

# On the Joys of Missing Data

Todd D. Little,<sup>1</sup> PhD, Terrence D. Jorgensen,<sup>1</sup> MS, Kyle M. Lang,<sup>1</sup> MA, and E. Whitney G. Moore,<sup>2</sup> PhD

<sup>1</sup>Center for Research Methods and Data Analysis, University of Kansas, and <sup>2</sup>Department of Kinesiology, Health Promotion, and Recreation, University of North Texas

All correspondence concerning this article should be addressed to Todd D. Little, PhD, Center for Research Methods and Data Analysis, University of Kansas, Lawrence, KS 66045, USA. E-mail: yhat@ku.edu

Received February 28, 2013; revisions received and accepted May 30, 2013

We provide conceptual introductions to missingness mechanisms—missing completely at random, missing at random, and missing not at random—and state-of-the-art methods of handling missing data—full-information maximum likelihood and multiple imputation—followed by a discussion of planned missing designs: Multiform questionnaire protocols, 2-method measurement models, and wave-missing longitudinal designs. We reviewed 80 articles of empirical studies published in the 2012 issues of the *Journal of Pediatric Psychology* to present a picture of how adequately missing data are currently handled in this field. To illustrate the benefits of using multiple imputation or full-information maximum likelihood and incorporating planned missingness into study designs, we provide example analyses of empirical data gathered using a 3-form planned missing design.

**Key words** full-information maximum likelihood; missing data analysis; missingness mechanisms; multiple imputation; planned missing design.

Missing data are not problematic, per se—how we approach and treat missing data, on the other hand, can be highly problematic. In fact, all of the traditional methods that are still popular in pediatric research suffer from many drawbacks that introduce bias and lack of power (Enders, 2010; Graham, 2012; van Buuren, 2011). In contrast to the drawbacks of traditional approaches (e.g., listwise or pairwise deletion, mean substitution), modern treatments for missing data such as multiple imputation (MI) and full-information maximum likelihood (FIML) easily accommodate valid inferences when the study design includes careful consideration of the reasons for why data might go missing. We provide a conceptual introduction to the steadily growing literature on modern missing data treatments. Admittedly, the devil is in the details of how to best minimize any deleterious effects of missing data, and we provide citations to ample resources where these details are carefully described. In this article, therefore, we only highlight the many ways that missing data can and should be addressed in pediatric research.

When a sample of participants is used to make inferences about a population of persons, three issues arise

when data are missing: power, bias, and recoverability. How much power has been lost? Are the remaining data now a biased representation of the population? Can the power loss be recovered? Can the bias be recovered? With modern treatments for missing data, the answer to each of these questions depends on which of three possible mechanisms (Rubin, 1976) is responsible for the missing data. All missing data happen because of one of these three mechanisms, and in any data collection endeavor, all three mechanisms are likely to be involved to some degree.

## Mechanisms of Missing Data

Rubin (1976) classified the possible ways that data could go missing. With traditional approaches to missing data, statistical analyses are conducted under the assumption that all the absent data were due to a missing completely at random (MCAR) process. This assumption is rarely tenable in uncontrolled environments. However, there are certain situations in which the MCAR mechanism could be encountered in practice. Consider the case of a longitudinal classroom-based study of prosocial behaviors among

elementary school students. If some of the students are not measured for a given wave because they were kept home owing to illness, that missingness would be MCAR. Because the students' illness is unrelated to any of the variables in the study, their nonresponse can be considered a random sample of the complete data. With planned missing data designs, MCAR is guaranteed to be the mechanism for the parts of the data frame that are missing by design. When MCAR is the mechanism, no bias is introduced, so only power is impacted. In this regard, the point estimates from the observed data are on average no different from what they would have been had there not been any missing data. The standard errors (SE) around the estimates, however, are larger because of the reduced sample size. Modern missing data treatments recover most of the power that is lost with the MCAR mechanism and keep the unbiased nature of the point estimates uninfluenced. The MCAR mechanism is, therefore, an ideal reason for missingness, because the impact of it is readily remedied with modern treatments.

The second mechanism, termed missing at random (MAR), refers to missingness that is due to a predictable reason and, therefore, becomes a random effect that is easily estimated. In a study of sexual activity among high school students, for example, highly religious students may feel uncomfortable with the subject of the investigation and choose not to answer questions related to their sexual behaviors. In this case, it is the subjects' religious convictions which are motivating the nonresponse. If a measure of religiosity is included in the survey instrument, then that item can be included as an auxiliary variable (i.e., a variable included in the model only to predict the missing values) during the subsequent missing data analysis. By proactively accounting for likely causes of missingness in this way, the estimated model parameters are adjusted to accurately reflect the original population values. The MCAR and MAR mechanisms are both referred to as ignorable missing data mechanisms because bias is either nonexistent (MCAR) or recoverable (MAR) and power is restored when a modern treatment is used.

The third mechanism of missing data is missing not at random (MNAR). This mechanism occurs when the missingness on a given variable is caused by the subjects' levels of that variable. One possible example may occur in a study of adolescent tobacco use if heavy smokers do not report the number of cigarettes they smoke. In this example, regularly engaging in an illegal activity may instill a fear of reprisal that prevents frequent smokers from reporting their activities. Thus, it is the number of cigarettes smoked is causing nonresponse on the smoking frequency item. In MNAR situations, the cause of the missingness cannot be

used to correct the parameter estimates for the bias that occurs due the reason(s) for missingness. Careful planning of a study can reduce any potential impact of an MNAR mechanism either by including direct measures of the potential causes of missingness or by including reasonable proxies or known correlates of the causes of such missingness. Including these proxies as auxiliary variables in the subsequent missing data analysis can correct some of the bias introduced by the nonresponse (Enders, 2010; Schafer, 1997). If such variables cannot be included, this mechanism will result in bias; however, the bias would not be any greater than traditional complete case or listwise deletion methods. Recent evidence indicates that even when the data are MNAR, a modern approach can recover some of the MNAR bias (Collins, Schafer, & Kam, 2001). Enders (2010, chapter 10) elaborates on additional models for MNAR data—the selection model and the pattern mixture model—both of which (in very different ways) take into account the joint distribution of the observed data and the probability that the data are missing. Modern missing data treatments under MAR follow the logic of intent to treat features of randomized designs (i.e., measure potential reasons for missingness and conduct the analyses under the MAR assumption).

As we have indicated, there are two modern approaches to addressing missing data. The first method is MI and the second approach is FIML estimation. We provide conceptual nontechnical introductions to these two approaches before turning to the topic of planned missing data designs.

## Multiple Imputation

Imputing missing data involves making a copy of the original data set and replacing missing values with plausible estimates of what those values would have been, had they been observed (Rubin, 1987). Old imputation techniques (mean substitution, regression imputation) result in two kinds of bias. The estimates themselves (e.g., correlations, group mean differences) will usually be too large or too small unless the values are missing completely at random (MCAR; i.e., the missingness is not related to any other variables whatsoever), which is a difficult assumption to defend unless the data are missing at random by design. The significance tests associated with those estimates are also biased because the SEs are reduced by introducing artificial certainty in the estimate.

MI is frequently misinterpreted as “making up data”—it is not, but mean substitution is. The mean of observed data may not be an arbitrarily made up value, but replacing every missing value with the mean is an

expression of the assumption that one can be absolutely certain that this single value is exactly what would have been observed if the observation were not missing. Mean substitution and regression imputation introduce no noise, no margin of error, and no variability around the plausible estimate of the missing value. Consequently, although the estimated mean of observed data is unchanged by mean substitution of missing data, the *SE* artificially shrinks because no deviations from the mean exist among those substitutions. Deviations from the mean are the foundation of estimating variability in the population (i.e., sum of squares), and such an estimate is too small if all missing observations are assumed to have zero deviation from the mean (or predicted value, in the case of regression imputation).

A technique called stochastic regression imputation gets around this by adding some random noise (i.e., *stochasticity*) to the predicted values, thus introducing the variability seen in real observed data. This makes the estimate of the missing value much more plausible. But how certain can one be in a single estimate of what a missing value might truly have been? MI essentially solves this problem the same way stochastic regression imputation does, but goes further by calculating several plausible estimates of a missing value instead of a single estimate. The rationale behind this multiplicity is similar to that for large samples in general.

Suppose a researcher is interested in finding the degree of belief among high-school children in a certain school district that condom use prevents the spread of HIV. Lack of time and funding prevents interviewing all 4,000 children in the district, so a sample is gathered from within the school district to draw an inference about the population from which it was drawn. In an extreme case, suppose a sample of  $N=1$  indicated 7 on a 10-point scale of support for condom use (where higher numbers indicate greater support). How much confidence should we have that this is the average degree of support among all children in the district? The next sample of  $N=1$  will likely yield very different results than the initial sample. The principle that larger random samples yield more certainty about estimates is discussed extensively even in introductory texts on research methods and statistics. MI operates on the same principle. Any substitution for a missing value is only one among many plausible substitutions, and our estimate of the missing information is more robust when many plausible values are sampled.

For instance, we might not know precisely how Subject 42 would have responded, but based on the observed range of responses (say, normally distributed with  $M=6$  and  $SD=2$ ), we could narrow it down to a

plausible range from which to sample substitution values (i.e., sample random numbers from a normal distribution with  $\mu=6$  and  $\sigma=2$ ). Currently, best practices indicate between 20 and 100 imputations are sufficient to recover the missing information in most cases (Graham, Olchowski, & Gilreath, 2007), given that sufficient variables are included in the data set that are related to the missing values themselves or to the reasons for missingness. In our previous example, suppose the population of female students shows greater support for condom use than male students, and students from high-SES families show greater support than students from low-SES families. Knowing the gender and SES of Subject 42 allows us to narrow down the range of plausible values even further (e.g., random values from a normal distribution with  $\mu=8.5$  and  $\sigma=0.5$ ).

### Full-Information Maximum Likelihood

Suppose that we would like to forgo the complications introduced into our data analysis when we use MI. Is there a way that we can make unbiased inferences without going through the process of fitting multiple replicates of our analysis model and pooling the results? Well, if the statistical technique to be used can accommodate maximum likelihood estimation (e.g., SEM, multilevel modeling), then the answer to this question is: Yes, by using FIML to fit the hypothesized model a single time while maintaining the benefits of a principled missing data tool.

Though the underlying mathematical principles exceed the scope of this article, the conceptual framework underlying FIML estimation is relatively simple. Consider the case of many participants filling out the same survey (i.e., a standard cross-sectional design). Even if some of those participants do not complete the survey, we would still like to fit a model that will allow us to draw accurate conclusions about the entire sample. FIML estimation can help accomplish this goal by using the observed responses to supplement the loss of information due to the missing responses.

By way of illustration, consider your computer's monitor. The image on this monitor is made up of many rows of pixels, just as a data set is made up of many rows of responses. If a pixel dies in your monitor, you are still able to understand the image on the screen because you can use the information from surrounding pixels to infer what the undamaged image would be. This principle will hold true even if a relatively large proportion of the pixels in your monitor were to fail, so long as there are enough pixels remaining for you to "connect the dots" and

extrapolate the implied complete image from the partial image on your screen. Similarly, FIML estimation uses what are known as casewise log-likelihoods to achieve an analogous effect when used to fit a statistical model to incomplete data. By using only what is known from the observed data, FIML can infer what the whole model should look like without needing to know what the missing responses would truly be. In this way, just as your eye can look at a damaged computer monitor and still understand how the complete image would appear, FIML can be applied to an incomplete data set to produce estimates that correctly describe the entire sample.

Numerous simulation studies have verified that when correlates of reasons for MAR missingness are measured and included in the analysis or imputation model, FIML and MI yield unbiased estimates of both parameters and their SEs (see example simulations in Enders, 2010, and Schafer & Graham, 2002).

Regardless of the mechanism of missingness, researchers should use the most principled technique available for their question of interest. When missing is MAR, for example, ad hoc approaches that attempt to treat missing data without considering their underlying structure (e.g., listwise deletion, casewise deletion, mean substitution, conditional mean imputation, last point carried forward, etc.) should be avoided in all situations. When less than 5% of the data are missing, and these missing values are spread across a reasonable proportion of the variables of interest (i.e., no single variable is mostly unobserved), a less rigorous missing data tool can often be used. Under these circumstances, a single stochastic regression imputation or single EM imputation offer easily implemented, yet principled, ways to treat the missing data. If a maximum likelihood technique is used, a noninclusive implementation of FIML can also provide acceptable results in this situation. Once the proportion of missing data gets much higher than 5%, and certainly once it exceeds 10%, we recommend using inclusive implementations of MI or FIML, as described above.

### Planned Missing Data Designs

Planned missing data designs have been suggested for many years but only recently have they begun to percolate into the design choices of applied researchers. Much of the work conducted in pediatric psychology would benefit from the applications of planned missing designs (Graham, Taylor, Olchowski, & Cumsille, 2006). Three designs are particularly useful for applied pediatric researchers: The multiform questionnaire protocol, the

two-method measurement model, and the wave-missing longitudinal design. For all planned missing designs, the critical element of them is random assignment. With true random assignment, the missing data from these designs are, by definition, MCAR in nature. Recall that MCAR produces no bias in the estimated parameters of a given statistical model, only power is diminished. Also recall that the two modern approaches to missing data treatments restore the lost power. MCAR with a modern treatment is a truly win-win situation for applied pediatric researchers!

### Multiform Questionnaire Protocol

Rather than creating short-forms of different scales or eliminating constructs because of time constraints or concerns about burden and fatigue, a multiform design can be implemented. Multiform designs can also reduce respondent reactivity. For researchers conducting intervention studies, the control condition often will show improvements just by virtue of reacting to the questionnaire protocol. Reducing exposure to all items of a given construct reduces the reactivity to the construct as a whole. Increases in validity (Harel, Stratton, & Aseltine, 2011) coupled with reductions in the cost to participants (Enders, 2010; Graham, 2012) are the two primary benefits of a multiform design. With a multiform planned missing design, the analyzed data contain all the needed items and information when a modern treatment of the missing data is used.

The simplest multiform design that researchers should use is the three-form design. As the name implies, three different questionnaire forms are created and randomly assigned to participants. The key to a three-form design is assigning items to four different blocks or sets, which are designated X, A, B, and C. The X block contains items that are administered to all participants. The A, B, and C blocks are paired to create the three different forms of items:  $X + A + B$ ,  $X + A + C$ , and  $X + B + C$ . That is, one of the blocks of items in A, B, or C are intentionally not administered. Each form is about 75% of the length of a full protocol. More blocks of items can be generated and put together to create forms that have even fewer administered items. In fact, multiform designs can be used to generate forms that contain ~40% of the full battery items (i.e., each form has 60% missing data; Raghunathan & Grizzle, 1995).

The three-form version of a multiform design introduces around 25–30% missing data (depending on number of items assigned to A, B, and C blocks), but these data are



Table 1. Schematics for (a) Three-Form Planned Missing Design and (b) Four-Occasion Wave Missing Design

| Form | Common Set X | Variable Set A | Variable Set B | Variable Set C |
|------|--------------|----------------|----------------|----------------|
| 1    | 25% of items | 25% of items   | 25% of items   | Missing        |
| 2    | 25% of items | 25% of items   | Missing        | 25% of items   |
| 3    | 25% of items | Missing        | 25% of items   | 25% of items   |

| N  | Occasion of Measurement |         |         |         |
|----|-------------------------|---------|---------|---------|
|    | 1                       | 2       | 3       | 4       |
| 50 | Yes                     | Yes     | Yes     | Yes     |
| 50 | Yes                     | Yes     | Yes     | Missing |
| 50 | Yes                     | Yes     | Missing | Yes     |
| 50 | Yes                     | Missing | Yes     | Yes     |
| 50 | Missing                 | Yes     | Yes     | Yes     |

Note. Proportions of variables in each block need not exactly match this schematic.

Note. It is not strictly necessary to include a group that is measured at all occasions, but it can help stabilize parameter estimation (Graham, Taylor, & Cumsille, 2001).

MCAR because each of the three forms is randomly assigned to each participant. The top tier of Table 1 shows a schematic of the pattern of complete and missing data that results from using such a design. In assigning items to blocks, a number of considerations are involved. First, the X block, which is administered to all participants, typically will contain the essential demographic variables as well as key variables that are likely to predict the MAR mechanism. Although the intentional parts of the missing data are MCAR, nearly any study will also have additional missing information on top of the randomly controlled MAR data. In addition to these variables, we recommend that at least one item from each construct be included in the X block. This one item would be an indicator of the construct with the best item properties (i.e., the item with the highest loading from a CFA model of all items for a given construct). The rest of the variables associated with each construct would be evenly distributed across the A, B, and C blocks of items. Here, each of the A, B, and C blocks would contain one or more items from each construct if enough items for a given construct are available. If not enough items per construct exist, the pattern of assignment of items to constructs should (a) balance the number of items in each block as equally as possible, and (b) maximize the between-block correlations among the items.

The higher the between-block associations are in this design, the more efficient is the missing data recovery process, which leads to greater power and greater convergence rates when the data are analyzed. The multiform designs are optimal for large sample studies that rely on SEM procedures. Based on simulation work, the three-form design requires sample sizes of about 180 and greater to

achieve acceptable coverage and convergence (Jia et al., 2013). At sample sizes of this magnitude, SEM procedures can be used (see Little, 2013).

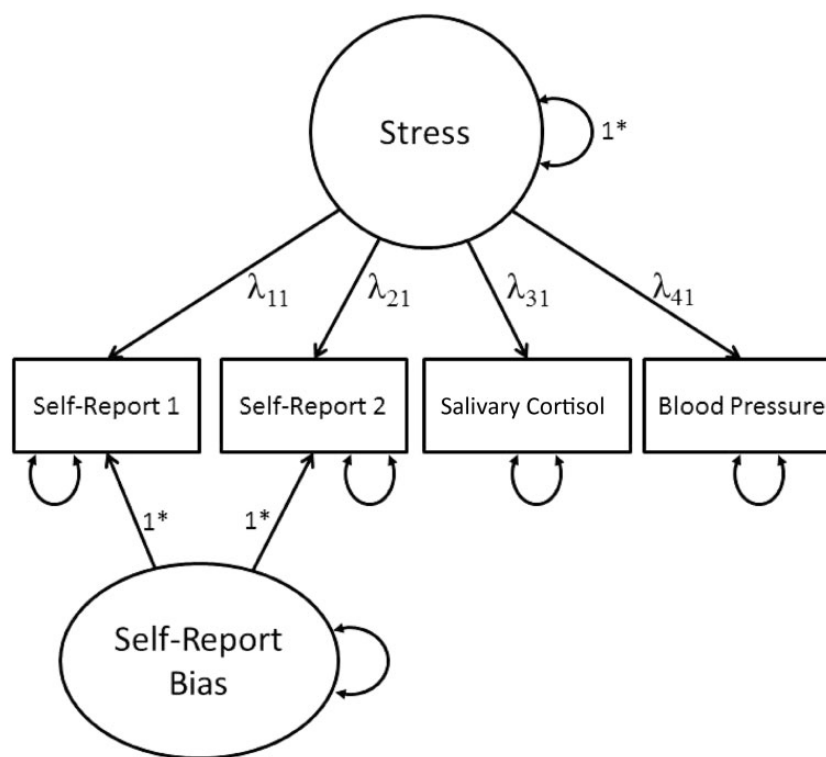
## Two-Method Planned Missing Design

Unlike the three-form design, which intentionally omits variables to reduce cost and burden, the two-method design is a way to increase the power of an otherwise underpowered study. The two-method design is ideally suited for contexts in which an expensive, but highly valid, method for assessing a construct is desired. If a cheaper and, by implication, less valid method of measuring the same construct exists, the two methods can be partnered together to dramatically increase the sample size while holding the costs of a given study constant. In addition to the requirement that two methods of measuring the same construct exists, the two-method planned missing data design is also predicated on a multivariate measurement model to represent the construct of interest. As with multiform designs, sample sizes for these designs need to be large enough to support the estimation of latent constructs. Figure 1 is a depiction of a two-method analysis of stress as represented by the gold standard of cortisol measured using two assays and by a simple self-report questionnaire of perceived stress. The items of this self-report measure are parceled into three indicators.

## Wave-Missing Longitudinal Design

The multiform questionnaire protocol can be administered in a longitudinal design, but there may be little reduction in the cost of obtaining measurements on every occasion, even if the battery of measurements were shorter. A rationale similar to the multiform protocol underlies a wave-missing longitudinal design, in which complete data are obtained from participants on each occasion of measurement, but participants are randomly assigned to be measured only on a subset of those occasions (Graham, Taylor, & Cumsille, 2001).

Suppose there are four waves of measurement (once every 3 months) across which one wishes to estimate the change trajectory of coping during the first year following cancer diagnosis. It is unlikely that all participants would participate on all occasions, and it would be difficult to know the mechanism of missingness. Measuring a subset of the full sample on each occasion would reduce costs and participant burden, making it less likely that a subject would not be measured for some reason other than random assignment (Harel, Stratton, & Aseltine, 2011).



**Figure 1.** Path diagram illustrating a two-method planned missing design.

Graham, Taylor, and Cumsille (2001) present several similar design possibilities, the simplest being to divide the sample into five parts, four of which would not be measured at Wave 1, 2, 3, or 4 (see bottom tier of Table I). One subset would be measured at all occasions, which Graham et al. (2001) explain is not a requirement for these designs, but it can help to have a small subset of complete-data cases to “help with estimation of higher-order partial correlations” (p. 341). This would result in one fifth (20%) of data points MCAR by design, but greater proportions of missingness (and by implication, reduction in cost and participant burden) can be achieved—even when adding more measurement occasions! For example, with five waves of measurement, one can divide the sample into 11 parts and assign 10 of those subsets to be measured on only three of five occasions (i.e., there are 10 possible combinations of three waves: 123, 124, 125, 134, 135, 145, 234, 235, 245, and 345). With the last subset being measured on all occasions, this would result in more than one third (36.4%) of data points MCAR by design.

### Review of Recently Published Articles

Of the 113 articles published in the 2012 issues of the *Journal of Pediatric Psychology*, we reviewed 80 empirical studies after excluding meta-analyses, review articles,

introductory and theoretical articles, commentaries, and qualitative studies. Missing data resulted in decreased sample size ranging from one or two no-shows to as much as 65% missing from the original sample (see Karlson et al., 2012). Among the 80 reviewed studies, only 45 (56.25%) mentioned missing data explicitly in the text or a table of descriptive statistics. Of those 45, only three mentioned testing whether the missingness was related to other variables, justifying their MAR assumption; eight additional studies stated that MAR or MCAR was assumed to be the missingness mechanism; and seven merely cited attrition as the reason for missingness, without elaborating on the mechanism. The remaining 27 of those who mentioned missing data did not discuss the mechanism, although two of those 27 discussed the possibility that missingness could have been related to other variables (i.e., an MAR mechanism). Of the 45 studies that mentioned missing data, 20 involved repeated measurements, and 15 of those 20 (i.e., 75% of studies with repeated measures that mentioned missing data) attributed at least some part of the missing data to attrition.

Of the 45 studies that mentioned missing data, nine made no mention of how missing data were handled, so we can assume that their analyses included only complete cases. Along with these nine, 17 additional studies

explicitly used listwise deletion, so 57.8% (26 out of 45) appear to have analyzed only a subsample of their data, operating on the assumption that the subsample was as representative of the population(s) as the full sample (i.e., MCAR mechanism)—an unlikely scenario. Some studies used more than one method to handle missing data—such as mean imputation for variables with few missing values, then subsequent listwise deletion—so the combination of the 26 studies that use listwise deletion and the following studies that use other methods form a sum greater than 45. Six studies used some form of single-imputation method (i.e., replacing missing values with a single value rather than a distribution of values as in MI): Three used mean substitution, two used a single imputation using the EM algorithm, and one used hot-deck imputation. Details about various deletion and imputation methods can be found in Schafer and Graham (2002), Chapter 2 of Graham (2012), and Chapter 2 of Enders (2010).

Only two studies used MI, and 11 studies used FIML, so less than one third (28.9%) of studies that mentioned missing data appear to have used state-of-the-art methods for handling missing data. FIML analyses seem to have been used as often as they were because seven of them were latent variable models in Mplus (Muthén & Muthén, 2012), which uses FIML as the default method to handle missing data. Likewise, the other four FIML analyses were multilevel longitudinal models (i.e., with multiple observations nested within subjects) fit with SPSS, HLM, or SAS, for which FIML (or REML, which is similar to FIML but maximizes the likelihood of the residual covariance matrix rather than the full matrix) is also the default estimator. Among the reviewed articles in which missing data were well handled using various analyses, interested readers can review Moran et al. (2012) for an example of maximum likelihood in a multilevel regression, Cushing and Steele (2012) for an example of a single imputation using the EM algorithm before a multiple-group SEM, and Herge et al. (2012) for an example of FIML in a SEM.

Of the 80 studies under discussion, none used a planned missing design of any kind. Any study whose materials include multiple-item scales would benefit from a multiform questionnaire protocol by reducing participant burden, likely increasing the validity and response rates of their measurements. Likewise, any study whose measures are proxies for others that would be more valid or reliable, but more expensive, invasive, or time-consuming, would benefit from a two-method planned missing design. Some studies, however, would benefit more than others; namely, studies that collect measurements across three or more

questionnaires could increase their sample size by assigning smaller subsets of those items to a larger number of people, and studies whose measurements have biological counterparts (which may be more common among pediatric and other medical studies than it would be among behavioral studies in general) would benefit from administering such measures to a small sample while collecting proxies from a much larger one.

Based on such criteria and the information available in each article, we ascertain that at a bare minimum, nine of these 80 studies would have benefited noticeably from using a multiform questionnaire protocol, 11 would have benefited from a two-method planned missing design, and three would have benefited from a wave-missing longitudinal design. Thus, more than a quarter (28.75%) of the studies under review would have benefited from at least one of the planned missing data designs. Again, this is an underestimate, given that the majority of studies used at least one scale of some kind, which could be partitioned into a multiform protocol. Examples of constructs under investigation which could benefit from a two-method approach include biological variables such as asthma severity, stress, glycemic control, pain, and neurological assessments, as well as behavioral and environmental variables such as teasing, substance (ab)use, and parental reports of children's pain and postinjury symptoms, which could be supplemented by a small subset of complementary measures on the children themselves.

## Empirical Example

To compare how traditional and state-of-the-art methods may be used to handle missing data, we provide a demonstration of a set of analyses using empirical data that were gathered using a three-form planned missing design. These are real data, not simulated data, so we have no absolute point of reference for what the true estimates should be, but in our introduction we have provided sources demonstrating that modern methods are the gold standard for handling missing data. Rather, our empirical example and accompanying SPSS and R syntax (see online appendix) are provided as tools to assist researchers who are new to these methods, using analytic methods we found to be commonly used in our review of 2012 articles.

## Theory

Given the high rates of adolescent physical inactivity, increasing students' positive experiences and attitudes regarding physical education (PE) is one important

approach available to increase adolescents' physical activity levels (Division of Adolescent and School Health & CDC, 2010). Thirty years of Achievement Goal Perspective Theory (AGPT) research conducted by educational and sport/exercise psychologists has consistently shown that experiencing a task-involving climate is related with reporting greater effort, enjoyment, and focus on mastery goals rather than performance goals (Roberts, 2012; Smith, Smoll, & Cumming, 2009; Weigand & Burton, 2002). A task-involving climate emphasizes individual improvement, effort, and cooperative learning (Seifriz, Duda, & Chi, 1992). A basic AGPT tenet is that the class climate experienced by students will influence their goal orientation (e.g., definition of success; Nicholls, 1989). Individuals reporting a high task goal orientation view success in terms of their learning and improvement, regardless of their skill level and the skill level of those around them (Nicholls, 1989). For the purposes of this study's examples, we will focus on the task-involving class climate's ability to predict student's future task goal orientation and PE satisfaction. Individuals' level of task goal orientation in some studies has been significantly lower for males than females; therefore, gender was controlled for in the analyses (Walling & Duda, 1995). When adolescents experience a task-involving climate in PE, they are also more likely to report a desire to continue being physically active (Nicholls, 1989; Roberts & Treasure, 2012), which has been related to not just improved weight status, but positive mental health benefits as well (Goldfield, Adamo, Rutherford, & Murray, 2012; Loth, Mond, Wall, & Neumark-Sztainer, 2010).

## Method

To allow time for the students' perceptions of the PE class motivational climate to develop, the students were surveyed at the semester midpoint and the week before the end of the semester (~6 weeks later). The surveys were administered using a three-form planned missing design (Moore, 2011). The students completed the surveys anonymously; their responses were matched using the class period, birthdate, and gender information that they provided. Responses were measured on a scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*).

The surveys were randomly distributed to the students in each class period. Therefore, the survey version (i.e., form XAB, XAC, or XBC) each student completed was not pre-assigned. During the first measurement wave, 33% completed survey Version 1, 34% completed Version 2, and 33% completed Version 3. During the second measurement

wave 35% completed Version 1, 31% completed Version 2, and 34% completed Version 3. A total of 563 secondary students' survey data are used in these examples.

## Measures

The 27-item Perceived Motivational Climate in Exercise Questionnaire (PMCEQ) was used to assess the students' perceptions of the PE class' motivational climate (Huddleston, Fry, & Brown, 2012). The task-involving subscale is composed of 14 items. The stem was worded to be specific to students' PE class setting, for example: "In this PE class, the teacher emphasizes always trying your best."

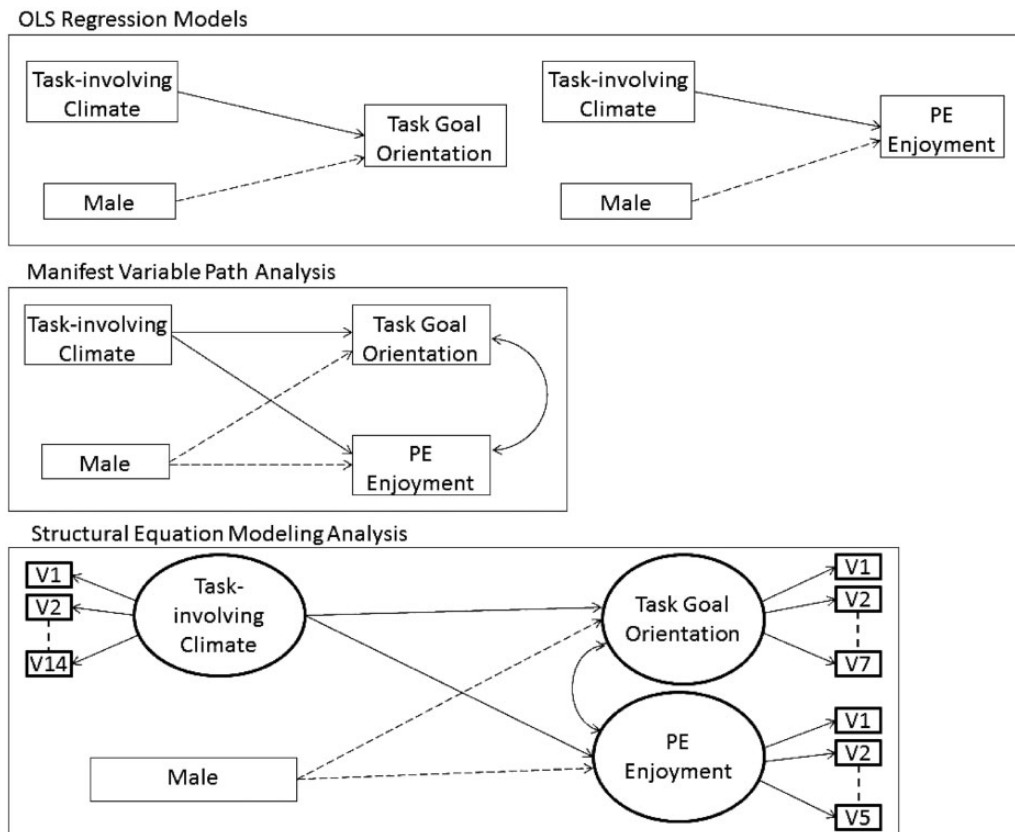
The 13-item Task and Ego Orientation in Sport Questionnaire (TEOSQ) was used to assess the students' self-reported goal orientations (Duda, 1989; Duda & Nicholls, 1992). The task goal orientation subscale is composed of seven items. The stem was worded to be specific to the PE class setting, for example: "In this PE class, I feel successful when something I learn makes me want to practice more." The five-item Intrinsic Satisfaction Sport Scale (ISSS) was developed by Duda and Nicholls (1992) to assess students' self-reported intrinsic enjoyment of sport. The stem was worded to be specific to the PE class setting, for example: "In this PE class, I usually find time flies."

## Analytic Model

Based on the review of the prior year's analytic approaches described in *Journal of Pediatric Psychology* articles, the most commonly performed analyses and modern counterparts were conducted to illustrate the different methods available for handling missing data, and the impact these analyses may have on results. Annotated syntax files are available at [http://quant.ku.edu/main/Supplemental\\_Materials](http://quant.ku.edu/main/Supplemental_Materials) for all analyses using IBM SPSS version 20 (and Amos for SEM) and R 2.15.2 (R Core Team, 2012). For analyses of multiple imputed data sets, we conducted 100 imputations in SPSS (for imputation in R, we used the Amelia II package; Honaker, King, & Blackwell, 2011). In addition to our online appendix, Graham (2012) includes an entire chapter on MI using SPSS, for the interested reader. These imputed data sets were then used in each of the three following analytic approaches: Ordinary least-squares (OLS) regression, path analysis, and SEM. Analysis models are depicted in Figure 2.

Regarding the assumptions behind various imputation procedures—which are discussed at length in Enders (2010) and Graham (2012)—our continuous variables showed no marked deviations from normality, and our categorical variables (e.g., sex, grade level) were handled





**Figure 2.** Three analysis models.

appropriately. SPSS offers two MI procedures, but the default (and more realistic choice) is to use chained equations, which use separate regression equations that are appropriate for imputing each variable based on its data type: Linear regression for continuous variables; binomial, multinomial, or ordinal logistic regression for categorical variables; and Poisson regression for count variables. The assumptions behind these equations are the same as in any analysis context (e.g., for linear regression, that the errors are independently normally distributed with the same variance), but there is no requirement that the raw data are multivariate normal. In R, Amelia uses a bootstrapped EM algorithm (EMB), which assumes the variables are multivariate normally distributed. However, it will make appropriate transformations to variables that are indicated to be nominal (in our case, gender) or ordinal (in our case, grade and class period, included in the imputation model but not the analysis model), so that these variables will be approximately continuous and normal (Honaker et al., 2011).

An OLS regression analysis was conducted for each outcome variable (i.e., task goal orientation and enjoyment). These regressions were conducted using averages of the available scale items. Then we used path analysis,

which allowed for the naturally occurring correlation between task goal orientation and enjoyment to be modeled, while simultaneously testing the predictive strength of the task-involving climate and gender for both outcomes: Task goal orientation and enjoyment. This benefit is a feature of path analysis and SEM regardless of whether any missing data are present. Finally, SEM analysis was conducted once using MI and once using FIML, which is the default approach to handling missing data by Amos. The default in the R package lavaan (Rosseel, 2012), used for path analysis and SEM, is listwise deletion, but FIML is easily requested. The auxiliary variables used to provide additional information for the FIML estimator were the following scale scores from the first wave of data collection: Caring climate, ego-involving climate, task goal orientation, ego goal orientation, and enjoyment. The scale scores were used from the first measurement period in order not to overwhelm the FIML estimator with auxiliary variables, yet to inform the process as much as possible (see Enders, 2010, regarding the use of auxiliary variables in FIML and imputation models). An additional benefit of conducting the analysis in an SEM framework was the ability to account for item measurement error (i.e., residual variance) when modeling latent variables.

Table II. *Parameter Estimates for Each Predictor and Outcome Using Four Missing-Data Methods*

| Dependent Variable  | Analysis Type | Missing Data Method | Task-Involving Climate |              |         | Gender (Male–Female) |              |         |
|---------------------|---------------|---------------------|------------------------|--------------|---------|----------------------|--------------|---------|
|                     |               |                     | <i>b</i>               | <i>SE(b)</i> | $\beta$ | <i>b</i>             | <i>SE(b)</i> | $\beta$ |
| Enjoyment of PE     | OLS           | LD <sup>a</sup>     | 0.288***               | 0.071        | 0.221   | –0.164               | 0.094        | –0.095  |
|                     | OLS           | PD <sup>a</sup>     | 0.279***               | 0.072        | 0.212   | –0.113               | 0.092        | –0.068  |
|                     | OLS           | MI                  | 0.142**                | 0.049        |         | –0.042               | 0.048        |         |
|                     | PA            | MI                  | 0.269***               | 0.005        | 0.205   | –0.094***            | 0.006        | –0.056  |
|                     | PA            | FIML                | 0.270***               | 0.067        | 0.218   | –0.101               | 0.084        | –0.059  |
|                     | SEM           | MI                  | 0.202***               | 0.004        | 0.198   | –0.141***            | 0.008        | –0.066  |
|                     | SEM           | FIML                | 0.223***               | 0.069        | 0.220   | –0.141               | 0.121        | –0.060  |
| Mastery orientation | OLS           | LD <sup>a</sup>     | 0.360***               | 0.067        | 0.288   | –0.136               | 0.089        | –0.082  |
|                     | OLS           | PD <sup>a</sup>     | 0.354***               | 0.068        | 0.279   | –0.171*              | 0.087        | –0.106  |
|                     | OLS           | MI                  | 0.155**                | 0.045        |         | –0.060               | 0.042        |         |
|                     | PA            | MI                  | 0.338***               | 0.005        | 0.266   | –0.154***            | 0.006        | –0.096  |
|                     | PA            | FIML                | 0.338***               | 0.063        | 0.282   | –0.156*              | 0.080        | –0.095  |
|                     | SEM           | MI                  | 0.285***               | 0.005        | 0.273   | –0.219***            | 0.008        | –0.101  |
|                     | SEM           | FIML                | 0.332***               | 0.071        | 0.310   | –0.221               | 0.120        | –0.100  |

Note. OLS = ordinary least-squares regression; PA = path analysis; SEM = structural equation modeling; LD = listwise deletion; PD = pairwise deletion; MI = multiple imputation; FIML = full-information maximum likelihood; *b* = unstandardized coefficient;  $\beta$  = standardized coefficient (interpreted as a partial correlation coefficient). Standardized coefficients unavailable from combined MI results of OLS in SPSS. Total *N* = 563.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001.

<sup>a</sup>LD and PD resulted in a sample size of *n* = 318 complete cases on all predictors and outcomes.

## Results and Discussion

The results of the example analyses were in line with AGPT. Specifically, the students' perceptions of a task-involving climate at the first measurement significantly predicted their later reported task goal orientation and enjoyment in PE (see results in Table II). Additionally, being a male was less predictive of assuming a task goal orientation and of reporting enjoyment in PE. Lastly, the students' self-reported enjoyment and task goal orientation were significantly and strongly correlated. The pattern of results is similar for both outcomes.

Table II makes clear not only the bias that is present when missing data are not handled properly but also the impact bias can have on substantive conclusions from the results. Listwise deletion (i.e., using only complete cases on all predictors and outcomes) resulted in inflated parameter and *SE* estimates for the male regression weight for both outcome variables. The impact was that these parameters were reported to be nonsignificant, reflecting the loss of power and precision from omitting cases from analysis. Similar results were found using pairwise deletion (i.e., using different subsamples of complete cases for each outcome's separate OLS regression). The results from these example analyses illustrate the inherent bias when listwise or pairwise deletion is used to analyze a potentially nonrepresentative subsample from the full sample.

Although the OLS regression conducted with MI had improved *SE* estimates, the model still did not account for

the strong correlation present between the two outcome variables (i.e., task goal orientation and enjoyment). Additionally, this analysis was still unable to account for measurement error. The benefit of this analysis and the other analyses conducted using the imputed data was that the relationships between all the variables collected could be used to inform the imputation process.

The path and SEM analyses conducted using FIML and the MI data illustrate how similar the results can be when using these two modern approaches to handling missing data. The parameter estimates for these two techniques with either the path analysis or the SEM analysis approach were usually identical to the second decimal place. The differences in results of these two approaches were with the *SE* estimates. The *SE* estimates from MI were lower than from FIML, which is understandable given the much larger number of variables available to inform the MI algorithm. This illustrates the strength of MI when used as intended, for large data sets (Rubin, 1987). If the estimation model had been more complex, the FIML estimation would have benefited from having a greater amount of information included in the model that could inform the estimation process. Alternatively, if the data set had not included so many other variables that were not included in the analysis or as auxiliary variables, then the *SE* estimates from FIML would have been closer to those from MI. The path analysis technique modeled all four manifest variables' (e.g., scale scores) relationships simultaneously, including the strong correlation between the outcome

variables. This resulted in the reported relationships better representing the interrelationships that occur in reality. The SEM analysis had the additional benefit of being able to account for the item-level measurement error for the three constructs measured by scales (i.e., task-involving climate, task orientation, and enjoyment).

## Conclusions

In this article, we highlight the joys of missing data. Rather than ruing the existence of missing data, embracing the power and efficiency of modern treatments for missing data allows researchers to regain the upper hand when data go missing. Embracing modern treatments for missing data also allows researchers to go beyond the traditional paradigms for data collection by incorporating intentionally missing data designs. Such designs are woefully underutilized by researchers today. The unfounded skepticism that restricts their use is based in the lack of exposure to the statistical theory that underlies modern missing data treatments. Today's computers and software can readily and easily handle both planned and unplanned missing data. Any resistance to using these modern approaches is misguided and fundamentally hinders the progress of scientific inquiry. Carefully planning for the unplanned missing information and using the power of intentionally missing data will lead to far greater generalizability for future research than the mountains of biased and underpowered research of preceding centuries. We hereby offer a loud and confident battle cry for a paradigm shift in our science.

## Funding

This research was funded in part by a grant (NSF0066969) from the National Science Foundation (T. Little & W. Wu, co-PIs), and in part by the Center for Research Methods and Data Analysis (T. Little, Director), University of Kansas.

*Conflicts of interest:* None declared.

## References

- Brown, T. C., Fry, M. D., & Little, T. D. (2013). The psychometric properties of the perceived motivational climate in exercise questionnaire. *Measurement in Physical Education & Exercise Science*, 17, 22–39.
- 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.
- Cushing, C. C., & Steele, R. G. (2012). Psychometric properties of *sizing me up* in a community sample of 4<sup>th</sup> and 5<sup>th</sup> grade students with overweight and obesity. *Journal of Pediatric Psychology*, 37, 1012–1022.
- Division of Adolescent and School Health, & CDC. (2010). Trends in the Prevalence of Physical Activity National YRBS: 1991-2009. HHS (Ed.), *National Trends: YRBS 1991-2009*. Atlanta, GA: HHS & CDC.
- Duda, J. (1989). Relationship between task and ego orientation and the perceived purpose of sport among high school athletes. *Journal of Sport & Exercise Psychology*, 11, 318–335.
- Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology*, 84, 290–299.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.
- Goldfield, G. S., Adamo, K. B., Rutherford, J., & Murray, M. (2012). The effects of aerobic exercise on psychosocial functioning of adolescents who are overweight or obese. *Journal of Pediatric Psychology*, 37, 1136–1147.
- Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer. doi:10.1007/978-1-46144018-5\_1
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really necessary? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8, 206–213.
- Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in the analysis of change. In L. M. Collins, & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 335–353). Washington, DC: American Psychological Association.
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, 11, 323–343.
- Harel, O., Stratton, J., & Aseltine, R. (2011). *Designed missingness to better estimate efficacy of behavioral studies* (Technical Report 11-15). Storrs, CT: Department of Statistics, University of Connecticut.
- Herge, W. M., Streisand, R., Chen, R., Holmes, C., Kumar, A., & Mackey, E. R. (2012). Family and youth factors associated with health beliefs and health outcomes in youth with Type 1 diabetes. *Journal of Pediatric Psychology*, 37, 980–989.

- Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45, 1–47. Retrieved from <http://www.jstatsoft.org/v45/i07>.
- Huddleston, H., Fry, M. D., & Brown, T. C. (2012). Corporate fitness members' perceptions of the environment and their intrinsic motivation. *Revista de Psicología del Deporte*, 21, 15–23.
- Jia, F., Moore, E. W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (2013). *Planned missing data design on small sample size: How small is too small?* Manuscript submitted for publication.
- Karlsen, C. W., Leist-Haynes, S., Smith, M., Faith, M. A., Elkin, T. D., & Megason, G. (2012). Examination of risk and resiliency in a pediatric sickle cell disease population using the Psychosocial Assessment Tool 2.0. *Journal of Pediatric Psychology*, 37, 1031–1040.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford.
- Loth, K. A., Mond, J., Wall, M., & Neumark-Sztainer, D. (2010). Weight status and emotional well-being: Longitudinal findings from Project EAT. *Journal of Pediatric Psychology*, 36, 216–225.
- Moore, E. W. G. (2011). *Planned missing data survey design*. Retrieved from [crmda.ku.edu](http://crmda.ku.edu).
- Moran, L. M., Taylor, H. G., Rusin, J., Bangert, B., Dietrich, A., Nuss, K. E., . . . Yeates, K. O. (2012). Quality of life in pediatric mild traumatic brain injury and its relationship to postconcussive symptoms. *Journal of Pediatric Psychology*, 37, 736–744.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.
- Nicholls, J. G. (1989). *The competitive ethos and democratic education*. Cambridge, MA: Harvard University Press.
- R Core Team. (2012). *R: A language and environment for statistical computing (version 2.15.2) [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>.
- 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
- Roberts, G. C., & Treasure, D. C. (2012). *Advances in motivation in sport and exercise* (3rd ed.). Champaign, IL: Human Kinetics.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. Available from <http://lavaan.ugent.be/>.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581–592.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York, NY: Chapman Hall.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177.
- Seifriz, J. J., Duda, J. L., & Chi, L. (1992). The relationship of perceived motivational climate to intrinsic motivation and beliefs about success in basketball. *Journal of Sport & Exercise Psychology*, 14, 375–391.
- Smith, R. E., Smoll, F. L., & Cumming, S. P. (2009). Motivational climate and changes in young athletes' achievement goal orientations. *Motivation & Emotion*, 33, 173–183.
- van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: CRC Press.
- Walling, M. D., & Duda, J. L. (1995). Goals and their associations with beliefs about success in and perceptions of the purposes of physical education. *Journal of Teaching in Physical Education*, 14, 140–156.
- Weigand, D., & Burton, D. (2002). Manipulating achievement motivation in physical education by manipulating the motivational climate. *European Journal of Sport Science*, 2, 1–14.