomial coefficient structural parameters and threshold measurement parameters in the missingness process (available only when MNAR-PP LTA is fit).

Parameter	MNAR missingness mechanism			MAR missingness mechanism		
	Fit MNAR-PP LTA			Fit MNAR-PP LTA		
	Pop. value	Avg. Est.	%ARB	Pop. value	Avg. Est.	%ARB
$\alpha_{q2=1}$	−4.00	−4.083	2.07	−1.08	−1.078	0.19
$\alpha_{q3=1}$	−3.25	−3.349	3.04	−0.47	−0.473	0.66
$\beta_{q2=1\|c_1^y=1}$	2.75	2.792	1.52	0	0.003	*
$\beta_{q2=1\|c_1^y=2}$	2.00	2.038	1.90	0	0.005	*
$\beta_{q2=1\|c_2^y=1}$	2.50	2.553	2.12	0	−0.019	*
$\beta_{q2=1\|c_2^y=2}$	2.00	2.036	1.80	0	−0.032	*
$\beta_{q3=1\|c_2^m=1}$	0.50	0.513	2.54	0.50	0.508	1.54
$\beta_{q3=1\|c_2^y=1}$	2.75	2.782	1.17	0	0.003	*
$\beta_{q3=1\|c_2^y=2}$	1.75	1.767	0.99	0	0.002	*
$\beta_{q3=1\|c_3^y=1}$	2.75	2.836	3.11	0	0.000	*
$\beta_{q3=1\|c_3^y=2}$	1.75	1.827	4.41	0	−0.022	*
$\upsilon_{m1t\|q_t=1}$	−0.76	−0.760	0.01	−0.76	−0.761	0.16
$\upsilon_{m2t\|q_t=1}$	−0.46	−0.458	0.48	−0.46	−0.457	0.59
$\upsilon_{m3t\|q_t=1}$	−0.92	−0.921	0.05	−0.92	−0.920	0.01
$\upsilon_{m4t\|q_t=1}$	−0.60	−0.600	0.08	−0.60	−0.599	0.17
$\upsilon_{m5t\|q_t=1}$	−0.65	−0.651	0.15	−0.65	−0.653	0.42
$\upsilon_{m1t\|q_t=2}$	1.62	1.622	0.12	1.62	1.623	0.17
$\upsilon_{m2t\|q_t=2}$	2.25	2.250	0.01	2.25	2.251	0.06
$\upsilon_{m3t\|q_t=2}$	1.77	1.773	0.18	1.77	1.770	0.02
$\upsilon_{m4t\|q_t=2}$	2.50	2.501	0.06	2.50	2.504	0.17
$\upsilon_{m5t\|q_t=2}$	2.00	2.004	0.19	2.00	2.004	0.18

* Cannot calculate %ARB for this cell because the population parameter is 0 and we cannot divide by 0.
MNAR-PP LTA missing-not-at-random parallel-process LTA, *Pop. value* population parameter value, *MAR* missing-at-random, *Avg. Est.* average estimate.

4.2. Results

Average estimates for multinomial coefficients and thresholds are given in Tables 2, 3, 4. They were converted into probabilities provided in supplementary material (Online Appendix Tables OA.3-OA.6). For instance, in the outcome process, multinomial coefficient results from Table 2 were used to compute the probabilities $\tau_{k2|k_1}$, $\tau_{k3|k_2}$, π_{k_1}, π_{k_2}, and π_{k_3} given in Online Appendix Table OA.3. In the missingness process, many different probabilities can be constructed from multinomial coefficient estimates in Table 4, depending on the substantive interest of the researcher (see Sects 3.3, 3.4). Online Appendix Table OA.5 provides a subset of these probabilities: $\tau_{q2|k_1,k_2}$ and $\tau_{q3|k_2,k_3,q_2}$.

More than 99 % of samples converged without estimation problems. The result of main interest was %ARB for the multinomial coefficient structural parameters in the conventional LTA that pertain to timepoints ≥2 (since there was no missingness at $t = 1$ and thus little to no possibility for MNAR missingness to bias ω_{k_1}).

When the missingness mechanism is latent-state-dependent MNAR for missingness starting at $t = 2$, Table 2 shows that simply fitting a conventional LTA yields an average of 15 % ARB for structural parameters of the outcome process pertaining to timepoints ≥ 2; however, fitting a MNAR-PP LTA reduces this bias to <1 % on average. As expected, measurement parameters in the outcome process had less bias than structural parameters when fitting a conventional LTA under this latent-state-dependent MNAR—on average, 6 % ARB in Table 3. This bias was again reduced by fitting a MNAR-PP LTA (to 1 % ARB on average in Table 3).

When the missingness mechanism is actually MAR, for missingness starting at $t = 2$, mistakenly /unnecessarily fitting a MNAR-PP LTA resulted in equivalent parameter recovery to fitting a conventional LTA. Under MAR, both the fitted MNAR-PP LTA and conventional LTA yielded on average <1 % ARB for structural parameters, in Table 2, and measurement parameters, in Table 3.

Furthermore, Table 4 shows accurate recovery of missingness process parameters when fitting a MNAR-PP LTA. When missingness was MNAR, fitting the MNAR-PP LTA yielded average %ARB of 2 % for structural parameters of the missingness process and <1 % for measurement parameters of the missingness process. When missingness was MAR, multinomial slopes relating missingness and outcome processes were correctly estimated to be near-zero in the fitted MNAR-PP LTA (see Table 4).

Though not the focus of the simulation demonstration, for each estimated parameter, the average analytic standard error, empirical standard deviation, and standard error bias are provided in Online Appendix Tables OA.7, OA.8, and OA.9. Coverage of nominal 95 % interval estimates is provided in Online Appendix Tables OA.10, OA.11, and OA.12. Recall that the conventional LTA and MNAR-PP LTA both retain all available cases; thus, neither suffers from efficiency loss that would be expected due to listwise deletion. Standard error bias was similar across cells of the design–regardless of generating and fitted models. For measurement parameters of the outcome or missingness process %ARB for standard errors ranged from 3 to 4 %. For structural parameters of the outcome or missingness process, %ARB for standard errors ranged from 2 to 4 %.

4.3. Summary of Simulation Results

This simulation demonstration illustrated two points. (1) Latent-state-dependent MNAR missingness can induce an increase in bias for a conventional LTA, mainly in structural parameters, and this increase in bias can be averted by fitting a MNAR-PP LTA. (2) When missingness is truly MAR, mistakenly/unnecessarily fitting a MNAR-PP LTA can still allow LTA parameters to be recovered as well as in the conventional LTA, on average. That is, under MAR missingness, there was little change in outcome process parameter estimates between fitting a conventional LTA versus MNAR-PP LTA. This simulation suggests that if a researcher substantively suspects missingness to be dependent on time t or $t-1$ outcome states, a sensitivity analysis comparing estimates from a MNAR-PP LTA and a conventional (MAR) LTA model could be helpful.

Regarding generalizability, when fitting a conventional (MAR) LTA to the MNAR-PP LTA generated data, parameter bias would have been larger if, for instance, latent missingness states were more separated and/or latent outcome and missingness states were more highly related. Latent missingness states are more separated when there are greater differences in the probability of missingness for item j, $\rho_{mjt|q_t}$, across latent missingness states at time t.

4.4. Extensions

The focus of this simulation demonstration was on the two points mentioned in the above summary section. This subsection briefly considers four extension topics to the simulation demonstration: the selection of K and Q; the implications of misspecifying Q; the implications of fitting the MNAR-PP LTA when missingness state at time t depends on outcome state *only* at time t (not at time $t-1$); and an alternate N.

First we consider the selection of K and Q. A two-step model-building approach is commonly used in constructing parallel-process longitudinal mixtures in other contexts (e.g., Bray, Lanza, & Collins, 2010; Flaherty, 2008; Jackson, Sher, & Schulenberg, 2005; Nagin, 2005; Nagin & Tremblay, 2001; Sterba, 2013; Witkiewitz & Villarroel, 2009). In this two-step strategy, first, the number of latent states is selected in each process separately and, next, the best-fitting single-process models are combined into a final parallel-process model. Under the latent-state-dependent MNAR mechanism considered here, departures from MAR arise at the structural level, not at the measurement (i.e., within-state submodel) level. This fact could aid the performance of such two-step model-building strategies for the MNAR-PP LTA. Here we empirically examine recovery of correct K and Q when the best-fitting K and best-fitting Q are separately determined using single-process models. For this examination, we use data from the earlier simulation demonstration generated from a $K = 3$ $Q = 2$ MNAR-PP LTA. The best-fitting K and Q were separately determined from single-process models using a within-sample ranking of BIC and AIC (Akaike, 1974) (lower is better). $K = 2, 3,$ and 4 and $Q = 1, 2, 3,$ and 4 were considered.[2] Among the ≥ 495 samples that converged across all numbers of states, BIC preferred $K = 3$ and $Q = 2$ in $\geq$99 % of samples whereas AIC preferred $K = 3$ in 82 % of samples and $Q = 2$ in 76 % of samples (in remaining samples, AIC preferred $K = 4$ and $Q = 3$). In sum, the two-stage model-building approach shows promise for constructing MNAR-PP LTAs.

Second, we consider the implications of misspecifying the number of latent missingness states (Q) for recovering outcome process parameters. In the context of the simulation demonstration provided earlier, underspecifying Q (i.e., $Q = 1$) should provide results similar to assuming MAR. (Multinomial coefficients relating outcome and missingness processes cannot be estimated when $Q = 1$). Thus, underspecifying Q could interfere with recovering outcome process parameters. On the other hand, overspecifying Q would not necessarily interfere with recovering outcome process parameters because the overspecified Q, if estimable, could still approximate the dominant patterns of latent missingness state differences. We fit a misspecified $K = 3$, $Q = 3$ MNAR-PP LTA to data generated from the $K = 3$, $Q = 2$ MNAR-PP LTA from the original demonstration. For structural and measurement parameters of the outcome process, %ARB is provided in Tables 5 and 6 (left columns), and is computed using samples that converged without estimation errors (77 %). Average %ARB for structural and measurement parameters (1 %) was comparable to simulation results presented earlier.

Next, we consider consequences of fitting the MNAR-PP LTA under another misspecification of the missingness mechanism. Specifically, suppose that missingness state at time t depends on outcome state at time t but does *not* depend on outcome state at $t - 1$—even though such dependency is allowed by the MNAR-PP LTA. Tables 5 and 6 (right-side columns) provide results from fitting our $K = 3$, $Q = 2$ MNAR-PP LTA when, in actuality, missingness state at time t does not depend on outcome state at time $t - 1$. %ARB was comparable to the earlier presented simulation results (on average 1 %ARB) because here the true model is a more restrictive version of the MNAR-PP LTA being fit. $\beta_{q_2=1|c_1^y=1}$, $\beta_{q_2=1|c_1^y=2}$, $\beta_{q_3=1|c_2^y=1}$, and $\beta_{q_3=1|c_2^y=2}$ paths were estimated at near-zero values (on average, .008, .011, −.009, and .022, respectively). In sum, the MNAR-PP LTA displays some robustness when the specified missingness submodel is more general than the latent-state-dependent mechanism which generated the missingness.

Finally, we consider an alternative number of cases, $N = 1{,}000$, that is also commonly used in LTA applications (e.g., recently Lee & Vondracek, 2014; Meier et al., 2013; Soto-Ramirez et al., 2013; Williford et al., 2014). We kept the same generating parameters and missingness proportions as in the original simulation. Results relevant to the two demonstration points of our original simulation are tabled in the Online Appendix (Tables OA.1 and OA.2). We found

[2] The $Q = 1$ state model necessarily has a different structure; it simply consists of estimating a set of missingness indicator thresholds for that single class.

TABLE 5.
Simulation extension under two alternative misspecifications of the missingness process: percent absolute relative bias (%ARB) for multinomial coefficient structural parameters in the outcome process.

Parameter	Pop. value	MNAR missingness mechanism			
		Fit MNAR-PP LTA		Fit MNAR-PP LTA	
		Misspecification: in population, $Q = 3$		Misspecification: in population, q_t does not depend on k_{t-1}	
		Avg. Est.	%ARB	Avg. Est.	%ARB
$\omega_{k_1=1}$	−1.2	−1.200	0.03	−1.196	0.30
$\omega_{k_1=2}$	−0.9	−0.881	2.11	−0.888	1.36
$\alpha_{k_2=1}$	−1.2	−1.205	0.45	−1.200	0.02
$\alpha_{k_2=2}$	−0.9	−0.900	0.05	−0.905	0.57
$\alpha_{k_3=1}$	−1.2	−1.203	0.27	−1.202	0.17
$\alpha_{k_3=2}$	−0.9	−0.897	0.28	−0.905	0.57
$\beta_{k_2=1\|c_1^y=1}$	2.0	2.018	0.92	2.015	0.73
$\beta_{k_2=1\|c_1^y=2}$	1.0	1.001	0.14	1.002	0.17
$\beta_{k_2=2\|c_1^y=1}$	1.0	0.989	1.14	0.986	1.37
$\beta_{k_2=2\|c_1^y=2}$	1.5	1.531	2.08	1.526	1.73
$\beta_{k_3=1\|c_2^y=1}$	1.75	1.767	0.94	1.763	0.71
$\beta_{k_3=1\|c_2^y=2}$	1.0	0.988	1.18	1.011	1.12
$\beta_{k_3=2\|c_2^y=1}$	1.0	0.958	4.17	1.010	1.04
$\beta_{k_3=2\|c_2^y=2}$	1.5	1.525	1.65	1.523	1.56

LTA latent transition analysis, *MNAR-PP LTA* missing-not-at-random parallel-process LTA, *Pop. value* population parameter value, *MAR* missing-at-random, *Avg. Est.* average estimate.

a similar overall pattern of results to the original simulation, as follows. Regarding point (1), when missingness was MNAR, fitting a conventional LTA instead of a MNAR-PP LTA again, on average, entailed larger %ARB (3 times larger) for structural parameters of the outcome process and also larger %ARB for measurement parameters, on average. Regarding point (2), when missingness was MAR, structural and measurement parameters of the outcome process again had approximately the same amount of bias, on average, when fitting the conventional LTA versus MNAR-PP LTA. A difference compared to the original simulation demonstration was that more samples encountered estimation problems (such as singularity of the information matrix), and this was the case regardless of which model was fit (conventional LTA or MNAR-PP LTA). Samples generated with MAR missingness, rather than MNAR missingness, encountered fewer estimation problems. As with other kinds of MAR and joint MNAR models, a general point is that different combinations of rates of missingness, missingness mechanism parameters, and overall N could impact the chance of incurring estimation problems, even for a properly specified model.

5. Empirical Application

An empirical example is employed to illustrate the implementation and interpretation of the MNAR-PP LTA. In Sect. 5.1, we begin by describing substantive background for our research questions. Although our research questions can be addressed with a conventional LTA, we motivate the need for a sensitivity analysis comparing the conventional LTA and MNAR-PP LTA results due to a suspected latent-state-dependent MNAR missingness mechanism. In sensitivity analyses

TABLE 6.
Simulation extension under two alternative misspecifications of the missingness process: percent absolute relative bias (%ARB) for threshold measurement parameters in outcome process.

Parameter	Pop. value	MNAR Missingness mechanism			
		Fit MNAR-PP LTA		Fit MNAR-PP LTA	
		Misspecification: In population, $Q = 3$		Misspecification: In population, q_t does not depend on k_{t-1}	
		Avg. Est.	%ARB	Avg. Est.	%ARB
$v_{y1t\|k_t=1}$	−0.90	−0.907	0.74	−0.902	0.27
$v_{y2t\|k_t=1}$	−1.69	−1.716	1.54	−1.700	0.58
$v_{y3t\|k_t=1}$	−2.20	−2.216	0.71	−2.209	0.42
$v_{y4t\|k_t=1}$	−1.25	−1.245	0.41	−1.243	0.53
$v_{y5t\|k_t=1}$	−1.48	−1.491	0.74	−1.484	0.25
$v_{y1t\|k_t=2}$	0.32	0.332	3.60	0.328	2.34
$v_{y2t\|k_t=2}$	0.95	0.960	1.06	0.960	1.00
$v_{y3t\|k_t=2}$	−0.20	−0.191	4.27	−0.189	5.70
$v_{y4t\|k_t=2}$	0.15	0.152	1.35	0.150	0.27
$v_{y5t\|k_t=2}$	0.55	0.559	1.55	0.562	2.11
$v_{y1t\|k_t=3}$	2.10	2.106	0.31	2.109	0.45
$v_{y2t\|k_t=3}$	2.67	2.678	0.28	2.679	0.35
$v_{y3t\|k_t=3}$	1.80	1.809	0.52	1.806	0.34
$v_{y4t\|k_t=3}$	2.55	2.566	0.63	2.568	0.71
$v_{y5t\|k_t=3}$	2.30	2.303	0.12	2.300	0.01

LTA latent transition analysis, *MNAR-PP LTA* missing-not-at-random parallel-process LTA, *Pop. value* population parameter value, *MAR* missing-at-random, *Avg. Est.* average estimate.

comparing outcome process parameter estimates between a MAR versus joint MNAR model, little or no change is often viewed as consistent with a MAR mechanism (e.g., Feldman & Rabe-Hesketh, 2012; Verbeke et al., 2001; Xu & Blozis, 2011). But, since we can never completely rule out MNAR when analyzing empirical data, it is most conservative to view estimates with little change simply as robust to perturbation of the missingness assumptions. Greater change in outcome process parameter estimates between a MAR versus joint MNAR model may be due to the missingness mechanism being MNAR (as was the case in Sect. 4 simulation) but could possibly arise for other reasons when using empirical data (e.g., model misspecifications or influential case(s); Molenberghs & Verbeke, 2005).

5.1. Background and Research Questions

Conduct problems have been conceptualized as a stage-sequential latent construct wherein adolescents can progress to more severe or multi-domain delinquency through different sequences—for instance, sequences starting with authority resistance/drug use behaviors versus starting with property damage/shoplifting behaviors versus starting with bullying/fighting behaviors (see, e.g., Hinshaw, Lahey, & Hart, 1993; Lahey & Loeber, 1994; Loeber, Keenan, & Zhang, 1997; Moffitt et al., 2008). Other adolescents may stay in a stable state (or remit) instead of progressing to more severe behaviors. In this empirical example, interest lies in using LTA to investigate developmental stage-sequential change in a latent 'conduct problems' construct for late adolescents leaving the foster care system. Late adolescents emerging from the foster care system are at heightened risk of conduct problems (McMillen et al., 2005). However, little is known about the conduct problem symptom profiles exhibited by adolescents in foster care. Also, little is known about the proportion of adolescents from foster care who progress into more serious states as they enter young adulthood, or regress into less serious states. These research topics can be investigated with LTA.

5.2. Sample and Outcome Variables

Our empirical example analysis dataset consists of $N = 730$ late adolescents (age 18–19) who had participated in foster care in the Midwestern United States (see Courtney & Cusick, 2007 for further details). Starting in 2002, these late adolescents were interviewed about their delinquent behavior every two years, for 3 occasions. The conduct problems latent construct is here measured by $J = 8$ binary indicators: deliberate property damage, petty theft, selling drugs, hurting someone enough to require medical care, breaking and entering, group fighting, stealing $>$\$50, and pulling a gun or knife on someone.

5.3. Missing Data

Only 396 of the 730 cases had complete data for all three timepoints. 93 cases dropped out on all J items at $t = 2$. Also, 123 cases dropped out on all J items at $t = 3$; 58 cases had missingness on some—but not all—of the J items at a given timepoint; and 60 cases had intermittent missing data for all J items at $t = 2$. Substantively, latent-state-dependent nonignorable missingness was considered plausible for the following reasons. Persons in an elevated conduct problems latent state may be more likely to skip the interview or even drop out. Also, persons in a latent state displaying some illegal behavior (e.g., selling drugs) may be more likely to provide a pattern of intermittent item-specific missingness for sensitive questions related to the illegal behavior, due to social desirability concerns or fear of reprisal (Groves et al., 2002). To explore the impact of accounting for potential latent-state-dependent missingness, a sensitivity analysis was performed comparing estimates from a conventional LTA versus MNAR-PP LTA.

5.4. Analysis Plan

Our model-building approach for the MNAR-PP LTA was to, first, select the number of latent states using separate single-process models for the outcome process and missingness process. An alternative approach would be to specify a full parallel-process model and select K and Q simultaneously (but see Nagin & Tremblay, 2001, p. 26 for practical drawbacks of this alternative). Configural invariance was imposed for latent states over time in both processes (as in Collins & Lanza, 2010). Second, we tested for measurement invariance in each single-process model. Third, we combined the best-fitting single-process K-state outcome model and Q-state missingness model into a final parallel-process MNAR-PP LTA. BIC was used for model comparisons (following, e.g., Nagin, 2005). M*plus* 7.11 was employed for model fitting, using the EM algorithm. Multiple sets of random starts were used to decrease the possibility of local solutions. M*plus* syntax for the MNAR-PP LTA is provided in the Online Appendix.

5.5. Results and Sensitivity Analysis

For the $T = 3$ outcome process, $K = 4$ latent ('conduct problem') outcome states (BIC = 10166.8) were selected as better fitting compared to $K = 2$ (BIC = 10470.3), $K = 3$ (BIC = 10269.8), and $K = 5$ (BIC = 10224.4). Furthermore, for $K = 4$, measurement invariance was imposed because it led to better fit (BIC = 10166.8, from above) than when allowing for measurement non-invariance (BIC = 10340.1). For the missingness process (at timepoints 2 and 3), $Q = 2$ latent missingness states (BIC = 2563.1) and $Q = 3$ (BIC = 2563.5) fit nearly equally well, and much better than $Q = 4$ (BIC = 2630.2). Since the $Q = 2$ and $Q = 3$ BICs were

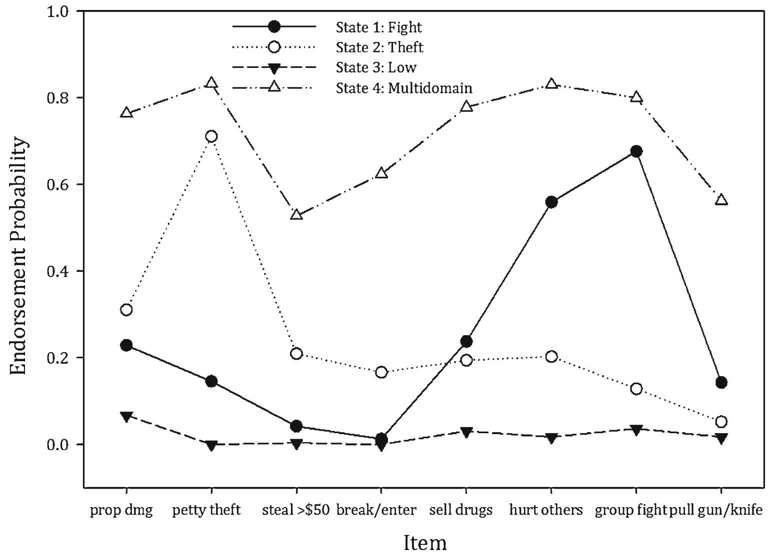

FIGURE 3.
Empirical example measurement parameter results (for both MNAR-PP LTA and LTA): plot of conduct outcome item endorsement probabilities per state at time t. *Notes* Measurement invariance was imposed, so this plot would look the same at $t = 1, 2, 3$. The marginal probability of membership in each state at at time t is given in Table 8 for both the MNAR-PP LTA and LTA.

essentially undifferentiable according to Raftery's (1995) criteria (see also Sterba & Pek, 2012), AIC was consulted to see if these two models could be discriminated based on generalizability; AIC strongly preferred $Q = 3$ (AIC = 2416.567) to $Q = 2$ (AIC = 2475.8). $Q = 3$ was selected. Furthermore, for $Q = 3$, measurement invariance was imposed because it led to better fit (BIC = 2563.5, from above) than when allowing for measurement non-invariance (BIC = 2670.0).

The $K = 4$, $Q = 3$ MNAR-PP LTA was constructed by combining the best-fitting outcome and missingness process models described above. First, we describe the measurement parameter results from the MNAR-PP LTA. Next, we describe the structural parameter results from the MNAR-PP LTA. Finally, we compare these results to the conventional LTA. For ease of interpretation, results are reported after converting thresholds and multinomial coefficient estimates into item probabilities and latent state/transition probabilities, using formulas presented earlier.

5.5.1. Defining Latent Outcome States in the Conventional LTA versus MNAR-PP LTA The conduct item endorsement probabilities for the conventional LTA were virtually identical to the MNAR-PP LTA (differing by an average of $< .01$ [range: 0-.025]); this consistency is expected under a latent-state-dependent MNAR mechanism, which mainly can affect structural parameters. Thus, only one plot of conduct item endorsement probabilities per outcome latent state at time t is provided for both models, in Fig. 3. In Fig. 3, State 4 (labeled "*multi*") has high endorsement probabilities for all conduct items. State 3 (labeled "*low*") is nondelinquent. State 2 (labeled "*theft*") had high endorsement probability for theft. State 1 (labeled "*fight*") had high endorsement probabilities for fighting and hurting others. Hence, States 1 and 2 each manifest single-domain delinquency (*theft* versus *fight*) whereas State 4 manifests multi-domain delinquency.

5.5.2. Defining Latent Missingness States for MNAR-PP LTA Missingness latent State 3 (here, labeled "*drug*") had an elevated item-specific missingness probability, .40, for one sensitive

TABLE 7.
Empirical example sensitivity analysis results. Missingness process structural parameter probabilities from the MNAR-PP LTA.

Element probability		MNAR-PP LTA		
		Miss	Present	Drug
π_{q2}		.22	.75	.03
		Miss	Present	Drug
π_{q3}		.31	.66	.03
		Miss	Present	Drug
$\tau_{q3\|q2}$	Miss	.59	.38	.03
	Present	.22	.75	.03
	Drug	.54	.46	.00
		Miss	Present	Drug
$\tau_{q2\|k2}$	Fight	.35	.57	.08
	Theft	.21	.75	.04
	Low	.11	.89	.00
	Multi	.59	.41	.00
		Miss	Present	Drug
$\tau_{q3\|k3}$	Fight	.47	.41	.13
	Theft	.67	.30	.02
	Low	.17	.82	.01
	Multi	.20	.80	.00

MNAR-PP LTA missing-not-at-random parallel-process latent transition analysis, *Miss*, *present*, and *drug* are names given to latent missingness states in the text. *Fight, theft, low,* and *multi* are names given to latent outcome states in the text.

item (drug dealing), but lower missingness probabilities for all other items (.04–.20). Persons in this state may be involved with drugs at time t but are hesitant to report this due to social desirability concerns or fear of criminal prosecution. Missingness State 2 (here labeled "*present*") had $\leq$.01 missingness probability for all J items and State 3 (here labeled "*miss*") had $>$.99 missingness probability on all J items.

5.5.3. Structural Parameters of the Missingness Process for MNAR-PP LTA Table 7 provides latent state and transition probabilities for the missingness process of the MNAR-PP LTA. Table 7 marginal probabilities (π_{q_2} and π_{q_3}) show that, from times 2 to 3, the marginal probability increases for the *miss* state (.22 to .31), decreases for the *present* state (.75 to .66), and remains stable for the social-desirability *drug* state. Table 7 transition probabilities $\tau_{q_3|q_2}$ show that there is greatest stability in membership in the *present* state (those *present* at time 2 have .75 probability of being *present* at time 3), whereas the social-desirability *drug* state is least stable.

To conserve space, Table 7 does not list all conditional probabilities relating missingness and outcome processes from Eq. (6). Instead, we show $\tau_{q_2|k_2}$ and $\tau_{q_3|k_3}$ (calculated using formulas in Table 1, row 3) and give some examples of how these change when also conditioning on other latent states. Table 7 shows that persons in the *low* conduct problems state at time t have the highest probability of being in the *present* state at time t: .89 at time 2 ($\tau_{q_2=2|k_2=3}$) and .82 at time 3 ($\tau_{q_3=2|k_3=3}$). The latter probability would increase from .82 to .86 for those who were in the *low* conduct state at *both* times 2 and 3 *and* the *present* state at time 2 ($\tau_{q_3=2|k_2=3,k_3=3,q_2=2}$).

TABLE 8.
Empirical example sensitivity analysis results. Outcome process structural parameter probabilities: LTA versus MNAR-PP LTA.

Element probability		Conventional LTA				MNAR-PP LTA			
		Fight	Theft	Low	Multi	Fight	Theft	Low	Multi
π_{k_1}		.37	.24	.29	.10	.35	.26	.29	.10
π_{k_2}		.24	.11	.61	.03	.28	.12	.54	.06
π_{k_3}		.11	**.09**	**.77**	.02	.14	**.19**	**.65**	.02
$\tau_{k_2 \mid k_1}$	Fight	.53	.01	.44	.02	.62	.02	.35	.01
	Theft	.00	.35	**.64**	**.01**	.00	.33	**.52**	**.14**
	Low	.00	.01	.98	.01	.00	.03	.97	.00
	Multi	**.48**	.22	.09	.22	**.59**	.18	.06	.18
$\tau_{k_3 \mid k_2}$	Fight	.45	.01	.51	.03	.49	.04	.45	.02
	Theft	.00	**.44**	**.48**	.08	.00	**.58**	**.36**	.06
	Low	.00	.04	.97	.00	.00	.12	.88	.00
	Multi	.16	**.52**	**.14**	.18	.13	**.80**	**.00**	.07

When a probability differs by $\geq$.10 across models, both values are bolded.
LTA latent transition analysis, *MNAR-PP LTA* missing-not-at-random parallel-process LTA. *Fight, theft, low,* and *multi* are names given to latent outcome states in the text.

In contrast, persons in each delinquent conduct state (*multi*,*theft*, or *fight*) at time t have elevated probabilities of being in the *miss* missingness state at time t. For instance, members of the *multi* conduct state at time 2 have a .59 probability of being in the *miss* state at time 2—compared with .11 for those in the *low* conduct state. Similarly, members of the *theft* conduct state at time 3 have a .67 probability of being in the *miss* state at time 3—compared with .27 for those in the *low* conduct state. This probability of being in *miss* at $t = 3$ would, for instance, increase from .67 to .84 for those who were in the *theft* state at *both* $t = 3$ and $t = 2$ and in the *miss* state at $t = 2$. Finally, those in the *fight* outcome state consistently have the highest probability of being in the *drug* missingness state. In sum, Table 7 suggests an overall pattern where membership in latent conduct states evidencing delinquent behavior is associated with membership in latent missingness states that have higher probabilities of nonresponse for all or some items. Next we investigate the impact that allowing for these associations across missingness and outcome states has on the structural parameters of substantive interest, in the outcome process.

5.5.4. Structural Parameters of the Outcome Process for Conventional LTA versus MNAR-PP LTA Table 8 juxtaposes latent outcome state probabilities from the Conventional LTA (left column) versus MNAR-PP LTA (right column). For ease of visualization, when a probability differs by $\geq$ .10 across models, both values are bolded in Table 8.

Regarding marginal probabilities, the time 1 latent conduct state probabilities (π_{k_1}) are very similar across models in Table 8, reflecting the lack of missing data at time 1. At times 2 and 3, both models agree that the *low* state is most prevalent, followed by *fight, theft,* and then *multi*. However, at times 2 and 3, MNAR-PP LTA usually implies higher marginal probabilities (π_{k_2}, π_{k_3})

for delinquent conduct states (particularly *theft*), whereas the conventional LTA implies a higher probability for the *low* conduct state.

Regarding transition probabilities ($\tau_{k_2|k_1}$ and $\tau_{k_3|k_2}$) in Table 8, both models agree that the vast majority of persons starting in the *low* conduct state stay in the *low* conduct state. However, compared to the conventional LTA, the MNAR-PP LTA implies a lower probability of transferring from any delinquent state into the *low* state. Instead, the MNAR-PP LTA implies more persons transfer from *multi* to *fight* from $t = 1$ to 2 and more persons transfer from *multi* to *theft* from $t = 2$ to 3; the MNAR-PP LTA also implies more stability from $t = 2$ to 3 in *theft* state membership.

5.5.5. Summary of Sensitivity Analysis Results In sum, the above analysis suggests that current and previous membership in delinquent states (*fight*, *theft*, or*multi*) is associated with membership in missingness states with higher nonresponse probabilities for some/all items (*miss, drug*). To summarize answers to our substantive research questions, conduct problems in foster care adolescents could be characterized by two single-domain delinquency states (*fight* and*theft*), one multi-domain delinquency state (*multi*), and a nondelinquent state (*low*). If foster care youth are in the *low* state in late adolescence, their conduct is unlikely to worsen through early adulthood. It is more common for adolescents exhibiting single-domain delinquency (*theft* or *fight*) to progress to *multi*-domain delinquency or regress to *low* across early adulthood, but rare for foster care youth to improve from *multi*-domain to *low* conduct problems states across a 2-year span. Some key LTA results were robust to varying the missingness assumptions between MAR (a conventional LTA) versus latent-state-dependent MNAR (a MNAR-PP LTA), but other results were sensitive. In particular, the definition of *fight, theft, multi,* and *low* latent outcome states in Fig. 3, the rank order of latent outcome state prevalences at time t, and the stability of *low* state membership over time for those starting in *low* were robust. However, the prevalence of the *theft* state and the rate of transition into and out of the *theft* state were particularly sensitive to missingness assumptions. Overall, due to the extent and nature of missing data, we are less sure about the developmental course of theft for foster care participants, as compared to other domains of delinquent behavior.

6. Discussion

When missingness is nonrandom (MNAR), exclusively modeling the outcome-generating mechanism risks parameter bias. Joint models can be used to represent the dependency between the outcome-generating process and missingness mechanism. In the psychology literature, there has been increasing application of joint MNAR models where the outcome-generating mechanism is specified as a growth trajectory model (including multilevel growth, growth mixture, or groups-based trajectory models; Enders, 2011; Feldman & Rabe-Hesketh, 2012; Hedeker & Gibbons, 1997; Lu et al., 2011; Muthén et al., 2011). However, another popular longitudinal outcome model used by psychologists—LTA—has not been extended into a joint MNAR model for accommodating intermittent missingness and dropout. This is an important gap because LTA is often used to model stage-sequential change in sensitive, antisocial, delinquent, and victimization constructs (e.g., Bair-Merritt et al., 2012; Cleveland et al., 2012; Cochran et al., 2013; Hopfer et al., 2013; Lanza & Collins, 2008; Mackesy-Amiti et al., 2014; Williford et al., 2014) which may present a greater risk of MNAR missingness (Groves et al., 2002; Little et al., 2012).

In this *Case Study*, we presented a parallel-process MNAR model for LTA (MNAR-PP LTA) in which missingness indicators at time t define latent missingness states at time t and these latent missingness states at time t are allowed to depend on latent outcome states at times t and $t - 1$ (as well as on missingness state at $t-1$). The MNAR-PP LTA is a kind of shared parameter joint MNAR model since $\mathbf{m}_{it}$ and $\mathbf{y}_{it}$ are conditionally independent given current and previous latent outcome states. We used an empirical investigation of stage-sequential change in conduct problems among foster care adolescents to illustrate the utility of the MNAR-PP LTA. In this empirical example (Sect. 5), previous and current membership in high conduct problem states (*theft, fight, multi*) predicted membership in missingness states having high nonresponse probabilities for all/some items (*miss*, *drug*). This finding is consistent with sensitivity analyses results from other studies where participants currently and previously experiencing greater psychiatric problems and/or disability were more likely to have missing responses (e.g., Beunckens, Molenberghs, Thijs, & Verbeke, 2007; Haviland et al., 2011; Kurland & Heagerty, 2004; Muthén et al., 2011; Roy, 2003). This finding could be used to improve study design; for instance, psychologists may want to tailor retention efforts to adolescents who are at greatest risk of conduct problems.

After accounting for these latent state dependencies between outcome and missingness processes using a MNAR-PP LTA, we compared results to a conventional LTA. The MNAR-PP LTA implied higher marginal probabilities for delinquent conduct states at times 2 and 3 and lower transition probabilities from these delinquent conduct states into the *low* problems state. Hence, our sensitivity analysis revealed that the conventional LTA may overestimate the proportion of foster care youth who end up in a *low* conduct problems state. Importantly, overestimating how many adolescents desist from delinquent behaviors could lead policymakers to allocate fewer therapeutic resources to support the well-being of adolescents as they transition out of state care and into adulthood.

Five strengths of the MNAR-PP LTA are summarized here. (1) The MNAR-PP LTA allows LTA parameters for the outcome process ($\boldsymbol{\theta}_y$) to be interpreted as in the conventional LTA, which facilitates using a sensitivity analysis to assess how their estimates change between conventional LTA versus MNAR-PP LTA. (2) The MNAR-PP LTA flexibly accommodates a realistic combination of intermittent missingness and/or dropout, some of which may occur for all J items or for just item j. (3) Although the MNAR-PP LTA, like other MNAR models, embodies only one possible MNAR mechanism out of many, it accommodates a substantively plausible MNAR mechanism for stage-sequential outcome processes: latent-state-dependent nonignorable missingness. As demonstrated in the Sect. 4 simulation, fitting a conventional LTA in the presence of latent-state-dependent nonignorable missingness biases primarily structural parameters of the LTA, and this bias can be reduced by applying MNAR-PP LTA. (4) When missingness is truly MAR (unbeknownst to the researcher) mistakenly/unnecessarily fitting a MNAR-PP LTA may have little downside; the Sect. 4 simulation demonstrated that under this circumstance, MNAR-PP LTA can recover LTA parameters as well as does the conventional LTA, on average. (5) The MNAR-PP LTA can be implemented in available commercial software (see Online Appendix syntax).

6.1. Conceptualizing the MNAR-PP LTA within a sensitivity analysis framework

Other joint MNAR models share the MNAR-PP LTA's ability to recover outcome process parameter estimates—when MAR missingness holds—and also reduce bias—when their stipulated MNAR missingness mechanism holds (e.g., National Research Council, 2010; Gottfredson, Bauer, & Baldwin, 2014; Yang & Maxwell, 2014; Zhang & Wang, 2012). Nevertheless, the prevailing recommendation in the missing data literature is not that one should fit a joint MNAR model by default, without ever fitting a MAR model. Instead, joint MNAR models are recommended for use in comparison with MAR models, in the context of sensitivity analyses (e.g., Beunckens et al., 2007; Enders, 2011; Feldman & Rabe-Hesketh, 2012; Molenberghs & Kenward, 2007; Little & Rubin, 1999; Lu et al., 2011; Mallinckrodt et al., 2003; Muthén et al., 2011; National Research Council, 2010; Roy, 2003; Roznitsky et al., 1998; Schafer & Graham, 2002; Verbeke et al., 2001; Xu & Blozis, 2011).

One reason for this recommendation is as follows. If missingness is nonignorable, fitting a seriously misspecified joint MNAR model may (e.g., Gottfredson et al., 2014; Yang & Maxwell,

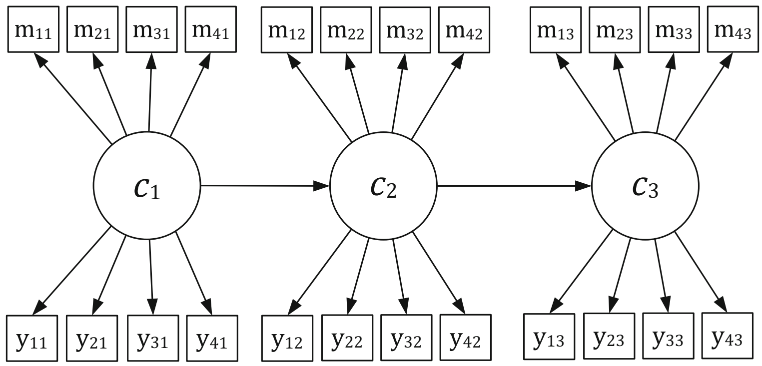

FIGURE 4.
An alternative MNAR specification: missing-not-at-random *single*-process latent transition analysis (MNAR-SP LTA) model. Shown where missingness starts at time 1. See Sect. 6.2 for limitations of this MNAR-SP LTA relative to the MNAR-PP LTA from Fig. 2 Panel B. *Notes* In this path diagram, $T = 3$, $J = 4$, and $K = Q$. *SP* single process, *PP* parallel process.

2014) but is not guaranteed to (e.g., Minini & Chavance, 2003; Yang & Maxwell, 2014) provide better estimates than fitting a MAR model. As summarized by Molenberghs and Verbeke (2005, p. 575) "A sensible compromise between blindly shifting to MNAR models or ignoring them altogether is to make them a component of a sensitivity analysis. It is important to consider the effect on key parameters….In many instances a sensitivity analysis can strengthen one's confidence in the MAR model". Such sensitivity analyses typically consider the MAR model as the "primary analysis" (Little et al., 2012, p. 1358) and use scientifically plausible joint MNAR model results to form an "envelope of conclusions" (Carpenter, Pocock, & Lamm 2002, p. 1049) around the MAR results. This approach was used in our empirical example.

6.2. Alternatives to the MNAR-PP LTA

As part of their own sensitivity analysis, researchers may be interested in fitting not one but several LTA models that each allow for nonignorable missingness under alternative assumptions. As discussed previously in Sect. 3.1, there are key limitations to specifying outcome-dependent selection or random-coefficient-dependent joint MNAR models in the context of LTA. Instead, one possibility is to investigate the sensitivity of outcome submodel parameter estimates to alternative specifications of the missingness submodel in the MNAR-PP LTA. For instance, researchers could specify a different choice of Q, as done in Sect. 4.4. Also, the MNAR-PP LTA specification in Eqs. (6) and (11) could be extended to allow missingness state at time t to depend on latent outcome states before $t - 1$ (e.g., $t - 2$), providing that there was a substantive rationale to do so. Alternatively, the missingness process could be extended from first- to second-order Markov in the MNAR-PP LTA (although this does sacrifice parsimony).

Another possibility is to consider a different shared parameter MNAR model which is actually a special case of the MNAR-PP LTA. In this special case, $\mathbf{m}_{it}$ and $\mathbf{y}_{it}$ would depend on the exact *same* latent states at time t—rather than $\mathbf{m}_{it}$ and $\mathbf{y}_{it}$ each depending on different latent states, which are then associated at the structural level. A path diagram of this single-process model (here termed MNAR-SP LTA) is provided in Fig. 4. When $Q_t = K_t$, this special case would arise by placing constraints $\tau_{q_t|k_{t-1},k_t,q_{t-1}} = \tau_{q_t|k_t}$, and $\tau_{q_t|k_t} = 1$ in the Eq. (11) MNAR-PP LTA, when $q_t = k_t$. Researchers could compare results between MNAR-PP LTA and MNAR-SP LTA.

However, this special case MNAR-SP LTA has the following three disadvantages compared with the MNAR-PP LTAs presented earlier. First, this special case model is more restrictive than the Eq. (11) MNAR-PP LTA in that it requires $Q_t = K_t$, which may not always hold in practice. (This requirement was, for instance, not met in the empirical example.) Second, when missingness begins at $t \geq 2$, the special case MNAR-SP LTA encounters conceptual difficulties that are not encountered when using the Eq. (6) MNAR-PP LTA. Had missingness begun at time 2, this MNAR-SP LTA would have a different set of indicators of state membership at time 1 (i.e., $\mathbf{y}_{it}$) versus after time 1 (i.e., $\mathbf{y}_{it}$ and $\mathbf{m}_{it}$). In turn, this would violate measurement invariance of latent outcome states and complicate the substantive interpretation of latent transition probabilities. That is, we would no longer be able to interpret large diagonal elements of the transition probability matrix as indicating "stability" in state membership, since the states change in substantive meaning over time (e.g., at time 1, state 1 would represent a particular pattern of outcome behaviors but, at time 2, state 1 would represent a pattern of missingness *and* outcome behaviors). Third, even if missingness begins at time 1, the MNAR-SP LTA does not allow outcome process parameters to be interpreted in the same way as a conventional LTA. For instance, the conventional LTA provides initial and transition probabilities for latent outcome states, which are the focus of substantive interest. However, the MNAR-SP LTA provides these probabilities for latent states representing a pattern of *both* outcomes and missingness.[3] In contrast, the proposed MNAR-PP LTA allows researchers to interpret outcome parameters in the same way as with the conventional LTA. Because of these three disadvantages, this special case MNAR-SP LTA was not a focus of the current article, but it could potentially be useful in particular substantive settings. Therefore, we provide software syntax for one way to fit the MNAR-SP LTA in the Online Appendix.

Note that this special case MNAR-SP LTA has conceptual similarities to Haviland et al. (2011). In a growth trajectory modeling framework, they allowed missingness indicators as well as univariate repeated measure outcomes to depend on the exact same latent classes.

6.3. *Future Directions and Limitations*

Several potential extensions of this study and of the MNAR-PP LTA could serve as topics for future research. First, the MNAR-PP LTA in Eqs. (6) and (11) could be extended to accommodate ordinal rather than binary outcomes. Second, future simulation research could compare alternative model-building strategies (e.g., picking K and Q in separate preliminary single-process models versus simultaneously in a full MNAR-PP LTA). Third, future simulation research could assess the performance of the MNAR-PP LTA under other conditions and model misspecifications.

As another extension, observed covariates can be added to the structural submodels of the MNAR-PP LTA to predict previous or current outcome state membership and/or missingness state membership. Observed covariates can also be added to the measurement submodels of the MNAR-PP LTA to predict outcome indicators within outcome state and/or missingness indicators within missingness state. These interesting possibilities for including covariates represent a straightforward extension of the logistic parameterization of the MNAR-PP LTA presented here. We simply need to add observed predictors (and accompanying slopes) to our multinomial or binary logistic regression equations (e.g., see Eqs. 2, 3, 5, 7–9) which we had already been using to calculate item, initial latent state, and latent transition probabilities. After doing so, these probabilities for person i will be conditional also on person i's values on the observed covariates.

6.4. *Conclusions*

This article was motivated by an examination of stage-sequential change in a discrete latent conduct problems construct, where a latent-state-dependent nonignorable missingness mechanism

[3] Suppose the MNAR-*SP* LTA results indicated that a particular state at time t evidenced a high probability of violent behavior and a high probability of missingness. Persons might obtain a similarly high posterior probability of assignment to that MNAR-*SP* LTA state simply by endorsing many aggressive items (even if they had no missingness), or by having many missing responses (even if they endorsed no aggressive items). Yet, substantively, researchers would want to be able to distinguish among such persons.

was suspected. We developed a joint MNAR model (MNAR-PP LTA) for accommodating this missingness mechanism, demonstrated aspects of its performance via simulation, and illustrated its interpretation in a sensitivity analysis for our empirical example. We hope this study increases psychologists' understanding of how alternative missingness assumptions can impact conclusions about stage-sequential latent change. We encourage further study and development of joint MNAR models for LTA, given psychologists' frequent application of LTA to assess change in illicit, risky, and delinquent behaviors.

References

Albert, P. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, *56*, 602–608.

Albert, P., & Follmann, D. (2003). A random effects transition model for longitudinal binary data with informative missingness. *Statistica Neerlandica*, *57*, 100–111.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Baggio, S., Studer, J., Deline, N' Goran, Dupuis, M., Henchoz, Y., Mohler-Kuo, M., et al. (2014). Patterns and transitions in substance use among young Swiss men: A latent transition analysis approach. *Journal of Drug Issues*. doi:10.1177/0022042614526996.

Bair-Merritt, M., Ghazarian, S., Burrell, L., & Duggan, A. (2012). Patterns of intimate partner violence in mothers at-risk for child maltreatment. *Journal of Family Violence*, *27*, 287–294.

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

Bray, B., Lanza, S., & Collins, L. (2010). Modeling relations among discrete developmental processes: A general approach to associative latent transition analysis. *Structural Equation Modeling*, *17*, 541–569.

Cain, A., Epler, A., Steinley, D., & Sher, K. (2012). Concerns related to eating, weight, and shape: Typologies and transitions in men during the college years. *International Journal of Eating Disorders*, *45*, 768–775.

Carpenter, J., Pocock, S., & Lamm, C. (2002). Coping with missing values in clinical trials: A model based approach applied to asthma trials. *Statistics in Medicine*, *21*, 1043–1066.

Catts, H., Compton, D., Tomblin, J., & Bridges, M. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*, 166–181.

Chen, B., Yi, G., & Cook, R. (2011). Progressive multi-state models for informatively incomplete longitudinal data. *Journal of Statistical Planning and Inference*, *141*, 80–93.

Cleveland, M., Lanza, S., Ray, A., Turrisi, R., & Mallett, K. (2012). Transitions in first-year college student drinking behaviors: Does pre-college drinking moderate the effects of parent- and peer-based intervention components? *Psychology of Addictive Behaviors*, *26*, 440–450.

Cochran, G., Field, C., & Caetano, R. (2013). Changes in classes of injury-related risks and consequences of risk-level drinking: A latent transition analysis. *Journal of Behavioral Health Services & Research*,. doi:10.1007/s11414-013-9378-3.

Collins, L., & Lanza, S. (2010). *Latent class and latent transition analysis*. Hoboken, NJ: Wiley.

Collins, L., & Wugalter, S. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, *27*, 131–157.

Cole, B., Bonetti, M., Zaslavsky, A., & Gelber, R. (2005). A multistate Markov chain model for longitudinal, categorical quality-of-life data subject to non-ignorable missingness. *Statistics in Medicine*, *24*, 2317–2334.

Cook, E., Pflieger, J., Connell, A.,& Connell, C. (2014). Do specific transitional patterns of antisocial behavior during adolescence increase risk for problems in young adulthood? *Journal of Abnormal Child Psychology*. doi:10.1007/s10802-014-9880-y.

Courtney, M., & Cusick, G. (2007). *Crime during the transition to adulthood: How youth fare as they leave out-of-home care in Illinois, Iowa and Wisconsin*. [Computer file].

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

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

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.

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

Flaherty, B. P. (2008). Testing the degree of cross-sectional and longitudinal dependence between two discrete dynamic processes. *Developmental Psychology*, *44*, 468–480.

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.

Groves, R., Dillman, D., Eltinge, J., & Little, R. (2002). *Survey nonresponse*. New York: Wiley.

Haviland, A., Jones, B., & Nagin, D. (2011). Group-based trajectory modeling extended to account for nonrandom participant attrition. *Sociological Methods & 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.
Hinshaw, S., Lahey, B., & Hart, E. (1993). Issues of taxonomy and comorbidity in the development of conduct disorder. *Development and Psychopathology*, *5*, 31–49.
Hopfer, S., Hecht, M., Lanza, S., Tan, X., & Xu, S. (2013). Preadolescent drug use resistance skill profiles substance use, and substance use prevention. *Journal of Primary Prevention*, *34*, 395–404.
Ibrahim, J. G., Chen, M.-H., Lipsitz, S., & Herring, A. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association*, *100*, 332–346.
Jackson, K., Sher, K., & Schulenberg, J. (2005). Conjoint developmental trajectories of young adult alcohol and tobacco use. *Journal of Abnormal Psychology*, *114*, 612–626.
Jansen, I., Hens, N., Molenberghs, G., et al. (2006). The nature of sensitivity in missing not at random models. *Computational Statistics & Data Analysis*, *50*, 830–858.
Kroesen, M. (2014). Modeling the behavioral determinants of travel behavior: An application of latent transition analysis. *Transportation Research*, *65*, 56–67.
Kurland, B., & Heagerty, P. (2004). Marginalized transition models for longitudinal binary data with ignorable and non-ignorable drop-out. *Statistics in Medicine*, *23*, 2673–2695.
Lahey, B., & Loeber, R. (1994). Framework for a developmental model of oppositional defiant disorder and conduct disorder. In D. K. Routh (Ed.), *Disruptive behavior disorders in childhood*. New York: Plenum.
Langeneine, R., & van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods and Research*, *18*, 416–441.
Lanza, S., & Collins, L. (2008). A new SAS procedure for latent transition analysis: Transitions in dating and sexual risk behavior. *Developmental Psychology*, *44*, 446–456.
Lee, S., & Song, X.-Y. (2003). Maximum likelihood estimation and model comparison for mixtures of structural equation models with ignorable missing data. *Journal of Classification*, *20*, 221–255.
Lee, B., & Vondracek, F. (2014). Teenage goals and self-efficacy beliefs as precursors of adult career and family outcomes. *Journal of Vocational Behavior*, *85*, 228–237.
Lin, H., Liu, D., & Zhou, X. (2010). A correlated random-effects model for normal longitudinal data with nonignorable missingness. *Statistics in Medicine*, *29*, 236–247.
Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112–1121.
Little, R. J. A. (2008). Selection and pattern-mixture models. In G. Fitzmaurice, et al. (Eds.), *Advances in longitudinal data analysis* (pp. 409–431). London: CRC Press.
Little, R.J.A., D'Agostino, R., Cohen, M., Dickerson, K., Emerson, S., & Farrar, J. et al. (2012). The prevention and treatment of missing data in clinical trials: Special report. *The New England Journal of Medicine*, 1355–1360.
Little, R. J. A., & Rubin, D. B. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models: Comment. *Journal of the American Statistical Association*, *94*, 1130–1132.
Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: Wiley.
Liu, X., Waternaux, C., & Petkova, E. (1999). Influence of human immunodeficiency virus infection on neurological impairment: An analysis of longitudinal binary data with informative drop-out. *Applied Statistics*, *48*, 103–115.
Loeber, R., Keenan, K., & Zhang, Q. (1997). Boys' experimentation and persistence in developmental pathways toward serious delinquency. *Journal of Child and Family Studies*, *6*, 321–357.
Lu, Z. L., Zhang, Z., & Lubke, G. (2011). Bayesian inference for growth mixture models with latent class dependent missing data. *Multivariate Behavioral Research*, *46*, 567–597.
Lu, Z. L., & Zhang, Z. (2014). Robust growth mixture models with non-ignorable missingness: Models, estimation, selection, and application. *Computational Statistics & Data Analysis*, *71*, 220–240.
Mackesy-Amiti, M., Ouellet, L., Finnegan, L., Hagan, H., Golub, E., Latka, M., et al. (2014). Transitions in latent classes of sexual risk behavior among young injection drug users following HIV prevention intervention. *AIDS Behavior*, *18*, 464–472.
Mallinckrodt, C., Sanger, T., Dube, S., DeBrota, D., Molenberghs, G., Carroll, R., et al. (2003). Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry*, *53*, 754–760.
McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.
McMillen, J., Zima, B., Scott, L., et al. (2005). Prevalence of psychiatric disorders among older youths in the foster care system. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*, 88–95.
Meier, M., Caspi, A., Houts, R., Slutske, W., Harrington, H., Jackson, K., et al. (2013). Prospective developmental subtypes of alcohol dependence from age 18 to 32 years: Implications for nosology, etiology, and intervention. *Development and Psychopathology*, *25*, 785–800.
Minini, P., & Chavance, M. (2003). Sensitivity analysis of longitudinal normal data. *Statistics in Medicine*, *23*, 1039–1054.
Moffitt, T. E., Arseneault, L., Jaffee, S., et al. (2008). DSM-V conduct disorder: Research needs for an evidence base. *Journal of Child Psychology & Psychiatry*, *49*, 3–33.
Molenberghs, G., & Kenward, M. G. (2007). *Missing data in clinical studies*. Chichester: Wiley.
Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. New York: Springer.
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.
Muthén, L.K., Muthén, B.O. (1998–2014). Mplus 7 users guide. Los Angeles, CA: Muthén & Muthén.
Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard Press.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods*, *6*, 18–34.
National Research Council (2010). The prevention and treatment of missing data in clinical trials. In *Panel on handling missing data in clinical trials. Division of Behavioral and Social Sciences and Education*. Washington, D.C.: National Academies Press.
Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte-Carlo simulation study. *Structural Equation Modeling*, *14*, 535–569.
Nylund, K., Muthén, B. Nishina, A., Bellmore, A., Graham, S. (2006). Stability and instability of peer victimization during middle school: Using latent transition analysis with covariates, distal outcomes, and modeling extensions. www.statmodel.com.
Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.
Reboussin, B. A., Reboussin, D. M., Liang, K., & Anthony, J. C. (1998). Latent transition modeling of progression of health risk behaviors. *Multivariate Behavioral Research*, *33*, 457–478.
Rizopoulos, D., Verbeke, G., Lesaffre, E., & Vanrenterghem, Y. (2008). A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, *64*, 611–619.
Rodgers, S., Ajdacic-Gross, V., Muller, M., Hengartner, M., Holtforth, M., Angst, J., et al. (2013). *The role of sex on stability and change of depression symptom subtypes over 20 years: A latent transition analysis*. European Archives of Psychiatry: DOI. doi:10.1007/s00406-013-0475-3.
Rotnitzky, A., Robins, J., & Scharfstein, D. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association*, *93*, 1321–1339.
Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, *59*, 829–836.
Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.
Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.
Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
Soto-Ramirez, N., Ziyab, A., Karmaus, W., Zhang, H., Kurukulaaratchy, R., Ewart, S., et al. (2013). Epidemiologic methods of assessing asthma and wheezing episodes in longitudinal studies: Measures of change and stability. *Journal of Epidemiology*, *23*, 399–410.
Sterba, S. K. (2013). Understanding linkages among mixture models. *Multivariate Behavioral Research*, *48*, 775–815.
Sterba, S.K. & Gottfredson, N. (2014). Diagnosing global case influence on MAR versus MNAR model comparisons. Structural Equation Modeling, In press.
Sterba, S. K., & Pek, J. (2012). Individual influence on model selection. *Psychological Methods*, *17*, 582–599.
Ten Have, T., Kunselman, A., Pulkstenis, E., & Landis, J. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, *54*, 367–383.
Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., & Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics*, *3*, 245–265.
Tofighi, D., & Enders, C. (2007). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Greenwich, CT: Information Age.
Velicer, W., Martin, R., & Collins, L. (1996). Latent transition analysis for longitudinal data. *Addiction*, *91*, S197–S209.
Verbeke, G., Lessafre, E., & Spiessens, B. (2001). The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, *35*, 419–434.
Williford, A., Boulton, A., & Jenson, J. (2014). Transitions between subclasses of bullying and victimization when entering middle school. *Aggressive Behavior*, *40*, 24–41.
Witkiewitz, K., & Villarroel, N. (2009). Dynamic association between negative affect and alcohol lapses following alcohol treatment. *Journal of Consulting and Clinical Psychology*, *77*, 633–644.
Witkiewitz, K. (2008). Lapses following alcohol treatment: Modeling the falls from the wagon. *Journal of Studies on Alcohol and Drugs*, *69*, 594–604.
White, T., & Erosheva, E. (2013). Using group-based latent class transition models to analyze chronic disability data from the National Long-Term Care Survey 1984–2004. *Statistics in Medicine*, *32*, 3569–3589.
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.
Xu, S., & Blozis, S. (2011). Sensitivity analyses of mixed models for incomplete longitudinal data. *Journal of Educational and Behavioral Statistics*, *36*, 237–256.
Yang, M., & Maxwell, S. (2014). Treatment effects in randomized longitudinal trials with different types of nonignorable dropout. *Psychological Methods*, *19*, 188–210.
Yang, X., Shoptaw, S., Nie, K., Liu, J., & Belin, T. (2007). Markov transition models for binary repeated measures with ignorable and nonignorable missing values. *Statistical Methods in Medical Research*, *16*, 347–364.
Zhang, Z., & Wang, L. (2012). A note on the robustness of a full Bayesian method for nonignorable missing data analysis. *Brazilian Journal of Probability and Statistics*, *26*, 244–264.

Manuscript Received: 21 FEB 2014
Published Online Date: 20 FEB 2015