A Proof-of-Concept Simulation of the Accelerated Longitudinal Planned Missing Design for Latent Panel Modeling

By

Luke McCune

Submitted to the Department of Psychology and the
Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Master of Arts

Committee members

Pascal Deboeck, Chairperson

Wei Wu

Carol Woods

Date defended: June 20, 2014

The Thesis Committee for Luke McCune certifies
that this is the approved version of the following thesis :

A Proof-of-Concept Simulation of the Accelerated Longitudinal Planned Missing Design for Latent Panel Modeling

Pascal Deboeck, Chairperson

Date approved: June 20, 2014

Abstract

Longitudinal planned missing, represented in the literature by the time-lag model (McArdle & Woodcock, 1997) and the cohort sequential design (Nesselroade & Baltes, 1979), has been thus far restricted to growth modeling and often does not fully utilize the benefits of the planned missingness by estimating a full-longitudinal model. The accelerated longitudinal design may serve as a more flexible and powerful alternative. This study presents a test of the accelerated longitudinal design in a simulated latent panel modeling framework to examine the method's appropriateness for contexts untestable using traditional longitudinal planned missing designs. Three-, four-, and five-cohort models are tested, using a continuum of sample sizes and cohort effect sizes. Results indicate that factor loadings, factor variances, and stabilities across time are replicated well, while characteristics and relationships of the means (i.e., manifest intercepts, latent means, and especially cohort differences) show low efficiency relative to the full sample case. In general, the technique is recommended when no cohort effects are expected, though more expansive research into other possible modeling situations should follow.

List of Figures

List of Tables

Introduction

In studying developmental processes, such as early communication (e.g.,Greenwood et al., 2013), adolescent social and emotional development (see Durlak et al., 2011), and cognitive decline in older adults (e.g., Small et al., 2012, collection of longitudinal data is a necessity for drawing unambiguous conclusions about the process of change and stability (Singer & Willett, 2003). Further, correctly controlled longitudinal data can offer a strong basis for making casual conclusions (Gerring, 2005). Nevertheless, collecting longitudinal data can prove a substantial obstacle to applied researchers in terms of tracking and following up with participants, minimizing attrition, and compensating participants for their continued participation in the study. Findings ways to maximize the efficiency of longitudinal data collection by reducing or eliminating these obstacles is therefore of great value to furthering the study of human development. One such approach, as the focus of this study, involves the use of planned missing data to reduce data collection burdens on developmental researchers. This specific approach, deemed the accelerated longitudinal design, collects multiple cohorts across a short span of time in order to make conclusions across a longitudinal age span. The purpose of this study is to investigate the accuracy and precision of model estimation, specifically for latent panel models, of this particular planned missing data design.

Planned Missing Data

The idea of planned missing efficiency designs for studies (i.e., of sampling subsets of items from different subsets of respondents) started with the matrix sampling techniques discussed by Shoemaker (1973). The goal of this study design was to improve response rates and response quality by reducing the number of items administered to each participant. While matrix sampling has substantial limitations (e.g., it can only be used to estimate item means), it planted the seed for a variety of planned missing methods, including the two-method design, the three-form design, the time-lag design, and, of most importance to this study, the accelerated longitudinal design.

Much of the development of planned missing data designs has focused on cross-sectional (i.e., single time-point) studies. Specifically, research has focused on the two-method planned missingness design (Graham et al., 2006) and the three-forms design (Graham et al., 1996a; Allison & Hauser, 1991). The two-method planned missing design involves administering a cheap but typically biased or error-prone measure of a construct to all participants and administering a more expensive but relatively error-free measure of the same construct to only a small random subset of participants. Using both measures in this way not only allows researchers to cut costs by reducing the number of expensive measures given, but, in a latent modeling framework, it also allows the researcher to extract and model the variance in the cheap measure that is due to error. This extraction of measurement error creates a cleaner common construct among both the cheap and expensive measures (Graham et al., 2006).

The three-forms design, on the other hand, makes no distinction between measures but still has participants respond to only a subset of measures. As per its namesake, the three-forms design has three different but overlapping sets of items, with each participant being randomly assigned to receive one set or another (Graham et al., 1996b). Each of these three forms usually contains a set of variables common to each form (typically demographics, important dependent variables, or auxiliary variables used in predicting unplanned missingness) along with two out of three possible variable sets. For example, Hecht et al. (2003) used a three-forms design, measuring all adolescent participants on demographic items and key variables related to alcohol and drug use while randomly splitting the remaining 23 psychosocial questions between three separate forms. Any given participant was asked to respond to only two of the three forms, shortening the length of the questionnaire, reducing fatigue and allowing the researchers to collect higher-quality data.

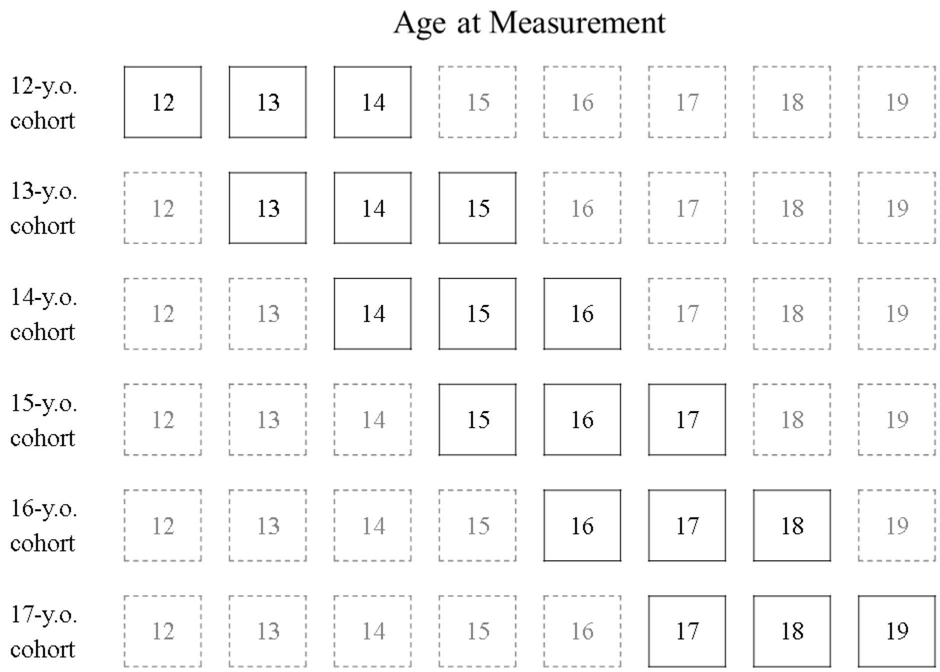
Longitudinal Planned Missing Designs

Longitudinal planned missing designs have received much less attention, comparatively, than their cross-sectional counterparts, and are generally much more specific to certain questions or situations. One such example is the time-lag model developed by McArdle & Woodcock (1997). The design is simple: after an initial data collection, groups of participants are randomly assigned to one other data collection point. So some participants might be measured at Time 1 and Time 2, while others at Time 1 and Time 3 and still others at Time 1 and Time 4, and so on. While this data collection scheme does allow for more valid estimates of practice effects, since the practice effects is neither renewed by nor confounded with a single retest time point, it is harder to estimate patterns of growth more complex than linear models (McArdle et al., 2002).

Focused more on estimating growth rather than accounting for practice effects, the accelerated longitudinal design provides a useful, if underutilized, method of data collection for longitudinal researchers. The basic design was introduced as the "convergence" model by Bell (1953), and made popular by Nesselroade & Baltes (1979) as the cohort-sequential design. This design specifies that several different cohorts are measured over an identical span of time. What results is a diagonally-arranged dataset, wherein each age cohort has data spanning a number of time-points across a particular age range, with all other possible points serving as planned missing for that particular cohort (see Figure 1). Such a design not only shortens the span of time across which a researcher would need to collect data, but the planned missingness component also reduces the cost of the study by decreasing the total number of collected observations. Developmental studies have found that using the cohort-sequential design in a multiple group framework (see (McArdle & Anderson, 1990; McArdle & Hamagami, 1992) closely approximates results found when using a single-cohort true longitudinal design (Duncan et al., 1996). Additionally, having fewer observations per cohort reduces the amount of attrition, while having multiple cohorts also allows history effects (i.e., differences based on historical events experienced by the cohorts) to be separated from the effect of age (Miyazaki & Raudenbush, 2000).

The modeling strategy most commonly used with the accelerated longitudinal planned missing design is multiple group growth modeling, wherein growth trajectories for each cohort (across that particular cohort's age range) are estimated in separate but simultaneous models (Duncan et al., 1996). With the proper parameter constraints on the loadings of the individual growth models, each group then represents an individual segment of the growth across all sampled ages, overlapping with other cohorts depending on the age-span and number of cohorts collected. Estimates like the

Figure 1: Representation of accelerated longitudinal collection scheme



Note. Solid lines denote collected data. Dashed lines denote planned missing data. Figure based on Kofler et al. (2011).

average latent slope for these separate growth models can then be constrained to equality between cohorts to examine the convergence of evidence for a unifying growth trend. One major issue with this approach to using accelerated longitudinal data is that it necessitates the use of multiple-group modeling, such that each group is estimated separately; this requirement greatly increases the sample size necessary to estimate such a model compared to a single-group approach (Muthén, 1989). A goal of this study is to eliminate the need for a multiple-group analysis and instead make use of MIMIC modeling, wherein groups are represented by dummy-coded predictors of latent means (see Hancock, 2001), in order to reduce the sample size requirements for accelerated longitudinal modeling. As long as measurement invariance between groups can be assumed for the latent variables, the MIMIC modeling approach provides a technique that is even more efficient and effective than the multiple-group approach.

Latent Panel Modeling

Another prominent drawback of typical implementations of the accelerated longitudinal design is its restriction to growth modeling. While latent growth modeling is useful for measuring trends across time, it is difficult to model stability across time (Little et al., 2006). Latent growth models are also not able to model relationships between individual timepoints, precluding the ability to include cross-lagged paths and mediation relationships. These points of weakness for latent growth modeling are the points of greatest strength for latent panel modeling, and will be features of focus for this simulation study.

Latent panel modeling is based on means and covariance structures (MACS) analysis and structural equation modeling (SEM) made popular in the social sciences by Joreskog, Sorbom, Magidson, and Cooley (1979) and Bollen (1989). Expanding on factor analysis matrix algebra, the SEM equations include different matrices to account for mean information and regression pathways:

$$\Sigma = \Lambda(\mathbf{I}-\beta)^{-1}\Phi(\mathbf{I}-\beta)^{-1\prime}\Lambda' + \Theta \tag{1}$$

where Σ represents the n x n symmetric covariance matrix of the manifest variables, where n equals the number of manifest variables; Λ represents an n x m matrix of loadings for the latent variables on the manifest variables, where m equals the number of latent variables; $\mathbf{I}$ is an m x m identity matrix used to scale the regression pathways; β represents an m x m matrix of latent regression pathways; Φ represents a symmetric m x m matrix of the covariances between factors; and Θ represents a symmetric n x n matrix of manifest residual variances and covariances (i.e., manifest information not explained by the latent characteristics and relationships). Note that this equation is a simplification where all manifest variables are dependent (i.e., y-side) and there are no independent manifest predictors (i.e., x-side variables).

Further, since mean information is also modeled in SEM, parameters are estimated to recreate the manifest level mean vector:

$$\mathrm{Y} = \Lambda(\mathbf{I}-\beta)^{-1}\mathrm{A} + \mathrm{T} \tag{2}$$

where Y represents the n x 1 manifest variable mean vector; A represents the m x 1 vector of latent means and intercepts; and T represents the n x 1 vector of manifest variable intercepts. Again, this is a simplification where all manifest variables are dependent (i.e., y-side).

The inclusion of the β matrix lies as the main difference between latent growth modeling and latent panel modeling. At least when measuring a single construct across time, latent growth modeling does not make use of the β matrix whereas latent panel modeling estimates "stabilities" (i.e., autoregressive pathways from latent factors at one timepoint to the next). Even when expanding to multiple latent variables, latent regressions are typically only estimated between higher-level growth factors, if at all, rather than between latent factors at individual timepoints as in latent panel modeling. The significance of latent panel modeling's reliance on the β matrix plays into the model's flexibility with measuring timepoint-to-timepoint relationships and its ability to answer categorically different questions versus those answered by latent growth modeling.

Latent panel modeling is crucial in identifying relations among individual time-points on important substantive phenomena measured via multiple indicators. These relationships can be autoregressive (i.e., the same construct correlated with itself across time), cross-lagged (i.e., a construct at one time-point regressed on a different construct at another time-point), or even mediation relationships (i.e., the effect of one construct on another as carried through a third construct) (Little et al., 2007). Specific examples of where latent panel modeling has been used to study steady-change constructs genetic effects on alcohol abuse (van Beek et al., 2012) and BMI (Silventoinen et al., 2011), and Mathew effects in reading (Bast & Reitsma, 1997). Given that latent panel modeling answers distinctly different questions than latent growth modeling and steps should be taken to make the convenience and efficiency of the accelerated longitudinal design applicable to sets of research questions uniquely answered by latent panel modeling.

Multiple Imputation

The other major disadvantage of the prior research using accelerated longitudinal data is that the multiple group framework means that the planned missing data component is ignored in favor of fitting models only to the (planned) complete data within each group. Whereas in the cross-sectional planned missing designs the missing completely at random (MCAR) data imposed by the design is accounted for by modern missing data techniques in order to maximize efficiency and maintain optimal power, the longitudinal planned missing designs tend to remain using complete data analytic techniques to work around the planned missingness instead of utilizing it to its fullest extent. If the planned missing longitudinal data could be accounted for by modern missing data techniques, true longitudinal data structures could be realized such that each participant would be treated as having data across the full span of timepoints by utilizing data from other cohorts to infer the participant's missing scores.

Unfortunately, full information maximum likelihood (FIML), a popular and powerful modern missing data method that uses a separate likelihood function for each different pattern of missingness (Arbuckle, 1996), is not appropriate for the pattern of missingness in accelerated longitudinal designs. This restriction is due to the requirement of the accelerated longitudinal design that some ages have no coverage with other ages; for example, in the 6-cohort design illustrated in Figure 1, there are no rows in the dataset which contain observations at both age 12 and age 19. Since relationships between these ages are represented in the model, and because there are no likelihood functions which would include both ages, there is a disconnect between the model and the data that cannot be remedied by FIML estimation.

One modern missing data technique that is appropriate for addressing the accelerated longitudinal design is multiple imputation. Multiple imputation, proposed by Rubin (1987), involves the repeated prediction and insertion of plausible values based on other available data. Earlier EM imputation (see Dempster et al., 1977) only created one set of complete data, thereby overestimating the precision of the resulting estimates. Multiple imputation was created to account for the variability due to missingness, pooling estimates across imputations while simultaneously accounting for the between-imputation variance imposed by the missingness itself. The most popular form of multiple imputation uses Bayesian methods of prediction based on a multivariate normal distribution with uninformative priors to create a Markov chain based on initially-random parameter draws (Schafer, 1997). This technique is commonly referred to as Markov Chain Monte Carlo imputation and is generally recommended for imputation of multivariate normal data. Since this technique does not depend on separate functions for each pattern of missingness, the lack of coverage between certain ages within the accelerated longitudinal design is no hindrance for multiple imputation. Therefore, it is the modern missing data method of choice for this study.

Accelerated Longitudinal Planned Missing Design

Making use of multiple imputation to account for the planned missingness within the accelerated longitudinal design, allowing the then-complete data to be fit to a latent panel model to assess factor stability across time while a MIMIC modeling approach accounts for between-cohort differences is the specific focus of this simulation. Such an approach allows for more varied models over growth curves and would increase the power of the analysis by making many fewer estimates and separations than the multiple-group approach. Further, whereas previous modeling (e.g., Duncan et al., 1996) has depended on the invariance of growth across cohorts in order to make inferences about overall longitudinal trajectories, the accelerated longitudinal planned missing design approach investigated here would allow researchers to delve deeper into cohort effects (i.e., differences in means between cohorts) through MIMIC modeling without sacrificing the ability to make true longitudinal inferences.

No studies to date have used or investigated this rethinking and expansion of the accelerated longitudinal design. The current simulation, then, represents a proof-of-concept to show that this technique is viable, accurate, and efficient for answering longitudinal questions. It explores the efficiency gained by using the accelerated longitudinal planned missing design for latent panel modeling, and it tested the effects of low sample size or strong cohort effects on parameter estimation under the design. Specifically, this Monte Carlo simulation study investigated the efficiency and accuracy of parameter estimation when fitting accelerated longitudinal data to a latent panel MIMIC model, varying sample size, number of cohorts, and the strength and direction of cohort differences.

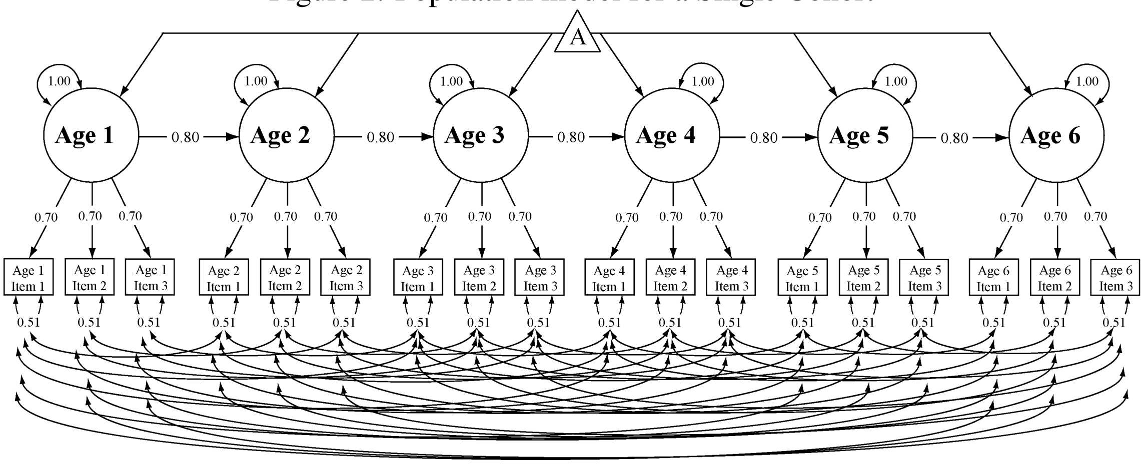
Method

The Monte Carlo simulation consisted of 10,000 replications across three separate cohort conditions. Each replication began with a data generation phase, wherein samples of randomly varying size and randomly varying cohort effects were drawn from a population multiple-group latent panel model to represent the multiple cohorts across time. Data generation was followed by analysis of the full sample for the particular replication according to the specified MIMIC model, including extraction of relative bias and efficiency for each parameter family. Missing data was imposed on the sample in a manner matching the appropriate accelerated longitudinal design, and the missing data was then multiply imputed. The multiply imputed dataset was analyzed with the same MIMIC model as the full sample, and parameter bias and efficiency were calculated as well as relevant statistics comparing the full sample design to the accelerated longitudinal design. These procedures are each detailed below.

Data Generation

Several points of consideration were involved in developing the population model for use in the simulation, including the age span, the number of cohorts, the magnitude of cohort effects, and the sample size ranges to investigate. The age span was fixed at 6, so to allow for multiple even distributions of timepoints across cohorts and in order to isolate the effect of the number of overlapping ages between cohorts. The number of cohorts was either 3, 4, or 5, representing data collected across 4, 3, or 2 ages, respectively. The 3, 4, and 5 cohort designs represent a 33%, 50%, and 67% reduction in the total number of observed datapoints and in the overall length of the theoretical study being conducted. While this more categorical, fixed nature is necessary for representing the number of cohorts given the nature of its scale and how it drastically changes the structure of the data, representing the effect of cohort differences and sample sizes was included in a more continuous, random fashion. The methods of including and varying the cohort effects and sample size are delineated below, as part of the specific data generation process.

Figure 2: Population model for a Single Cohort



Note. "A" represents the generated cohort effect parameter.

The structural equation model from which data were generated for each replication consisted of 6 ages, with a single latent variable at each age. Each latent variable was identified by 3 indicators, allowing for a just-identified model for each factor (Little et al., 2002a). Given the conceptual foundation of collecting multiple cohorts, separate equivalent latent panel models were created in a multiple-group framework to act as the generating model. To represent cohort differences, the latent means of all of the latent variables were given values from a uniform distribution ranging from -0.8 (representing a strong negative cohort effect) to 0.8 (representing a strong positive cohort effect) for each cohort except the last cohort. The last cohort, in this case, represented the baseline group with latent means at zero to serve as a reference point. Since the latent means represented the only population parameters that were different between groups, meaning all other parameters were invariant between groups, the MIMIC models used in the analysis of the data (see Analysis section below) could adequately recreate the population structure despite the difference in specification. For a graphical presentation of the population model for a single cohort, please see Figure 2.

The parameters within the population latent panel model correspond to common values typically used in most simulation research regarding such structural equations. The loadings for each indicator were set at 0.7, meaning that the latent variable associated with each indicator accounted for 49% of the variance for each indicator. This percentage lies near the average variance-accounted-for by many social science articles (see Shafer, 2006; Peterson, 2000), and has been used previously in seminal simulations (e.g., Kenny & McCoach, 2003; Bandalos, 2002; Curran et al., 1996). Keeping the loadings equal between all indicators and across all ages also assumes tau-equivalence (i.e., equal item weighting in determining the construct) and weak measurement invariance over age. Both of these assumptions allowed for more control over the model and elimination of sources of error that are extraneous to the research questions at hand. In order to further maintain simplicity at the item level, item residual variances were set to 0.51 and intercepts to 0, allowing the indicators to correspond to a standard normal distribution with a mean of 0 and total variance of 1. To account for correlated measurement error for the same item across age, a simplex structure (Little et al., 2006) was fit to the θ matrix, such that the residual error is correlated at 0.30 between adjacent ages; 0.30^2, or 0.09, between ages that are two age-units apart; 0.30^3, or 0.027, between ages that are three age-units apart, etc. Latent stabilities also corresponded to a strict simplex structure (also referred to as a lag 1 autoregressive model; Jöreskog, 1970), such that each time-point was regressed on the previous time-point with a β of 0.8.

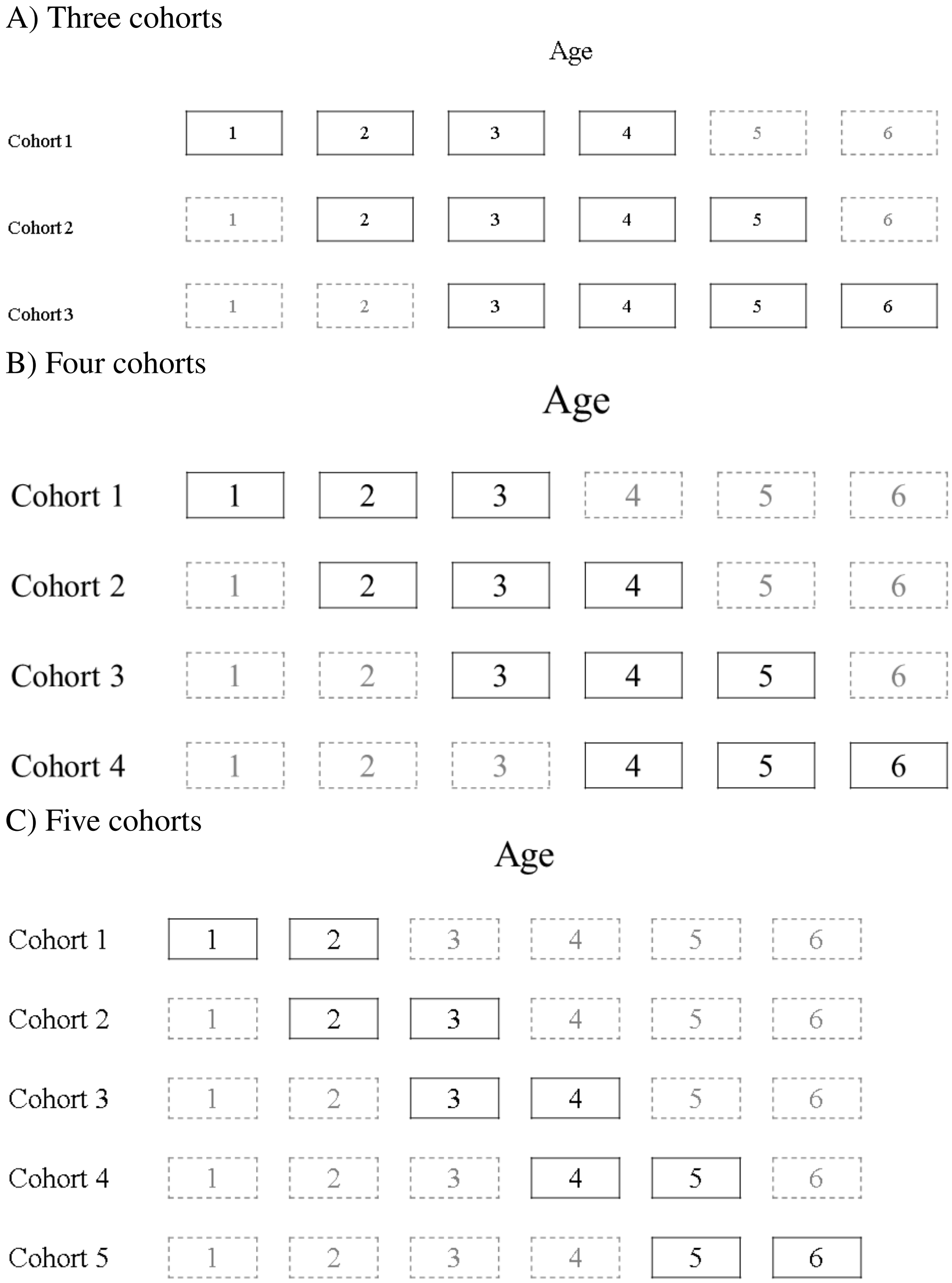
For each replication, the estimates for the population model for each group (including the varying cohort effect size) were used in Equations 1 and 2 to derive the population covariance matrix and mean vector, respectively. Sample data for each group were drawn from a multivariate distribution with the population characteristics using the `rmvnorm` function from the `mvtnorm` package (Genz et al., 2013; Genz & Bretz, 2009), using a random draw from a uniform distribution varying between 50 to 1000 in order to identify the sample size per cohort. These sample size boundaries represent the general range that has been used in previous simulations regarding missing data in structural equation modeling (Enders, 2004; Graham, 2003; Enders & Bandalos, 2001). Once generated from the population covariance matrix and mean vector, the data could then be analyzed and assessed.

Analysis

Each generated sample was analyzed according to its respective SEM MIMIC model (see Figure 4 for an example) using maximum likelihood estimation through the `sem` function in the `lavaan` package (Rosseel, 2012). As required for unambiguous testing for latent characteristics (Little et al., 2007), strong invariance across time was enforced in the estimated model. With the loadings and intercepts constrained to be equal across ages, the variances and means of the latent constructs at Age 2 through 6 were allowed to be estimated while the variance and mean at the first ages were fixed to 1 and 0, respectively, to set the latent scale. Therefore, variances at subsequent ages represent proportional increases or decreases versus the first age and the means at subsequent ages represent standard deviations from the first age. To detect cohort differences, dummy coded indicators corresponding to cohort were included as singly-indicated latent constructs, with unconstrained regression effects on each of the latent age variables. These constructs function identically to their corresponding manifest variables (see Little et al., 2002b), though they are more easily incorporated into the `lavaan` modeling procedure. Finally, latent autoregressive pathways and manifest residual variances and covariances were allowed to be freely estimated across ages in order to allow for possible differences in accuracy and efficiency of estimating latent stabilities and manifest errors between different timepoints.

For the accelerated longitudinal data, certain data cells based on the number of cohorts (see Figure 3) were changed to missingness to mimic the data collection scheme of the accelerated longitudinal approach. The resulting data with missing constituted the accelerated longitudinal sample. After missing data was imposed, 100 MCMC multiple imputations were generated using the `amelia` function in the `Amelia II` package (Honaker et al., 2011) using all of the focal manifest variables as well as the cohort dummy codes as predictors of missing values. The procedure used, by default, sufficient statistics without specified priors from the dataset with missing as starting values to begin the multiple imputation. The multiple imputations were then analyzed according to the estimation model and the results pooled using Rubin's (1987) rules via the `sem.mi` function implemented in the `semtools` package (Pornprasertmanit et al., 2013).

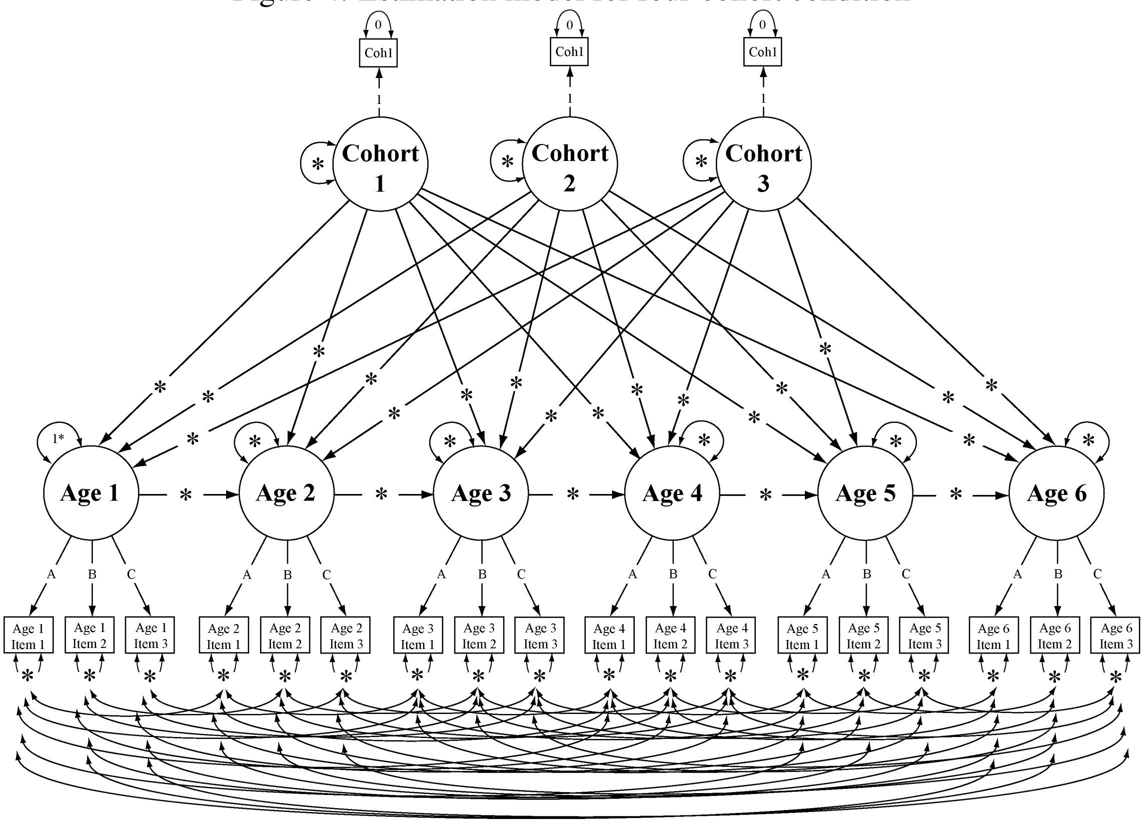
Figure 3: Accelerated longitudinal data structures



Note. Boxes with solid, black lines represent data is present for given cohort and age, while dashed, grey boxes represent imposed missingness by design.

For each set of results, the difference of the estimated parameters versus the replication's population parameters constituted parameter bias caused by sampling, the MIMIC approach, and/or the accelerated longitudinal design. In order to put the biases of separate parameters on similar scales, each calculated bias was scaled by the corresponding population parameter to represent rel-

Figure 4: Estimation model for four-cohort condition



Note. Asterisks indicate free values. Letters indicate values constrained to equality. Correlated residuals modeled

ative bias of the sample. Estimates of relative bias then represent the percentage that the estimated parameters are over- or under-estimated versus their ideal, population values. To effectively test the bias introduced by the accelerated longitudinal planned missing design specifically, the relative difference in bias between the accelerated longitudinal design and the full sample was calculated by dividing the difference in average bias for a particular parameter group (e.g., λ) by the actual value of the population parameter, again putting the measure (this time, difference in bias) of separate parameters onto similar scales.

While systematic (i.e., directional) bias is adequately assessed via average parameter bias (and thus the relative difference in average bias), one risk of using this statistic is that issues are masked when positive *and* negative bias are involved. In these cases, the efficiency of the measure is useful in determining the magnitude of bias, regardless of direction. Calculated as the mean square error (MSE), or the average of the squared deviation of individual parameter estimates from their population values, it generally serves as an estimate of the precision of the estimated parameter. The ratio of the efficiency of the full sample estimates against the efficiency of the accelerated longitudinal estimates represents the relative efficiency of the accelerated longitudinal design, where lower values indicate lower precision for the planned missing design in comparison to the full sample.

Results

A very small proportion of replications were discarded due to non-convergence either in estimation or in imputation; 2, 1, and 0 replications for the Three, Four, and Five Cohort conditions, respectively, were discarded out of the 10,000 replications for each condition. Additionally, replications which produced inadmissible parameter estimates (i.e., $\pm 400\%$ relative bias versus population values) were excluded, as it is expected that applied researchers would regard such estimates as blatantly false Enders & Bandalos (2001). Only 1-2% of the replications in each cohort condition violated the criteria, allowing for 9897, 9890, and 9888 admissible replications for the Three, Four, and Five Cohort conditions, respectively. From these admissible replications, data regarding relative bias and efficiency, as well as the comparative statistics of relative difference in bias and relative efficiency were summarized and assessed by condition.

Relative Bias and Efficiency

Across conditions, descriptive statistics related to the relative bias and efficiency showed very good estimation of factor loadings, factor variances, and stabilities (see Table 1). Each of these families of estimates were accurately and precisely estimated between both the full-sample and accelerated longitudinal designs, with all conditions having less than 6% bias relative to their population values and substantially low MSEs. Together, these results indicate that the variance/covariance structure of the latent factors is adequately represented by both the full-sample and accelerated longitudinal designs.

While manifest intercepts also showed very low relative bias on average (below 1% across all conditions), the variance across samples was exceedingly high. The same pattern of low average bias but high sampling variability affected the latent means, as well. Specifically, when two cohort effects are strong in a consistent direction (i.e., both positive or both negative), the efficiency of the manifest and latent mean information for the full-sample data tended to decrease dramatically (see Figures 5 and 6 for examples). However, the variance was much smaller when the cohort effects were mixed in terms of their direction or when one or more were near zero, as represented by the lighter-colored diagonal in Figures 5 and 6. For the accelerated longitudinal design, on the other hand, the variance for the manifest intercepts was much more a function of the magnitude of the first cohort's effect (see Figure 7 for an example) and the variance for the latent means was only associated with the magnitude of the penultimate cohort's effect (e.g., Cohort 3 in the four cohort design) (see Figure 8 for an example). In short, while the manifest and latent mean information was estimated with no directional bias on average by either full-sample or accelerated longitudinal designs, samples with strong cohort effects tended to be problematic in the estimation of these parameters for both designs.

Estimates of manifest residual errors and residual covariances, as well as estimates of the cohort effects themselves, were produced for the full-sample data with very low average bias (less than 5%) and high precision (MSEs of 0.010 and below), meaning that correlated measurement error and cohort differences were adequately represented using the MIMIC model with full-sample data. In contrast, the accelerated longitudinal design generally underestimated the manifest residual errors and error covariances (by about 20%) and the cohort effects (from 73.5% on average for three cohorts to 89.9% on average for five cohorts). The low MSE values indicate that there was low variance of the manifest residual errors and error covariances, and that the estimates tended to be closer to the population values from the underestimated average rather than further away. However, the alternatively high MSE values for the cohort effects denote low efficiency for the estimation of cohort differences in the accelerated longitudinal design.

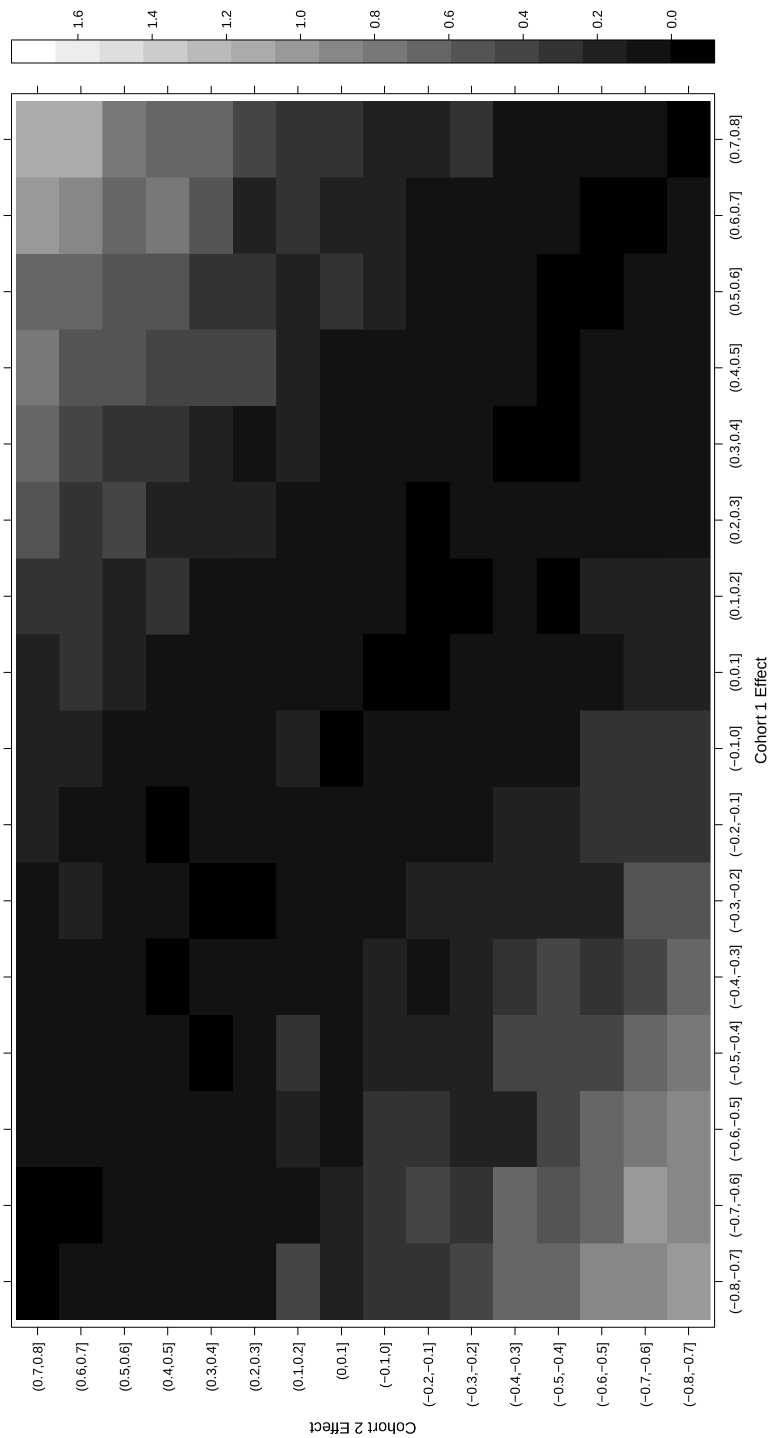
Table 1: Relative Bias and MSE Across Conditions

Estimates		Three Cohorts RB	Three Cohorts MSE	Four Cohorts RB	Four Cohorts MSE	Five Cohorts RB	Five Cohorts MSE
		Average *(95% ECI)*	Median *(95% ECI)*	Average *(95% ECI)*	Median *(95% ECI)*	Average *(95% ECI)*	Median *(95% ECI)*
Factor Loadings	Full-Sample	-0.003	0.000	-0.002	0.000	-0.003	0.000
		(*-0.071, 0.058*)	(*0.000, 0.003*)	(*-0.059, 0.050*)	(*0.000, 0.002*)	(*-0.055, 0.044*)	(*0.000, 0.002*)
	Accelerated Longitudinal	-0.004	0.000	-0.007	0.001	-0.011	0.001
		(*-0.116, 0.100*)	(*0.000, 0.010*)	(*-0.123, 0.096*)	(*0.000, 0.010*)	(*-0.130, 0.093*)	(*0.000, 0.010*)
Manifest Intercepts[1]	Full-Sample	-0.002	0.107	0.005	0.164	-0.006	0.211
		(*-0.880, 0.872*)	(*0.000, 0.913*)	(*-1.095, 1.125*)	(*0.000, 1.571*)	(*-1.290, 1.272*)	(*0.001, 2.128*)
	Accelerated Longitudinal	0.003	0.077	0.001	0.080	0.000	0.078
		(*-0.539, 0.540*)	(*0.000, 0.321*)	(*-0.543, 0.541*)	(*0.000, 0.323*)	(*-0.542, 0.540*)	(*0.000, 0.323*)
Manifest Residual Variances/ Covariances	Full-Sample	-0.005	0.000	-0.020	0.000	-0.012	0.000
		(*-1.187, 1.186*)	(*0.000, 0.003*)	(*-1.085, 1.009*)	(*0.000, 0.002*)	(*-0.948, 0.887*)	(*0.000, 0.002*)
	Accelerated Longitudinal	-0.218	0.001	-0.184	0.001	-0.205	0.001
		(*-1.115, 0.674*)	(*0.000, 0.006*)	(*-0.699, 0.276*)	(*0.000, 0.006*)	(*-0.482, 0.011*)	(*0.000, 0.005*)
Latent Variances	Full-Sample	0.003	0.004	0.001	0.003	0.002	0.003
		(*-0.126, 0.157*)	(*0.001, 0.044*)	(*-0.112, 0.130*)	(*0.000, 0.031*)	(*-0.098, 0.120*)	(*0.000, 0.026*)
	Accelerated Longitudinal	0.015	0.009	0.023	0.010	0.051	0.013
		(*-0.196, 0.285*)	(*0.001, 0.126*)	(*-0.193, 0.303*)	(*0.001, 0.136*)	(*-0.168, 0.358*)	(*0.002, 0.175*)
Latent Means[1]	Full-Sample	-0.001	0.148	0.006	0.237	-0.006	0.322
		(*-0.996, 1.004*)	(*0.004, 1.194*)	(*-1.237, 1.266*)	(*0.010, 1.976*)	(*-1.469, 1.456*)	(*0.017, 2.765*)
	Accelerated Longitudinal	-0.003	1.390	0.003	2.829	-0.002	3.568
		(*-0.386, 0.383*)	(*0.071, 4.465*)	(*-0.541, 0.537*)	(*0.162, 8.615*)	(*-0.500, 0.490*)	(*0.152, 13.814*)
Stabilities	Full-Sample	0.000	0.001	0.001	0.000	0.001	0.000
		(*-0.034, 0.035*)	(*0.000, 0.005*)	(*-0.028, 0.031*)	(*0.000, 0.004*)	(*-0.026, 0.029*)	(*0.000, 0.003*)
	Accelerated Longitudinal	0.002	0.001	0.000	0.001	-0.009	0.002
		(*-0.047, 0.053*)	(*0.000, 0.012*)	(*-0.049, 0.055*)	(*0.000, 0.013*)	(*-0.060, 0.048*)	(*0.000, 0.015*)
Cohort Effects	Full-Sample	0.002	0.006	0.009	0.006	-0.001	0.006
		(*-0.516, 0.540*)	(*0.001, 0.046*)	(*-0.468, 0.558*)	(*0.002, 0.044*)	(*-0.531, 0.510*)	(*0.002, 0.043*)
	Accelerated Longitudinal	-0.735	0.611	-0.811	0.652	-0.899	0.601
		(*-1.041, -0.441*)	(*0.034, 1.698*)	(*-1.284, -0.345*)	(*0.089, 1.610*)	(*-1.378, -0.401*)	(*0.106, 1.511*)

Note. RB = relative bias; MSE = mean square error; ECI = empirical confidence interval

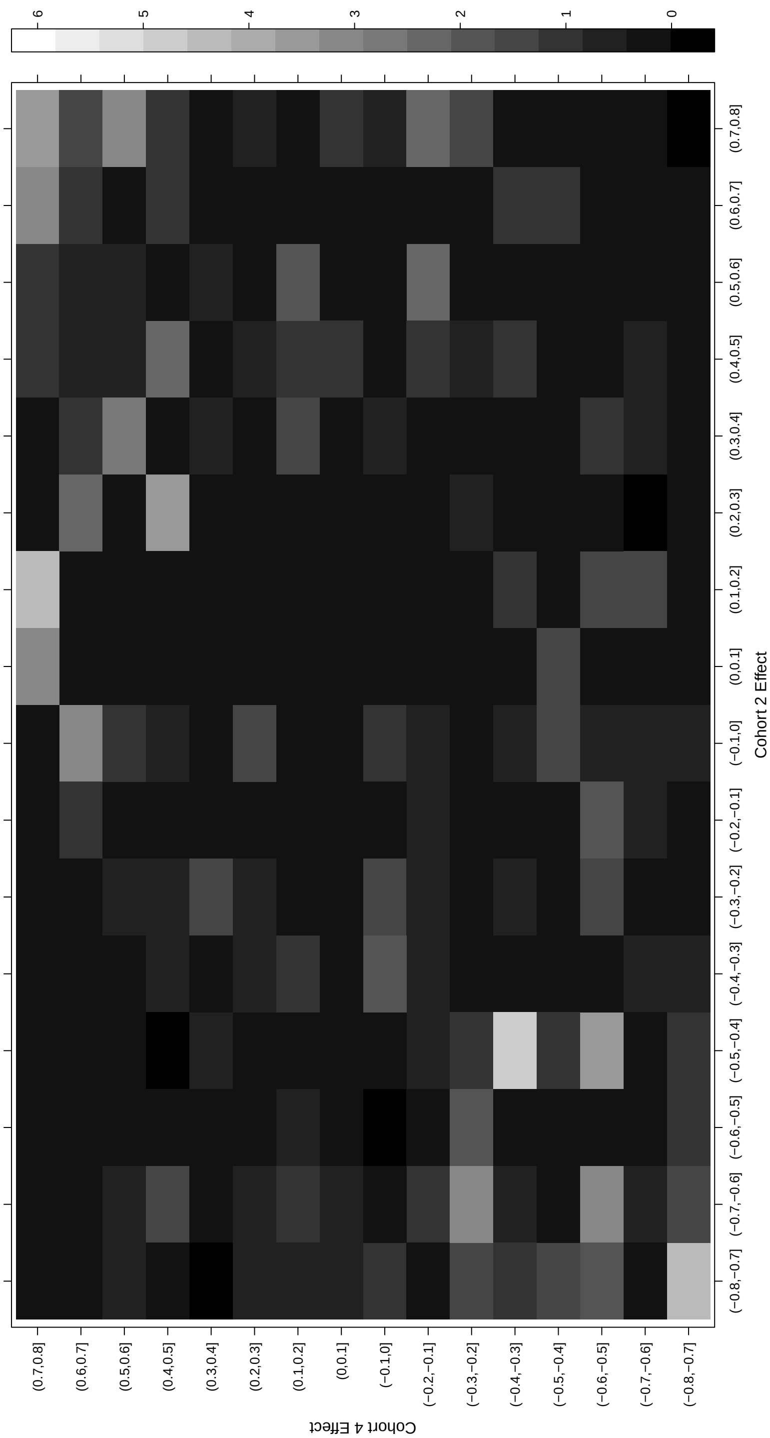
[1]Since the population vector contains only zeroes, relative bias is not calculable. Instead, absolute bias is used.

Figure 5: Manifest Intercept MSE across Both Cohort Effects for Three Cohort Condition with Full-Sample Data



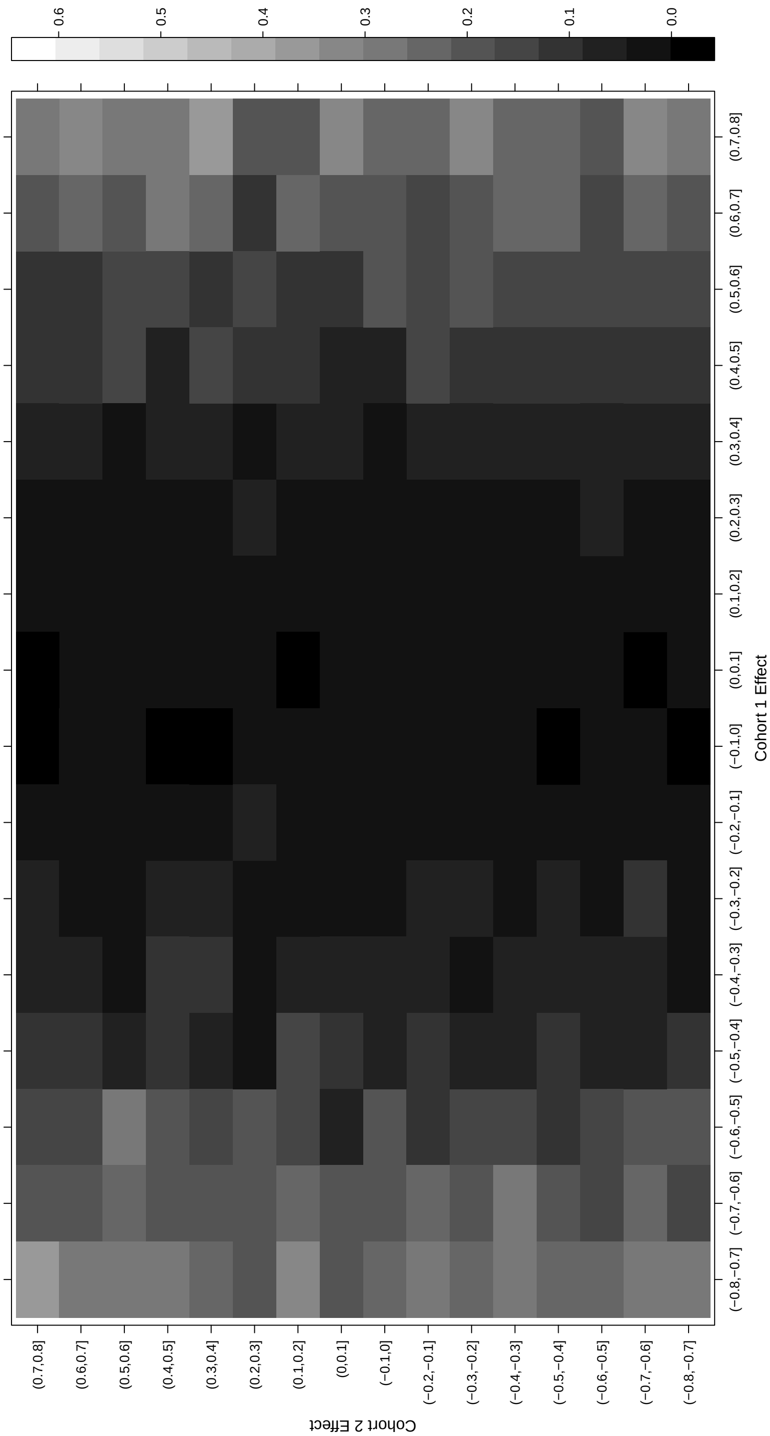
Note. Lighter shades represent higher MSE/lower efficiency.

Figure 6: Latent Mean MSE across Cohort Effects for Cohorts 2 and 4 for Five Cohort Condition with Full-Sample Data



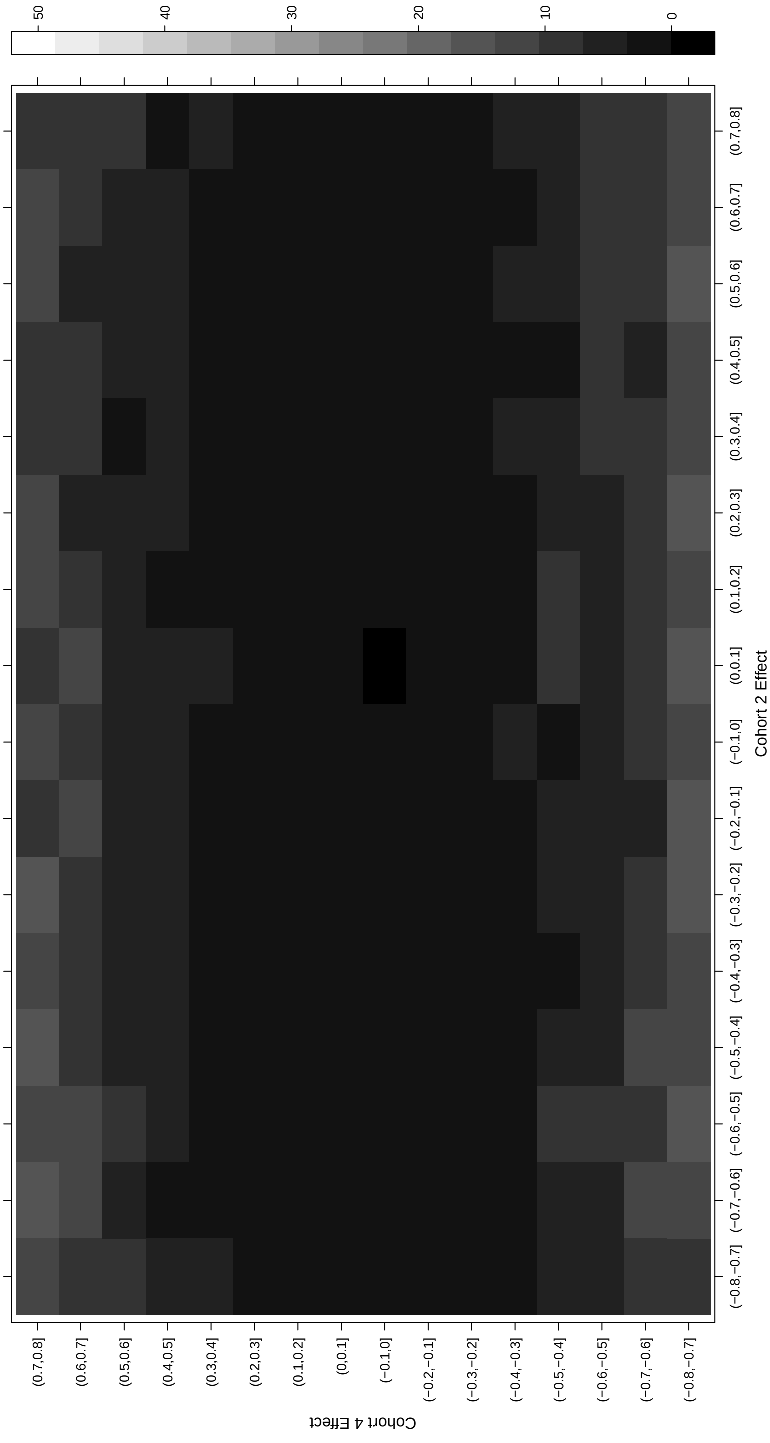
Note. Lighter shades represent higher MSE/lower efficiency.

Figure 7: Manifest Intercept MSE across Both Cohort Effects for Three Cohort Condition with Accelerated Longitudinal Data



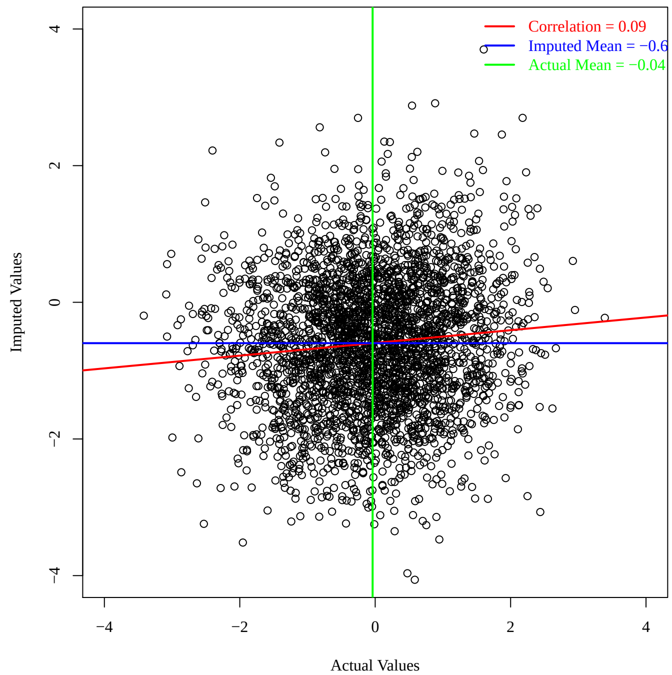
Note. Lighter shades represent higher MSE/lower efficiency.

Figure 8: Latent Mean MSE across Cohort Effects for Cohorts 2 and 4 for Five Cohort Condition with Full-Sample Data



Note. Lighter shades represent higher MSE/lower efficiency.

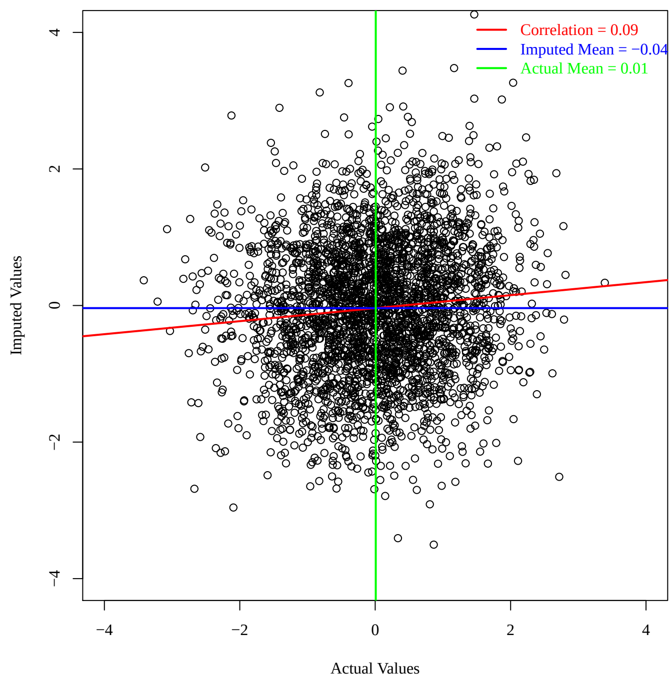
Figure 9: Actual vs. Imputed Values for the First Variable in a Four Cohort Replication



Upon further inspection of the cohort effects for the accelerated longitudinal design, it was found that estimates of cohort effects on the latent variable corresponding to the first age were consistently estimated near 0 across cohorts, magnitude of population cohort effects, and sample sizes, while cohort effect estimates at other ages varied considerably across replications. The source of the consistent zero estimation at the first age, as well as the general underestimation of the cohort effects, was attributed to the negative bias of the imputed values versus the actual values in each corresponding replication's full-sample dataset (see Figure 9 for an example). In turn, the negative bias imposed by the imputation was connected with the magnitude of the cohort effects themselves, since replications where the cohort effects were zero or near zero had much lower bias in the imputed values (see Figure 10 for an example). As a result, replications with cohort effects at or near zero produced the least bias in the estimation of cohort effects for the accelerated longitudinal design.

Attempts to mitigate or eliminate the bias created by the cohort effects on the imputations were

Figure 10: Actual vs. Imputed Values for the First Variable in a Four Cohort Replication with No Cohort Effects



unsuccessful. These strategies included alterations to the population model (e.g., adding a baseline positive linear trend across ages, mean differences between cohorts only at the first age or only at age available for a given cohort in the accelerated longitudinal design), alterations to the analysis model (e.g., constraining the effects of a given cohort to be equal across ages, eliminating strong invariance constraints across ages), and alterations to the imputations (e.g., imputing the data according to a tall, multilevel structure rather than a wide, multivariate structure, including generated auxiliary variables to bolster the imputation procedure). While changing the imputation method from multivariate normal imputation in Amelia II (Honaker et al., 2011) to Multivariate Imputation by Chained Equations (MICE) in the mice package(van Buuren & Groothuis-Oudshoorn, 2011) did eliminate the imputation bias at the first age, the procedure merely introduced the same phenomenon of underestimation to the last age. This difference between imputation methods was most likely caused by the visit sequence of mice compared to the multivariate imputation method of Amelia II, since mice sequentially imputes each variable in the order they were entered (i.e., from indicator 1 at Time 1 to indicator 3 at Time 6) while Amelia II imputes the data all at once without regard to order. Therefore, neither specific population conditions nor modeling constraints nor imputation changes were sufficient to correct the underestimation of present cohort effects; only by not having cohort effects in the population were cohort differences (or, rather, the lack of cohort differences) accurately estimated.

Comparison of Full-Sample to Accelerated Longitudinal Design

A more direct comparison of the full-sample design to the accelerated longitudinal design showed that, for factor loadings and latent stabilities, the accelerated longitudinal approach was no more biased than the full-sample design but generally showed wide variance in efficiency. For other estimates, however, the accelerated longitudinal design was substantially more biased and less efficient than the full-sample approach (see Table 2). Results for the factor loadings showed that there was no directional bias introduced by the accelerated longitudinal approach versus the full-sample approach, but there were strong and varied efficiencies. The empirical 95% confidence intervals of the relative efficiencies (RE) illustrated this instability across replications, covering between a 92.5% reduction in variance due to the accelerated longitudinal approach (i.e., RE = 13.294) to an increase in variance of 83.3 times (i.e., RE = 0.012), using the three cohort results as an example. These wild swings in efficiency were not linked to either sample size or cohort effects; their variance was simply due to chance fluctuations across replications (see Figures 11 & 12 for examples). Therefore, while the accelerated longitudinal loadings are accurate on the whole, each individual loading may not be trustworthy.

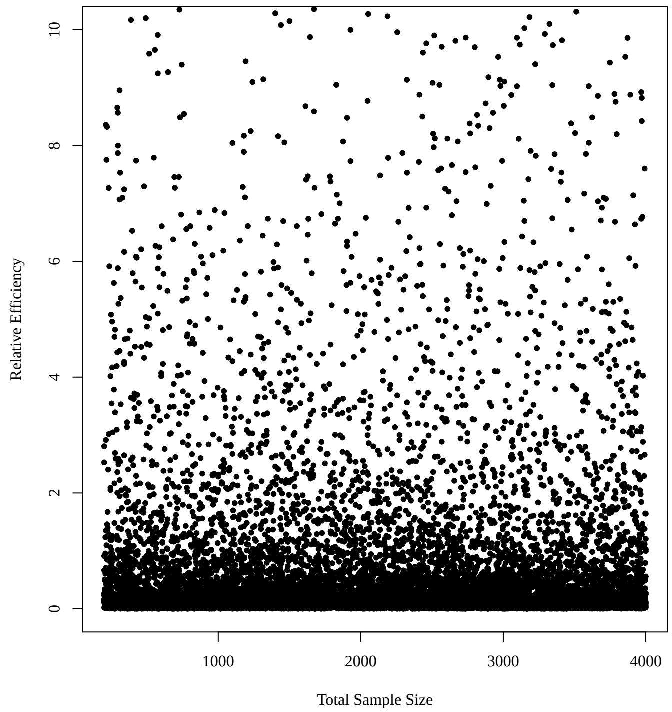
The manifest intercepts and, to a smaller extent, the latent variances, means, and stabilites also exhibited a wide variance in relative efficiency; some replications showed much greater efficiency for the accelerated longitudinal estimates while estimates from other replications demonstrated strong support for the full-sample approach. As with the factor loadings, this variance in efficiency did not show a substantial association with cohort effect sizes or sample size and could only be attributed to random sampling differences. These random sampling differences did not tend to

Table 2: Relative Difference in Bias and Relative Efficiency Between Designs

	Three Cohorts		Four Cohorts		Five Cohorts	
	RDB	RE	RDB	RE	RDB	RE
Estimates	Average (*95% ECI*)	Median (*95% ECI*)	Average (*95% ECI*)	Median (*95% ECI*)	Average (*95% ECI*)	Median (*95% ECI*)
Factor Loadings	-0.002 (*-0.090, 0.087*)	0.380 (*0.012, 13.294*)	-0.005 (*-0.103, 0.085*)	0.281 (*0.008, 11.765*)	-0.009 (*-0.062, 0.061*)	1.582 (*0.004, 10.734*)
Manifest Intercepts	–	1.963 (*0.004, 278.74*)	–	2.558 (*0.008, 403.33*)	–	130.334 (*0.008, 593.66*)
Manifest Residual Variances/ Covariances	-0.212 (*-1.532, 1.053*)	0.533 (*0.321, 0.859*)	-0.164 (*-1.273, 0.965*)	0.441 (*0.251, 0.779*)	-0.192 (*-1.165, 0.734*)	0.418 (*0.211, 0.755*)
Latent Variances	–	0.444 (*0.052, 3.087*)	–	0.313 (*0.030, 2.716*)	–	0.374 (*0.017, 1.828*)
Latent Means	–	0.176 (*0.002, 0.677*)	–	0.098 (*0.003, 1.558*)	–	0.508 (*0.003, 4.089*)
Stabilities	0.001 (*-0.032, 0.040*)	0.535 (*0.094, 2.555*)	0.000 (*-0.039, 0.044*)	0.390 (*0.058, 2.203*)	-0.010 (*-0.054, 0.040*)	0.361 (*0.034, 1.483*)
Cohort Effects	-0.737 (*-1.220, -0.296*)	0.011 (*0.002, 0.260*)	-0.820 (*-1.438, -0.204*)	0.010 (*0.002, 0.122*)	-0.898 (*-1.609, -0.208*)	0.023 (*0.002, 0.119*)

Note. RDB = relative difference in bias; RE = relative efficiency; ECI = empirical confidence interval

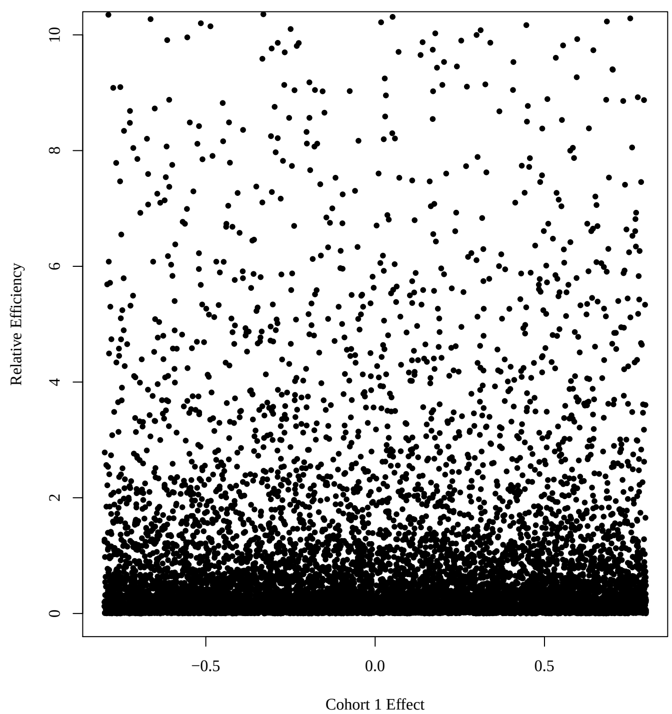
Figure 11: Relative Efficiency of Factor Loadings across Sample Size for Four Cohort Condition



create more average bias in the stabilities, however. Across conditions, the relative difference in bias for the stabilities tended to be less than 5% in either direction, meaning the accelerated longitudinal design provided stability estimates that were as accurate, on average, as the full-sample approach.

The accelerated longitudinal design was not as accurate as the full-sample approach, however, in estimating residual variances and covariances or in estimating cohort effects. For each, there was more underestimation versus the full-sample approach, and this underestimation was especially consistent for cohort effects. Specifically, the 95% empirical confidence intervals show that the cohort effects could be expected to be between 29.6% to 122% more biased on average in the accelerated longitudinal design (using the three cohort condition as an example), while the manifest residual variances and covariances could be expected to be underestimated by 1.5 times to overestimated by 1.05 times (again, using the three cohort condition as an example). The consistent underestimation present in the cohort effects also substantially reduced the accelerated

Figure 12: Relative Efficiency of Factor Loadings across Cohort 1 Effect Size for Four Cohort Condition



longitudinal design's efficiency in estimating such pathways, such that the accelerated longitudinal approach was generally between 0.2% and 15.4% as efficient as the full-sample design across conditions according to the 95% empirical confidence intervals. Estimation of residual variances and covariances, on the other hand, was between 24% and 81% as efficient across conditions as the full-sample design. Therefore, while the accelerated longitudinal approach produced biased estimates on average for manifest residuals and cohort effects, the residuals and residual covariances at least were close enough to maintain a moderate level of efficiency versus the full-sample design.

Discussion

Overall, the prospects of using the accelerated longitudinal planned missing design for latent panel modeling are mixed according to this simulation study. The results for the variance and covariance components of the latent variables (represented by the factor loadings, factor variances, and stabilities) were very promising. These parameters were accurately estimated by the accelerated longitudinal design, producing no more bias on average than the corresponding full-sample approach, and they had low variance overall. While there was high variance in the relative efficiency of the accelerated longitudinal design's loadings, factor variances, and stabilities versus the full-sample approach, the miniscule magnitude of the MSEs themselves show that the estimates have very low variance in an absolute sense even if they are less than ideal in a relative sense. Practically speaking, using the accelerated longitudinal design to efficiently collect longitudinal data should provide accurate, precise conclusions about the communalities between variables (via the loadings) (e.g., the composition of verbal and nonverbal communication across early childhood; Blaga et al., 2009) and stability of the focal construct across time (e.g., relationship of job performance with itself across time; Hülsheger et al., 2010).

Estimation of mean information (i.e., manifest intercepts and latent means) was also unbiased on average, though large amounts of variance for these estimates within both the full-sample and accelerated longitudinal approaches is less than encouraging for any single given application.

Given small or nonexistent cohort effects, however, the efficiency of both approaches becomes more acceptable. Thus studies looking to examine strong/metric longitudinal invariance (e.g., establishing consistency of a physical activity enjoyment scale across time; Mullen et al., 2011) or the difference of latent means across time (e.g., difference in peer victimization across grade levels; Williford et al., 2012) can and should make use of the more cost- and time-efficient accelerated longitudinal approach as long as strong cohort effects are not expected.

Regardless of cohort effects, the accelerated longitudinal approach also had some issues in accurately estimating residuals and residual covariances. Underestimated by about 20% on average, these parameters would under-represent the amount of variance due to measurement error, both time-specific (e.g., weather's effect on negative affect, as a barrier to physical activity; Geller et al., 2012) and common across time (e.g., specific variance associated only with autonomy in a model about positive school experiences; Stiglbauer et al., 2013). However, given the high relative efficiency versus the full-sample design and the low MSEs, the average underestimation does not seem too severe as to substantially hinder the residuals. Further, since residual variances and covariances are typically included only to provide cleaner, better estimates of latent characteristics and relationships (Kline, 2011), and are not typically part of substantive hypotheses or conclusions, the impact of some minor underestimation in such parameters is minimal from a practical standpoint.

The underestimation present for the cohort effects, on the other hand, does present a considerable obstacle for utilizing the accelerated longitudinal design with latent panel modeling. Although this study found that cohort effects will be accurately estimated when they are not present in the population, departures from this specific circumstance (e.g., when children's birth cohort/peer group influences their baseline level of aggression; Stanger et al., 1997) lead to consistent and substantial underestimation caused by biased imputations. Additionally, different population conditions (aside from absent cohort effects), model specifications, or imputation methods did not appear to correct this parameter bias. The silver lining to this dilemma of inaccurate cohort effects is that, even with strongly biased estimates of cohort differences, the cohort effects are extracted by the dummy codes and do not appear to bias the stabilities or latent means. In other words, the accurate estimation of other important latent characteristics depends on the possibly inaccurate estimation of the cohort effects. While future studies may perfect aspects of the accelerated longitudinal design to better evaluate differences between cohorts, current applied researchers can still make use of the technique to effectively and efficiently answer questions about longitudinal stability and change.

Limitations and Future Research

In the interest of focusing attention on the performance of the accelerated longitudinal design under generally ideal circumstances, several restrictions were put on the simulation that should be examined and tested fully in future research. The first among these restrictions was the use of constant stabilities. While a consistent autoregressive effect across time does fit with an ideal simplex structure, exploring the effect of differential stabilities and extra lag effects could provide more information about the accelerated longitudinal design's use in circumstances where the latent panel model is less-than-perfect. For example, Eisenhower, Blacher, & Baker (2013) found significant one- and two-year autoregressive effects of child behavior problems, such that the effect of a two-year delay predicted a significant portion of variance in child behavior problems above a simple one-year delay. Testing the ability of the accelerated longitudinal design to estimate such relationships would assess its usefulness and flexibility for other researchers in need of efficient longitudinal data collection methods to fit their complex models of stability and change.

In addition to testing the accelerated longitudinal design's capacity to incorporate more nuanced autoregressive effects, priority should be given to identifying the ability of accelerated longitudinal latent panel models to detect model mispecification. Based on the previous example, would the model fit from the accelerated longitudinal approach be able to highlight an issue if only a one-year autoregressive were used, or would the increased variance due to the missingness make detection of a two-year autoregressive difficult? Researchers should also investigate the effects of forcing longitudinal invariance when it is violated, given the necessity of longitudinal invariance for inclusion and unambiguous testing of across-time latent characteristics (e.g., stabilities) (Kline, 2011). If longitudinal invariance cannot be safely assumed, or if parameters are missing that should be included, then the method should be powerful enough to provide the researcher with an adequate warning.

In the same way that this study chose to maintain simplicity with constant stabilities and correctly specified analysis models, so too was simplicity enforced by examining only one latent construct over time. A sizeable amount of latent panel research tends to consider multiple constructs across time, however, often with cross-lagged pathways to denote lagged effects of one construct on another (Little et al., 2007). A number of these studies also include mediational models (up to 65% of studies in some journals; Rucker, Preacher, Tormala, & Petty, 2011), testing an indirect chain of cross-lagged effects from one variable through another to an outcome. Given the popularity cross-lagged and mediational panel models and the prospective boon that the accelerated longitudinal approach could provide saavy researchers, it is critical that efforts be made to test the performance of the accelerated longitudinal design with multiple variables across time. If it proves useful under such circumstances, this longitudinal planned missing design could ease the burden of data collection for large percentages of developmental researchers.

Conclusions and Implications

With the goal of expanding the popular cohort sequential data design, this study proposed utilizing the design's planned missing data component with latent panel modeling. In this way, the efficiency gains realized by the planned missingness can be applied to more than the common implementations of multiple-group growth modeling, and can instead be used in assessing stability rather than overall growth. While strong cohort effects can cause issues both with their own estimation and estimation of manifest and latent mean information, estimates of the factor loadings, variances, and stabilities remain unbiased and generally efficient regardless of cohort effect size and direction. More work testing the design under other models and situations should follow these encouraging results, providing the flexibility and empirical support needed to inspire applied researchers to use the accelerated longitudinal planned missing design to make their studies fast, cheaper, and more efficient.

References

Allison, P. D. & Hauser, R. M. (1991). Reducing bias in estimates of linear models by remeasurement of a random subsample. *Sociological Methods & Research*, 19(4), 466–492.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9(1), 78–102.

Bast, J. & Reitsma, P. (1997). Mathew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research*, 32(2), 135–167.

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child development*, 24(2), 145–152.

Blaga, O. M., Shaddy, D. J., Anderson, C. J., Kannass, K. N., Little, T. D., & Colombo, J. (2009). Structure and continuity of intellectual development in early childhood. *Intelligence*, 37(1), 106–113.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological methods*, 1(1), 16.

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

Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods*, 1(3), 236.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child development*, 82(1), 405–432.

Eisenhower, A., Blacher, J., & Baker, B. L. (2013). Mothers' perceived physical health during early and middle childhood: Relations with child developmental delay and behavior problems. *Research in developmental disabilities*, 34(3), 1059–1068.

Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and psychological measurement*, 64(3), 419–436.

Enders, C. K. & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8(3), 430–457.

Geller, K. S., Nigg, C. R., Motl, R. W., Horwath, C. C., & Dishman, R. K. (2012). Transtheoretical model constructs for physical activity behavior are invariant across time among ethnically diverse adults in hawaii. *Psychology of sport and exercise*, 13(5), 606–613.

Genz, A. & Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Heidelberg: Springer-Verlag.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2013). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9996.

Gerring, J. (2005). Causation: A unified framework for the social sciences. *Journal of Theoretical Politics*, 17(2), 163–198.

Graham, J. W. (2003). Adding missing-data-relevant variables to fiml-based structural equation models. *Structural Equation Modeling*, 10(1), 80–100.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996a). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31(2), 197–218.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996b). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31(2), 197–218.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological methods*, 11(4), 323.

Greenwood, C. R., Walker, D., Buzhardt, J., Howard, W. J., McCune, L., & Anderson, R. (2013). Evidence of a continuum in foundational expressive communication skills. *Early Childhood Research Quarterly*, 28(3), 540–554.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and mimic approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, 66(3), 373–388.

Hecht, M. L., Marsiglia, F. F., Elek, E., Wagstaff, D. A., Kulis, S., Dustman, P., & Miller-Day, M. (2003). Culturally grounded substance use prevention: an evaluation of the keepin'it real curriculum. *Prevention Science*, 4(4), 233–248.

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45(7), 1–47.

Hülsheger, U. R., Lang, J. W., & Maier, G. W. (2010). Emotional labor, strain, and performance: Testing reciprocal relationships in a longitudinal panel study. *Journal of occupational health psychology*, 15(4), 505.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, 23(2), 121–145.

Jöreskog, K. G., Sörbom, D., Magidson, J., & Cooley, W. W. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.

Kenny, D. A. & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10(3), 333–351.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford press.

Kofler, M. J., McCart, M. R., Zajac, K., Ruggiero, K. J., Saunders, B. E., & Kilpatrick, D. G. (2011). Depression and delinquency covariation in an accelerated longitudinal sample of adolescents. *Journal of consulting and clinical psychology*, 79(4), 458.

Little, T. D., Bovaird, J. A., & Slegers, D. W. (2006). Methods for the analysis of change. *Handbook of personality development*, (pp. 181–211).

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002a). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural equation modeling*, 9(2), 151–173.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002b). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural equation modeling*, 9(2), 151–173.

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31(4), 357–365.

McArdle, J. J. & Anderson, E. (1990). Latent variable growth models for research on aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 21–44). San Diego, CA: Academic Press, third edition.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental psychology*, 38(1), 115.

McArdle, J. J. & Hamagami, F. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. *Experimental aging research*, 18(3), 145–166.

McArdle, J. J. & Woodcock, R. W. (1997). Expanding test–retest designs to include developmental time-lag components. *Psychological Methods*, 2(4), 403.

Miyazaki, Y. & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods*, 5(1), 44–63.

Mullen, S. P., Olson, E. A., Phillips, S. M., Szabo, A. N., Wójcicki, T. R., Mailey, E. L., Gothe, N. P., Fanning, J. T., Kramer, A. F., & McAuley, E. (2011). Measuring enjoyment of physical activity in older adults: invariance of the physical activity enjoyment scale (paces) across groups and time. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1), 1–9.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54(4), 557–585.

Nesselroade, J. R. & Baltes, P. B. (1979). *Longitudinal research in the study of behavior and development*. New York, NY: Academic Press.

Peterson, R. A. (2000). A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters*, 11(3), 261–275.

Pornprasertmanit, S., Miller, P., Schoemann, A., & Rosseel, Y. (2013). *semTools: Useful tools for structural equation modeling*. R package version 0.4-0.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*, volume 307. Wiley.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 5(6), 359–371.

Schafer, J. (1997). *Analysis of incomplete multivariate data*. London, England: Chapman\ & Hall.

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, ces-d, hamilton, and zung. *Journal of clinical psychology*, 62(1), 123–146.

Shoemaker, D. M. (1973). *Principles and procedures of multiple matrix sampling*. Oxford, England: Ballinger.

Silventoinen, K., Kaprio, J., & Yokoyama, Y. (2011). Genetic regulation of pre-pubertal development of body mass index: a longitudinal study of japanese twin boys and girls. *Behavior genetics*, 41(2), 234–241.

Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, England: Oxford University Press.

Small, B. J., Dixon, R. A., McArdle, J. J., & Grimm, K. J. (2012). Do changes in lifestyle engagement moderate cognitive decline in normal aging? evidence from the victoria longitudinal study. *Neuropsychology*, 26(2), 144.

Stanger, C., Achenbach, T. M., & Verhulst, F. C. (1997). Accelerated longitudinal comparisons of aggressive versus delinquent syndromes. *Development and Psychopathology*, 9(1), 43–58.

Stiglbauer, B., Gnambs, T., Gamsjäger, M., & Batinic, B. (2013). The upward spiral of adolescents' positive school experiences and happiness: Investigating reciprocal effects over time. *Journal of school psychology*, 51(2), 231–242.

van Beek, J. H., Kendler, K. S., de Moor, M. H., Geels, L. M., Bartels, M., Vink, J. M., van den Berg, S. M., Willemsen, G., & Boomsma, D. I. (2012). Stable genetic effects on symptoms of alcohol abuse and dependence from adolescence into early adulthood. *Behavior genetics*, 42(1), 40–56.

van Buuren, S. & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3), 1–67.

Williford, A., Boulton, A., Noland, B., Little, T. D., Kärnä, A., & Salmivalli, C. (2012). Effects of the kiva anti-bullying program on adolescents' depression, anxiety, and perception of peers. *Journal of abnormal child psychology*, 40(2), 289–300.