# Oh %&$#!, they cut my funding:

# Using 'post hoc' planned missing data designs to salvage longitudinal research

Yi Feng & Gregory R. Hancock

University of Maryland



Author Note

Yi Feng*, Department of Human Development and Quantitative Methodology, University of Maryland, 1228 Benjamin Building, 3942 Campus Drive, College Park, MD 20742-1115. E-mail: yifeng94@umd.edu

Gregory R. Hancock, Department of Human Development and Quantitative Methodology, University of Maryland, 1230 Benjamin Building, 3942 Campus Drive, College Park, MD 20742-1115. Email: ghancock@umd.edu

*Corresponding author

Acknowledgments

We are most grateful for the thoughtful comments of Dr. Maria A. Gartstein of Washington State University and Dr. Natasha Cabrera of the University of Maryland in the planning and execution of this work.

Supplementary Online Resources

`simPM`'s home page with source code and tutorials: https://yifengedms.github.io/simPM/

## Abstract

Having one's funding cut in the course of conducting a longitudinal study has become an increasingly real challenge faced by developmental researchers. The main purpose of the current work is to propose 'post hoc' planned missing (PHPM) data designs as a promising solution in such difficult situations. The current study discusses general guidelines that can be followed to search for viable PHPM designs within a given budget restriction. Illustrative examples across different longitudinal research contexts are provided, each showing how PHPM data designs can help salvage longitudinal studies when an unexpected funding cut occurs mid-study. With the illustrative examples it is also shown how developmental researchers can conveniently identify viable designs using the R package `simPM`.

**Oh %&$#!, they cut my funding:**

**Using 'post hoc' planned missing data designs to salvage longitudinal research**

Researchers who are interested in studying developmental processes typically follow a sample of individuals longitudinally, measuring them repeatedly over an extended period of time. Such multi-wave panel designs are popular in psychological research as they enjoy much more analytical flexibility than their cross-sectional counterparts (Little, Card, Preacher, & McConnell, 2009; Lynn, 2009). Depending upon the research questions of interest, the resulting longitudinal or repeated measures data could be examined using, for example, autoregressive models (Bollen & Curran, 2004; Jöreskog, 1970), latent growth models (Hancock, Harring, & Lawrence, 2013; Meredith & Tisak, 1990; Preacher, Wichman, MacCallum, & Briggs, 2008), or latent difference score models (Grimm, Mazza, & Mazzocco, 2016; McArdle & Hamagami, 2001).

Multi-wave longitudinal studies, especially those extending over multiple years, are often very costly, however, and thus largely impossible without substantial support from external sources. When following children over time, for example, costs associated with gathering the necessary data can be prohibitive, not to mention the logistical challenges of securing and retaining participants throughout the study. For these reasons, a congratulatory communication from the funding agent is met by the developmental researcher, especially the relatively early-career scholar, with elation, relief, validation, and determination.

Unfortunately, not only is it exceedingly difficult to secure funding for longitudinal studies, in fact nowadays there is no guarantee that the funding is secure even after being awarded. Each year, federal agencies must submit their annual budget request to receive funding through congressional appropriations. For a variety of internal and external reasons, the allocation and appropriation of federal monies may differ from year to year. For example, since 2005 the annual budget for the National Institutes of Health's Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) has actually dropped five times relative to the prior year (in 2006, 2007, 2012, 2013, and 2014), by as much as $38,803,000 in a single year (Congressional Budget Justifications, n.d.). As a result, too common is the modern tale of the researcher who, after meticulous planning regarding a study's design, analysis, and sample size, is informed by the sponsor that changes in federal budgets necessitate that funds for the current study must be reduced – *after the study is already underway*. That is, after securing participants, getting consent forms, and starting to gather data for the early waves of the longitudinal investigation, the researcher may now be faced with an ultimatum: convince the funding agency that the study can still be conducted under the new budget constraints without compromising the inferential integrity, rigor, and reproducibility of the study or risk the remainder of the project being terminated altogether. And now, the joy experienced upon the original news of funding is instantly replaced by a sense of panic and uncertainty whose reverberations may be felt by the researcher well beyond the context of the current investigation.

Unexpected funding cuts as described above, especially those occurring after the start of a study, can put researchers in an exceedingly challenging situation. Options include eliminating non-core personnel (e.g., research assistants), reducing costs for existing personnel (e.g., limiting conference travel), and rethinking the amount of data to be gathered. Regarding the latter, researchers might consider trying to maintain the initial design — usually a complete data design — but with a smaller sample size or fewer waves of data collection. Doing so, however, will most likely risk a considerable drop in the power of key statistical tests. Alternatively, they could try to use planned missing (PM) data methods (e.g., Mistler & Enders, 2012; Rhemtulla &

Hancock, 2016) in which they strategically collect only partial data from each of the participants for the remainder of the study, possibly even enabling them to keep the original sample size as well as total number of measurement occasions. Such designs implemented by necessity after the start of a longitudinal study, referred to herein as 'post hoc' PM (PHPM) data designs, are the focus of this work.

This paper is organized as follows. We first review the definition and features of PM designs, focusing on the longitudinal context. Next, we discuss the implementation of PHPM designs mid-study, in particular how to search for optimal designs. We then introduce an R package, `simPM`, that automates a Monte Carlo simulation-based procedure to search the space of PHPM designs for possible options. Three illustrative examples are presented, demonstrating how the proposed search strategies automated by `simPM` can be applied in a range of longitudinal research scenarios. We conclude with a discussion of the limitations and promise of PHPM designs, particularly in cases of reduced research funding.

## PM Designs

As suggested by the name, in PM designs some of the data are purposefully planned to be missing. That is, the researchers themselves determine – prior to data collection –which parts of the data are to be left missing, typically by assigning participants completely at random to different strategically-chosen missing data configurations. PM designs can have some clear practical advantages over complete data designs, the most obvious being a lower financial cost of data collection. As a powerful example, in some research scenarios involving biomarkers the cost of data collection can be so prohibitive that researchers can rarely afford a sample size large enough to achieve sufficient statistical power when a complete data design is used (e.g., Luedtke, Sadikova, & Kessler, 2019). In such cases, PM designs can be advantageous as they permit researchers to collect data from a larger total sample without increasing the cost of data collection, while maintaining adequate levels of power as well. Another advantage of PM designs involves reducing participants' response burden. For some studies with long instruments or multiple repeated measures, participant fatigue can result in less trustworthy responses, higher nonresponse rates, and attrition. PM designs can therefore be a useful option for maintaining response validity as each participant is no longer required to provide complete data, and for reducing nonrandom missingness and attrition in the data that are solicited. Additionally, in studies where participants are measured repeatedly over time, practice (or carry-over) effects can potentially confound the ability to assess real change in the traits or ability levels. PM designs have been shown to perform well in terms of disentangling the effects of time and practice, as well as reducing the bias in estimating latent means over time (Jorgensen et al., 2014).

Besides the practical advantages, PM designs also have attractive statistical properties. As researchers have complete control over the missingness, the probability of missing data thus does not depend on the data or other variables, making the data missing completely at random (MCAR; for more technical details regarding missing data mechanisms, refer to Enders, 2010; Little & Rubin, 2002; Rubin, 1976). MCAR data are benign in the sense that they do not introduce bias in parameter estimates, and modern missing data techniques such as full information maximum likelihood (FIML) can be appropriately applied (e.g., Enders, 2010, 2013; Graham, Hofer, & MacKinnon, 1996; Mistler & Enders, 2012). In theory, however, missing data means reduced covariance coverage (numbers of cases for each pair of variables in order to model their covariance), and hence a loss of information that can impact statistical power and estimation efficiency (Mistler & Enders, 2012; Rhemtulla, Savalei, & Little, 2016). Therefore, for PM designs to work effectively, thereby preserving enough information to yield sufficient statistical power, careful consideration is required as to which parts of the data may be missing and how to allocate participants strategically to different missing data patterns.

Indeed with careful planning, PM designs have shown great potential in longitudinal studies across different research scenarios (e.g., Galbraith, Bowden, & Mander, 2017; Garnier-Villarreal, Rhemtulla, & Little, 2014; Graham, Taylor, Olchowski, & Cumsille, 2006; Jorgensen et al., 2014; Little & Rhemtulla, 2013; Rhemtulla & Hancock, 2016; Rhemtulla & Little, 2012). Methodological literature has suggested that certain PM designs can result in a relatively small loss of power; further, most PM designs studied have been found to outperform the corresponding complete data designs that use a reduced sample size (Graham, Taylor, & Cumsille, 2001; Mistler & Enders, 2012; Wu, Jia, Rhemtulla, & Little, 2016). For example, in the case of a linear latent growth model with five repeated measures, PM designs can yield smaller standard errors associated with the effect of a time-independent covariate on the linear latent slope factor, as compared to a complete case design with the same number of total data points (Graham et al., 2001). In a simulation study that focused on linear growth models and quadratic growth models with six repeated measures, Mistler and Enders (2012) reported the loss of power in assessing parameters associated with linear and quadratic growth factors to be relative small when the missingness was concentrated in the middle waves of data collection. Meanwhile, in the context of models with latent variables rather than measured variables as outcomes, so called second-order latent growth models (e.g., Hancock, Kuo, & Lawrence, 2001), Rhemtulla et al. (2014) showed that imposing missing data at the item level resulted in minimal loss in relative efficiency and statistical power for testing the structural model parameters as well as measurement model parameters, in comparison to wave-level missingness.

The findings from the above investigations certainly provide valuable information for researchers who wish to implement PM designs in practice, including those who plan to employ them during on-going longitudinal research when an unexpected funding cut occurs. That said, previous methodological studies only covered a few specific types of models, they focused on parameters that were somewhat arbitrarily chosen, and they compared the power and efficiency among a fairly small number of possible PM designs. As such their guidance may be limited in different research scenarios or when other types of longitudinal models are involved. Further, if we consider implementing PM designs after a funding cut occurs mid-study, obviously it will be impossible for researchers to alter anything that has already happened prior to the funding cut. For example, if the funding cut is announced after the first wave of data collection, the data points subject to missingness planning will only be limited to future waves after the first wave. Consequently, some of the suggested PM designs in previous studies may not even be feasible in this special case, thereby providing limited information toward seeking a PHPM salvage plan. Our goals for the current work include providing researchers with some useful guidelines, illustrations, and tools to assist in the search for viable and optimal PHPM designs, as detailed below.

## Searching for PHPM Designs

As discussed previously, in general it is important to find a PM design that preserves information and yields sufficient statistical power for testing focal model parameters. To achieve this goal, *a priori* power analysis (sample size planning) methods are used to make comparisons across a set of plausible PM designs. Although analytical derivations of asymptotic power based on noncentrality parameters is a possible option, Monte Carlo simulations can be used to evaluate the empirical power for PM designs in a more flexible and straightforward way, even when it involves complex models and nonnormal data (Hancock & French, 2013; Muthén & Muthén, 2002; Schoemann, Miller, Pornprasertmanit, & Wu, 2014). Therefore, we limit our discussions to Monte Carlo simulation methods in this paper. More specifically, we propose the following steps as a set of general guidelines when searching for viable and optimal PHPM designs as necessitated by, for example, a budget cut during a longitudinal study.

*Step 1*. Assess the design constraints after the funding cut.

*Step 2*. Given the design constraints, map out possible PHPM designs.

*Step 3*. Conduct a power analysis for each plausible PHPM design using Monte Carlo simulations.

*Step 4*. Evaluate the performance of plausible PHPM designs and make a selection based on predetermined criteria.

Each of the above steps is elaborated upon below in detail.

*Step 1*. Researchers must begin by carefully considering the practical and theoretical constraints on the allocation of planned missingness. Some practical constraints include, but are not limited to, the original sample size, the cost of each data point, the reduced available budget for data collection, the number of waves already completed, and the number of remaining future waves able to be manipulated in a PM design. Designs may also be constrained by substantive theoretical considerations. For example, there may be a 'gold standard' measure on which researchers wish to collect complete data from all the participants for future study, such as for age-norming purposes; therefore, such variables should be free of planned missing data. Additionally, researchers may be particularly interested in the functional form of a nonlinear growth trajectory, and thus need data at a minimum number of points to formally assess the nature of the change under study. In sum, getting all practical and theoretical elements on the table is essential for the redesign of the study using PHPM methods.

*Step 2*. Next, the space of all possible PHPM designs must be defined, filtering out those that are infeasible due to the constraints enumerated in Step 1. For longitudinal (or repeated measure) studies, there are two main governing structures for PM designs: wave-level planned missingness and item-level planned missingness (e.g., Rhemtulla et al., 2014). In wave-level PM designs, participants are randomly assigned to miss one or more entire waves of data collection. For example, in a 3-wave study with three measured variables at each wave, some participants may be assigned to provide no data on any of the measures at the second wave (see Figure 1a). On the other hand, in item-level PM designs, all participants can be measured at each wave of data collection, but each only provides data on a subset of measures. Within the 3-wave example, one possible item-level PM design would involve all subjects participating in all waves of data collection, but with each participant missing one measured variable within a specific wave (see Figure 1b). Of course more complex options are possible as well; Table S1 in the online supplementary materials shows, for example, a design with subjects strategically assigned to a total of 15 different missing data patterns. In the end, as it is unlikely that one specific type of design can always yield the best results across different research scenarios, including more candidate designs in the search space can increase the chance that a researcher will find a design with satisfactory performance. Once all possible designs are mapped out, researchers need to screen the designs based on the constraints considered at Step 1. For example, given the unit cost of each data point, the predetermined sample size (e.g., the original sample size), and the assignment of participants to missing data patterns, the cost of designs can be calculated and compared against the revised budget. Designs that cannot be afforded are excluded from further consideration.

*Step 3*. For each of the plausible designs identified in Step 2, Monte Carlo simulation methods may be used to assess their empirical power. For PM designs in general, the basic idea is to repeatedly draw a large number of random samples (e.g., 1000) from the population defined by a population model, where that model's parameter values can be reasonably justified through appeals to theory, prior literature, pilot work, or minimum effect sizes of interest. This process is no different from a power analysis process that should have preceded the original investigation, and indeed the original focal and peripheral model parameter values may still be valid. Within the simulation each random sample drawn is of the pre-specified size and, for reference, has complete data. Next, the planned missing data patterns within the PHPM design (determined at Step 2) are imposed on the complete data, yielding incomplete data with the desired patterns. Each random sample (with missingness) is then modeled using FIML, yielding significance tests for all model parameter estimates. Estimation results are collected over all (e.g., 1000) replications of the random draws; empirical power for testing each (nonzero) parameter may in turn be estimated as the proportion of replications yielding a statistically significant results at some prespecified α-level (e.g., .05). Other model fitting information may be collected as well, such as rates of model convergence. For more detail on power analysis using Monte Carlo simulation, readers are referred to Hancock and French (2013), Muthén and Muthén (2002), and Schoemann et al. (2014).

*Step 4*. Finally, the performance of all plausible PHPM designs must be evaluated based on the information obtained via Monte Carlo simulation. By comparing power for focal parameters across plausible designs, a subset of viable designs may be selected according to some predetermined criteria. For example, within the budget limit, PHPM designs that yield higher statistical power with regard to the focal parameters are preferred. And when the power levels are close across multiple designs, a PHPM design that costs less, yields a higher model convergence rate, and can be more easily managed and implemented given the practical considerations will typically be preferred.

## `simPM`: An Automated Search Tool for PM and PHPM Designs

The steps proposed above are generalizable and can be applied in a wide range of research scenarios, some of which will be illustrated with examples in the following sections. But even with the above guidelines, developmental researchers may find the work quite tedious and time-consuming to configure the possible designs, determine their eligibility, set up the simulations, and summarize the results over replications, especially when there is a large number of plausible PHPM designs to be evaluated. Wu and colleagues (2016) proposed a systematic PM search algorithm specifically for when linear and quadratic growth models are of interest, and they have automated the M*plus*-based simulation procedures in the `SEEDMC` package within the R software (R Core Team, 2018); unfortunately, it is not applicable for PHPM designs or other types of longitudinal models. Therefore, to better assist the researchers in implementing PHPM designs, an R package `simPM` (SIMulation-based power analysis for Planned Missing designs; Feng & Hancock, 2019) was developed to automate the search for optimal PM designs generally and for PHPM designs more specifically, such as in the case of an unexpected funding cut during a longitudinal study. In brief, `simPM` implements Steps 2-4 automatically through a single R function; researchers only need to supply the requested information by properly specifying the arguments in the function. The `simPM` package affords flexibility in many aspects, including the capacity to accommodate a wide range of longitudinal models involving repeated measures, as well as those with distal variables. As we will demonstrate in the following examples, `simPM` can help developmental researchers find viable and optimal PHPM designs that preserve power in longitudinal studies without needing to go through the steps proposed above manually.

A literature search within *Child Development* (up through July, 2019) revealed 268 published studies that employed latent growth models (LGMs) in their data analysis, and 87 fitting autoregressive and cross-lagged panel models to longitudinal data. Given the clear popularity of such models in developmental research, in the following section we provide an example of a linear LGM with a measured outcome, an example of a second-order linear LGM, and an example illustrating an autoregressive and cross-lagged panel model.

**Linear LGM**

Linear LGMs are very useful for examining individuals' growth trajectories in terms of both the reference level and the growth rate, and are widely used in longitudinal developmental studies. For example, Umaña-Taylor, Gonzales-Backen, and Guimond (2009) followed 323 Latino youths over 4 years, and used linear LGMs to assess yearly change in adolescents' self-esteem levels. Our first example is based on their study, with modifications made for purposes of clarity of illustration. Consider a group of researchers who are interested, like Umaña-Taylor et al., in studying the growth of youths' self-esteem. A longitudinal study has been proposed that extends over 4 years, with participants assessed on self-report self-esteem annually. The data will be analyzed using a linear LGM, as seen in Figure 2. Because the goal of the study is to examine the typical growth trajectory as well as the inter-individual trajectory variability, the model parameters of focal interest are the mean and variance of both the latent intercept and slope.

Unfortunately, after completing the first wave of data collection, the sponsor cites necessary agency-wide funding cuts and decides to implement a 30% reduction in funding for remaining waves. As a result, in order to assess whether the study can still be carried out, and in turn convince the funding agent of such, the researchers need to find a design that yields sufficient power but costs no more than the reduced budget.

As the first step to identifying viable PHPM design, the researchers need to carefully assess the design constraints (Step 1). As discussed earlier, some of the most important practical issues to consider at this step include the original sample size, the unit cost of data points, the available budget for future waves of data collection, the number of waves already completed, and the number of remaining future waves. For this proposed longitudinal study, the researchers have recruited 323 participants, a sample size that was determined based on a pre-study power analysis. There were four waves of data collection originally intended, and we will assume that the unit cost of collecting one data point at each wave is $5, $5, $10, and $15, respectively. Their initial budget for data collection was funded based on a complete data design, which costs $11,305 in total, assuming the unit cost of data points is constant across individuals within a wave. The first wave of data collection was completed at a cost of $1,615 (= $5 × 323 participants), leaving $9,690 (= $11,305 – $1,615) for the remaining three waves. Due to the mid-study funding cut of 30%, however, the available funding now decreases to $6,783 (i.e., 70% of $9,690). As a result, any viable PHPM design should cost no more than this amount. [As a side note, the amount of money budgeted for this study, while perhaps seeming small, is just used as an example (Indeed, any amount chosen as an example will, in time, seem small).]

After assessing the design constraints at Step 1, we can proceed with Step 2 to map out the space of possible PHPM designs for this longitudinal study. In this case, there is only one measured self-esteem score at each wave of data collection, thus rendering wave-level missing and item-level missing designs equivalent. For the possible designs in this scenario, we could randomly assign participants to be measured at two of the remaining three waves (i.e., missing one wave) or at only one of the remaining waves (i.e., missing two waves). We can also assign some proportion of the participants to provide complete data across the three remaining waves of data collection as originally planned. This proportion is determined by the researchers. By having a (small) proportion of the participants provide complete data across the three remaining waves of data collection, we hope to ensure adequate covariance coverage, reduce non-convergence problems, and improve estimation efficiency (Mistler & Enders, 2012; Rhemtulla et al., 2016). Once the possible missing designs drawing from combinations of these options are mapped out, we can determine the cost of each design and keep only those that are within the budget limit ($6,873). Next, we will assess the empirical power for each of the plausible designs using Monte Carlo simulations (Step 3). As with any Monte Carlo simulation procedures, we first need to define the population model and analysis model, as was seen in Figure 2. Whereas in actual practice the population values used for model parameters in a power analysis would be those from the planning of the original study, in this example the population values for the model parameters come from the estimates obtained by fitting the linear LGM to the summary statistics presented in Umaña-Taylor et al. (2009), simply for illustration purpose. The last step, then, is to compare power across the plausible designs, and to select the PHPM design that outperforms the other candidates based on pre-determined criteria (Step 4).

As mentioned earlier, these steps do not need to be done manually, as `simPM` can take care of the search process through a single function. The R code used to search for an optimal PHPM design in this illustrative example is provided in online supplementary Appendix A. As seen in the code, once the relevant information is supplied, `simPM` will automatically map out all the plausible wave-level missing designs, conduct simulation-based power analysis for each design, and select an optimal design. For this example, the optimal wave-level PHPM design is presented in Figure 3. As suggested by the missing pattern plot, about 20% of the 323 participants were assigned to provide complete data for all the following waves, while the remaining 80% were randomly assigned to one of three different missing data patterns. For participants in planned missing conditions, each would be missing two of the three remaining waves by design. This PHPM design costs $4,530, about two-thirds of the remaining $6,873 budget. Based on 1000 replications, this design yields an empirical power estimate of 1.000 for testing the mean of the latent intercept factor, the mean of the latent slope factor, and the variance of the latent intercept factor, and 0.999 for testing the variance of the latent slope factor. Additionally, the model convergence rate indicates convergence in 92% of the 1000 replications. Thus, by implementing this PHPM design, the researchers can complete the data collection portion of their longitudinal study with a lower budget, while maintaining satisfactory statistical power for testing the focal parameters. In fact, if the researchers needed to cut even more of the data collection budget, perhaps to compensate for more challenging cuts elsewhere in the study, the sample size could be lowered even further while still maintaining an acceptable level of power for all focal parameter tests (e.g., .80).

**Second-order linear LGM**

In longitudinal developmental studies, it is not uncommon for researchers to collect data from multiple sources, such as self-report measures, behavioral observations, biomarkers, and so forth. For example, the Fragile Families & Child Wellbeing Study (Goldberg & Carlson, 2014) follows a cohort of children born in large U.S. cities between 1998 and 2000, interviewing both the mother and father at each wave of data collection as well as collecting data from other sources such as teacher surveys and child saliva samples. Another example of multivariate-multi-occasion assessment is the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD-SECCYD; NICHD Early Child Care Research Network, 2001), where a cohort of children was followed across several years, with reports from mothers, fathers, and teachers obtained at each wave of data collection. In scenarios such as these, researchers may wish to examine intraindividual change using a second-order latent growth model, where the growth outcome is a multi-indicator latent construct that is theoretically free of measurement error (e.g., Hancock et al., 2001).

The example offered here draws from a study by Grimm and Ram (2009) that used the above NICHD-SECCYD data. The authors modeled longitudinal change in externalizing behavior from children in two-parent heterosexual households, where latent externalizing behavior was indicated by child behavior reports provided by the mother, the father, and the teacher. Suppose for this example that a complete-case longitudinal investigation has been funded and initiated, intending to gather problematic behavior information from all three sources for 1,135 children across four waves, each collected at grades 1, 3, 4, and 5. The researchers plan to fit a second-order linear LGM as shown in Figure 4, where the three indicators of the latent externalizing behavior construct $\eta$ across each of the four waves are the responses from the mother (*M1-M4*), the father (*F1-F4*), and the teacher (*T1-T4*), correspondingly. Because the researchers are interested in studying growth trajectories of children's problematic behaviors, the parameters of focal research interest are the mean and variance of both the latent intercept and slope. Unfortunately, however, after the first wave of data collection, the funding agency announces a 30% reduction in the remaining funding. The researchers are now asked to provide a new research plan showing how the study can be continued without compromising the scientific rigor and statistical power; if this is not possible, or if the plan is otherwise unsatisfactory, then the study will be terminated with only the first wave collected. They turn to PHPM designs to explore potential options to salvage the study.

Following the general guideline proposed previously, the researchers start by enumerating the design constraints (Step 1). In this study, the original complete-case sample size determined by *a priori* power analysis methods was $n$=1,135. The unit cost of one source's data point is $5 at the first two waves, $10 at the third wave, and $15 at the fourth wave of data collection; therefore, it will cost $15 to collect complete data (from all three sources) for one child at the first and second waves, $30 at the third wave, and $45 at the fourth wave – a total of $105 per child for complete data. The initial total budget granted for data collection is $119,175 (= $105 × 1,135 children). The researchers have received and spent the full $17,025 (= $15 × 1,135) needed to pay for the first wave of complete data collection, thus leaving $102,150 (= $119,175 – $17,025) intended for the remaining three waves. After the funding agency announces the 30% reduction in the remaining funding prior to the second wave of data collection, the available budget comes down to $71,505 (= 70% of $102,150) for the future waves of data collection. In addition to the financial constraint, theoretical considerations may play a limiting role too. As an example, imagine that the researchers need to collect complete data for mothers' report at the third wave for a side study focusing on a specific developmental juncture. Now, having articulated the above constraints, the researchers can continue with Step 2 to map out all possible wave-level as well as item-level PHPM designs, compute the cost of each design, and keep the plausible ones given the constraints. They can then proceed with Step 3 to run Monte Carlo simulations for each plausible design to obtain the empirical power. For this example, the population model is presented in Figure 4, with simulation parameter values set to the estimates obtained by fitting the model to the summary statistics reported by Grimm and Ram (2009). To properly fit the second-order linear LGM, the intercepts of the first-order latent factors were constrained to zero, the mother's report was used as the scale indicator across time with its intercepts also constrained to zero, the other measured indicators' loadings and intercepts were estimated and constrained equal over time, and corresponding variable residuals were covaried across time (see Hancock et al., 2001). Step 4 involves comparisons among plausible designs, with an optimal design selected.

For this illustrative example, we again chose to automate Steps 2 through 4 using `simPM`, using the R code provided in online supplementary Appendix B. We started with the search for the optimal wave-level missing design. As the reader may have already noticed, however, one obvious drawback of wave-level PHPM designs is that the choices of plausible designs are usually slim, especially when the number of waves is limited. In this case, to stay within the revised budget the researchers could have 10% of the participants provide complete data and the remaining 90% randomly assigned to miss two of the three remaining waves (Figure 5a). The empirical power estimated for this design over 1000 replications, as shown in Table 1, is 1.000 for testing the means of the latent intercept and slope factors and the latent intercepts variance. However, it is obviously underpowered with regard to the test of the latent slope variance (0.421).

Considering that the results for the wave-level PHPM designs are not entirely satisfactory, we continued the search with item-level PHPM designs. By default, `simPM` searches for optimal item-level missing designs that have equal number of missing data points across the patterns (e.g., Figure 1b and Table S1). The optimal item-level PHPM design selected by `simPM` is presented in Figure 5b, where 90% of the participants are randomly assigned to one of the 126 missing data patterns and each is missing four measured indicators. The empirical power yielded by this design is summarized in Table 1, and has much better performance than the wave-level missing design. However, this item-level missing design is likely infeasible from an administrative standpoint, given that there are fewer than 10 participants being assigned to each of the 126 missing data patterns.

Given the practical concerns, we therefore introduce and automate another search strategy for item-level PHPM designs in `simPM`—*forward selection*. In forward selection, the missing data patterns are assembled sequentially in the search for viable and optimal designs. Comparisons are first made among missing data patterns each with only one missing indicator. More specifically, for each pattern under comparison, the parameter with the lowest power level will be identified (e.g., ${\sigma_\beta}^2$) and its empirical power will be recorded; the data patterns are then rank-ordered based on this weakest power level, selecting the pattern (or design) ranked the highest power. Next, designs with two unique missing data patterns are compared, the first being the optimal one-indicator-missing pattern already identified and the second being the missing data pattern with any two missing indicators. Again, the design ranked the highest with regard to the weakest power is selected. Next, we would make comparisons and select the optimal design with three unique missing data patterns, where the third pattern contains three missing indicators. This procedure continues until it reaches the missing data pattern with the maximum possible number of missing indicators, or until it includes the desired number of missing data patterns. In this way, the research has full control over the number of unique missing data patterns in the PHPM design, helping to ensure the revised study's manageability. This procedure will still result in an item-level missing design, but it will have different numbers of missing data points in each pattern within the design.

This forward selection procedure is fully automated in `simPM`, using the R code provided in online supplementary Appendix B. For illustration purposes, in this case we specified a maximum number of five missing data patterns, and that we wish to collect complete data for mothers' report at the third wave. Given these constraints, the program automatically ran simulation-based power analyses for 218 designs; the optimal PHPM design returned by forward selection is shown in Figure 5c, where 95% of the participants are randomly assigned to one of the five missing data patterns (and the remaining 5% providing complete data across all remaining waves). The empirical power for testing the focal parameters using this design is shown in Table 1. Of the designs considered, this is the overall best with regard to power, and can be implemented at a cost of 97.6% of the remaining reduced budget (i.e., spending $69,802 of the available $71,505). This would therefore be a reasonable PHPM design choice to propose for moving forward.

**Autoregressive and cross-lagged model**

Autoregressive and cross-lagged models constitute another common analytical approach in developmental studies, as they are useful for testing reciprocal relations and mediation effects that feature prominently in many developmental theories (Selig & Little, 2012). As an example, De Laet et al. (2014) measured 586 children annually for three years, using an autoregressive and cross-lagged model strategy to examine the longitudinal reciprocal relations among teacher-child support, sociometric popularity, and perceived popularity throughout late childhood. Inspired by this study, suppose as an example that some developmental researchers are interested in studying the connections between peer relationships and teacher-child relationships. They have been funded for a longitudinal panel study following 1000 children for three years, annually measuring two measures of peer relationships (self-perceived popularity and peer-perceived popularity) and one measure of teacher-child relationships (teacher-child support), with the intention of fitting an autoregressive cross-lagged relating all three measures as seen in Figure 6. In this study, assume a primary interest in the cross-lagged path coefficients predicting the peer-perceived popularity from the self-perceived popularity measured at the previous time point ($b_{PS1}$ and $b_{PS2}$), as well as in the cross-lagged paths predicting teacher-child support from the peer-perceived popularity measured at the previous time point ($b_{TP1}$ and $b_{TP2}$). Unfortunately, as with the two previous illustrative examples, before the second wave of data collection begins the funding agency decides to cut their remaining funds by 30%. In order to come up with a compelling revised plan, the researchers turn to PHPM designs.

Starting with Step 1's enumeration of design constraints, it is known that the unit cost of a single measure is \$5, \$5, and \$10 at each of the three waves of data collection, respectively, making the original complete-case design budget for gathering three measures at each of three waves equal to \$60,000. In the first wave they have utilized the full amount of \$15,000 (= \$5 × 3 measures × 1000 participants), leaving \$45,000 for the remaining two waves. As a result of the funding cut, to continue with their longitudinal study and fulfill their research goals the researchers need to implement a new design that yields sufficient power without exceeding the new budget of \$31,500 (= 70% of \$45,000). Step 2 through Step 4 can be manually accomplished following the guideline detailed above. As always, it is required that researchers specify the population model and analysis model prior to Monte Carlo simulations. For this example, the population model is the autoregressive and cross-lagged model shown in Figure 6, with population values for the model parameters specified (for illustrative purposes) as the parameter estimates after fitting this model to the sample summary statistics presented in De Laet et al. (2014).

As in previous examples, we used the R package `simPM` to search for viable designs; our code appears in online supplementary Appendix C. For this illustrative example, as in the linear LGM example, 10% of the participants were assigned to provide complete data in the remaining two waves of data collection in order to help ensure the covariance coverage, reduce the non-convergence problems, and improve the estimation efficiency. The optimal plausible wave-level PHPM design (which is also the only possible wave-level missing design in this case) is shown in Figure 7a, with 90% of the participants assigned to missing one of the remaining two waves of data collection. As shown in Table 2, this wave-level PM design has unacceptable power for detecting the majority of the focal parameters.

Without getting a satisfactory result from the wave-level missing designs, we continued to search for an optimal item-level PHPM design using `simPM`. By running the R code provided in online supplementary Appendix C, `simPM` proceeded (by default) to search for an item-level PHPM design that has an equal number of missing data points across the patterns: it automatically maps out five possible item-level PM designs (i.e., within each design, all the missing data patterns contain an equal number of missing data points $k$; $k = 1,\ldots, 5$), implements Monte Carlo simulations, and then compares the power across the four plausible designs that are within the budget limit. The optimal item-level PHPM design in this case is presented in Figure 7b, in which 90% of the participants are randomly assigned to one of 15 missing data patterns, each missing two measured indicators, and with the remaining 10% providing complete data. This design costs precisely 100% of the remaining available budget. Over the 1000 replications, as seen in Table 2 this design yields better power for testing the focal parameters compared to the wave-level PM design, although the test for coefficient $\boldsymbol{b}_{PS1}$ is still underpowered (0.678). Therefore, we continued to search for another possible item-level PHPM design using forward selection in `simPM`, which would yield item-level missing designs with unequal numbers of missing data points across the patterns. Another feature of forward selection is that we can specify how many missing data patterns we wish to deal with. For this illustrative example, suppose we specified a total of three missing data patterns that can be implemented fairly easy. The optimal item-level PHPM chosen by forward selection is presented in Figure 7c. For this design, 10% of the participants are assigned to provide complete data and the remaining 90% are randomly assigned to one of the three missing data patterns. The cost of this design is $31,500, 100% of the remaining available budget. Over the 1000 replications, this design is shown to yield satisfactory empirical power for testing all the focal parameters (Table 2), with the previously concerning parameter now having a test whose power is an acceptable 0.819.

## Discussion

Honestly, we wish this paper weren't necessary. We wish that we lived in a world where well-designed longitudinal studies were always funded precisely as budgeted. Shifting fiscal climates, however, have evolved a different reality, one fraught with many challenges as we have discussed. Such mid-study changes to external support also precipitate challenges of researcher conduct. For example, should researchers plan their studies, with or without *a priori* planned missing data, aiming at 'just the right amount' of power for all focal parameters concerned? Doing so would seem to be honorably frugal, as we were trained to be in our design of studies, and yet is vulnerable to being unable to absorb potential changes in the funding stream. On the other hand, should researchers request money for higher sample sizes than statistical power would technically warrant, padding designs as an insurance policy in the event of a mid-study funding decrease but risking not being funded in the first place due to the study's increased cost? This is the modern developmental researcher's conundrum.

The current work does not, indeed cannot, resolve the conundrum. We have tried, however, to help the researchers who find themselves in this situation, by giving them guidelines and tools to help, so to speak, throw out a bit of the bathwater while still keeping the baby. Through the implementation of PHPM designs, as we have illustrated with three software-aided examples as templates for the applied researcher, longitudinal research might be able to be salvaged when an unexpected funding cut occurs. In fact, designs of this type not only help to address financial issues, but also carry collateral benefits as well, including mitigating response burden on participants, reducing potential practice effects over repeated measures, and lessening the likelihood of attrition, while ensuring that data are gathered in a strategic way as to inform the model parameters most critical to the developmental researcher. Also, importantly, PHPM designs can be handled with standard structural equation modeling software (e.g., M*plus*, `lavaan`), and even made considerably easier with the `simPM` R software package utilized here.

Besides minimizing the tedium of enumerating plausible PM designs, running simulations, and comparing results across models, the `simPM` tool has additional unique advantages. To begin with, it can be used in a wide range of research scenarios involving repeated measures, and is not limited to a specific set of models. In this paper, we employed `simPM` with three longitudinal models commonly used in developmental research, but it can certainly accommodate many other models as well. Additionally, it offers great flexibility in specifying the features of PM designs. For example, researchers can specify on which measured variable(s) they want to collect complete data, and all candidate PM or PHPM designs will be generated accordingly. The tool also allows the unit cost of data points to differ across measures and across time points, which is important as cost can vary greatly for data to be collected through a survey, an interview, a lab observation, or a biomarker, as well as at different ages.

Of course, PHPM designs have limitations, as does any *a priori* power analysis procedure. First and foremost, as a speculative procedure it is only as good as the information upon which the speculation is based. That is, if the population parameters that seed the process are founded on faulty theory or prior studies that are not entirely on-point, the power estimates and in turn the selection of viable and optimal designs can be compromised. Additionally, in order to evaluate and compare the expense of each PHPM design, we require information on the unit cost of data points. Breaking the cost of data collection down to the level of measured variables, however, is not always so straightforward. If, for example, the researchers believe the unit cost of data collection can only be defined at the wave level, then wave-level PM designs may be the only option as item-level PM designs might not be helpful for lowering the cost.

With regard to `simPM` specifically, it also has room to grow. One of the challenges in implementing PM designs in general is that unplanned missing data and attrition are still likely to occur, which will impact the statistical power and possibly even the parameter estimates. The `simPM` tool is not currently able to accommodate this planning; rather, one would need to conduct sensitivity analyses with a subset of viable PHPM designs within, say, M*plus* or `lavaan` directly, to gauge the overall impact of this additional layer of missingness and to aid in final PHPM design selection. Finally, because the power analysis for PHPM designs is conducted after one or more waves of data collection are completed, we may wish to use the actual data gathered thus far to refine the population parameter values that seed the power analysis. These and other exciting developments are already in the works for the next version of `simPM`. But even before that time, we believe that the methods and tools outlined in the current paper can prove enormously helpful to researchers who find themselves in the unfortunate position of having to make adjustments in order to continue their important lines of developmental research.

# References

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models a synthesis of two traditions. *Sociological Methods & Research*, *32*, 336-383. doi:10.1177/0049124103260222

Congressional Budget Justifications. (n.d.). Retrieved July 12, 2019 from https://www.nichd.nih.gov/about/budget/CJ

De Laet, S., Doumen, S., Vervoort, E., Colpin, H., Van Leeuwen, K., Goossens, L., & Verschueren, K. (2014). Transactional links between teacher-child relationship quality and perceived versus sociometric popularity: A three-wave longitudinal study. *Child Development*, *85*, 1647-1662. doi:10.1111/cdev.12216

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

Enders, C. K. (2013). Analyzing structural equation models with missing data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.) (pp. 493-519). Charlotte, NC: Information Age Publishing, Inc.

Feng, Y., & Hancock, G. R. (2019). *simPM: SIMulation-based power analysis for Planned Missing designs*. R package version 0.0.0.9000. Retrieved from https://yifengedms.github.io/simPM/

Galbraith, S., Bowden, J., & Mander, A. (2017). Accelerated longitudinal designs: An overview of modelling, power, costs and handling missing data. *Statistical Methods in Medical Research*, *26*, 374-398. doi:10.1177/0962280214547150

Garnier-Villarreal, M., Rhemtulla, M., & Little, T. D. (2014). Two-method planned missing designs for longitudinal research. *International Journal of Behavioral Development*, *38*, 411-422. doi:10.1177/0165025414542711

Goldberg, J. S., & Carlson, M. J. (2014). Parents' relationship quality and children's behavior in stable married and cohabiting families. *Journal of Marriage and Family*, *76*, 762-777. doi: 10.1111/jomf.12120

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. doi:10.1207/s15327906mbr3102_3

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, D. C.: American Psychological Association. doi:10.1037/10409-011

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. doi:10.1037/1082-989x.11.4.323

Grimm, K. J., & Ram, N. (2009). A second-order growth mixture model for developmental research. *Research in Human Development*, *6*, 121-143. doi:10.1080/15427600902911221

Hancock, G. R., & French, B. F. (2013). Power analysis in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.) (pp. 117-159). Charlotte, NC: Information Age Publishing, Inc.

Hancock, G. R., Harring, J. R., & Lawrence, F. R. (2013). Using latent growth models to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course (2nd ed.)* (pp. 309-341). Charlotte, NC: Information Age Publishing, Inc.

Hancock, G. R., Kuo, W. L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*, 470-489. doi:10.1207/s15328007sem0803_7

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, *57*, 239-251. doi:10.2307/2334833

Jorgensen, T. D., Rhemtulla, M., Schoemann, A., McPherson, B., Wu, W., & Little, T. D. (2014). Optimal assignment methods in three-form planned missing data designs for longitudinal panel studies. *International Journal of Behavioral Development*, *38*, 397-410. doi:10.1177/0165025414531094

Little, R., & Rubin, D. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley. doi:10.1002/9781119013563

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*, 199-204. doi:10.1111/cdep.12043

Little, T. D., Card, N. A., Preacher, K. J., & McConnell, E. (2009). Modeling longitudinal data from research on adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., pp. 15-54). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9780470479193.adlpsy001003

Luedtke, A., Sadikova, E., & Kessler, R. C. (2019). Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clinical Psychological Science*, *7*, 445-461. doi:10.1177/2167702618815466

Lynn, P. (Ed.). (2009). *Methodology of longitudinal surveys*. Wiley Series in Survey Methodology. Chichester, UK: John Wiley & Sons. doi:10.1002/9780470743874

McArdle, J. J., & Hamagami, F. (2001). Linear dynamic analyses of incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *Methods for the analysis of change* (pp. 139-175). Washington, DC: American Psychological Association. doi:10.1037/10409-005

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122. doi:10.1007/bf02294746

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* (pp. 742-754). New York, NY: Guilford Press.

Muthén, L. K. & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 599-620. doi:10.1207/s15328007sem0904_8

NICHD Early Child Care Research Network. (2001). Nonmaternal care and family factors in early development: An overview of the NICHD Study of Early Child Care. *Journal of Applied Developmental Psychology*, *22*, 457-492. doi:10.1016/s0193-3973(01)00092-2

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Rhemtulla, M., & Hancock, G. R. (2016). Planned missing data designs in educational psychology research. *Educational Psychologist*, *51*, 305-316. doi:10.1080/00461520.2016.1208094

Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of Cognition and Development*, *13*, 425-438. doi:10.1080/15248372.2012.717340

Rhemtulla, M., Savalei, V., & Little, T. D. (2016). On the asymptotic relative efficiency of planned missingness designs. *Psychometrika*, *81*, 60-89. doi:10.1007/s11336-014-9422-0

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581-592. doi:10.2307/2335739

Schoemann, A. M., Miller, P., Pornprasertmanit, S., & Wu, W. (2014). Using Monte Carlo simulations to determine power and sample size for planned missing designs. *International Journal of Behavioral Development*, *38*, 471-479. doi:10.1177/0165025413515169

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265-278). New York, NY, US: The Guilford Press.

Umaña-Taylor, A. J., Gonzales-Backen, M. A., & Guimond, A. B. (2009). Latino adolescents' ethnic identity: Is there a developmental progression and does growth in ethnic identity predict growth in self-esteem? *Child Development*, *80*, 391-405. doi:10.1111/j.1467-8624.2009.01267.x

Wu, W., Jia, F., Rhemtulla, M., & Little, T. D. (2016). Search for efficient complete and planned missing data designs for analysis of change. *Behavior Research Methods*, *48*, 1047-1061. doi:10.3758/s13428-015-0629-5

Table 1

*Summary of the Optimal PHPM Designs for the Second-Order Linear LGM Example*

| | Focal Parameter | Wave-level PHPM | Item-level PHPM | Item-level PHPM via Forward Selection |
|---|---|---|---|---|
| Empirical Power | $\mu_\alpha$ | 1.000 | 1.000 | 1.000 |
| | $\mu_\beta$ | 1.000 | 1.000 | 1.000 |
| | $\sigma_\alpha^2$ | 1.000 | 1.000 | 1.000 |
| | $\sigma_\beta^2$ | 0.421 | 0.790 | 0.940 |
| Cost of the Design | | $40,950 | $61,830 | $69,802 |
| Convergence Rate | | 0.702 | 0.891 | 0.930 |
| Number of Missing Data Patterns | | 3 | 126 | 5 |
| Percentage of Completers[a] | | 10% | 10% | 5% |

*Note*: [a] Completers are the participants who are assigned to provide complete data on all the measures across all the future waves of data collection.

Table 2

*Summary of the Optimal PHPM Designs for the Autoregressive and Cross-Lagged Panel Model Example*

| | Focal Parameter | Wave-level PHPM | Item-level PHPM | Item-level PHPM via Forward Selection |
|---|---|---|---|---|
| Empirical Power | $b_{PS1}$ | 0.644 | 0.678 | 0.819 |
| | $b_{PS2}$ | 0.787 | 0.865 | 0.848 |
| | $b_{TP1}$ | 0.339 | 0.806 | 0.854 |
| | $b_{TP2}$ | 0.905 | 0.936 | 0.931 |
| Cost of the Design | | $24,750 | $31,500 | $31,500 |
| Convergence Rate | | 1.000 | 1.000 | 1.000 |
| Number of Missing Data Patterns | | 2 | 15 | 3 |
| Percentage of Completers[a] | | 10% | 10% | 10% |

*Note*: [a] Completers are the participants who are assigned to provide complete data on all the measures across all the future waves of data collection.

**(a) An example of wave-level PHPM designs**

| | Wave 1 | | | Wave 2 | | | Wave 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Measure A | Measure B | Measure C | Measure A | Measure B | Measure C | Measure A | Measure B | Measure C |
| Pattern 1 | | | | | | | | | |
| Pattern 2 | | | | | | | | | |

**(b) An example of item-level PHPM designs**

| | **Wave 1** | | | **Wave 2** | | | **Wave 3** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Measure A | Measure B | Measure C | Measure A | Measure B | Measure C | Measure A | Measure B | Measure C |
| Pattern 1 | | | | | | | | | |
| Pattern 2 | | | | | | | | | |
| Pattern 3 | | | | | | | | | |
| Pattern 4 | | | | | | | | | |
| Pattern 5 | | | | | | | | | |
| Pattern 6 | | | | | | | | | |

*Figure 1*. Examples of wave-level PHPM designs and item-level PHPM designs after wave 1 data collection is completed. Grey shade indicates PHPM data.

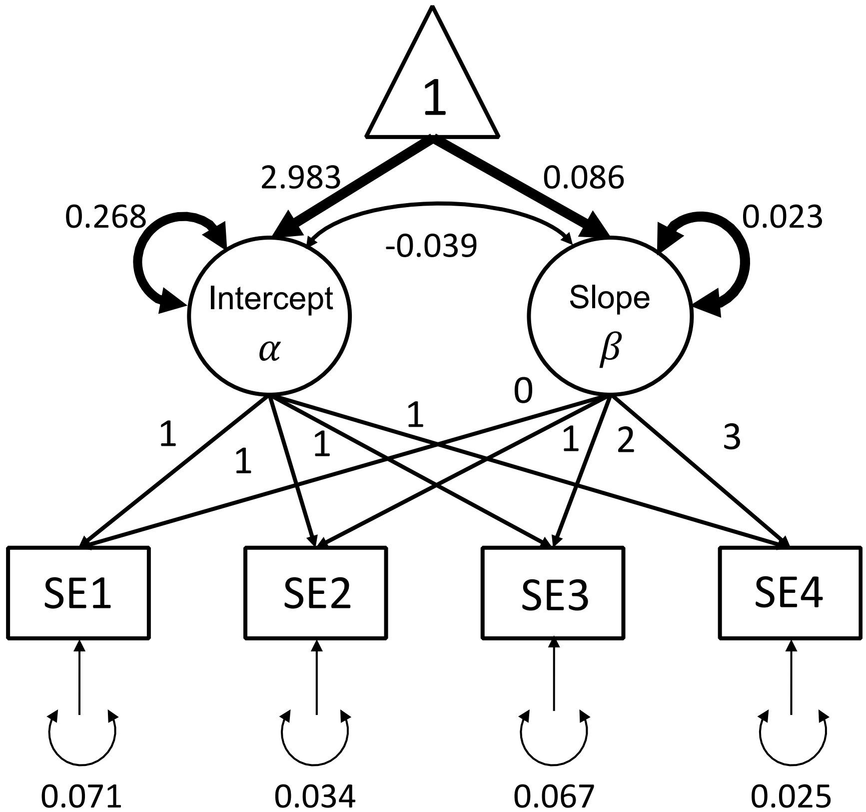


*Figure 2*. Linear LGM assessing yearly growth in youths' self-esteem. The focal path coefficients are shown in bold. Parameter values used for simulation were obtained by fitting the LGM to the summary statistics provided by Umaña-Taylor et al. (2009).

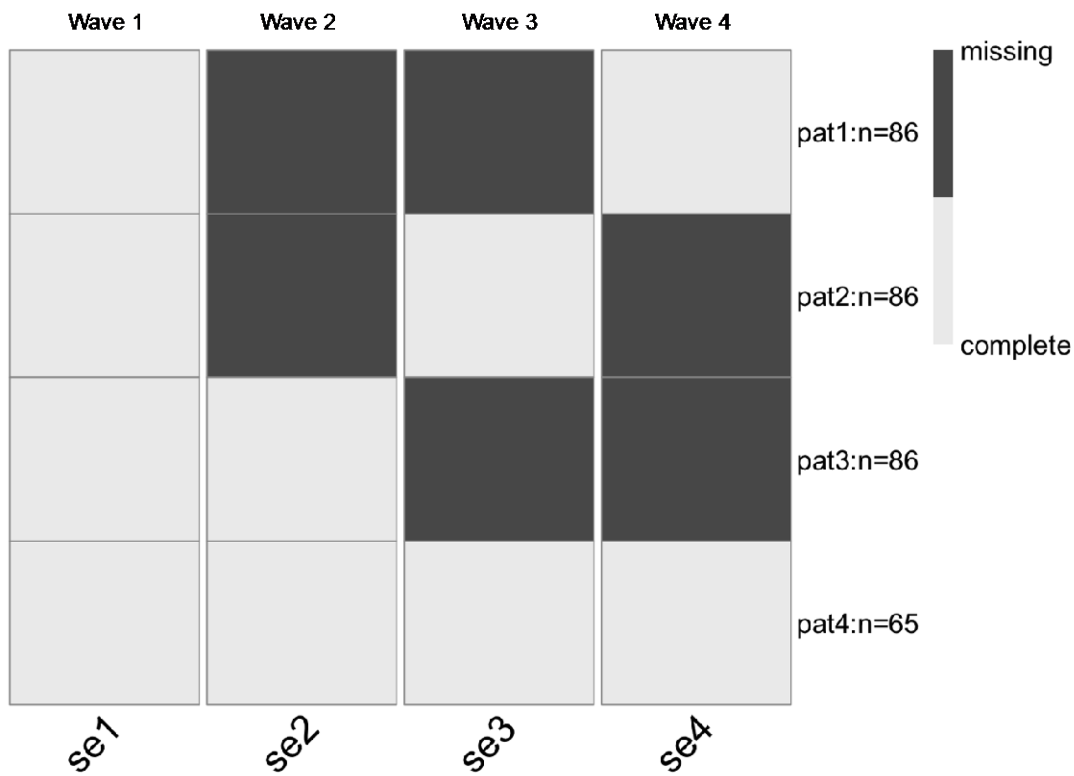


*Figure 3*. Optimal wave-level PHPM design for the linear LGM example (se = self-esteem, pat = pattern).

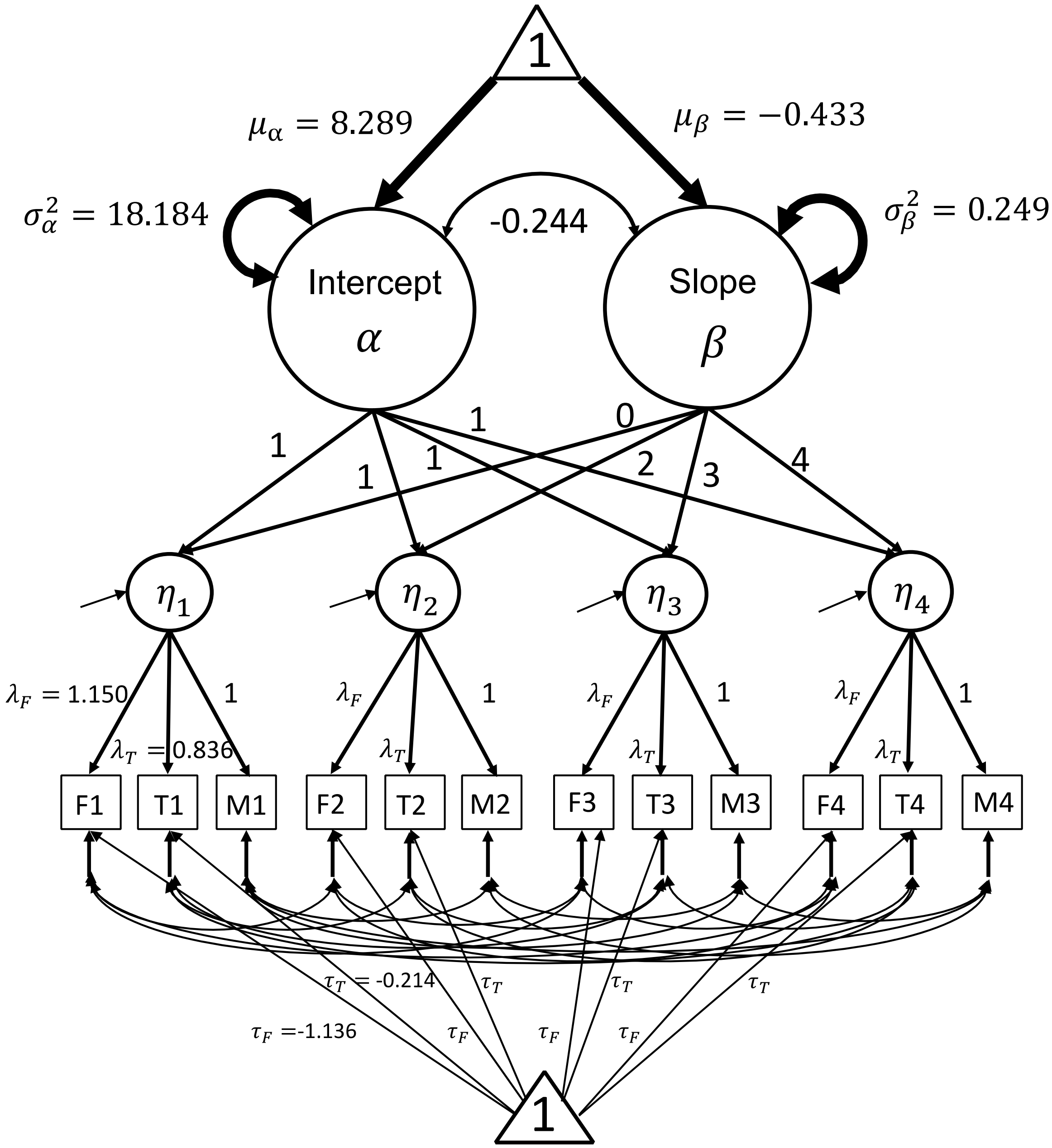


*Figure 4*. Second-order linear LGM assessing youths' problem behaviors. The focal path coefficients are shown in bold. The four waves of data were each collected at grades 1, 3, 4, and 5.

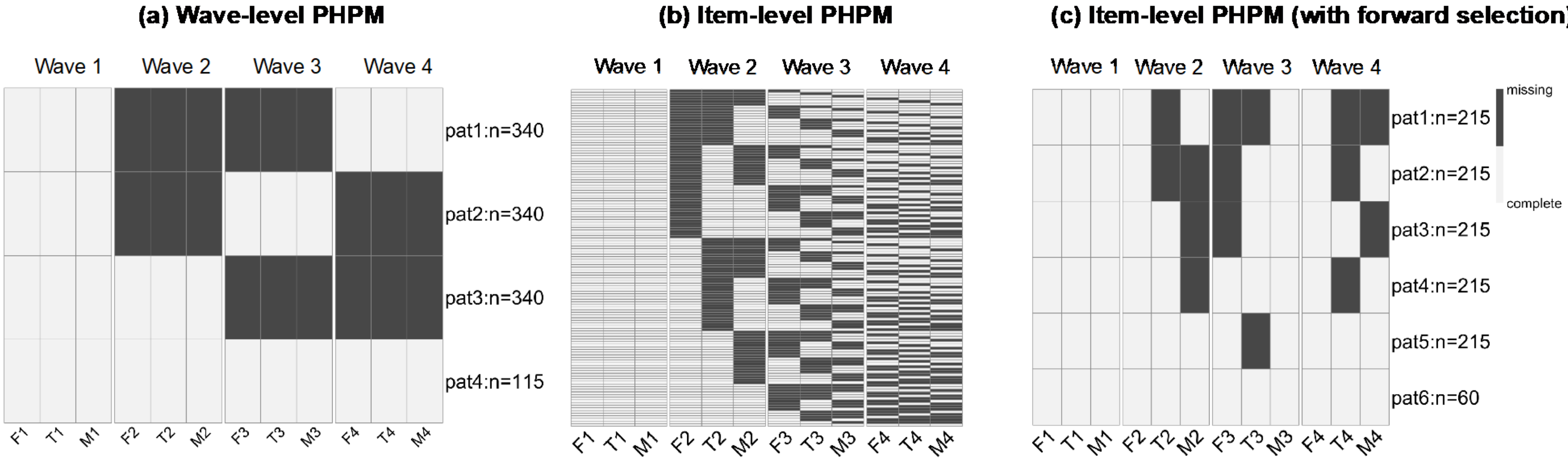


*Figure 5*. Optimal PHPM designs for the second-order linear LGM example (F = father, T = teacher, M = mother, pat = pattern). (a) Optimal wave-level PHPM designs. (b) Optimal item-level PHPM design: there are 127 participants being assigned to provide complete data across the remaining 3 waves. The rest of the participants are equally assigned to one of the 126 unique missing data patterns ($n$ = 8 for each pattern). (c) Optimal PHPM design obtained via forward selection.

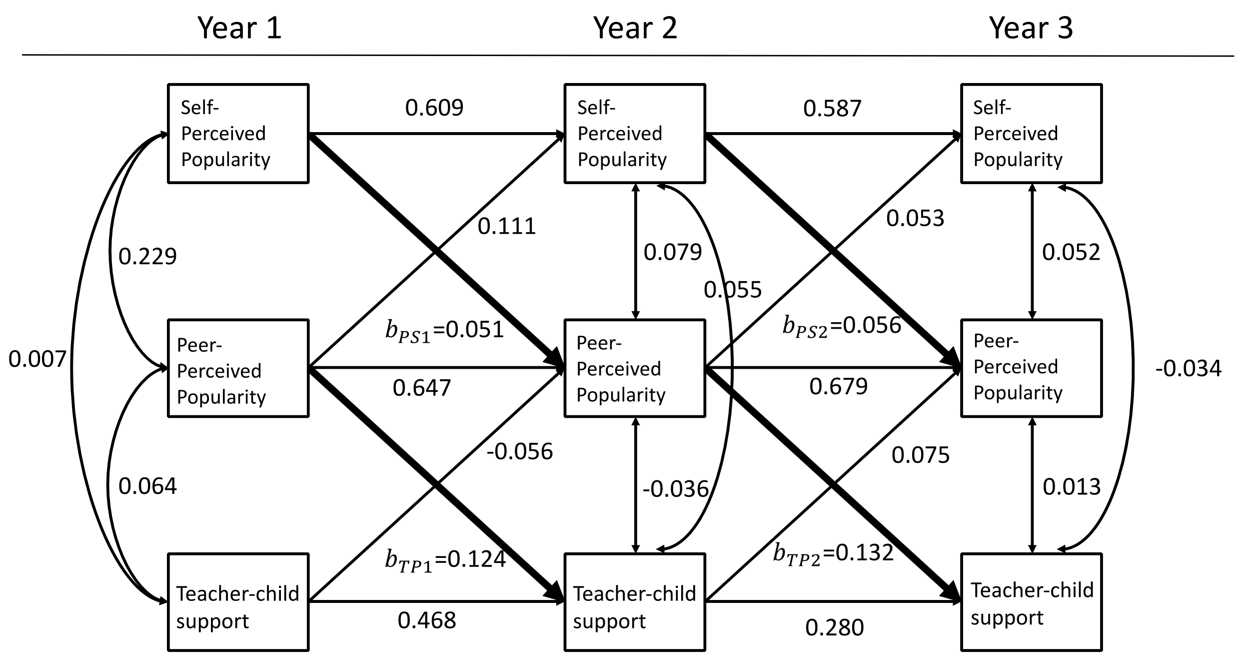


*Figure 6*. Autoregressive and cross-lagged panel model examining the longitudinal reciprocal relations between children's peer popularity and teacher-child support over three years. The focal path coefficients are shown in bold. Parameter values used for simulation were obtained by fitting the model to the summary statistics provided by De Laet et al. (2014). The residuals of the endogenous variables that were measured at the same wave were allowed to covary.

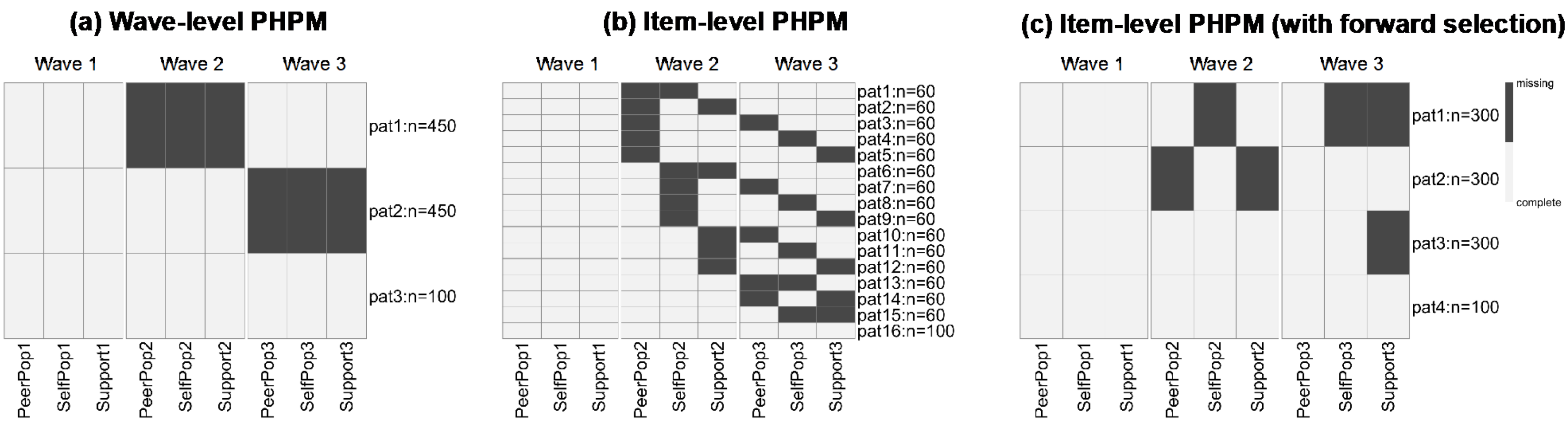


*Figure 7*. Optimal PHPM designs for the autoregressive and cross-lagged panel model example (PeerPop = peer-perceived popularity, SelfPop = self-perceived popularity, Support = teacher-child support, pat = pattern). (a) Optimal wave-level PHPM designs; (b) Optimal item-level PHPM design; (c) Optimal PHPM design obtained via forward selection.

Table S1

*Example of Item-level PHPM Designs After Wave 1 Data Collection Is Completed*

| | **Wave 1** | | | **Wave 2** | | | **Wave 3** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Measure A | Measure B | Measure C | Measure A | Measure B | Measure C | Measure A | Measure B | Measure C |
| Pattern 1 | | | | | | | | | |
| Pattern 2 | | | | | | | | | |
| Pattern 3 | | | | | | | | | |
| Pattern 4 | | | | | | | | | |
| Pattern 5 | | | | | | | | | |
| Pattern 6 | | | | | | | | | |
| Pattern 7 | | | | | | | | | |
| Pattern 8 | | | | | | | | | |
| Pattern 9 | | | | | | | | | |
| Pattern 10 | | | | | | | | | |
| Pattern 11 | | | | | | | | | |
| Pattern 12 | | | | | | | | | |
| Pattern 13 | | | | | | | | | |
| Pattern 14 | | | | | | | | | |
| Pattern 15 | | | | | | | | | |

## Appendix A

## R code used in the linear LGM

```
popModel <- '

i=~1*se1+1*se2+1*se3+1*se4
s=~0*se1+1*se2+2*se3+3*se4

se1~0*1
se2~0*1
se3~0*1
se4~0*1

se1~~0.071*se1
se2~~0.034*se2
se3~~0.067*se3
se4~~0.025*se4

i~2.983*1
s~0.086*1

i~~0.268*i+(-0.039)*s
s~~0.023*s

'


analyzeModel <- '

i=~1*se1+1*se2+1*se3+1*se4
s=~0*se1+1*se2+2*se3+3*se4

se1~0*1
se2~0*1
se3~0*1
se4~0*1

se1~~se1
se2~~se2
se3~~se3
se4~~se4

i~1
s~1

i~~i+s
s~~s

'
```

```
# wave missing design
wave.ex1=simPM(popModel=popModel,
        analyzeModel=analyzeModel,
        VarNAMES=c("se1","se2","se3","se4"),
        Time=4,
        Time.complete=1,
        k=1,
        pc=0.2,
        pd=0,
        costmx=c(5,10,15),
        n=323,
        nreps=1000,
        focal.param=c("i~1","s~1","i~~i","s~~s"),
        complete.wave=NULL,
        eval.budget=T,
        rm.budget=30*323*0.7,
        distal.var=NULL,
        seed=12345,
        engine="l",
        methods="wave")

summary.opt(wave.ex1)
summary(wave.ex1$opt.output)

plotPM(wave.ex1,Time=4,k=1,labels=F,col=c("gray96","gray35"))
```

Appendix B

R code used in the second-order linear LGM example

```
#population model

popModel='
EXB1=~1.150*F1+0.836*T1+1*M1
EXB2=~1.150*F2+0.836*T2+1*M2
EXB3=~1.150*F3+0.836*T3+1*M3
EXB4=~1.150*F4+0.836*T4+1*M4


interc=~1*EXB1+1*EXB2+1*EXB3+1*EXB4
slope=~0*EXB1+2*EXB2+3*EXB3+4*EXB4

interc~~-0.244*slope

interc~8.289*1
slope~-0.433*1

interc~~18.184*interc
slope~~0.249*slope

EXB1~~1.084*EXB1
EXB2~~1.777*EXB2
EXB3~~1.457*EXB3
EXB4~~1.700*EXB4

T1~-0.214*1
T2~-0.214*1
T3~-0.214*1
T4~-0.214*1

M1~0*1
M2~0*1
M3~0*1
M4~0*1

F1~-1.136*1
F2~-1.136*1
F3~-1.136*1
F4~-1.136*1


M1~~23.886*M1
F1~~17.737*F1
T1~~55.074*T1
M2~~20.223*M2
F2~~8.941*F2
T2~~66.698*T2
M3~~16.905*M3
F3~~13.922*F3
T3~~61.995*T3
M4~~19.324*M4
F4~~13.410*F4
```

```
T4~~71.127*T4

F1~~4.256*F2+7.040*F3+5.737*F4
F2~~5.440*F3+3.590*F4
F3~~6.165*F4

T1~~23.603*T2+24.666*T3+23.168*T4
T2~~35.213*T3+29.648*T4
T3~~33.815*T4

M1~~12.975*M2+11.153*M3+11.683*M4
M2~~12.219*M3+11.332*M4
M3~~11.807*M4

'

# analysis model

analyzeModel='


EXB1=~NA*F1+a*F1+b*T1+1*M1
EXB2=~NA*F2+a*F2+b*T2+1*M2
EXB3=~NA*F3+a*F3+b*T3+1*M3
EXB4=~NA*F4+a*F4+b*T4+1*M4

interc=~1*EXB1+1*EXB2+1*EXB3+1*EXB4
slope=~0*EXB1+2*EXB2+3*EXB3+4*EXB4

interc~~slope

interc~1
slope~1
interc~~interc
slope~~slope

EXB1~~EXB1
EXB2~~EXB2
EXB3~~EXB3
EXB4~~EXB4

F1~c*1
F2~c*1
F3~c*1
F4~c*1

T1~d*1
T2~d*1
T3~d*1
T4~d*1

M1~0*1
M2~0*1
M3~0*1
M4~0*1

F1~~F1
```

```
F2~~F2
F3~~F3
F4~~F4

T1~~T1
T2~~T2
T3~~T3
T4~~T4

M1~~M1
M2~~M2
M3~~M3
M4~~M4

F1~~F2+F3+F4
F2~~F3+F4
F3~~F4

T1~~T2+T3+T4
T2~~T3+T4
T3~~T4

M1~~M2+M3+M4
M2~~M3+M4
M3~~M4
'


# wave level missing

wave.ex4=simPM(popModel=popModel,
               analyzeModel=analyzeModel,
VarNAMES=c("F1","T1","M1","F2","T2","M2","F3","T3","M3","F4","T4","M4"),
               Time=4,
               Time.complete=1,
               k=3,
               pc=0.1,
               pd=0,
               costmx=c(rep(5,3),rep(10,3),rep(15,3)),
               n=1135,
               nreps=1000,
focal.param=c("interc~1","slope~1","interc~~interc","slope~~slope"),
               complete.wave=NULL,
               eval.budget=T,
               rm.budget=90*1135*0.7,
               distal.var=NULL,
               seed=12345,
               engine="l",
               methods="wave"
               )

summary.opt(wave.ex4)
summary(wave.ex4$opt.output)

plotPM(wave.ex4,Time=4,k=3,labels=F,col=c("gray96","gray35"))
```

```
# item level missing

item.ex4=simPM(
  popModel,
  analyzeModel,
  VarNAMES=c("F1","T1","M1","F2","T2","M2","F3","T3","M3","F4","T4","M4"),
  distal.var = NULL,
  n=1135,
  nreps=1000,
  seed=12345,
  Time=4,
  k=3,
  Time.complete=1,
  costmx=c(5,5,5,10,10,10,15,15,15),
  pc=0.1,
  pd=0,
  focal.param=c("interc~1","slope~1","interc~~interc","slope~~slope"),
  eval.budget=T,
  rm.budget=90*1135*0.7,        # remaining available budget
  complete.var=NULL,
  engine="l",
  methods="item"
)

summary.opt(item.ex4)
summary(item.ex4$opt.output)
plotPM(item.ex4,Time=4,k=3,row.names=F,labels=F,col=c("gray96","gray35"))


#item-level missing using forward selection

forward.ex4=simPM(
  analyzeModel = analyzeModel,
  popModel=popModel,
  VarNAMES=c("F1","T1","M1","F2","T2","M2","F3","T3","M3","F4","T4","M4"),
  distal.var=NULL,
  n=1135,
  nreps=1000,
  seed=123321,
  Time=4,
  k=3,
  Time.complete=1,
  costmx=c(rep(5,3),rep(10,3),rep(15,3)),
  pc=0.05,
  pd=0,
  focal.param=c("interc~1","slope~1","interc~~interc","slope~~slope"),
  max.mk=5,                 # maximum number of missing data patterns
  eval.budget=T,
  rm.budget=1135*(15+30+45)*0.7,
  complete.var="M3",
  engine="l",
  methods="forward"
)
summary.opt(forward.ex4)
summary(forward.ex4$opt.output)

plotPM(forward.ex4,Time=4,k=3,labels=F,col=c("gray96","gray35"))
```

# Appendix C

R code used in the autoregressive and cross-lagged panel model example.

```
# set up the population model
popModel <- '


SelfPop2~0.609*SelfPop1+0.111*PeerPop1
SelfPop3~0.587*SelfPop2+0.053*PeerPop2

PeerPop2~0.051*SelfPop1+0.647*PeerPop1+(-0.056)*Support1
PeerPop3~0.056*SelfPop2+0.679*PeerPop2+0.075*Support2

Support2~0.468*Support1+0.124*PeerPop1
Support3~0.280*Support2+0.132*PeerPop2

PeerPop1~~0.229*SelfPop1+0.064*Support1
SelfPop1~~0.007*Support1

PeerPop2~~0.079*SelfPop2+(-0.036)*Support2
SelfPop2~~0.055*Support2

PeerPop3~~0.052*SelfPop3+0.013*Support3
SelfPop3~~(-0.034)*Support3


#means
PeerPop1~3.273*1
SelfPop1~0.048*1
Support1~2.905*1

PeerPop2~1.247*1
SelfPop2~(-0.343)*1
Support2~1.193*1

PeerPop3~0.814*1
SelfPop3~(-0.161)*1
Support3~1.451*1

#variances

PeerPop1~~0.447*PeerPop1
SelfPop1~~1.010*SelfPop1
Support1~~0.540*Support1

PeerPop2~~0.238*PeerPop2
SelfPop2~~0.607*SelfPop2
Support2~~0.432*Support2

PeerPop3~~0.211*PeerPop3
SelfPop3~~0.380*SelfPop3
Support3~~0.424*Support3

'
```

```
# set up the analysis model
analyzeModel <- '

SelfPop2~SelfPop1+PeerPop1
SelfPop3~SelfPop2+PeerPop2

PeerPop2~SelfPop1+PeerPop1+Support1
PeerPop3~SelfPop2+PeerPop2+Support2

Support2~Support1+PeerPop1
Support3~Support2+PeerPop2

PeerPop1~~SelfPop1+Support1
SelfPop1~~Support1

PeerPop2~~SelfPop2+Support2
SelfPop2~~Support2

PeerPop3~~SelfPop3+Support3
SelfPop3~~Support3


#means
PeerPop1~1
SelfPop1~1
Support1~1

PeerPop2~1
SelfPop2~1
Support2~1

PeerPop3~1
SelfPop3~1
Support3~1

#variances

PeerPop1~~PeerPop1
SelfPop1~~SelfPop1
Support1~~Support1

PeerPop2~~PeerPop2
SelfPop2~~SelfPop2
Support2~~Support2

PeerPop3~~PeerPop3
SelfPop3~~SelfPop3
Support3~~Support3

'
```

```
# wave-level missing

wave.ex2=simPM(popModel=popModel,analyzeModel=analyzeModel,

VarNAMES=c("PeerPop1","SelfPop1","Support1","PeerPop2","SelfPop2","Support2",
"PeerPop3","SelfPop3","Support3"),
        distal.var = NULL,
        n=1000,
        nreps=1000,
        seed=1234,
        Time=3,
        k=3,
        Time.complete=1,
        costmx=c(5,5,5,10,10,10),
        pc=0.1,
        pd=0,
focal.param=c("PeerPop2~SelfPop1","Support2~PeerPop1","PeerPop3~SelfPop2","Su
pport3~PeerPop2"),
        eval.budget=T,
        rm.budget=45*1000*0.7,         # remaining available budget
        complete.wave=NULL,
        engine="l",
        methods="wave")

summary.opt(wave.ex2)
summary(wave.ex2$opt.output)

par(mfrow=c(1,1),oma=c(0.05,0.05,0.05,0.05),adj=0.5)
plotPM(wave.ex2,Time=3,k=3,labels=F,col=c("gray96","gray35"), fontsize_col=12)


## item-level missing designs

item.ex2=simPM(
  popModel,
  analyzeModel,
VarNAMES=c("PeerPop1","SelfPop1","Support1","PeerPop2","SelfPop2","Support2",
"PeerPop3","SelfPop3","Support3"),
  distal.var = NULL,
  n=1000,
  nreps=1000,
  seed=12345,
  Time=3,
  k=3,
  Time.complete=1,
  costmx=c(5,5,5,10,10,10),
  pc=0.1,
  pd=0,
focal.param=c("PeerPop2~SelfPop1","Support2~PeerPop1","PeerPop3~SelfPop2","Su
pport3~PeerPop2"),
  eval.budget=T,
  rm.budget=31500L,        # remaining available budget 45*1000*0.7
  complete.var=NULL,
  engine="l",
  methods="item"
)
summary.opt(item.ex2)
```

```
summary(item.ex2$opt.output)
plotPM(item.ex2,Time=3,k=3,labels=F,col=c("gray96","gray35"), fontsize_col=12)


#item-level PM designs using forward selection
forward.ex2=simPM(
  popModel,
  analyzeModel,
VarNAMES=c("PeerPop1","SelfPop1","Support1","PeerPop2","SelfPop2","Support2",
"PeerPop3","SelfPop3","Support3"),
  distal.var = NULL,
  n=1000,
  nreps=1000,
  seed=12345,
  Time=3,
  k=3,
  max.mk=3,
  Time.complete=1,
  costmx=c(5,5,5,10,10,10),
  pc=0.1,
  pd=0,
focal.param=c("PeerPop2~SelfPop1","Support2~PeerPop1","PeerPop3~SelfPop2","Su
pport3~PeerPop2"),
  eval.budget=T,
  rm.budget=31500L,
  complete.var=NULL,
  engine="l",
  methods="forward"
)

summary.opt(forward.ex2)
summary(forward.ex2$opt.output)
plotPM(forward.ex2,Time=3,k=3,labels=F,col=c("gray96","gray35"),
fontsize_col=12)
```