

Planned Missing Data Designs for Developmental Researchers

Todd D. Little and Mijke Rhemtulla

University of Kansas

ABSTRACT—*Planned missing data designs allow researchers to collect incomplete data from participants by randomly assigning participants to have missing items on a survey (multiform designs) or missing measurement occasions in a longitudinal design (wave missing designs) or by administering an intensive measure to a small subsample of a larger dataset (two-method measurement designs). When these designs are implemented correctly and when missingness is dealt with using a modern approach, the cost of data collection is lowered (sometimes dramatically) and reduced participant burden may result in higher validity as well as lower rates of unplanned missing data. In reviewing these planned missing designs, we briefly describe results of ongoing research on bias and power associated with each.*

KEYWORDS—*planned missing data; missing by design; intentionally incomplete data; multiform design; three-form design; two-method measurement; wave missing*

Todd D. Little and Mijke Rhemtulla, Center for Research Methods and Data Analysis, University of Kansas.

Support for this project was provided by NSF Grant 1053160 to the first author (Wei Wu, co-PI), a Banting postdoctoral fellowship from the Social Sciences and Humanities Research Council of Canada to the second author, and the Center for Research Methods and Data Analysis at the University of Kansas (Todd D. Little, director). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the funding agencies.

Correspondence concerning this article should be addressed to Todd Little, Center for Research Methods and Data Analysis, University of Kansas, 1425 Jayhawk Blvd, Watson Library, 470, Lawrence, KS 66045; e-mail: yhat@ku.edu.

© 2013 The Authors

Child Development Perspectives © 2013 The Society for Research in Child Development
DOI: 10.1111/cdep.12043

Planned missing (PM) designs allow researchers to collect incomplete data from participants by randomly assigning them to have missing items (e.g., multiform designs), missing measurement occasions (e.g., wave missing designs), or missing measures (e.g., two-method measurement designs). These designs have several benefits: (a) shortening surveys or assessments reduces the burden on participants, leading to higher quality data; (b) shortened surveys allow more items in a study, increasing the breadth of constructs, and (c) the cost of data collection declines.

We first describe three types of PM designs—multiform, wave missing, and two-method measurement—developmentalists can use in planning or continuing a study. Because PM designs require the use of modern missing data methods (e.g., multiple imputation [MI] or maximum likelihood estimation), we briefly review these methods and their repercussions for both planned and unplanned missing data.

PM DESIGNS FOR DEVELOPMENTAL RESEARCH

Multiform Designs

Multiform designs (also called split-questionnaire, partial-questionnaire, and split-ballot designs) reduce the number of items (e.g., questions, test items) each participant responds to by creating multiple forms that each contain a subset of the total items to be assessed (Graham, Hofer, & MacKinnon, 1996; Graham, Taylor, Olchowski, & Cumsille, 2006; Raghunathan & Grizzle, 1995; Sirotnik & Wellington, 1977; Thomas, Raghunathan, Schenker, Katzoff, & Johnson, 2006; Wacholder, Carroll, Pee, & Gail, 1994). Participants are randomly assigned one of the created forms. In the three-form design, for example, all items are allocated to one of four blocks: X, A, B, and C (see Table 1). Each form is then composed of the X block plus two of the three remaining blocks, so that all participants respond to items in the X block, and two thirds of participants respond to the items in the A–C blocks (Graham et al., 1996). If 10% of items are assigned to the X block and 30% each to the A–C blocks, each participant completes a survey that is 30% shorter than the original.



Table 1
Three-Form Design

Form	Block			C
	X	A	B	
1	1	1	1	0
2	1	1	0	1
3	1	0	1	1

Note. 1 = items in block are included on form design; 0 = items in block are not included on form design. The number of items in each block can differ, though typically the A-C blocks should be of approximately equal length to equate the total length of each form.

In designing the study, the most important and most informative items should be placed in the X block to minimize lost information (Graham et al., 2006). For example, in a three-form design evaluating a drug-prevention intervention in middle school students, the X block contained 23 items querying nine demographic characteristics and 14 key variables surrounding drug use and norms, whereas other items (e.g., assessing psychosocial characteristics) were distributed across the A–C blocks (Hecht et al., 2003). When constructing a three-form design, scales measuring a construct should be divided across the four blocks, with the most central item or items belonging to the X block. For example, a seven-item loneliness scale might contain the item, “I feel lonely,” which would be a good candidate for the X block. The other six items—such as “I feel left out” and “I feel that no one likes me”—would be distributed evenly across the A, B, and C blocks. Any one participant would therefore respond to five of the seven items and all items would have at least two thirds of the sample responding to them.

Benefits of the multiform design include reducing the time needed to administer and code the protocol, and more importantly, reducing the burden placed on each participant. With the reduced fatigue and burden of these designs, the data may be more valid, effects may be stronger, and participants may respond to more items (Harel, Stratton, & Aseltine, 2012).

Variations on multiform designs include increasing the number of forms to counterbalance the order of administration, and creating more blocks (e.g., six blocks, resulting in 10 or more forms) to introduce more planned missingness. For example, the 10-form, six-block design can easily accommodate a 50% reduction in the item burden on participants if each participant is assigned to receive the X block plus two additional blocks (Graham et al., 2006). The number of blocks used may be influenced by the full set of items. If several constructs are measured by scales that contain more than 10 items each, putting a few key items in the X block and one to two items each in the remaining blocks may be reasonable. In contrast, constructs that are measured by a small number of items are not easily divided across six blocks.

For surveys that are administered electronically, it is unnecessary to create a fixed number of forms. Instead, items could be

randomly assigned to participants. In this case, it is still good practice to have an X block of items that is seen by all participants as well as to constrain the randomization of the remaining items so that every participant sees at least some items measuring every construct. For example, using the example of the loneliness scale, every participant would see the item “I feel lonely” plus a randomly selected set of two additional items from that scale. This practice ensures that each participant provides sufficient information about every construct.

Power

Power loss is a pressing question because parameter estimates are estimated less efficiently in the presence of missing data. Less efficient estimation means that parameters’ standard errors will be bigger, and thus the power to test them (e.g., to test whether they are different from zero) will decline. Characterizing the effect that a particular pattern of missingness will have on parameter estimates is not straightforward, but some trends are clear. Efficiency suffers most when estimating parameters that involve variables from different item sets (Rhemtulla, Jia, Wu, & Little, 2013). For example, the efficiency of a regression of a variable in the C block on one in the B block is much lower than if both items are in the C block or if one of the items is in the X block. In contrast, when constructs are measured by multiple items and these items are spread across item sets (e.g., the loneliness example given above), the efficiency loss for parameters involving these constructs tends to be quite small when missingness is imputed at the item level (Gottschall, West, & Enders, 2012). When scale items are distributed across blocks and the data are analyzed using a structural equation model, factor loadings can have low efficiency, but structural parameter estimates (e.g., regression paths among latent constructs) tend to be highly efficient (Rhemtulla et al., 2013).

Wave Missing

Longitudinal wave missing designs assign participants to one or more omitted occasions. For example, a design with monthly measurements for 6 months could assign each participant to be missing one measurement occasion (wave), resulting in one-sixth planned missing observations. Several familiar longitudinal designs can be construed as PM designs. For example, in a cross-sequential design, several cohorts of participants are measured longitudinally (Little, 2013); cohorts of 4-, 5-, and 6-year-olds at Wave 1 might be measured for 4 consecutive years until they are 7, 8, and 9 years old, respectively. This design can be seen as a missing data design where the youngest group is missing data at 8–9 years, the middle group is missing data at 4 and 9 years, and the oldest group is missing data at 4–5 years. Thus, a 6-year span is measured in just 4 years. Similarly, the developmental time-lag model (McArdle & Woodcock, 1997) begins with a design in which every participant is measured twice but the time between measurements varies. These data can be arrayed longitudinally where each lag

between measurements is a potential measurement occasion and each participant has data at the first time point plus one other time point (e.g., a participant with a 4-month time lag has complete data at Occasions 1 and 5, and missing data on Occasions 2, 3, 4, and 6). In this way, complete two-time-point data are transformed into PM multi-time-point data that can be analyzed as a growth curve model, for example.

As mentioned, a wave missing PM design randomly assigns participants to have particular waves omitted (see Table 2). The number and pattern of missing waves can be optimized for the kind of model and research question of interest. For example, a wave missing design can be optimized for a latent growth curve model analysis (see Hogue, Pornprasertmanit, Fry, Rhemtulla, & Little, 2013; Mistler & Enders, 2012; Graham, Taylor, & Cumsille, 2001; Rhemtulla et al., 2013, for details). Wave missingness can also be combined with item-level missingness using a multiform design at each measurement occasion.

Power

As with multiform design, the missing data patterns in wave missing designs influence the amount of information available to estimate parameters of interest. Few studies have examined the extent of power loss in wave missing designs, and these have looked exclusively at growth curve models. These studies found that (a) mean levels of latent intercepts and slopes do not suffer much efficiency loss (Mistler & Enders, 2012), (b) individual variability in latent intercepts and slopes are much less efficient when wave missingness is imposed relative to a complete data design (Rhemtulla et al., 2013), and (c) the power to detect the effect of a fully observed grouping variable (e.g., intervention vs. control) on a latent slope can be very great with planned missingness (Graham et al., 2001). In longitudinal designs, item-level missingness within a time point (e.g., using a multiform design at each occasion) is less detrimental to efficiency than wave missingness (Rhemtulla et al., 2013); however, the cost savings of wave level missing can be much greater. Such findings highlight the need to weigh the costs of lowered efficiency

Table 2
Wave Missing Design

% of sample	Measurement occasion				
	1	2	3	4	5
10	1	1	1	1	1
10	1	1	1	0	0
10	1	1	0	1	0
10	1	0	1	1	0
20	1	1	0	0	1
20	1	0	1	0	1
20	1	0	0	1	1

Note. 1 = participants are measured on this occasion; 0 = participants are not measured on this occasion. This particular wave missing design was shown by Graham, Taylor, and Cumsille (2001) to result in the highly efficient estimates of the effect of group membership on the linear slope.

against the costs of additional data collection in optimizing a PM design.

Two-Method Design

The two-method design is a remarkably effective way to leverage modern treatments for missing data to powerfully test critical hypotheses. This design is intended for situations when researchers face a choice between two very different measures of a construct. The first is considered a gold standard and is typically expensive or time consuming to collect. The second is inexpensive and quick, but contaminated by systematic measurement bias. For example, numerical ability in childhood might be assessed using either an in-person test (e.g., the Wechsler Objective Numerical Dimensions test [WOND]; Wechsler, 1996), which is time intensive but accurate, or a paper-and-pencil math test that can be administered easily and cheaply but is contaminated by bias related to children's written test-taking skills. Research that relies on the gold standard alone can be underpowered because the cost of the measure tends to limit sample size. Research that relies on the inexpensive measure alone suffers from diminished validity.

To use the two-method measurement design, the inexpensive measure is administered to the entire sample (which must be large enough to reliably estimate a structural equation model), and the gold standard is administered to a random subsample of those participants (see Table 3). Both measures are coded so that multiple indicators are available from each; for example, the WOND can be coded as two separate subscales, and any paper-and-pencil test or self-report measure with more than a single item can be summarized into multiple groups of variables (i.e., parcels; see Little, Rhemtulla, Gibson, & Schoemann, 2013). The reason for using multiple variables from each measure is that this design uses a latent variable model to quantify the measurement error and bias in the inexpensive measure and remove it from the focal construct.

The two-method measurement design can be extended longitudinally, with the gold standard measure included in a subset of measurement occasions. If the systematic bias in the inexpensive measure is expected to be unstable over time, the gold standard should be included more than once (Garnier-Villareal, Rhemtulla, & Little, 2013).

Table 3
Two-Method Measurement Design

Group	Method	
	Gold standard	Inexpensive or biased
Subset of <i>N</i>	1	1
Remainder of <i>N</i>	0	1

Note. 1 = participants receive measure; 0 = participants do not receive measure. The subset size should be determined based on the reliability of the measures for the two methods and the degree of correlation between them.

Analysis

A latent variable model can use the gold standard measure to separate the two sources of variance in the inexpensive (biased) measure by modeling a common factor that represents the shared variance between the two measures and a bias factor for the inexpensive measure (Graham et al., 2006). This bias factor for the inexpensive measure may contain meaningful information. For example, written test-taking skill, as a construct in a model, can be used to unconfound other constructs measured by written tests within the same model. In a manifest variable framework, the gold standard variable can simply be treated as a single measure of the construct of interest, with the inexpensive measure included as an auxiliary variable during imputation or model estimation.¹ Auxiliary variables are variables that correlate with other variables that have missing information or with the missingness itself (e.g., a variable that predicts which values are likely to be missing). This method will allow some of the missing information on the gold standard to be recovered from the inexpensive measure.

Power

Parameters estimated using the latent variable modeling approach (described in Graham et al., 2006) tend to be much more efficient than if only the gold standard is used (on a small sample) and more valid than if only the inexpensive measure is used. Simulation studies that examined the optimal ratio of gold standard to total sample size (Graham et al., 2006) indicate that if the gold standard measure is used on its own with the inexpensive measure as an auxiliary variable, efficiency will also be greater than if the inexpensive measure is not included. However, to our knowledge, the degree of efficiency gain from including this auxiliary variable has not been tested (see Collins, Schafer, & Kam, 2001, on the benefits of auxiliary variables).

Unplanned Missingness

Nearly all research contexts produce unplanned missing data, even when a PM design is used. We offer two recommendations for dealing preemptively with unplanned missing data. First, when analyzing power of a PM design, use rates of missing data from previous research to estimate the amount of unplanned missing data that may arise above the planned missingness. (for a good explanation of how to compute power with missing data, see Enders, 2010; see also Mistler & Enders, 2012; Schoemann, Miller, Pornprasertmanit, & Wu, in press). Second, include variables that are related to the reasons that missingness arises (e.g., conscientiousness, socioeconomic status), as well as variables that may be related to the variables with missing values. These auxiliary variables can be invaluable in attenuating bias and loss of efficiency due to missing data (Collins et al., 2001).

MODERN TOOLS TO DEAL WITH MISSING DATA

A PM design requires modern treatments for missing data. Techniques such as listwise or pairwise deletion (also known as complete case analysis) can work when the total amount of missing data is less than a few percent. However, with any appreciable amount of missing data, these methods result in substantially biased parameter estimates and incorrect standard errors (Graham, 2009). The modern treatments are either full information maximum likelihood estimation (FIML) or MI.

The decision to use FIML or MI is often a matter of convenience: FIML is the default estimator in current structural equation modeling software (e.g., Mplus, LISREL), whereas MI is more commonly used on data sets with a large number of variables before a particular analysis model is chosen. FIML is a model-based technique that estimates parameters given a particular model using all the observed data in a single step. MI is a two-step technique in which each missing value is filled in with a set of m imputed values, resulting in m complete data sets (m should be at least 20; Graham, Olchowski, & Gilreath, 2007; Schafer & Graham, 2002). The analysis of interest is done on every imputed data set and the results are combined across imputations according to Rubin's Rules (Rubin, 1987). By these rules, parameter estimates are averaged across imputations, whereas standard errors are composed of two sources of variance: the average standard error across imputations (within-imputation variance) and the variability in parameter estimates across imputations (between-imputation variance). Although combining results across imputations involves extra steps, common software packages (e.g., SAS, Mplus, SPSS) automate the steps. Both methods are equally appropriate and typically lead to the same results. Both methods also allow the inclusion of *auxiliary variables*, which improves the efficiency of estimation and reduces bias (Collins et al., 2001).

The advantages of using FIML or MI instead of deletion depend on the *missingness mechanism*, that is, the cause of missing data. When data are missing completely at random (MCAR), missingness does not depend on either the observed or missing values. PM designs conform to the MCAR assumption because participants are randomly assigned to a missing data pattern. When missingness is related to the values in the data, the missingness mechanism is either missing at random (MAR; when the probability of missingness is not predicted by the missing values themselves after accounting for observed values) or missing not at random (MNAR; when the probability of missingness is contingent on the missing values even after controlling for the observed values). Missing data that are not planned, for example, those that arise due to nonresponse or attrition, are unlikely to be MCAR.

When missingness is MCAR, modern methods increase the *efficiency* of parameter estimates compared to deletion methods.

¹Thanks to an anonymous reviewer for suggesting this approach.

Because modern methods use every observation, parameter estimates have smaller standard errors and therefore narrower confidence intervals than methods that do not use all the available data. When missingness is MAR, modern methods additionally assure that parameter estimates are *unbiased*, whereas deletion methods produce biased estimates under MAR (under MCAR, deletion methods produce unbiased estimates). If missingness is MNAR, the only methods for preventing bias are complex models that require making strong untestable assumptions (see Enders, 2010, 2011).

CONCLUSIONS

Planned missing designs are becoming more common in developmental research, particularly as research budgets tighten and modern missing data methods become more accessible. Ongoing research continues to flesh out the boundary conditions in their use and the overall ramifications of these designs in terms of cost, bias, validity, and power. As research continues to produce concrete recommendations, we expect to see more planned designs effectively implemented. Careful planning is essential. The PM designs are effective when optimized for a given project. Based on thoughtful power analyses, the cost savings of these designs can often be used to increase the number of participants to offset the expected loss in power. All PM designs can yield increased validity because of reduced burden on participants. The three designs highlighted here are not mutually exclusive: A given longitudinal study could easily include all three design elements. Given the potential benefits of these designs and the unequivocal statistical theory that underlies them, we encourage developmentalists to embrace them as the new paradigm for longitudinal studies as developmental science moves forward.

REFERENCES

- 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. doi:10.1037/1082-989X.6.4.330
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.
- Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*, 1–16. doi:10.1037/a0022640
- Garnier-Villareal, M., Rhemtulla, M., & Little, T. D. (2013). *Two-method planned missing designs for longitudinal research*. Manuscript submitted for review.
- Gottschall, A. C., West, S. G., & Enders, C. K. (2012). A comparison of item-level and scale-level multiple imputation for questionnaire batteries. *Multivariate Behavioral Research, 47*, 1–25.
- Graham, J. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530
- 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., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*, 206–213. doi:10.1007/s11121-007-0070-9
- Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing-data designs in analysis of change. In L. M. Collins & A. 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. doi:10.1037/1082-989X.11.4.323
- Harel, O., Stratton, J., & Aseltine, R. (2012). *Designed missingness to better estimate efficacy of behavioral studies* (Technical Report 11–15). Storrs: Department of Statistics, University of Connecticut.
- Hecht, M. L., Marsiglia, F. F., Elek, E., Wagstaff, D. A., Kulis, S., Dustman, P., et al. (2003). Culturally grounded substance use prevention: An evaluation of the keepin' it REAL curriculum. *Prevention Science, 4*, 233–248.
- Hogue, C. M., Pornprasertmanit, S., Fry, M. D., Rhemtulla, M., & Little, T. D. (2013). *Planned missing data designs for spline growth models in salivary cortisol research*. Manuscript submitted for review.
- 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*. doi:10.1037/a0033266
- McArdle, J. J., & Woodcock, R. W. (1997). Expanding test–retest designs to include developmental time-lag components. *Psychological Methods, 2*, 403–435. doi:10.1037/1082-989X.2.4.403
- 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. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association, 90*, 54–63. doi:10.1080/01621459.1995.10476488
- Rhemtulla, M., Jia, F., Wu, W., & Little, T. D. (2013). *Planned missing designs to optimize the efficiency of latent growth parameter estimates*. Manuscript under review.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177. doi:10.1037/1082-989X.7.2.147
- Schoemann, A. M., Miller, P., Pornprasertmanit, S., & Wu, W. (in press). Using Monte Carlo simulations to determine power and sample size for planned missing designs. *International Journal of Behavioral Development*.

- Sirotnik, K., & Wellington, R. (1977). Incidence sampling: An integrated theory for “matrix sampling”. *Journal of Educational Measurement, 14*, 343–399. doi:10.1111/j.1745-3984.1977.tb00050.x
- Thomas, N., Raghunathan, T. E., Schenker, N., Katzoff, M. J., & Johnson, C. L. (2006). An evaluation of matrix sampling methods using data from the National Health and Nutrition Examination Survey. *Survey Methodology, 32*, 217–231.
- Wacholder, S., Carroll, R. J., Pee, D., & Gail, M. H. (1994). The partial questionnaire design for case-control studies. *Statistics in Medicine, 13*, 623–634.
- Wechsler, D. (1996). *Wechsler objective numerical dimensions*. London, UK: Psychological Corporation.