International Journal of Behavioral Development

Published by:
SAGE
http://www.sagepublications.com

On behalf of:

International Society for the Study of Behavioral Development

Additional services and information for *International Journal of Behavioral Development* can be found at:

Email Alerts: http://jbd.sagepub.com/cgi/alerts

Subscriptions: http://jbd.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Jul 31, 2014

OnlineFirst Version of Record - Jan 23, 2014

What is This?

Planned missing designs to optimize the efficiency of latent growth parameter estimates

Mijke Rhemtulla,[1] Fan Jia,[2] Wei Wu,[2] and Todd D. Little[3]

Abstract

We examine the performance of planned missing (PM) designs for correlated latent growth curve models. Using simulated data from a model where latent growth curves are fitted to two constructs over five time points, we apply three kinds of planned missingness. The first is item-level planned missingness using a three-form design at each wave such that 25% of data are missing. The second is wave-level planned missingness such that each participant is missing up to two waves of data. The third combines both forms of missingness. We find that three-form missingness results in high convergence rates, little parameter estimate or standard error bias, and high efficiency relative to the complete data design for almost all parameter types. In contrast, wave missingness and the combined design result in dramatically lowered efficiency for parameters measuring individual variability in rates of change (e.g., latent slope variances and covariances), and bias in both estimates and standard errors for these same parameters. We conclude that wave missingness should not be used except with large effect sizes and very large samples.

Keywords

planned missing designs, latent growth curves, three-form design, wave missingness, longitudinal planned missingness

*This article accepted during Marcel van Aken's term as Editor-in-Chief.

Planned missing data designs allow researchers to reduce the testing burden on participants, leading to higher-quality data with less unplanned missingness and smaller fatigue and practice effects (Harel, Stratton, & Aseltine, 2012). Planned missingness can be applied to many complex models resulting in no added bias and minimal power loss. In the present paper, we apply both three-form planned missingness, where participants are assigned to miss a subset of items at every time point (Graham, Hofer, & MacKinnon, 1996; Graham, Hofer, & Piccinin, 1994; Graham, Taylor, Olchowski, & Cumsille, 2006; Hansen et al., 1988), and wave missingness, where participants are assigned to miss a subset of measurement occasions (Graham, Taylor, & Cumsille, 2001) to simulated latent growth curve data. We model latent growth curves of two constructs, allowing the intercepts and slopes to covary. We examine the results in terms of bias and efficiency of the estimates of these latent growth parameters—the means, variances, and covariances among the latent intercepts and slopes.

Latent growth curve models

Latent growth curve models (LGMs), (e.g., McArdle & Epstein, 1987; McArdle & Nesselroade, 2003; Preacher, Wichman, MacCallum & Briggs, 2008) are some of the most popular models for longitudinal data, because they allow researchers to simultaneously model group trends in change over time as well as individual variability in those trends. In a very basic linear latent growth curve, observed data from a single construct are gathered at a number of fixed time points from a group of people. Two latent variables—an intercept and a linear slope—capture participants' initial levels (where 'initial' typically refers to the first time point, but can be placed anywhere that makes sense) and their rates of linear change

over time. The mean intercept and mean slope describe the average intercept and the average rate of change; for example, a positive mean slope means that, on average, participants increase over time in the construct that is being measured. The variance of the intercept and slope reflect constant individual differences between people, and individual differences in the rates of change, respectively. For example, a large slope variance means that people's individual growth trajectories vary widely in their steepness.

The linear model can be modified to include non-linear trajectories by including quadratic and cubic trends, or by allowing the shape of the slope to be freely estimated rather than fixed in linear increments (Meredith & Tisak, 1990). This approach is called a latent basis curve, and it is appropriate whenever there is not a strong theoretical reason to assume that change in a construct is linear, or when a particular parametric shape does not fit the data well.

The construct(s) whose development are being modeled can be represented either by a single observed variable at each time point (e.g., a test score or scale score) or by a latent variable with several observed indicators. The latent variable approach is called a multivariate LGM, or a "curve-of-factors" model (McArdle, 1988) because the growth curve is overlaid on a series of latent factors that represent the construct over time instead of single observed

[1] University of Amsterdam, The Netherlands
[2] University of Kansas, USA
[3] Texas Tech University, USA

Corresponding author:
Mijke Rhemtulla, Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands.
Email: mrhemtulla@uva.nl
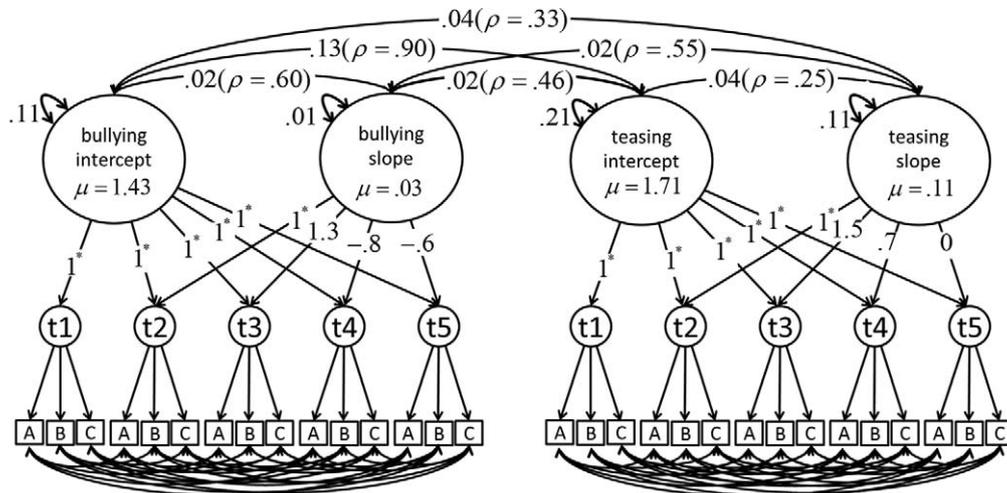
Figure 1. The data generating and analysis model. A, B, and C refer to items or item parcels/subscales from each of the three-form missing sets. Each observed variable is missing one-third of observations. Strong invariance is imposed on the model such that factor loadings are constrained to be equal across time points, and the intercepts of variables A to C are constrained to be equal across time points in estimation. The effects-coded method is used for model identification, which means that the set of loadings for each latent *t* variable is constrained to average 1.0 and the set of observed variable intercepts averages zero. The loadings and residual variances of the ten latent *t* variables were then estimated with the constraints. Intercepts of the latent variables *t1–t5* are 0 in the population, and fixed to 0 in estimation. Asterisks denote fixed parameter values.

variables. An important advantage of this approach is that measurement error is removed at the level of the underlying latent factors, so estimates of growth curve parameters can be estimated with greater precision (Hertzog, Lindenberger, Ghisletta, & von Oertzen, 2006; Hertzog, von Oertzen, Ghisletta, & Lindenberger, 2008).

A conceptual benefit of LGMs is that they allow researchers to study relations among individual differences in rates of change over time across different constructs (e.g., Curran, Stice, & Chassis, 1997; Rhemtulla & Tucker-Drob, 2011). For example, LGMs allow researchers to examine not only the correlation between children's math and reading aptitude at fixed points in time, but to examine whether individual children's rates of development in math are correlated with their rates of development in reading: Are quick math learners also quick reading learners?

The model we use in this paper is a two-construct multivariate LGM based on the parameter values reported in Little (2013, Figure 8.7). The model parameters were estimated from Espelage's data (see Espelage, Rao, Little, and Rose, 2011), in which bullying and homophobic teasing were measured at five time points, each six months apart. In that study, an LGM was fitted to each construct using a latent basis model to represent the change trajectory. The latent intercepts and slopes of the two constructs were allowed to correlate, and phantom constructs were used to facilitate the interpretation of the relations between the latent intercepts and slopes. In the present paper, we simplified the model by excluding the phantom constructs. Otherwise, our model was identical to theirs (see Figure 1).

Item-level planned missingness

Several planned missingness (PM) designs have been proposed for cross-sectional data (Bunting, Adamson, & Mulhall, 2002; Graham, Hofer, & MacKinnon, 1996; Graham et al., 2006). These typically involve imposing item-level missingness such that some or all participants are randomly assigned to respond to only a subset of items. For example, the three-form design (Hansen et al., 1988; Graham

et al., 1996; Graham et al., 1994; Graham et al., 2006), allows researchers to insert 25–33% missing data by assigning participants to complete a subset of all the survey items. In this design, surveys are divided into four subsets including a common set (X, which every participant completes) and three partial sets (A, B, and C, which two-thirds of participants complete). By giving each participant a smaller number of items to complete, participants are less likely to skip items, reducing the amount of non-random missingness (Harel et al., 2012).

Item-level missingness can be imposed in longitudinal data collection by assigning participants to respond to a subset of items at each measurement occasion. Recent studies (e.g., Jorgensen et al., in press; Jia et al., in press) have studied the performance of item-level missingness by imposing the three-form design at each time point in longitudinal panel models, where the predictive relations between constructs over time are of interest (e.g., longitudinal mediation). To our knowledge no studies have examined the effect of item-level planned missingness in the context of latent growth curve models. This omission is important because missingness can have very different effects on parameters of different models (e.g. Rhemtulla & Savalei, 2012; Rhemtulla, Savalei, & Little, under review). For example, when three-form missingness is imposed, the power of factor loadings, means, variances, and regression coefficients are decreased to a different extent. As such, three-form missingness may have better results for some models (e.g., confirmatory factor analysis models) than others (e.g. regression models).

Wave-level planned missingness

McArdle and colleagues (McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002; McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009; McArdle & Woodcock, 1997) have proposed several models that capitalize on missing data in longitudinal designs. For example, the accelerated time-lag design (McArdle & Woodcock, 1997) begins with a two time-point design where the amount of time between two testing occasions varies across

participants (as is frequently the case with two time-point data). The data are then arrayed into multiple time points such that every participant has complete data at the first measurement occasion and at one later occasion (determined by the particular time lag between testing occasions for each participant), with missing data at every other occasion. This layout creates a high proportion of missing data, as each participant is measured at just two time points out of many, but the resulting model describes continuous growth over time, and can even be used to quantify and remove practice effects from the growth trend. This design is an example of *wave-level* missingness, because participants are missing entire waves of data collection. An additional benefit of this kind of design is that it can drastically reduce the cost of data collection compared to a design where all participants are measured at each time point.

Graham et al. (2001) compared the cost efficiency of wave-level missingness designs to those with complete data at every wave, holding the total number of data points collected equal (i.e. they compared missing data designs to complete data designs with a smaller sample size). Beginning with a univariate linear latent growth curve model with five measurement occasions, they simulated several different wave-missing designs where each participant was missing up to three waves. They found that when missingness was concentrated in the middle waves (waves 2–4), the power to detect a significant regression path from an observed grouping variable to the latent linear slope was higher than when missingness was distributed evenly across all five waves. Moreover, they found that the efficiency of that regression coefficient was higher for any planned missing design than for a comparable complete data design that had the same total number of data points (e.g., a wave-missing design with 50% missing data produced more efficient regression estimates than a complete data design with a 50% smaller sample size).

Mistler and Enders (2012) described a follow-up to this finding where they examined the efficiency of the mean of the latent growth parameter in a six time-point linear growth curve model in two different wave-missing designs. In the first, each participant was missing two out of six occasions, with missingness distributed equally across all six time points. In the second, each participant was missing two out of the middle four occasions, with complete data at the first and last occasions. Both these conditions were compared to a complete data design with a reduced sample size to equate the total number of data points across the three designs. Power to detect the latent growth mean was highest for the second planned missing design (complete end points, 92%), and lower for the complete data—reduced sample design (83%) as well as the first planned missing design (81%). Power to detect the mean of a quadratic growth trend had the same pattern: highest for the second planned missing design (87%), and slightly lower for a complete data—reduced sample design (85%) and lowest for the first planned missing design (81%).

Both Graham et al (2001) and Mistler and Enders (2012) only examined power for a single parameter in latent growth models. In the present study we investigated the efficiency of a range of other model parameters in latent growth curves, including factor loadings, means and variances of the latent intercepts and slopes, and covariances between slopes.

We imposed three different planned missing designs to measure the effect of item-level and wave-level missingness on the bias and efficiency of latent growth parameter estimates. First, three-form missingness was applied to data at each wave. Second, we imposed the wave-level missingness that Graham et al. (2001) found to be most efficient, with 50% of the sample missing each of

waves 2–4, and 30% of the sample missing wave 5. Finally, we combined item-level and wave-level missingness in the same design.

Method

Population

Complete, multivariate normal data were generated from means and covariance matrices implied by the model shown in Figure 1 using population values from Little (2013) and Espelage et al. (2011). The model examined associations between latent intercepts and slopes of bullying and homophobic teasing. The value of the correlation between the two slopes (.55 in the original data) was one of the parameters we varied in the simulation; all other means, variances, and covariances were held constant at the values displayed in the figure (from Little, 2013). Each of the ten latent variables representing the two constructs at five time points was indicated by three observed variables. In addition, strong factorial invariance (Meredith, 1993) was specified by fixing the observed variables' loadings and intercepts equal across five time points. For example, the population values of loadings for bullying were {0.865, 1.122, 1.103} at all time points, and the intercepts were {0.417, –.111, –.036}. Strong factorial invariance is required when there is a measurement model at each time point to ensure that the definition of the construct does not change over time (see Little, 2013). Residual variances were not fixed to be equal across time (i.e., strict invariance was not imposed), and residual variances of the observed indicators were allowed to covary across the same items over time. Values of factor loadings, residual variances, and residual covariances are omitted from Figure 1 for clarity.

For each construct, the latent intercept had fixed loadings of 1.0 on the five first-order latent variables. The growth trends of the two constructs were identified by fixing the slope loading to 0 at the first occasion and to 1 at the second occasion; the slope loadings were freely estimated at the third through fifth occasions. The population values for these last three occasions were {1.3, –0.8, –0.6} and {1.5, 0.7, 0}, for the bullying and teasing slopes, respectively. The means of the first-order latent variables are fixed to zero. As a result, the mean-level information about the constructs is completely represented by the latent intercepts and slopes. Means and variances of the growth parameters are in Figure 1.

Within each type of planned missingness, we examined 15 conditions that were formed by crossing two factors: three levels of the latent slope correlation, $\rho_{S1,S2} = \{.1, .3, .55\}$, and five levels of sample size, $n = \{100, 300, 500, 800, 1000\}$. We were particularly interested in varying the latent slope correlation because it is arguably the parameter of greatest substantive interest (e.g., Rhemtulla & Tucker-Drob, 2011), and previous research has found that it is typically estimated with very low power (Hertzog et al., 2006). Thus, it is important to study the effects of planned missingness designs on this parameter in particular to discover under what conditions (e.g., with what size sample) it can be precisely estimated.

One thousand data sets were generated in each condition. Data were simulated in R (R Core Team, 2013) using the simsem package (Pornprasertmanit, Miller, & Schoemann, 2013) and then analysed using Mplus 7.0 (Muthén & Muthén, 1998–2011).

Missing data

Missingness was imposed according to three different planned missing designs (see Figure 2). In the *three-form design* (e.g.,
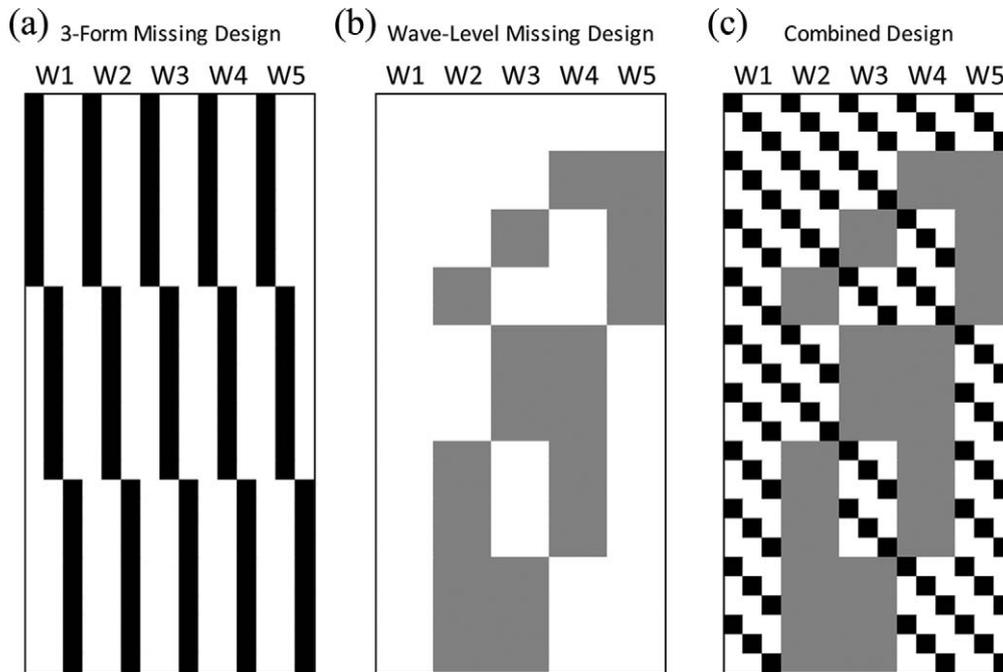
Figure 2. Planned missing data patterns. Each rectangle represents a data matrix where white refers to complete data, black to three-form missingness, and grey to wave-level missingness. In (a) the three-form design, every participant is missing one-third of variables at each wave. In (b) the wave-level missing design, every participant has complete data at the first wave and 90% of participants are missing two subsequent waves, such that 50% of observations are missing in waves 2–4, and 30% of observations are missing at wave 5. In (c) the combined design, the two forms of missingness are overlaid.

Graham et al., 2006), each indicator of each construct at each occasion was designated to come from a different item set, and therefore to have a different pattern of missingness. The three indicators of each construct within each time point were assigned to the three item sets (A to C), so were missing a non-overlapping one-third of data. This missingness was imposed by simply having the first third of the simulated participants be missing A items, the second third missing B items, and the last third missing C items. The common X set (i.e., items given to all participants) was not used in this simulation for simplicity.

Second, we imposed the *wave-missingness design* shown in the middle panel of Figure 2. This planned missing design was proposed by Graham et al. (2001), who found that it led to the greatest efficiency in estimating linear slopes, primarily because complete data are concentrated in the first and final waves, with a large amount of missing data in the middle waves. Because our model does not fix the growth coefficients to be linear, it is possible that Graham et al.'s pattern is not the ideal design for our model. However, there are still two practical reasons that this design might be preferred. First, it includes complete data at the initial wave, which allows researchers to ascertain complete demographic data from the same time point from every participant, which is desirable for many reasons (e.g., providing descriptive statistics and testing factorial invariance based on a demographic variable such as sex are much easier with complete data; and complete demographic data may provide important information about subsequent unplanned missing data). Second, this design imposes fewer missing data at the final wave (wave 5), which may counteract some of the effects of attrition. That is, attrition over the course of five waves may contribute to missing data over and above the planned missing data design, so it may be prudent to introduce less planned missingness at the last time point. This wave-missing design has 36% planned missing data.

Finally, the *combined design* superimposed the two types of missingness. The wave missingness design was implemented and, on those occasions where the entire wave was not missing, a three-form design was employed. The combined design led to a total of 57% missing data. Missingness was imposed using the sim-sem package.

Analysis

Simulated datasets were analysed using the model in Figure 1 in Mplus 7.0. The latent basis coefficients (i.e., slope loadings), means and variances of slopes and intercepts, and covariances among slopes and intercepts were freely estimated. The effects coding method was used for model identification (Little, Slegers, & Card, 2006), which means that the set of loadings for each first-order latent variable was constrained to average 1.0 and the set of observed variable intercepts averaged zero. Little et al. argued that this method of identification results in factor means and variances that are more interpretable, because they reflect the average across indicators, rather than an arbitrarily chosen single indicator (as in the marker variable identification approach).

Strong measurement invariance over time was imposed for each construct. The loadings and residual variances of the ten first-order latent variables were then estimated with these constraints. Residual variances of the 30 observed variables were freely estimated, and the pattern of residual covariance shown in Figure 1 was also freely estimated. All covariances not shown in Figure 1 were constrained to zero. Intercepts of the 10 latent factors were fixed to zero, and the intercepts of the observed variables were freely estimated (df = 362).

Missing data were dealt with using full-information maximum likelihood (FIML). FIML is a model-based approach that deals with

missingness in the same step that model parameters are estimated. Using FIML therefore necessitates using item-level data in the model (unlike multiple imputation, where items can be summed to produce scale scores after imputation and before analysis). However, it is not necessary that every item be included as a separate variable in the model; items in the same set will have the same pattern of missing data and these can be summed or averaged to form an item parcel. Using this trick, a long scale that is divided into item sets for a planned missing design can be parceled into three or four variables (one for each item set; see Little, Rhemtulla, Gibson, & Schoemann, 2013, for a justification of parcel use).

Outcome measures

For every condition we examined five outcomes: convergence rate, parameter bias, standard error bias, relative efficiency, and power. We describe each outcome below.

Convergence. Convergence is the proportion of generated datasets that result in a proper solution. When FIML in Mplus failed to converge upon a stable solution or when a solution produced an impossible value (e.g., negative variance estimates), we coded it as a convergence failure. In addition, some replications converged on solutions that were technically admissible but clearly uninterpretable, such as standard error estimates in the hundreds. Such results are also estimation failures because they are sufficiently extreme that results are untrustworthy.

We removed outliers using the following procedure. First, any replication that produced a standard error estimate (on any parameter) greater than 10 was flagged as an outlier and removed. This initial step allowed us to remove the most extreme outliers that could strongly affect our estimates of bias. This first step removed between 0 and 63 replications out of 1000 across conditions (mean = 11.0). Second, we computed the mean and standard deviation of each parameter estimate and each standard error estimate across replications within each condition, and used these values to construct an outlier cut-off that was 10 SD away from the mean. Any replication that had an estimate or SE outside this boundary was removed. This step removed between one and 40 replications per condition (mean = 14.6). On average, this procedure removed 2.6% of replications per condition. In the end, removing outliers had a barely perceptible effect on any of our results, except that some of the observed bias was more extreme with outliers left in.

Smaller sample sizes, more missing data, and more complicated models (e.g., those with more constraints) tend to produce higher rates of nonconvergence. We predicted that smaller samples and greater missing data rates would result in poor convergence, as the curve-of-factors model is a quite complex model.

Parameter bias. Parameter bias is defined as

$$ PRB = \frac{(\bar{\hat{\theta}} - \theta)}{\theta} $$

where θ is the population parameter value (e.g., a slope variance) used in data generation and $\bar{\hat{\theta}}$ is the average estimated parameter value across all converged replications for a given condition (Collins, Shafer, & Kam, 2001; Graham, 2009). Positive bias reflects estimates that are higher than expected, and negative bias reflects estimates that are lower than expected. Absolute values of parameter bias less than .05 are considered negligible

(Hoogland & Boomsma, 1998). We did not predict substantial bias in any condition, because our estimating model was the true population model, and the missing data were randomly distributed so that they should not introduce any systematic bias.

Standard error bias. Standard error bias (SEB) is the degree to which standard errors accurately reflect the standard deviation of parameter estimates

$$ SEB = \frac{(\overline{\widehat{SE}} - ESE)}{ESE} $$

where ESE is the empirical standard error (i.e., the standard deviation of parameter estimates)

$$ ESE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\hat{\theta}_i - \theta \right)^2} $$

and n is the number of converged replications. The standard error bias is considered acceptable if its absolute value is less than .10 (Hoogland & Boomsma, 1998).

Relative efficiency. Relative efficiency (RE) reflects the amount of information loss that occurs as a result of missing data. It is computed as a ratio of sampling variances (i.e., squared standard errors) of the complete data estimates to the missing data estimates

$$ RE = \frac{ESE_{\theta,\ complete}^2}{ESE_{\theta,\ incomplete}^2} $$

where $ESE_{\theta,\ complete}^2$ is the squared empirical standard error of the parameter with complete data, and $ESE_{\theta,\ incomplete}^2$ is the squared empirical standard error of the parameter from the planned missing design. RE ranges from 0 to 1, where higher values indicate a more efficient estimate and, like any proportion, it can also be interpreted as a percentage. RE can be used to compute 'effective sample size' by multiplying by the original sample size. For example, if a parameter has RE of .6, and sample size is 100, the estimate has the same power that it would have had with no missing data and a sample size of $100 \times .6 = 60$. By the same logic, to have the same power as the complete data design, we would need to increase the sample size to $n^* = 100/.6 = 167$ (Savalei & Rhemtulla, 2012). Because RE is a ratio of sampling variances, it itself is invariant to sample size.

It may be relevant to compare the efficiency of parameter estimates in the PM design to the efficiency of a complete data design with a reduced sample (Graham et al., 2001). For example, a PM design with 36% missing data can be compared to a complete data design with 36% fewer participants. This type of comparison allows a direct comparison of efficiency *per observation*. A value of efficiency relative to a reduced sample design ('adjusted relative efficiency') can be obtained by taking a ratio of RE to the percentage of complete data; for example, if RE = .80 with 36% missing data, then adjusted relative efficiency is $RE_{adj} = .8/(1 - .36) = 1.25$. In this case, we could conclude that the PM estimate is 25% more efficient per observation than a complete data design.

Power. Although RE reflects efficiency loss due to missing data, it is not directly related to power. If an effect size is large enough, it is possible that the power to detect an effect with a cut-off of $\alpha = .05$ will be very high, even with a large amount of missing data. In that case, it might not matter that a particular planned missing design is only half
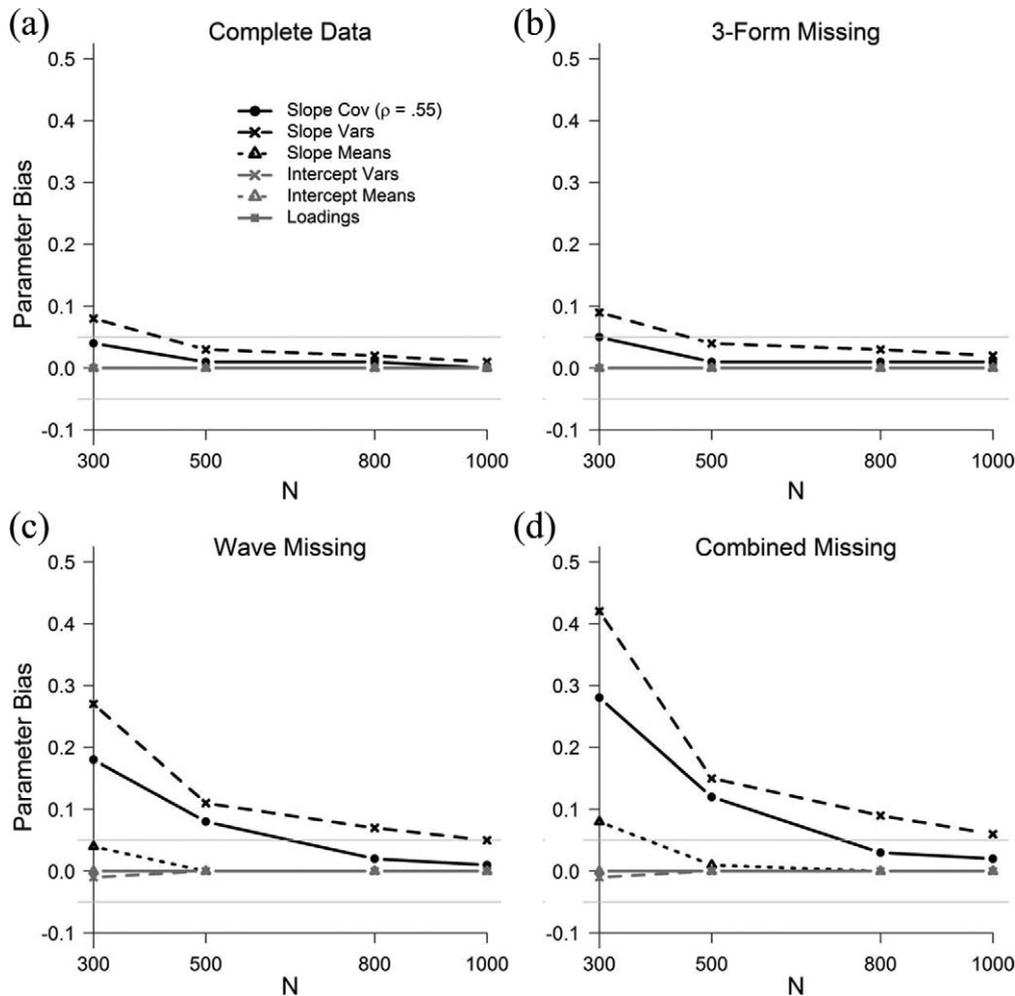
Figure 3. Parameter bias. Where lines are not visible, they are overlapping around bias = 0. (a) Complete data; (b) three-form missing; (c) wave missing; (d) combined missing. Horizontal lines at −.05 and .05 reflect bounds for acceptable bias.

as efficient as the complete data design, if it still has sufficient power. For this reason, we look at the proportion of replications in which the Wald test statistic $Z = \hat{\theta}\big/\widehat{SE_{\hat{\theta}}}$ is significant at $\alpha = .05$.

Results

Three-form missing

Convergence. When $n = 100$, convergence rates ranged from 63–68%; when $n > 100$, convergence was always higher than 95%. Recall that in the three-form missing design, one-third of data were missing. $n = 100$ is already a small sample size for a latent curve-of-factors model (Fan, 2003; Hamilton, Gagne, & Hancock, 2003; Leite, 2007), and it surprised us that even 60% of replications converged with 33% missing data. Even with complete data, convergence rates were around 72% with $n = 100$. Convergence was helped by the fact that the analysis model was exactly true in the population, whereas in real data one has some degree of misfit. We do not discuss the results for $n = 100$ further, because these results can be quite skewed when 40% of the samples have been removed, and because the low convergence rate suggests not to use

a planned missing design for correlated latent growth curve models with an n so small.

Parameter bias. We examined bias of factor loadings, means and variances of the latent intercept and slope, and covariance between the latent slopes. Figure 3 shows standardized parameter bias for complete data and each of the planned missing designs. Estimates of the slope variance when $n = 300$ had bias ranging from .08 to .11 depending on the level of the slope correlation. As Figure 3 shows, this bias was only a few percent smaller in the complete data condition; that is, three-form missingness did not have a noticeable effect on parameter bias. For all other parameters and sample sizes, bias ranged from −.01 to .06.

Standard error bias. Standard errors were estimated accurately in all conditions, as bias was always within the −.10 to .10 range (Figure 4). The worst bias observed was when $n = 300$ and the population value of the slope correlation was .10; in this condition the SE bias of the latent slope covariance was −.07. Apart from this value, SE bias ranged from −.04 to .01.

Relative efficiency. Figure 5 displays RE for six parameter types by sample size for all planned missing designs. All parameters
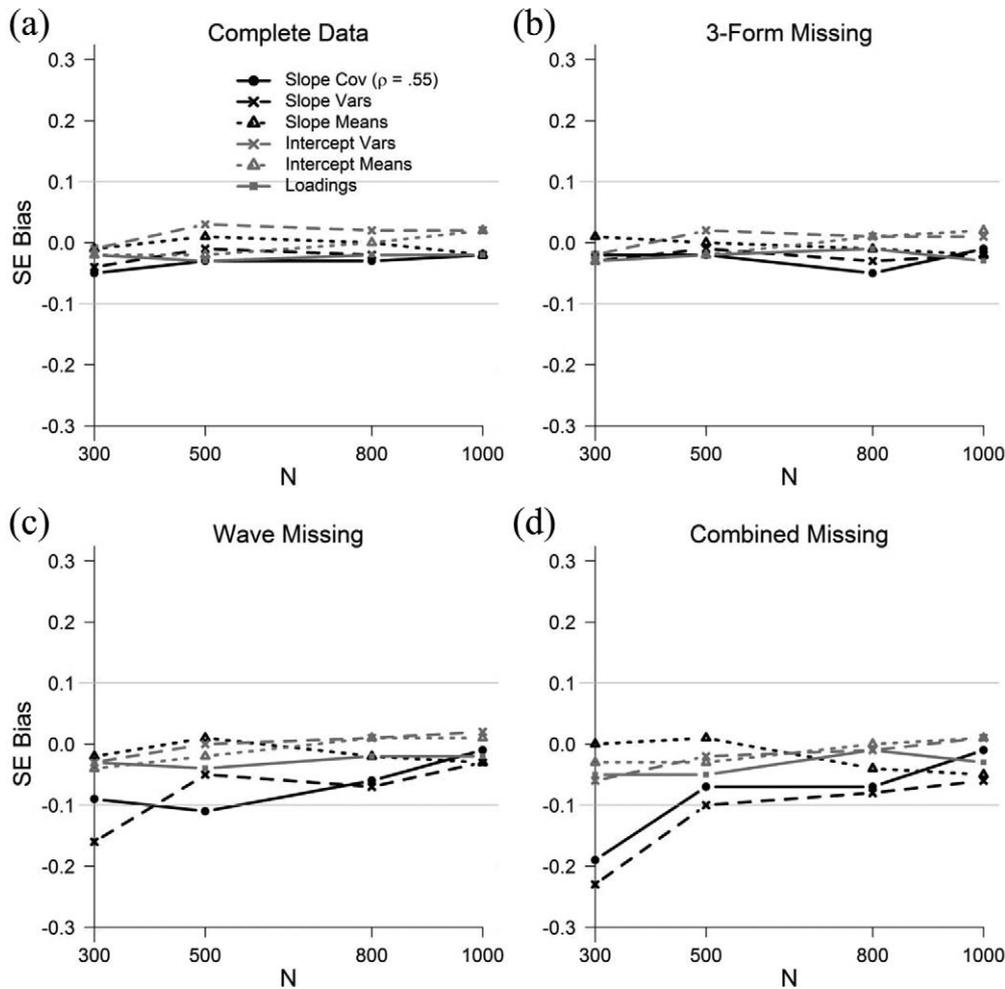
Figure 4. Standard error bias. (a) Complete data; (b) three-form missing; (c) wave missing; (d) combined missing. Horizontal lines at −.1 and .1 reflect bounds for acceptable bias.

are plotted for the condition where the population slope correlation is .55; other values of the slope correlation produced very similar results. Measurement model parameters (e.g., factor loadings) are most affected by the type of missingness imposed in the three-form design, with RE around 40%. Factor loadings are estimated based on the relations among the set of indicators. When all participants miss one indicator out of three and each pair of indicators has just one-third overlapping complete data, estimates involving the relations among these indicators become less precise.

In contrast, structural coefficients such as latent means, variances, and covariances all had RE higher than .80. These parameters are estimated based on the relations among the constructs over time. In the three-form design, every participant provides data at each time point, so estimates involving relations among constructs over time can be estimated with high precision. RE of .80 means that these parameters were less than 20% less efficient than complete data, even though 33% fewer data were collected in the three-form design. To directly compare the efficiency of the PM parameters to a complete design with reduced sample size, we can compute adjusted relative efficiency when RE = .80: $RE_{adj} = .80/(1 − .33) = 1.2$. In other words, these parameters are all at least 20% more efficient, *per piece of data*

collected, than a complete data design. The grey horizontal lines in Figure 5 reflect the RE of a reduced sample complete data design.

Power. Figure 6 displays power for each parameter type by sample size for complete data and each planned missing data design. With complete data, power to detect loadings and means and variances of the latent intercept is 100% at all sample sizes. Power to detect slope means is about 70% when $n = 300$ and reaches almost 100% when $n = 500$. Power to detect slope variances ranges from 55% when $n = 300$ to 94% when $n = 1000$. Power to detect the covariance between slopes is affected by both sample size and by the true slope correlation: when $\rho = .1$, power is virtually zero and only reaches 13% when $n = 1000$; when $\rho = .3$, power begins at 10% ($n = 100$) and reaches 83% ($n = 1000$); when $\rho = .55$, power begins at 40% ($n = 100$) and reaches 99% when $n = 800$. When three-form missingness is applied (top right panel of Figure 6), power is on average 4.5% less than with complete data (0–12% less).

Wave missingness

Convergence. Convergence rates with the wave-missing design were lower than with the three-form design. With $n = 100$,
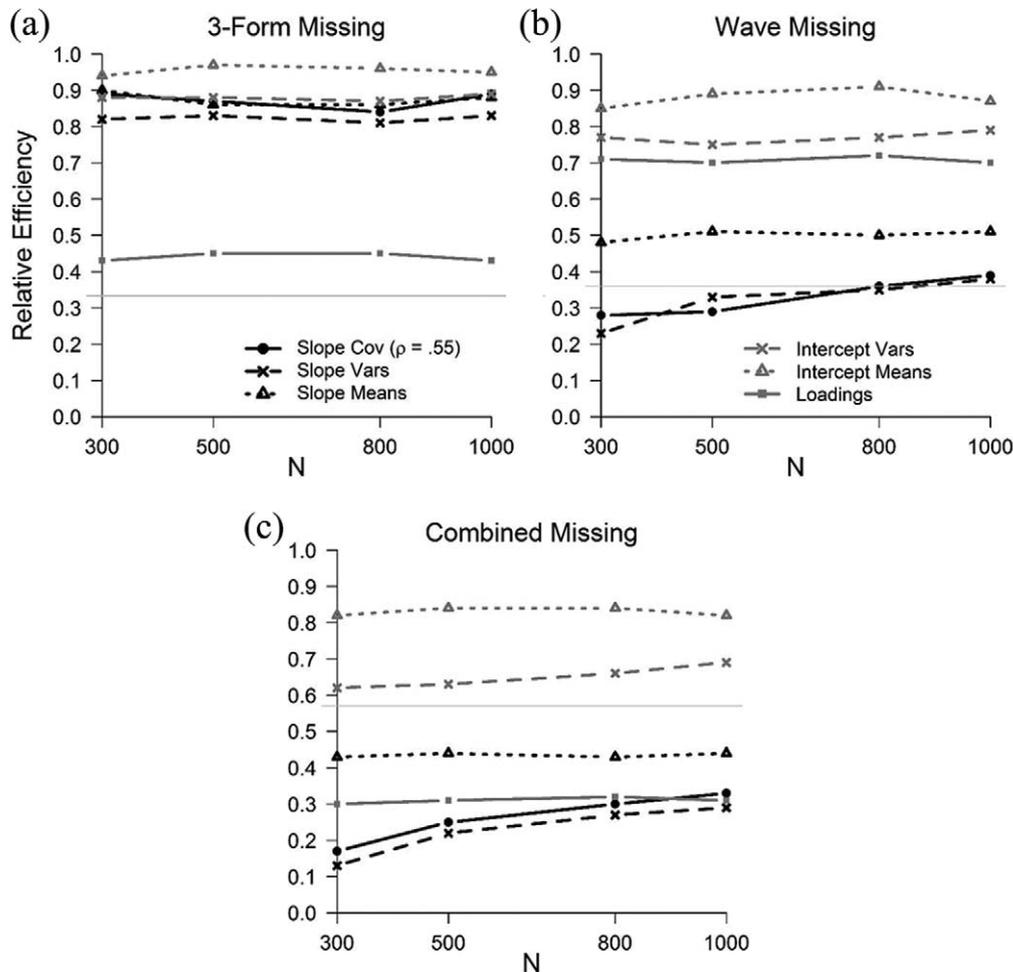
Figure 5. Relative efficiency. (a) Three-form missing; (b) wave missing; (c) combined missing. Horizontal lines at .33, .36, and .57 reflect the percentage missing values for three-form, wave, and combined missing, respectively.

convergence was less than 5%, and with $n = 300$, this rose to 80%. These convergence rates suggest that wave-level missingness is more problematic for model fitting than three-form missingness, because the two designs have roughly equal rates of missing data. The wave-missing design imposes 36% missing data, only slightly higher than the 33% imposed in the three-form design, yet convergence rates suffer much more. At $n \geq 500$, convergence rates were above 90%.

Parameter bias. Bias of factor loadings and means of latent intercepts and slopes was less than .01 at every sample size (see Figure 3). Estimates of the slope variance showed bias greater than 5% at all levels of n, and more than 10% bias when n was 500 or smaller. The degree of bias of the slope covariance increased with smaller values of the slope covariance. For example, when $n = 300$ and $\rho = .55$, slope covariance bias was .18, which rose to .54 when $\rho = .1$ (not shown). The degree of standardized bias, however, was not uniform across the two slopes, but instead it was much higher in estimates of the variance of the first latent slope (bullying) than the second slope (teasing). This difference is due to differences in the population values of the two slope variances. The population slope variance of bullying was .01, which value is in the denominator of the

standardized bias computation, so absolute bias of just .005 translates into 50% bias. In contrast, the population slope variance of teasing was .11, so an estimate would have to be off by .55 to result in 50% bias. While the standardized bias metric helps to make bias comparable across parameters, in the case of very small population values it may overstate the degree of bias.

Standard error bias. SE bias was negligible with the following exceptions: when $n \leq 500$, the SE of the slope covariance was underestimated by 9–17% and the SE of the slope variance was underestimated by 16–21%, with greater bias corresponding to smaller values of the slope covariance (Figure 4). In addition, when $n = 500$ and $\rho = .55$, SE of the slope covariance had −11% bias, and when $n = 500$ and $\rho = .55$, SE of the slope variance had −10% bias. SE bias appeared on the same parameters where parameter estimate bias appeared, that is, slope variances and covariances, particularly the variance of the bullying slope. As with parameter bias, SE bias tended to be more pronounced at small sample sizes and when the slope correlation was smaller.

Negative SE bias means that confidence intervals around parameter estimates will fail to capture the true parameter value 95% of the time. This type of bias also results in liberally biased significance tests (i.e., higher than nominal rates of type I error).
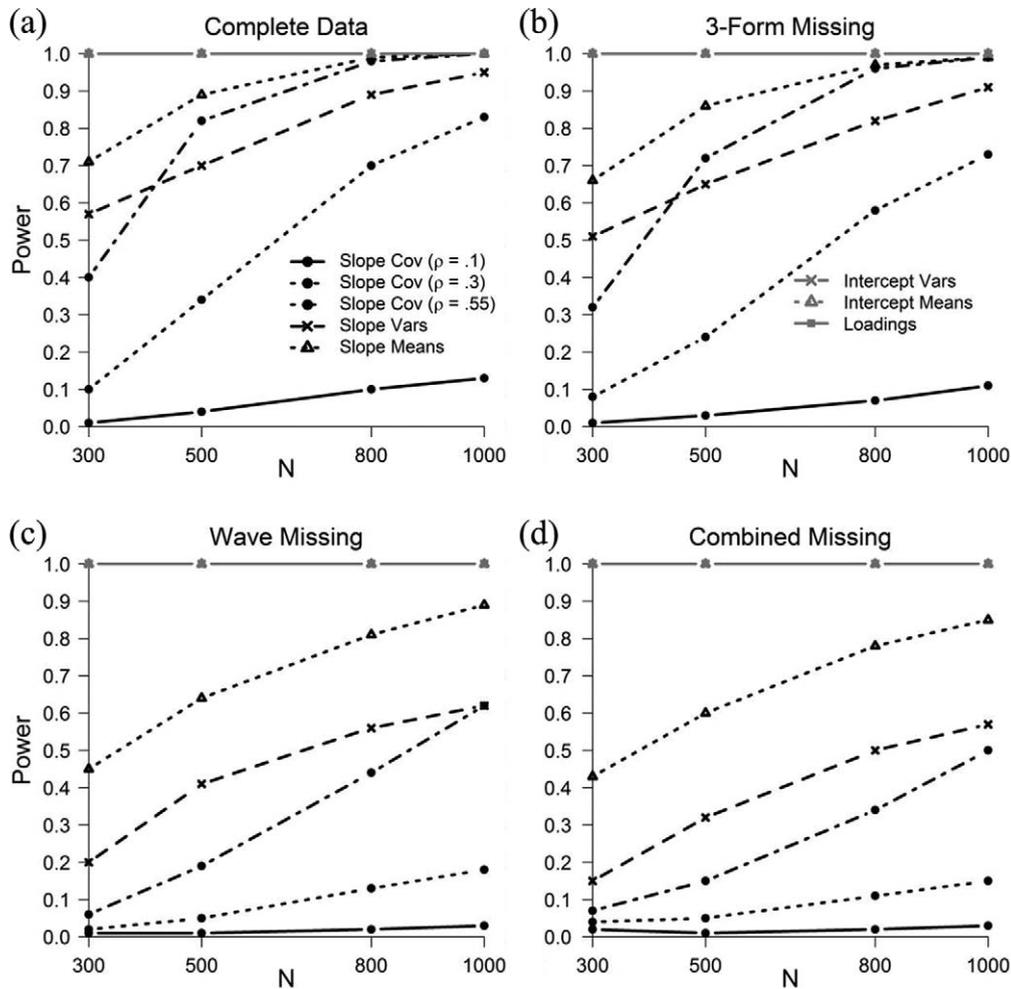
Figure 6. Power. (a) Complete data; (b) three-form missing; (c) wave missing; (d) combined missing. Lines representing intercept variances, intercept means, and loadings overlap at power = 1.0.

Underestimated SEs are particularly problematic when the parameter estimates themselves are biased, as is appearing in these results—a substantially biased parameter estimate coupled with a too-small standard error can result in confidence intervals that virtually never contain the parameter value that is being estimated. For all parameters, as sample size increased, SE bias shrunk to 0.

Relative efficiency. The pattern of RE in the wave-missing design differed substantially from the three-form design (see Figure 5). In the three-form design, measurement model parameters (e.g., factor loadings) had poor RE but structural model parameters (e.g., latent intercept and slope variances) had high RE. In the wave-missing design, missingness happens at the level of measurement occasions rather than items, so the structural model parameters that involve estimating relations across constructs over time suffer more efficiency loss.

Means and variances of the latent intercepts displayed the highest RE at around .88 and .77 respectively (RE_{adj} = 1.38 and 1.40, respectively) across all sample sizes, and factor loadings were also very efficiently estimated at RE of .71 (RE_{adj} = 1.11). These values represent a net increase in efficiency per piece of data. In contrast, RE for the variances and covariances among the latent growth parameters ranges from .14 to .39, representing a substantial loss in efficiency due to the wave-missing design (RE_{adj} = 0.22 to 0.61). These

values increase slightly with increasing *n*, which is unexpected but may reflect additional variability in parameter estimates that results from greater instability of the model at small sample sizes.

Power. Power loss compared to the three-form and complete data designs can be seen in Figure 6. As with complete data, power to detect significant factor loadings and means and variances of the latent intercepts is 100% for every value of *n*. Unlike with complete data or the three-form design, power of every other parameter type does not approach 100% even when *n* = 1000. In particular, the latent slope covariance does not achieve 50% power until its population value is .55 and *n* = 1000. In addition, latent slope variances reach about 60% power when *n* = 1000. One of the principal reasons for carrying out a latent growth curve model is to detect significant individual variability in the rate of change over time. These results suggest that imposing wave missingness may make this goal significantly more difficult to achieve.

Combined design

The combined design applied both wave-level missingness as well as item-level missingness for all non-missing time points, resulting in 57% missing data.

Convergence. The combined design showed slightly worse convergence than the wave-missing design. With $n = 100$, no models converged. With $n = 300$, convergence was 65%. It was not until $n \geq 500$ that convergence reached 90%. We present all results for $n \geq 300$ but we caution the reader that the combined model may simply have too much missing data for it to be recommended with n smaller than 500.

Parameter bias. The combined design showed the same pattern of parameter bias as the wave-missing design, only more pronounced (see Figure 3). Latent slope variances, especially that of the bullying construct, showed high positive bias at $n = 300$ and continued to be positively biased at every sample size studied. The latent slope covariance showed substantial bias at $n = 300$ (ranging from 28–85%, compared to 18–54% with only wave missingness), which decreased to 1–12% bias at $n = 500$.

Standard error bias. In the combined design, there was again substantial negative bias of the slope covariance parameter at the smallest sample size ($n = 300$), where bias ranged from –19% to –26%. In addition, the latent slope variances bias ranged from –23% to –25%. These values continue the trend seen in the wave-missing design but are more dramatic. At the next larger sample size, $n = 500$, SE bias of both these parameters dropped to within the 10% range.

Relative efficiency. The combined missingness design has missingness at both the item level (like the three-form design) and the occasion level (like the wave-missing design) and, as a result, both measurement parameters and structural parameters suffer substantial efficiency loss (see Figure 5). As noted earlier, in the combined missingness design just 43% of data points are collected (the other 57% are assigned to be missing) so RE values higher than .43 can be considered a net gain in efficiency per piece of data collected. In this design, only the means and variances of the intercepts resulted in a net efficiency gain per data point, with around .62 RE ($RE_{adj} = 1.44$). Means of the latent slopes had RE of 43% across all sample sizes, roughly equaling the number of complete data points ($RE_{adj} = 1.00$).

Factor loadings display low RE just as they did in the three-form missing design, but worse (RE is around 31%; $RE_{adj} = 0.72$), and variances and covariances of latent slopes display RE that varies from .08 to .33. As in the wave-missing design, RE is lower in smaller samples. When $n = 500$, RE is .22 for variances of latent slopes ($RE_{adj} = .51$), and .21 for the covariance between slopes ($RE_{adj} = .49$). The effect of these low RE values becomes apparent when we examine power.

Power. Power in the combined missingness design was again perfect for means and variances of the latent intercepts and for factor loadings. For other parameters, power was on average 4% lower than in the wave-missing design. The most substantial drop in power compared to wave missingness was the power to detect a slope covariance when the population correlation value was .55 and $n \geq 800$, when power was 10–12% lower. In general, though, adding item-level missingness on top of wave missingness did not much affect power. Compared to the three-form design, however, power was on average 27% lower, and as much as 62% lower for the slope covariance with larger n. As with the wave-missingness design, then, these results suggest that unless n is larger than the sample sizes examined here, the combined missingness design will

make it difficult to detect individual variability in rates of change and covariation among latent growth parameters.

Discussion

We investigated the behavior of parameter estimates in a correlated latent growth curve model using three planned missing designs: three-form missingness (missing items at each time point resulting in 33% missing data), wave missingness (missing measurement occasions resulting in 36% missing data), and both of these combined (resulting in 57% missing data). The results revealed that with sufficiently large sample sizes ($n = 300$ for three-form missingness and $n = 500$ for wave or combined missingness), convergence rates are high and most parameter estimates and standard errors show no substantial bias.

A primary concern about using planned missing designs is what effect they will have on the power to detect significant parameter values, such as individual variability in initial levels of a factor and in rates of change (i.e., latent intercept and slope variance), and covariation between rates of change across two parameters (i.e., latent slope covariance). We found that relative efficiency varied across types of parameters and planned missing designs. Structural parameters (e.g., means, variances, and covariances of latent intercepts and slopes) had high efficiency relative to a complete data design (RE) when missingness was imposed at the item level (i.e., in the three-form design), but much lower RE when it was imposed at the wave level (i.e., in the wave or combined missing designs). In contrast, measurement model parameters (e.g., factor loadings) had high relative efficiency in the wave-missing design.

The results reveal that imposing missing data at the item level using a three-form design has minimal effects on relative efficiency and power. In terms of all outcomes investigated here—convergence, parameter bias, standard error bias, relative efficiency, and power, the three-form design results looked very similar to complete data, despite containing only 66% of the data points in the complete data design. The only parameter that was substantially affected was factor loadings, which are typically not the focus of a latent growth curve model. Moreover, we have noticed that the effects coding method of identification is often less efficient than other methods of identification for evaluating the significance of item loadings and intercepts, so other identification methods (i.e., fixed loading, fixed factor variance) may have higher efficiency.

In contrast to item-level missingness, imposing missing data on entire measurement occasions can substantially diminish power for certain very important parameters. In particular, variances and covariances among slopes had dramatically diminished efficiency and very low power. This finding contrasts with Graham et al. (2001), who reported that any wave-missing design resulted in higher power per observation than complete data designs; however, their investigation focused on power to detect a regression coefficient predicting the latent slope from a fully observed grouping variable. It stands to reason that this parameter would have higher power in a planned missing design because only the dependent variable (slope variance) is affected by missingness. Based on results presented here, unless effect sizes are large (e.g., substantial slope variability is expected) or the sample size is larger than those examined in this paper (i.e., larger than 1000 participants), wave missingness is not recommended.

Limitations

The model examined here was based on one particular set of parameter estimates. Though we varied the strength of the correlation among latent slopes, we did not test the effects of varying factor loading strengths, or variances of latent intercepts and slopes. The slope variances in our model were small, and this surely affected the low power observed in all conditions. In a population with much higher rates of individual variability in rates of change, we may have concluded that even the combined missingness design could be recommended. Nonetheless, the results presented here give an idea of the scope of what RE and power associated with the two designs can look like for latent growth curve models.

We assessed power by examining the rejection rates of a Wald Z-test comparing the parameter values to zero. Hertzog et al. (2008) compared the power of several different methods for testing the significance of latent growth curve parameters and found that the most powerful method is the generalized variance test where a nested model likelihood-ratio test is used to compare the log-likelihood of the LGM to a constrained version where the latent slope variance and its covariances with other factors are constrained to zero (see also Gonzalez & Griffin, 2001). In the present model, the generalized variance test would result in a four-degree-of-freedom chi-square test, because four parameters (slope variance of one construct, and its covariance with the intercept of the same construct and the slope and intercept of the second construct) would be simultaneously constrained to zero. Hertzog et al. (2008) found that as the population value of these covariance parameters increases, the power of this generalized test can become much higher than that of the Wald test on the slope variance. We did not examine the performance of this generalized variance test in the present paper, and it is not clear to what extent it would have resulted in greater power than the Wald test.

The present model used latent basis curves, where the slopes can take on any shape. Although this model is often preferable to a linear growth model because it does not force a linear trajectory of change on phenomena that are frequently nonlinear, freeing those slope loadings may affect the power to detect individual variability in change over time. Mistler and Enders (2012) showed that a PM design with complete data at the first and last measurement occasions resulted in much greater power to detect the mean of a linear growth trend than a quadratic growth trend. This makes intuitive sense because, once the ends of a linear trajectory are anchored, data in between is not necessary; in contrast, anchoring the ends of a quadratic trajectory is not sufficient to identify the shape of a quadratic trend. Pilot work conducted during this investigation found that the relative efficiency of structural parameters in a linear latent growth curve model in the wave and combined missing designs was similar to the latent basis model used here, but it is nonetheless plausible that power could be higher when the shape of the growth trajectory is fixed to a parametric functional form.

Conclusion

These findings strongly support the use of item-level missingness (e.g., the three-form design) in longitudinal growth curve modeling. Support for wave-level or combined item-and-wave-level missingness is much more limited. However, wave level missingness has other potential benefits, which may sometimes trump the negative findings presented here. First, the cost savings of wave missingness may be much greater than those of item-level missingness; for example, the money saved by having 500 participants answer 75% of the items (using a three-form design) is unlikely to be as great as that saved by having to test only 350 participants at each wave (using a wave-missing design). But the relative efficiency results suggest that when a wave-missing design is employed, the efficiency loss outweighs the cost savings. Therefore, wave missingness employed in order to cut costs may not be a winning proposition. Second, wave-missing designs have the potential to reduce retest or practice effects that arise from repeated measurement over time. If retest effects are a concern, it is possible to minimize retest effects within the context of the three-form design by assigning different forms at each occasion (see Jorgensen et al., in press).

References

Bunting, B. P., Adamson, G., & Mulhall, P. K. (2002). A Monte Carlo examination of an MTMM model with planned incomplete data structures. *Structural Equation Modeling*, *9*, 369–389.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330–351.

Curran, P. J., Stice, E., & Chassis, L. (1997). The relation between adolescent alcohol use and peer alcohol use: A longitudinal random coefficients model. *Journal of Consulting and Clinical Psychology*, *65*, 130–140.

Espelage, D. L., Rao, M., Little, T. D., & Rose, C. A. (2011, August*). Linking bullying perpetration to homophobic name-calling during early adolescence*, Poster presented at the annual meeting of the American Psychological Association, Washington, DC.

Fan, X. (2003). Power of latent growth modeling for detecting group differences in linear growth trajectory parameters. *Structural Equation Modeling*, *10*, 380–400.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

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

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. In L. M. Collins & L. Seitz (Eds.), *Advances in data analysis for prevention intervention research* (pp. 13–63). Washington, DC: National Institute on Drug Abuse.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in the analysis of change. In L. M. Collins & A.G. Sayer (Eds.), *New methods for the analysis of change* (pp. 335−353). Washington, DC: American Psychological Association.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*, 323–343.

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods*, *6*, 258–269.

Hamilton, J., Gagne, P. E., & Hancock, G. R. (2003, April). *The effect of sample size on latent growth models*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Hansen, W. B., Graham, J. W., Wolkenstein, B. H., Lundy, B. Z., Pearson, J. L., Flay, B. R., & Johnson, C. A. (1988). Convergent and discriminant validity for assessment of skill in resisting a role play alcohol offer. *Behavioral Assessment*, *11*, 353–379.

Harel, O., Stratton, J., & Aseltine, R. (2012). Designed missingness to better estimate efficacy of behavioral studies. Technical Report (11-15), The Department of Statistics, University of Connecticut, CT.

Hertzog, C., Lindenberger, U., Ghisletta, P., & Oertzen, T. V. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *11*, 244–252.

Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling*, *15*, 541–563.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, *26*, 329–367.

Jia, F., Moore, E. W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (in press). Planned missing data design with small sample sizes: How small is too small? *International Journal of Behavioral Development*.

Jorgensen, T. D., Schoemann, A. S., McPherson, B., Rhemtulla, M., Wu, W., & Little, T. D. (in press). Assignment methods in the three-form planned missing design. *International Journal of Behavioral Development*.

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

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*. *18*, 285–300.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling, 13(1)*, 59–72.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *The handbook of multivariate experimental psychology*, *Vol. 2* (pp. 561–614). New York, NY: Plenum Press.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110–133.

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*, 115–142.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth-curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*, 126–149.

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. In J. Schinka, & W. Velicer (Eds.), *Comprehensive handbook of psychology, Vol. II: Research methods in psychology* (pp. 447–480). New York, NY: Pergamon.

McArdle, J. J., & Woodcock, R. W. (1997). Expanding test–retest designs to include developmental time-lag components. *Psychological Methods*, *2*, 403–435.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

Mistler, S. A., & Enders, C. K. (2012). Planned missing data designs for developmental research. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods*. New York, NY: Guilford Press.

Muthén, L. K., & Muthén, B. O. (1998–2013). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Pornprasertmanit, S., Miller, P., & Schoemann, A. (2013). simsem: SIMulated Structural Equation Modeling. R package version 0.5-3. Retrieved from http://CRAN.R-project.org/package=simsem

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oakes, CA: Sage Publications.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rhemtulla, M., & Savalei, V. (2012, May). Power in planned missing designs. Talk given at the Second Annual Modern Modeling Methods Conference, Storrs, CT.

Rhemtulla, M., Savalei, V., & Little, T. D. (2013, under review). On the asymptotic relative efficiency of planned missingness designs. *Psychometrika*.

Rhemtulla, M., & Tucker-Drob, E. M. (2011). Correlated longitudinal changes across linguistic, achievement, and psychomotor domains in early childhood: Evidence for a global dimension of development. *Developmental Science*, *14*, 1245–1254.

Savalei, V., & Rhemtulla, M. (2012). On obtaining estimates of the fraction of missing information from full information maximum likelihood. *Structural Equation Modeling*, *19*, 477–494.