ijbd

International Journal of Behavioral Development
1–11

DOI: 10.1177/0165025415618275
ijbd.sagepub.com
SAGE

Attrition in developmental psychology: A review of modern missing data reporting and practices

Jody S Nicholson,[1] Pascal R Deboeck,[2] and Waylon Howard[3]

Abstract

Inherent in applied developmental sciences is the threat to validity and generalizability due to missing data as a result of participant drop-out. The current paper provides an overview of how attrition should be reported, which tests can examine the potential of bias due to attrition (e.g., *t*-tests, logistic regression, Little's MCAR test, sensitivity analysis), and how it is best corrected through modern missing data analyses. To amend this discussion of best practices in managing and reporting attrition, an assessment of how developmental sciences currently handle attrition was conducted. Longitudinal studies ($n = 541$) published from 2009–2012 in major developmental journals were reviewed for attrition reporting practices and how authors handled missing data based on recommendations in the *Publication Manual of the American Psychological Association* (APA, 2010). Results suggest attrition reporting is not following APA recommendations, quality of reporting did not improve since the APA publication, and a low proportion of authors provided sufficient information to convey that data properly met the MAR assumption. An example based on simulated data demonstrates bias that may result from various missing data mechanisms in longitudinal data, the utility of auxiliary variables for the MAR assumption, and the need for viewing missingness along a continuum from MAR to MNAR.

Keywords

Longitudinal data, attrition, journal article reporting standards, missing data

Research questions in the psychological sciences often demand repeated measures over time to assess and predict developmental trends, typically spanning particular ages or developmental periods (Nesselroade & Ram, 2004). Repeatedly assessing participants over time involves addressing many challenges that may limit the completeness of observed data. Missing data occur among some variables within a particular measurement occasion (i.e., participant non-response) or among all variables across entire measurement occasions (i.e., attrition; Jeličić, Phelps, & Lerner, 2009). Participants dropping out of a study has been described as "virtually ubiquitous" in longitudinal research (Graham, 2009, p. 567), and often selective due to the process that produced missing data (Little & Rubin, 1989). Selective attrition may degrade the generalizability of statistical inferences by destroying the observed sampling distribution (Rubin, 1976). For example, the sample may change due to a certain type of individual selecting out of longitudinal follow-up (e.g., younger, less educated, more anxious) resulting in the inflation or suppression of true effects (Graham, 2009). When data are missing in this manner, statistical methods for longitudinal data analysis may lead to parameter bias, weaken generalizability, and compromise research resources (Collins, Schafer, & Kam, 2001; Graham, 2009; Schafer & Graham, 2002), making it critical for researchers to properly examine and report why data is missing.

Modern missing data methods such as multiple imputation (MI), which fills in each missing value with a set of m plausible values, and full-information maximum likelihood (FIML), which produces parameter estimates given all the observed data, are tools at the disposal of developmental scientists to try and counter the detrimental effects of missing data (Enders, 2010). When properly employed, these approaches increase the effective sample size relative to ad hoc approaches (e.g., single, nearest-neighbor, or mean substitution, pairwise or listwise deletion, stochastic regression imputation), reduce standard errors, and minimize bias related to participant non-response and attrition (Little & Rubin, 2002).

Importantly, the effectiveness of the MI and FIML procedures depends on the degree to which missingness (i.e., whether the data are missing or observed) is related to other variables in the data set and that the techniques are properly applied (Graham, 2012). That is, MI and FIML only yield unbiased parameter estimates when all variables that are causes or correlates of missingness are included in the missing data handling procedure (Collins et al., 2001; Enders, 2010). Furthermore, handling missing data may require specific procedures to retain any special features of the data. For example, failure to specifically consider interactive effects in the imputation phase, which may require a particular application of MI, may bias parameter estimates in the subsequent analysis phase (Enders, Baraldi, & Cham, 2014; Enders & Gottschall, 2011). Therefore, researchers are not only charged to report information about missing data and examine the mechanism by which data are missing, but also to properly employ these procedures to address missingness

[1] University of North Florida, USA
[2] University of Kansas, USA
[3] Seattle Children's Hospital Research Institute, WA, USA

Corresponding author:
Jody S Nicholson, University of North Florida, 1 UNF Drive, Jacksonville, FL 32224, USA.
Email: jody.nicholson8@gmail.com

(see APA, 2010; Graham, 2009, 2012; Hansen, Collins, Malotte, Johnson, & Fielding, 1985; Jeličić et al., 2009).

The purpose of the current paper is to highlight the importance of proper reporting, testing for mechanisms of missingness, and employing modern missing data techniques. First, the concerns of attrition and missingness in the context of the *MAR assumption* are presented, along with suggestions on how to report attrition. Furthermore, modern missing data techniques that can reduce bias potential due to missingness are discussed (i.e., MI and FIML). The presence of both of these, attrition reporting and modern missing data techniques, were examined in current reporting practices across a four-year span in two top developmental journals following the most recent publication of the *American Psychological Association Manual* (APA, 2010). To highlight the importance of proper implementation of modern missing data techniques, a series of Monte Carlo simulations compared the performance of various missing data handling strategies under different conditions of missingness and with the presence of auxiliary variables.

Attrition and Missingness

The extent to which missing data are capable of introducing bias depends on why the data are missing. Terms such as 'attrition' and 'dropout' are used to describe patterns of missing data over time rather than the reason for missingness. Because causes and correlates of missing data may be unknown, a distinction is made between three general processes that may cause missing data (and bias) in a given study including: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR; Little, 1992; Little & Rubin, 1989; Rubin, 1976).

The first mechanism, MCAR, describes the cases when the propensity for missingness is unrelated to both observed and unobserved explanatory variables (Little & Rubin, 1989). Given there are no variables measured or unmeasured that explain why missing values occur, missing observations can be regarded as the result of a completely random process. Most analytic techniques will be unbiased when missingness is due to this mechanism; substitution of mean values is one notable exception that can still produce biased inferences even with MCAR data (Enders, 2010). Though the MCAR mechanism will not typically result in biased parameter estimates, modern missing data handling procedures (e.g., MI and FIML) are often preferred over deletion techniques (e.g., listwise, pairwise deletion) to maximize statistical power (Enders, 2010; Graham, 2009). An important consideration in applied research is that the MCAR mechanism is unlikely unless implemented by design (Graham, 2009; Rubin, 1976).

In applied research most missingness is related to one or more observed or unobserved causes. When missingness is related to a variable that has been observed or is known, such as the length of time in a study, missing data are said to be MAR (Rubin, 1976). Contrary to its name, missing observations are not actually "missing at random." Rather, the missing observations are considered random after controlling for the variable on which the propensity for missingness depends (Little & Rubin, 2002). The MAR mechanism is more plausible than the MCAR mechanism because missingness is not assumed to be completely independent of all possible observed and unobserved variables. Rather, MAR assumes the propensity for missingness can be explained by other variables in the dataset. For example, participants with low socio-economic status (SES) may disproportionately leave longitudinal studies early, which (if ignored) may bias inferences towards representing primarily individuals with higher SES; as SES is predictive of the propensity for an observation to be missing, this variable can be used to correct the bias due to disproportional dropout and yield unbiased inferences relevant to the entire sample (Enders, 2010).

When the causes or correlates of missingness are unknown, unmeasured, or are simply not used to inform the missing data handling procedure, the data are said to be MNAR. In an intervention aiming to reduce depressive symptomatology, participants with early success may disproportionally begin to leave the study. If these participants leave before their depression is assessed again, the remaining participants with higher rates of depression may give the impression that the intervention had been unsuccessful. In this case, the variable on which the missing observations depend—depressive symptomatology—is unobserved for the individuals with missing observations. Because no observed variable predicting the missingness is available, analyses with MNAR data are unable to account for the process that generated missingness. The consequence is biased parameter estimates and inferences even when using modern missing data methods (Enders, 2010; Little, 1992; Little & Rubin, 2002). The effectiveness of algorithms handling missing data, such as MI and FIML, therefore depends on our ability to identify causes and correlates of missing observations. These causes and correlates allow researchers to assume the data are MAR, which is termed the *MAR assumption* (Enders, 2010; Graham, 2009; Little & Rubin, 1989).

In practice, the MCAR, MAR, and MNAR mechanism are not distinct entities; rather, most missing data situations involve a combination of causes based on the specific variables selected from the dataset (Graham, 2012). That is, the degree to which missing data may bias parameter estimates depends on the proportion of missing observations that can be described as an MNAR process. Unfortunately, researchers are typically unaware of the relative contribution of the MNAR mechanism and the MAR assumption (Enders, 2010; Peugh & Enders, 2004), which is evident when published research does not adequately report on missing data (Maloney, Johnson, & Zellmer-Bruhn, 2010; Jeličić et al., 2009).

Attrition: Reporting and Analytic Strategies for Examining MAR

Due to the important role of missing data in longitudinal research, best practice recommendations for reporting and handling missing data are essential to adequately interpret research findings (Graham, 2009). Clear and comprehensive reporting on missing data allows the reader to be aware of the percentage of the sample lost to attrition, pattern of attrition across time, and whether the final sample differs from the original sample (Hansen et al., 1985). Improper or vague reporting may make it difficult for the reader to assess potential threats to internal and external validity or could imply that the authors did not properly handle missingness analytically. For instance, incomplete descriptions of missing data limit a reader from assessing potential bias and threats to generalizability (Graham, 2009). Further, many popular statistical software programs default to deletion techniques such as pairwise and listwise deletion, which require the assumption data are MCAR and may dramatically decrease statistical power (Enders, 2010).

Current reporting standards in the psychological sciences have been outlined by the Journal Article Reporting Standards (JARS) group (APA, 2008), whose recommendations for reporting standards have been included in the appendix of the 6th edition of the Publication Manual of the American Psychological Association (APA, 2010, pp. 245–253). The missing data reporting standards outlined by JARS provides suggestions for missing data reporting in general and specific guidelines for addressing attrition (APA, 2008). According to JARS, when describing missing data in general one must report: (1) the percentage of the sample approached that participated in a study and the intended vs. actual sample size in the method section; (2) the flow through each stage of a study, percentage/frequency of missing data, causes of missingness, and method for addressing missing data in the results section; and (3) interpretation of results taking into account sources of bias and any threats to internal validity or generalizability during the discussion section. When addressing attrition specifically, the JARS report suggests the amount and possible cause of participant attrition must be reported in the method section. While this suggestion is made in the context of quasi-experimental designs, it is further specified that this standard could be adapted to other research designs (p. 845).

Research methodologists offer additional recommendations to those offered in the JARS report for describing attrition (Graham, 2009; Hansen et al., 1985; Jeličić et al., 2009). For example, in addition to adequately reporting attrition rates, methodologists suggest evaluating variables that are possible causes or correlates of attrition. This can be accomplished by comparing participants that remained in the study to those that dropped out across demographic and dependent/outcome variables using *t*-tests, Little's MCAR test, or logistic regression. When data are likely MNAR, methods such as sensitivity analyses can be used to gauge the stability of inferences (for in-depth discussion of how to perform these tests and present the results, see Carpenter & Kenward, 2013; Enders, 2010; Graham, 2012).

t-Tests. One of the most common methods for characterizing attrition mechanisms is to perform *t*-tests comparing the mean values on one or more covariates, predictor, or outcome variables for subjects who were and were not missing at the initial assessment (Dixon & Brown, 1983; Enders, 2010). To test whether missingness is related to observed or unobserved causes, a dummy variable is created based on participants having complete or missing data on a variable or having left or remained in the study (e.g., 0, incomplete data/left the study data; 1, complete data/remained in the study). Then, a series of independent samples *t*-tests are used to indicate whether the dummy variable leads to significant average differences among other variables of interest. If data are completely lacking in significant mean differences between study-related variables, the data could possibly be MCAR. While no mean differences may be apparent, however, these tests do not allow one to conclusively state that the data are MCAR. This is the case because unobserved variables could exist that may explain the missing patterns (Enders, 2010). Similarly, if only demographic variables are compared using *t*-tests, the researcher is not truly testing the MAR assumption, as it is necessary by the MAR definition to examine if patterns of missingess are related to the study's outcome variables. If a difference is found, including these variables as auxiliary variables in the subsequent imputation or analysis model may more reasonably approximate the MAR assumption because these variables help explain why the data are missing (Enders, 2010).

The primary limitations of this approach are those that are associated with *t*-tests. First, these tests are intended for use with normally distributed variables, or reasonable approximations thereof as *t*-tests are robust to slightly skewed data. These tests are not reasonable for testing when there is a substantial deviation from normality. When attrition is related to data that conforms to a multinomial distribution (e.g., nominal data), reasonable alternatives for trustworthy p-values are the chi-square test or a version of the generalized linear model. Second, the significance of these tests will depend on statistical power and are subject to alpha inflation related to multiple-comparison tests (Little, 1988). Additionally, these tests cannot readily include interactions between variables (e.g., anxiety predicts attrition within a particular treatment condition). Finally, it should also be noted that using *t*-tests presumes that a difference in means will occur when data are not related to a MCAR process. Enders (2010) highlights that MAR and MNAR processes can result in missing and non-missing groups with equal means on a variable of interest, but different variability, which the *t*-test approach would be unable to detect.

Little's MCAR test. Similar to the independent *t*-test approach, Little's MCAR test distinguishes whether missing observations are MCAR (null hypothesis) or dependent on other variables (MAR; alternative hypothesis). However, rather than conducting a large number of *t*-tests, Little (1988) provides a single, global test statistic by simultaneously testing for mean differences across all variables. This is accomplished by sorting the data into groups based on whether the variables are observed or missing across cases to detect missing data patterns. Variable means are then calculated for each subgroup and across groups for a particular variable. Little's MCAR test uses a chi-squared statistic to summarize the standardized mean difference between each variable's subgroup means and the overall mean. The mean differences are standardized because there are potentially large discrepancies in variable variances that would complicate any omnibus measure of mean differences (Enders, 2010). A significant chi-squared statistic would suggest a significant deviation in mean differences on one or more variables between subgroups and consequently a rejection of the null hypothesis that the data are MCAR.

There are several limitations of this method for assessing the cause of missingness. First, Little's MCAR test functions only as an omnibus evaluation of the MCAR mechanism. Therefore, unlike the independent *t*-test approach which evaluates individual predictors, a significant omnibus result does not provide potentially useful information regarding which variables are causes or correlates of missingness and should, therefore, be used as auxiliary variables (Enders, 2010). Additionally, methodologists caution against Type II errors that result in a failure to detect a MAR process due to low statistical power associated with Little's MCAR test (Enders, 2010; Lin, 2009; Thoemmes & Enders, 2007). Moreover, apparent support for the null hypothesis can result when missing observations are MNAR, not just MCAR.

Logistic regression. If whether or not participants left the study is utilized as a binary outcome, logistic regression can be used such that variables that potentially explain the missing status are included as predictors (Ridout & Diggle, 1991). Consequently, logistic regression can be both a global test of MAR as well as a test of individual predictors while controlling for the effects of other predictors (multivariate missingness; Jeličić et al., 2009).

The purpose in applying logistic regression would be to identify covariates and interactions between covariates that are predictive of missing observations. Identification of predictive covariates allows researchers to reasonably assume that data are MAR. The lack of predictive covariates would leave researchers in one of three conditions: (1) there is insufficient power to detect relevant covariates (i.e., Type II error); (2) data are MNAR and consequently analyses may not be trustworthy; or (3) data are MCAR (see Curran, Bacchi, Schmitz, Molenberghs, & Sylvester, 1998; Fielding, Fayers, & Ramsay, 2009). The first two conditions would make a researcher wary to proceed, and many consider the third condition untenable. Thus, a non-significant logistic regression may fail to indicate important covariates that may be influential for decreasing bias when missingness occurs.

Sensitivity analysis. Missing data analyses are complicated by the presence of MNAR data, and even modern missing data methods such as MI or FIML are not able to produce unbiased inferences when data are missing in this manner. Unfortunately, attrition that is dependent on the scores that would have been observed may be commonplace. While one cannot be assured of producing unbiased inferences, one thing that can be done if MNAR is a likely missing data mechanism is to examine how the pattern of results obtained would change under different, plausible MNAR scenarios. This approach is called sensitivity analysis.

It is possible to perform sensitivity analysis with both MI and FIML, although it is perhaps easier to imagine implementing in MI. For example, if one suspected an MNAR case where the most depressed people tended to drop from a study, one could impute values of depression. Then, on individuals missing depression scores, imputed values could be reduced by some amount (e.g., 1/2 standard deviation) to examine how the pattern of inferences change under such a condition. Sensitivity analysis does not provide a prescribed set of rules as to the alternative scenarios that should be tested, nor is this analysis a statistical test of MNAR. Under plausible MNAR conditions, however, the consistency of inferences from results can be assessed to provide insight on their trustworthiness (see Carpenter & Kenward, 2013; Enders, 2010, sections 10.16–10.20).

Techniques for Handling Missing Data

While logistic regression and comparison of individuals with missing and non-missing observations offers some possibility of characterizing the attrition mechanism, this does not serve as a tool to adequately address missing data. Advances in quantitative psychology and software have made the treatment of missing data easier than it has been in the past, as the tools for making valid inferences from incomplete data are more readily available to researchers. Two practical and common ways of addressing MAR data are Full Information Maximum Likelihood (FIML) and Multiple Imputation (MI; Enders, 2010; Graham, 2009). Both are based on statistical theory and still require an assumption of MAR and proper implementation in order for the resulting parameter estimates to be unbiased (Foster, Fang, & Conduct Problems Prevention Research Group, 2004; Schafer & Graham, 2002).

FIML is commonly employed in most structural equation modeling (SEM) packages (e.g., AMOS and Mplus; Arbuckle, 2005; Muthén & Muthén, 2008), although there is wide variability in how FIML is integrated as a default and some packages could still employ listwise deletion in some cases (e.g., missingness on covariates; Enders, 2010; Hox & Roberts, 2011). FIML is considered a model-based approach because missing data are handled within a single, iterative step where missing values are not imputed (Enders, 2010). Rather, model parameters and standard errors are directly estimated from the observed data of each individual case. Specifically, individual or case specific likelihood functions are generated and summed to obtain an overall discrepancy function (Enders, 2010). Said differently, FIML directly estimates a statistic and its associated standard error from the data. Consequently, FIML is considered a direct approach and is often referred to in the literature as direct-maximum likelihood (direct ML; Enders, 2010).

To clarify this approach, consider that FIML is similar to pairwise deletion in that both procedures use all the available data on a case wise basis (Enders, 2001, 2010). However, FIML does not simply perform a direct calculation of each covariance estimate from the available data. Rather, the FIML procedure relies on a probability density function to iteratively maximize the likelihood of the estimates employing, "... all of the information of the observed data, including mean and variance for the missing portions of a variable, given the observed portion(s) of other variables," (Wothke, 2000, p. 3). Consequently, FIML, unlike pairwise deletion, uses the appropriate sample size for standard error estimates (Allison, 2003; Enders, 2001) and is less likely to produce a non-positive definite matrix (Wothke, 1993). To not influence the interpretation of model parameters, auxiliary variables are incorporated into the analysis model of FIML (i.e., saturated correlates approach; see Graham, 2003).

While FIML is based on estimating the population parameters of hypothetical models, MI is a data-based approach that focuses on the observations. MI can be described as relating to three sequential steps including: (1) an imputation step; (2) an analysis step; and (3) a pooling step (Little & Rubin, 1989, 2002). The idea of the imputation step (i.e., estimating the imputation model) is to substitute each missing value with a regression-based estimate of that value using analysis and auxiliary variables (Enders, 2010). This is done more than once (e.g., 100 times; Enders, 2010; Graham, 2012) because a single imputation does not capture the uncertainty surrounding any particular imputed value. This process creates multiple copies of the original data set with each copy of the data set slightly varying on each of the imputed values (Enders, 2010).

Next, the analysis step (i.e., fitting the analysis model) is used to individually analyze a particular statistical model of interest for each of the imputed data sets using standard techniques for complete data. The analysis model is often different from the more general imputation model (Enders, 2010). For example, auxiliary variables are included in the imputation model but not in the analysis model. Finally, the pooling step is used to combine the results from all these repeated analyses into a single result (i.e., using Rubin's Rules; see Rubin, 1987).

Prior examination of how often imputation methods, as well as proper examination of mechanism for missingness, has suggested major need for improvement in our field. Jeličić et al. (2009) provided a review of the state of missing data from 100 randomly selected articles published between 2000 and 2006 across Child Development, Developmental Psychology, and Journal of Research on Adolescence. In the study, 81 of the 100 papers reported attrition specifically, with only 43% of these studies investigating differences between those who left the studies and those who remained (i.e., mechanism for missingness). Listwise deletion was noted for 47 studies, which equated to 82% of studies who had missing data with which to contend; seven studies used FIML and two studies used MI. The authors concluded that FIML and MI methods were not increasing across the period studied and speculated this may be due to the knowledge, software, and time required for implementing these procedures.

The current study re-examines reporting trends almost a decade after this review and immediately following the publication of the improved reporting standard. Specifically, the prevalence of attrition reporting, the type of checks made for mechanism for missingness, and prevalence of imputation methods were examined. To augment the importance of proper reporting and handling, a demonstration of the consequences of improper handling of missing data using simulated data is presented. The simulation compared how parameter estimates of data generated under different conditions of missingness compared with the parameter values used to generate the data (i.e., true values).

Attrition Reporting Review Method

Articles published in two major journals in developmental psychology, Child Development and Developmental Psychology, were reviewed across volumes published in 2009 and 2012 if the study was longitudinal (i.e., at least two repeated measures) and the authors reported on a research question of a longitudinal nature. Daily diary or burst data studies were excluded because the intensity of intra-individual studies may present different burden for attrition than longitudinal studies with less concentrated assessment time points. Based on this criteria, a total of 541 articles were identified; Child Development published 244 longitudinal studies and Developmental Psychology published 297 longitudinal studies from 2009–2012.

The first author and a graduate or undergraduate research assistant coded each article for information related to the quality of attrition reporting, investigation into the mechanism for missingness, and how missing data was handled. For quality attrition reporting, it was noted if "attrition" or "retention" were mentioned in the article via an electronic search and if a paragraph, flowchart, or section was devoted specifically to attrition and missing data. To investigate how authors evaluated if their data was MAR, it was noted if a comparison was made between those retained and those who left the study and whether comparisons were made on demographic or study variables, or both. The articles of authors who specifically referred to data in the context of MAR or MCAR were further examined for what type of test they used (*t*-test, Little's MCAR, logistic regression, specificity analysis), which imputation method they conducted, and if they used auxiliary variables. The rate of imputation strategies (i.e., FIML and MI) was also investigated in the entire sample.

Coding was conservative such that missing data values were assigned if vague wording in attrition reporting made interpretation difficult. For example, authors could report "no difference was found between those who remained in the study and those lost to attrition," but not report the exact variables that were compared. Therefore, whether they investigated study-related variables, or just demographic variables, could not be definitively assessed. In this case, a codebook devised a protocol for how to handle vague wording. Quality in coding was established by double coding 50% of the articles. Inter-rater reliability was assessed using Krippendorff's α (Hayes & Krippendorff, 2007); discrepancies between coders were resolved by the first author as a master coder. Besides maintaining an alpha-level above 0.85 for each variable coded, a bootstrapping sampling distribution of alpha was generated using 2000 draws to produce a 95% confidence intervals for α_{true}. The lower confidence interval for each item was maintained above 0.64; the probability that the reliability was less than the required minimum value of 0.80 was low (i.e., q maintained, on average, below 0.20).

Attrition Reporting Review Results

Table 1 provides results from attrition reporting across the four-year period, by journal, and for each year individually. From January 2009 to December 2012, authors did not demonstrate an increase in the prevalence of mentioning attrition in their papers, with just around half of authors mentioning attrition (46.8%) and a smaller portion providing more information on attrition in their method sections through paragraphs (26.1%) or sections and flowcharts (10.2%) elaborating on attrition. There was no systematic increase in attrition reporting or in the use of imputation methods across the four years reviewed and some decrease between years; for example, from 2009 to 2012, attrition reporting only increased by 0.1% and the use of FIML decreased by 0.1%.

From 2009–2012, a bit less than half of all authors made a comparison between demographic or study-related variables for those who left and those who remained in the study (46.9%), fewer examined both together (29.4%), or study-related variables specifically (33.3%), which has important implications for whether the MAR assumption is met. Out of the 541 articles, only 99 specifically mentioned mechanisms of missingness within the context of MAR or MCAR; a little more than half of these articles used a statistical test to examine if their data could be considered MCAR or MAR prior to imputation (51.5%; $n = 51$); of these studies, 22 utilized Little's MCAR test. Almost a third of the studies discussed the necessary assumption of MAR prior to imputation without actually formally testing the assumption. A small number of these studies that discussed mechanisms of missingness employed auxiliary variables (8.0%).

Simulated Data Example

Simulated data will help contextualize the importance of proper handling of missing data. As opposed to real data, simulated data allows for a demonstration of the effects of different mechanisms of missingness. In addition, the properties of the data are known, including the true relations between variables, making it possible to assess how close differing analyses of multiple samples come to representing the true values in the population. The general logic of a simulation study is to create a population with known characteristics, from which samples are drawn (i.e., paralleling data collection). The samples are then analyzed, and the compiled results are compared back to the population characteristics. The results are typically a summary of a large number of samples, so as to remove the variability inherent to any single sample. The focus of this simulation was to examine different types and combinations of missing data, and highlight the necessity of including auxiliary variables to meet the MAR assumption. Labels have been given to the variables to help readers contextualize the simulation, although the specific names are inconsequential.

Data were simulated to represent a population in which individuals were randomly assigned to one of three intervention groups (waitlist, counseling, home visits). The primary dependent variable (depression) was simulated for five occasions ($t = 0, 1, 2, 3, 4$) for each individual, representing a longitudinal study. An additional dichotomous variable (gender) and a normally distributed continuous variable (anxiety) were simulated at the initial occasion. Each individual was given a unique slope equal to a linear decrease in depression depending on their intervention group (waitlist = 0.0, counseling = 0.2, and home visits = 0.1 points over the five observations), plus their initial anxiety score ($\sim N(0,1)$), plus error (mean zero, variance equal to

Table 1. Attrition Reporting and Examination of Mechanism of Missingness in Child Development and Developmental Psychology from 2009–2013.

Attrition reporting	2009–2012	Comparison between journals DP vs. CD*	2009	2010	2011	2012
Sample size	541	—	124	109	150	158
Attrition mentioned	46.8% ($n = 253$)	48.1% vs. 45.1% ($p > 0.05$)	49.2% ($n = 61$)	42.2% ($n = 46$)	50.7% ($n = 76$)	44.3% ($n = 70$)
Elaboration provided in attrition paragraph	26.1% ($n = 141$)	26.9% vs. 25.0% ($p > 0.05$)	25.0% ($n = 31$)	30.3% ($n = 33$)	26.0% ($n = 39$)	24.1% ($n = 38$)
Flowchart or section devoted to attrition	10.2% ($n = 55$)	10.1% vs. 10.2% ($p > 0.05$)	12.1% ($n = 15$)	16.5% ($n = 18$)	6.0% ($n = 9$)	8.2% ($n = 13$)
Mechanisms for missingness						
Comparison made between those who left and those retained	46.4% ($n = 251$)	50.5% vs. 41.4% ($p = 0.02$)	49.2% ($n = 61$)	41.3% ($n = 45$)	54.0% ($n = 81$)	40.5% ($n = 64$)
Significant difference found in comparison	24.6% ($n = 133$)	60.7% vs. 41.2% ($p = 0.002$)	27.4% ($n = 34$)	20.2% ($n = 22$)	30.7% ($n = 46$)	19.6% ($n = 31$)
Demographic variables compared	40.1% ($n = 217$)	46.1% vs. 32.8% ($p = 0.001$)	47.6% ($n = 59$)	35.8% ($n = 39$)	43.3% ($n = 65$)	34.2% ($n = 54$)
Study-related variables compared	33.3% ($n = 180$)	36.7% vs. 29.1% ($p = 0.04$)	37.9% ($n = 47$)	25.7% ($n = 28$)	40.0% ($n = 60$)	28.5% ($n = 45$)
Demographic and study-related variables compared	29.4% ($n = 159$)	33.3% vs. 24.6% ($p = 0.02$)	32.3% ($n = 40$)	23.9% ($n = 26$)	36.0% ($n = 54$)	24.7% ($n = 39$)
Discussion of mechanism for missingness (MAR, MCAR, or MNAR)	18.3% ($n = 99$)	23.9% vs. 11.5% ($p < 0.001$)	21.8% ($n = 27$)	14.7% ($n = 16$)	19.3% ($n = 29$)	17.1% ($n = 27$)
Approach for handling missing data						
Multiple imputation	14.6% ($n = 79$)	13.5% vs. 16.0% ($p > 0.05$)	14.5% ($n = 18$)	12.8% ($n = 14$)	14.0% ($n = 21$)	16.5% ($n = 26$)
Full information maximum likelihood	41.2% ($n = 223$)	46.5% vs. 34.8% ($p = 0.004$)	41.9% ($n = 52$)	35.8% ($n = 39$)	44.4% ($n = 66$)	41.8% ($n = 66$)

Note. Comparison between journals provides the proportion of articles in each journal that showed evidence of the attrition/missing data item of interest in the table row, Developmental Psychology (DP) vs. Child Development (CD) using a chi-square goodness of fit. Developmental Psychology demonstrated better attrition reporting practices based on these comparisons.

20% of depression scores). Consequently, in this simulated population, more anxious individuals were more likely to show an increase in depression and those receiving counseling or home visits were more likely to show a decrease in depression. Gender was simulated using a random draw from a Bernoulli distribution with equal probability of success and failure such that the variable was unrelated to all other variables. Two-thousand samples, each consisting of 200 individuals, were drawn from the population.

Each of the 2000 samples was then degraded with missing observations in four different ways, resulting in 8000 additional samples. The missing observations were created to mimic attrition, and consequently once a missing value occurred all subsequent observations were also missing. All four missing data conditions resulted in 25% of the observations being labeled as missing. The four missing data conditions were:

A. Completely random attrition (MCAR)
B. An MAR condition (MAR) where the missingness mechanism is observed and high anxiety individuals were more likely to leave the study
C. An MNAR condition (MNAR) where individuals with the highest depression scores in the last wave were most likely to leave the study. In this condition, the missingness mechanism (highest depression scores) is not observed for the people with missing values
D. A mixed combination (MIX) with 2.5% MCAR, 15% MAR, and 7.5% MNAR

Across 2009–2012, FIML was most commonly used (see Table 1), which is automatically conducted in the Mplus (Muthén & Muthén, 2008) and AMOS software (Arbuckle, 2005). Because this was the most common missing data method employed, in the present simulation each of the samples were analyzed using FIML in Mplus (Muthén & Muthén, 2008). A latent growth curve model was used to describe changes in depression, allowing each individual to have a unique intercept and slope. The latent intercept and slope were predicted by two dummy coded variables: one indicating the presence of the "Counseling" condition, and the other a "Home Visit" condition; the average slope corresponded to the case when both dummy coded variables were equal to zero, that is the "Control" condition.

Analysis of all of the samples was conducted four different ways, mimicking ways in which a researcher might choose to analyze a particular sample. These analysis methods were: 1) analysis using listwise deletion (i.e., the default in many programs); 2) analysis using FIML and not specifying auxiliary variables ("No AUX"); 3) analysis using FIML and specifying an auxiliary

Table 2. Summary of Modeling Using Full Information Maximum Likelihood.

	Average Slope			Counseling Effect			Home Visits Effect		
	(Control Condition) = 0.000			(Relative to Control) = −0.200			(Relative to Control) = −0.100		
True Value	Mean Est.	Standard Deviation	Mean S.E.	Mean Est.	Standard Deviation	Mean S.E.	Mean Est.	Standard Deviation	Mean S.E.
Complete data*	0.001	0.134	0.137	−0.204	0.190	0.193	−0.102	0.196	0.193
Listwise Deletion									
MCAR	−0.006	0.309	0.287	−0.195	0.440	0.408	−0.100	0.431	0.407
MAR	−0.131	0.166	0.165	−0.207	0.232	0.234	−0.102	0.236	0.234
MNAR	−0.691	0.137	0.129	−0.097	0.177	0.178	−0.044	0.175	0.180
MIX	−0.378	0.154	0.155	−0.137	0.221	0.216	−0.065	0.220	0.217
No AUX (FIML)									
MCAR	−0.002	0.155	0.158	−0.201	0.223	0.223	−0.100	0.227	0.223
MAR	−0.107	0.163	0.161	−0.207	0.226	0.228	−0.102	0.230	0.228
MNAR	−0.588	0.139	0.128	−0.104	0.176	0.174	−0.047	0.173	0.176
MIX	−0.281	0.146	0.145	−0.153	0.206	0.202	−0.071	0.204	0.203
Unrelated AUX (FIML)									
MCAR	−0.002	0.156	0.158	−0.201	0.223	0.223	−0.100	0.227	0.223
MAR	−0.107	0.163	0.161	−0.207	0.227	0.228	−0.102	0.230	0.228
MNAR	−0.588	0.139	0.128	−0.104	0.177	0.174	−0.047	0.174	0.176
MIX	−0.282	0.147	0.145	−0.152	0.206	0.202	−0.070	0.205	0.203
Related AUX (FIML)									
MCAR	−0.001	0.148	0.151	−0.201	0.215	0.214	−0.100	0.217	0.214
MAR	−0.010	0.141	0.144	−0.205	0.200	0.203	−0.102	0.206	0.203
MNAR	−0.212	0.137	0.132	−0.169	0.180	0.179	−0.080	0.182	0.180
MIX	−0.076	0.141	0.139	−0.188	0.197	0.194	−0.092	0.199	0.194

Note: The mean estimates (Mean Est.), standard deviation of the mean estimates (Standard Deviation) and mean of the standard error estimates (Mean S.E.) across all samples are reported. AUX represents the use of Anxiety as an auxiliary variable in the analysis procedure. * With complete data, the same parameter estimates are found, regardless of the analysis procedure (listwise deletion, not using auxiliary variables, using an unrelated auxiliary variable, or using a related auxiliary variable). Full results of the simulation analysis are available on the journals' supplemental materials on their website.

variable that is unrelated to the attrition, (i.e. gender; Unrelated AUX); and 4) analysis using FIML and specifying an auxiliary variable that is related to the attrition in some of the missing data conditions (i.e. anxiety; Related AUX). Aside from use as an auxiliary variable, gender and anxiety were not otherwise used in the models. For each of the four analyses, the average estimates and standard deviations of the results for three parameters are presented: Average Slope (corresponding to the "Control" condition), Effect of Counseling, and the Effect of Home Visits. The values of the latter two parameters represent a difference in slopes relative to the "Control" condition.

The present simulation is based on the analysis of samples drawn from a single population. Typically in simulation studies the population is varied to examine how patterns of results change in response to variation in parameters (e.g., sample size, percentage of missing observations). Using one population, as in the present paper, limits the inferences that can be made. The patterns of results presented, however, are not specific to the values selected in the simulation, and the specific values are inconsequential for understanding the effects and analysis of data with attrition. The direction and magnitude of the bias and the degree of increase in variance estimates would be different for differing populations, but the present pattern of results is consistent with those reported in the missing data literature.

Data Simulation Results

In Table 2, the "True Value" line represents the true effects in the population that would ideally be recovered by the analyses. Analyses that, on average, produce deviations from these slope values indicate a systematic under- or over-estimation of the effect of treatment on the depression slopes (i.e., a presence of bias). The row labelled, "Complete Data," of Table 2 presents the average estimates and standard deviations when the data sets with no missing data are analyzed. The subsequent sections correspond to the results for the analysis of the MCAR, MAR, MNAR, and MIX datasets with listwise deletion, no auxiliary variables, an unrelated (to missingess) auxiliary variable (gender), and an auxiliary variable related to attrition (anxiety).

Listwise deletion tends to produce biased estimates of the slopes; with the exception of MCAR, the "Control" condition suggests a large decrease in depression over time when none is actually present (true slope equals zero). With no auxiliary variables that are predictive of the attrition, although FIML is used, the estimates in "No AUX" show similar bias as the model using listwise deletion. The inclusion of auxiliary variables that are not related to the missingness, such a gender ("Unrelated AUX"), also produced similar results.

When an auxiliary variable appropriate for predicting the missingness is included, as in the "Related AUX" condition, the MAR assumption is met for the data labeled "MAR." For the MAR data in "Related AUX" the estimated parameters are much less biased and nearly identical to the true values used to generate the data. As anxiety was correlated with the outcome of depression in this example, some mitigation of the bias occurred even when data are MNAR. The results for MNAR in part "Related AUX" are closer to the true values than with listwise deletion in the "No AUX" conditions, although some bias still persists. The mixed condition, which was a 2.5% – 15% – 7.5% mix of MCAR-MAR-MNAR, also shows substantial reduction in bias, although some bias still persist due to the MNAR attrition.

How missing data mechanisms and missing data analysis interact to affect the standard errors of estimates were not considered, which could have important effects on power. The benefits of modern missing data approaches for increasing power, over methods such as listwise deletion, have been demonstrated even in the case when data are MCAR (Enders, 2010; Graham, 2009); some evidence of this is apparent in Table 2, as the mean standard errors for MCAR data with listwise deletion tend to be much larger than the subsequent analyses with MCAR data. Full consideration of the effects on standard errors of the interaction between missing data mechanisms and missing data analysis, however, is beyond the scope of the present simulation.

Discussion

The recommendations made by the JARS group in 2008, and included in the 6th edition of the APA manual (available in July 2009), were not purported to be ground-breaking, but rather a clarification and expansion of practices that should already exist (APA, 2008). The committee of journal editors who originally devised the recommendations encouraged empirically based reviews of standards as "Not unlike the issues many psychologists study, the proposal and adoption of reporting standards is itself an intervention (p. 850)." Improvements in reporting and handling of missing data would be expected in the years following this *intervention*, but articles published in major developmental journals in a four-year period after these recommendations were published do not convey progress. In comparison to a review conducted at the beginning of the millennium, there was actually a decrease in attrition reporting (Jeličić et al., 2009), 81% vs. 46.8%, but about the same percentage of articles comparing those who left the study and those who remained (i.e., mechanism for missingness), 43% vs. 46.4%.

Authors may not completely understand the reason behind testing for differences between the baseline and final samples in the context of mechanisms for missingness. Sometimes anecdotal reports were provided (e.g., "participants were mainly lost due to scheduling conflicts or relocation") as rationale for why there was no systematic pattern in missingness. In fact, participants' scheduling conflicts or relocation could separate participants who leave and those who stay in a meaningful way and be indicative of data being MNAR if related to a variable of interest (e.g., marital conflict or academic success). Furthermore, articles also failed to adequately examine mechanisms for missingness by only examining differences in demographic variables and missing data. Analyses between demographic and study-related variables and attrition are needed to clearly document if data is likely to meet the MAR assumption (Graham, 2009; Hansen et al., 1985; Jeličić et al., 2009). The proportion of articles that put the comparison of those who left the study and those who remained in context of mechanisms for missingness (i.e., actually stating MAR or MCAR) was even smaller than those who compared those who left the study to those who remained on baseline variables. Even among these articles, many did not utilize an appropriate statistical test and only reported an assumption of MAR in order to meet the requirements of utilizing FIML or MI.

Best practices would be to properly report the pattern of missing data, ideally by devoting a paragraph or section to attrition and missing data in the method or results sections. Moreover, authors should discuss the mechanism for missingness, provide statistical evidence for the data meeting the MAR assumption, and discuss the potential bias and generalizability of the results in the discussion (see examples in Collins, Martino, Elliott, & Miu, 2011; Conduct Problems Prevention Research Group, 2011; Forget-Dubois et al., 2009; Garstein et al., 2010). Authors should also acknowledge in the limitation section of their paper that unmeasured variables could be related to the non-response indicator (resulting in data actually being MNAR), especially if there's a high rate of missing data (greater than 5%; Graham, 2009). Variables related to missingness should be included in the analysis or imputation model as the inclusion of auxiliary variables further reduces risk for biased results (Graham & Donaldson, 1993). In the articles reviewed, the inclusion of auxiliary variables was seldom done or reported (see examples in Belsky, Schlomer, & Ellis, 2012; Benner & Graham, 2009; Karna et al., 2011; Ponitz, McClelland, Matthews, & Morrison, 2009), perhaps due to a possible misperception that use of FIML, in itself, is sufficient to address missing data. As the simulation conveys, ignoring attrition results in biased estimates, especially in conditions of MNAR or mixed condition, even with FIML.

The simulation highlighted that including auxiliary variables, when related to the missingness (i.e., correlates or predictors of attrition or a dependent/outcome variable that is measured at a prior outcome), helps derive estimates that are closer to the true value even when missing data mechanisms may be complicated and mixed. The mixed condition provides a more realistic way for researchers to begin thinking of missing observations, rather than consider data as satisfying any single mechanism for missingness (Graham, 2009, 2012). Moreover, including an auxiliary variable irrelevant to data being MAR (i.e., gender) does not hurt the estimates under conditions of MCAR, MAR, and Mixed, suggesting it is probably better to include, rather than not, when the number of variables in question is small. The degree to which the effects of attrition can be mitigated through the use of FIML will therefore depend on the degree to which data can be shown to be MAR, and the inclusion of auxiliary variables will substantially reduce bias.

In practice, it is likely that the choice of auxiliary variables for a particular study are often based on convenience (i.e., some extra variables in the data set) or tradition (i.e., a set of auxiliary variables were used previously; Kreuter & Olson, 2011). It is preferable to choose auxiliary variables by intentionally collecting variables that are theoretically important to the response variables as well as to other key variables that might also contain missingness (Little & Vartivarian, 2005). Often auxiliary variables are selected that simply correlate highly with the analysis model variables (Enders, 2010; Graham, 2012). However, it may be difficult to establish a rule of thumb because auxiliary variables that correlate highly with outcome variables might not also relate highly to predictor variables. Even in the case where there is strong theory and the researcher has a deep understanding of the processes that might cause missing data, it seems unlikely that such information would lead to the selection of a specific set of auxiliary variables that correlate highly across all key analysis variables (Kreuter & Olson, 2011).

As Collins, Schafer, and Kam (2001) note, an inclusive strategy is recommended to reduce the chance of inadvertently omitting an important cause of missingness while allowing for noticeable gains in terms of increased power and reduced bias. Recent research, however, suggests including too many auxiliary variables with low sample sizes and weak associations among auxiliary and analysis variables may lead to bias and lower power (Hardt, Herke, & Leonhart, 2012; Thoemmes & Rose, 2014). The Principal Components Analysis auxiliary variable approach may be preferable in these circumstances (Howard, Rhemtulla, & Little, 2015). Regardless of the approach or number of auxiliary variables retained, the idea is to acknowledge that the MAR assumption is not automatically met and that an attempt should be made to more reasonably approximate it.

Even if missing data could be demonstrated to be MCAR, using imputation methods and including auxiliary variables will result in smaller standard errors, even though there is no difference in model estimates (Table 2). This suggests that even researchers assuming MCAR attrition may benefit from using modern methods for handling missing data, as the smaller standard deviation of parameter estimates across samples could result in more power. Authors could also improve their power and reduce bias by including indicators of non-responsiveness in their models, like the number of phone calls required to complete an interview, the length of previous interviews, or a baseline question asking, "What is the likelihood you will be able to complete this study or participate in future interviews?" (see Foster et al., 2004; Schafer & Graham, 2002).

The current paper has important implications on reporting and the consequences that can occur when not appropriately handling attrition under the conditions of MCAR or MAR. Articles were examined which were published in the four-year period after which the JARS recommendations were made; it is feasible due to the review process that some articles reviewed could have been accepted before these recommendations and the date of publication does not represent the data of acceptance. Furthermore, suggestions made by the JARS group on reporting standards in other areas may have improved practices in the field; the current paper just focused on one specific topic by focusing on missing data due to attrition. It is also important to acknowledge that the paper focused on comparing mean differences in individuals who remain in and leave a study (i.e., *t*-tests, Little's MCAR test, logistic regression, sensitivity analysis), when patterns of missingness may be predicted by variability between subjects even when mean differences are not evident (Raykov, Lichtenberg, & Paulson, 2012). The simulation used FIML as an imputation method to provide evidence for properly correcting for attrition using auxiliary variables. This method was chosen given it is the most used imputation method as evidenced by the review of the literature; MI would also have been a useful approach for correcting for missing data in this example. The interested reader is referred to other simulation studies that convey the detrimental impact of attrition reporting on parameter estimation (e.g., Demirtas & Schafer, 2003; Newman, 2003; Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006).

Conclusion

The evidence provided from a review of the field demonstrates that attrition reporting and the handling of missing data still has room for growth in the developmental scientists. To improve these reporting practices, researchers should turn to suggestions provided by methodologists in our field (Graham, 2009; Hansen et al., 1985) to supplement recommendations put in place by the JARS committee in the recent APA manual. It is also recommended to use preventative action during data collection to reduce the chances of attrition (Jeličić et al., 2009). Properly testing for mechanisms of missingness is imperative, and the article has outlined multiple ways authors can investigate whether their sample is biased due to missing data. When it is established that data can be assumed to be MAR, modern missing data techniques and auxiliary variables can be utilized to minimize biasing effects previously assumed, even with substantial attrition (Elobeid et al., 2009; Graham, 2009; Graham & Donaldson, 1993; Jeličić et al., 2009; Schafer & Graham, 2002).

Acknowledgements

The authors acknowledge Scott E Maxwell for early mentoring, guidance and encouragement on this project and would like to thank the undergraduate and graduate coders who provided countless hours towards the review of the literature presented in this article.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*, 545.

APA. (2010). *Publication manual of the American Psychological Association*. 6th ed. Washington, DC: American Psychological Association.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*, 839–851. DOI: 10.1037/0003-066X.63.9.839.

Arbuckle, J. L. (2005). *Amos 6.0 user's guide*. Chicago, IL: SPSS Incorporated.

Belsky, J., Schlomer, G. L., & Ellis, B. J. (2012). Beyond cumulative risk: Distinguishing harshness and unpredictability as determinants of parenting and early life history strategy. *Developmental Psychology*, *48*(3), 662–673. DOI: 10.1037/a0024454.

Benner, A. D., & Graham, S. (2009). The transition to high school as a developmental process among multiethnic urban youth. *Child Development*, *80*(2), 356–376. DOI: 10.1111/j.1467-8624.2009.01265.x.

Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*. Chichester, UK: John Wiley & Sons.

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*(4), 330–351.

Collins, R. L., Martino, S. C., Elliott, M. N., & Miu, A. (2011). Relationships between adolescent sexual outcomes and exposure to sex in media: Robustness to propensity-based analysis. *Developmental Psychology*, *47*(2), 585–591. DOI: http://dx.doi.org/10.1037/a0022563.

Conduct Problems Prevention Research Group (2011). The effects of the Fast Track preventive intervention on the development of conduct disorder across childhood. *Child Development*, *82*(1), 331–345. DOI: 10.1111/j.1467-8624.2010.01558.x.

Curran, D., Bacchi, M., Hsu Schmitz, S. F., Molenberghs, G., & Sylvester, R. J. (1998). Identifying the types of missingness in quality of life data from clinical trials. *Statistics in Medicine*, *17*, 739–756. DOI: 10.1016/s0197-2456(97)91175-2.

Demirtas, H., & Schafer, J. L. (2003). On the performance of random--coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in medicine*, *22*(16), 2553–2575.

Dixon, W. J. & Brown M. B. (Eds.) (1983). *BMDP statistical software manual*. Berkeley, CA: University of California Press.

Elobeid, M. A., Padilla, M. A., McVie, T., Thomas, O., Brock, D. W., Musser, B., . . . Allison, D. B. (2009). Missing data in randomized clinical trials for weight loss: Scope of the problem, state of the field, and performance of statistical methods. *PLoS One*, *4*, e6624. DOI: 10.1371/journal.pone.0006624.g003.

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, *8*(1), 128–141.

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

Enders, C. K., Baraldi, A. N., & Cham, H. (2014). Estimating interaction effects with incomplete predictor variables. *Psychological Methods*, *19*(1), 39–55.

Enders, C. K., & Gottschall, A. C. (2011). Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling*, *18*(1), 35–54.

Fielding, S., Fayers, P. M., & Ramsay, C. R. (2009). Investigating the missing data mechanism in quality of life outcomes: A comparison of approaches. *Health and Quality of Life Outcomes*, *7*(57). DOI: 10.1186/1477-7525-7-57.

Forget-Dubois, N., Dionne, G., Lemelin, J., Perusse, D., Tremblay, R. E., & Boivin, M. (2009). Early child language mediates the relation between home environment and school readiness. *Child Development*, *80*(3), 736–749. DOI: 10.1111/j.1467-8624.2009.01294.x.

Foster, E. M., & Fang, G. Y., & Conduct Problems Prevention Research Group. (2004). Alternative methods for handling attrition: An illustration using data from the fast track evaluation. *Evaluation Review*, *28*(5), 434–464. DOI: 10.1177/0193841X04264662.

Garstein, M. A., Bridgett, D. J., Rothbart, M. K., Robertson, C., Iddins, E., Ramsay, K., & Schlect, S. (2009). A latent growth examination of fear development in infancy: Contributions of maternal depression and the risk for toddler anxiety. *Developmental Psychology*, *46*(3), 651–668. DOI: http://dx.doi.org/10.1037/a0018898.

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

Graham, J. W. (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. (2012). *Missing data: Analysis and design*. New York: Springer.

Graham, J. W., & Donaldson, S. I. (1993). Evaluating interventions with differential attribution: The importance of non-response mechanisms and use of follow-up data. *Journal of Applied Psychology*, *78*, 119–128.

Hansen, W. B., Collins, L. M., Malotte, C. K., Johnson, C. A., & Fielding, J. E. (1985). Attrition in prevention research. *Journal of Behavioral Medicine*, *8*(3), 261–275. DOI: 10.1007/BF00870313.

Hardt, J., Herke, M., & Leonhart, R. (2012). Auxiliary variables in multiple imputation in regression with missing X: A warning against including too many in small sample research. *BMC Medical Research Methodology*, *12*(1), 184–197. DOI: 10.1186/1471-2288-12-184.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89.

Howard, W. J., Rhemtulla, M., & Little, T. D. (2015). Using principal components as auxiliary variables in missing data estimation. *Multivariate Behavioral Research 50*(3). 285–299. DOI: 10.1080/00273171.2014.999267.

Hox, J., & Roberts J. K. (Eds.) (2011). *Handbook of advanced multilevel analysis*. New York, NY: Routledge.

Jeličić, H., Phelps, E., & Lerner, R. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*, 1195–1199. DOI: 10.1037/a0015665.

Karna, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa antibullying program: Grades 4-6. *Child Development*, *82*(1), 311–330. DOI: 10.1111/j.1467-8624.2010.01557.x.

Kreuter, F., & Olson, K. (2011). Multiple auxiliary variables in nonresponse adjustment. *Sociological Methods & Research*, *40*(2), 311–332.

Lin, J. (2009). A probability based test for missing completely at random data patterns: Extending Little's missing completely at random test (Master's thesis). Retrieved from http://www.academia.edu

Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. DOI: 10.2307/2290157.

Little, R. J. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, *87*(420), 1227–1237. DOI: 10.1080/01621459.1992.10476282.

Little, R. J., & Rubin, D. B. (1989). The analysis of social-science data with missing values. *Sociological Methods & Research*, *18*, 292–326. DOI: 10.1177/0049124189018002004.

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. 2nd ed. Hoboken, NJ: Wiley-Interscience.

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology*, *31*(2), 161–168.

Maloney, M. M., Johnson, S. G., & Zellmer-Bruhn, M. E. (2010). Assessing group-level constructs under missing data conditions: A Monte Carlo simulation. *Small Group Research*, *41*(3), 281–307.

Muthén, L. K., & Muthén, B. (2008). *Mplus* (version 5.1). Los Angeles, CA: Muthén & Muthén.

Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, *1*, 9–29. DOI: 10.1080/15427609.2004.9683328.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, *6*(3), 328–362.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525–556.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, *45*(3), 605–619. DOI: 10.1037/a0015365.

Raykov, T., Lichtenberg, P. A., & Paulson, D. (2012). Mechanism in incomplete data sets: A multiple testing approach. *Structural Equation Modeling*, *19*, 399–408. DOI: 10.1080/10705511.2012.687660.

Ridout, M. S., & Diggle, P. J. (1991). Testing for random dropouts in repeated measurement data. *Biometrics*, *47*(4), 1617–1621. DOI: 10.2307/2531777.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. DOI: 10.1037/1082-989X.7.2.147.

Thoemmes, F., & Enders, C. K. (2007, April). *A structural equation model for testing whether data are missing completely at random*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Thoemmes, F., & Rose, N. (2014). A cautious note on auxiliary variables that can increase bias in missing data problems. *Multivariate Behavioral Research*, *49*, 443–459.

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, *76*(12), 1049–1064.

Wothke, W. (1993). *Nonpositive definite matrices in structural modeling*. Newbury Park, CA: Sage Publications Inc.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 219–240). Mahwah, NJ: Erlbaum.