Missing Data

Article *in* Organizational Research Methods · September 2014

DOI: 10.1177/1094428114548590

1 author:

Daniel A. Newman
University of Illinois, Urbana-Champaign

Article

Organizational Research Methods
2014, Vol. 17(4) 372-411

DOI: 10.1177/1094428114548590
orm.sagepub.com

Missing Data: Five Practical Guidelines

Daniel A. Newman[1]

Abstract

Missing data (a) reside at three *missing data levels of analysis* (item-, construct-, and person-level), (b) arise from three *missing data mechanisms* (missing completely at random, missing at random, and missing not at random) that range from completely random to systematic missingness, (c) can engender two *missing data problems* (biased parameter estimates and inaccurate hypothesis tests/inaccurate standard errors/low power), and (d) mandate a choice from among several *missing data treatments* (listwise deletion, pairwise deletion, single imputation, maximum likelihood, and multiple imputation). Whereas all missing data treatments are imperfect and are rooted in particular statistical assumptions, some missing data treatments are worse than others, on average (i.e., they lead to more bias in parameter estimates and less accurate hypothesis tests). Social scientists still routinely choose the more biased and error-prone techniques (listwise and pairwise deletion), likely due to poor familiarity with and misconceptions about the less biased/less error-prone techniques (maximum likelihood and multiple imputation). The current user-friendly review provides five easy-to-understand practical guidelines, with the goal of reducing missing data bias and error in the reporting of research results. Syntax is provided for correlation, multiple regression, and structural equation modeling with missing data.

Keywords

missing data, full information maximum likelihood (FIML), EM algorithm, multiple imputation, R syntax/R code

Statisticians (e.g., Little & Rubin, 2002; Schafer & Graham, 2002) recommend a few treatments for handling missing data (i.e., maximum likelihood and multiple imputation techniques), which are routinely ignored by researchers in psychology and management. Disregarding the advice of statisticians in this way is sometimes relatively harmless, but is sometimes quite harmful, depending on the amount of missing data, the pattern of missing data, and whether the data are missing in a strongly systematic (vs. weakly systematic or random) fashion. In order to advance statistical best practice while optimizing the trade-off between ease of implementation and likely degree of missing data bias and error, I offer five

[1]University of Illinois at Urbana-Champaign, Champaign, IL, USA

Corresponding Author:
Daniel A. Newman, Department of Psychology and School of Labor & Employment Relations, University of Illinois, 603 E. Daniel St., MC-716, Champaign, IL 61820, USA.
Email: d5n@uiuc.edu

practical guidelines and a decision tree for handling missing data. These guidelines address item-level, construct-level, and person-level missing data. These practical guidelines, if followed, would constitute a major step forward for rooting out missing data bias and error but would only require complex missing data treatments to be used in those cases where they are likely to yield the biggest payoffs.

The current presentation of missing data problems and methodological options partly recapitulates previous reviews of the topic (see Allison, 2002; Enders, 2001b, 2010; Graham, 2009; Little & Rubin, 1987, 2002; McKnight, McKnight, Sidani, & Figueredo, 2007; Newman, 2003, 2009; Schafer & Graham, 2002). Where I differ from these previous treatments, however, is in offering a pragmatic decision tree (Figure 1, described in the sections that follow) designed to aid in the selection of appropriate missing data techniques to address item-level missingness, construct-level missingness, and person-level missingness. Because I lack the space to thoroughly review all the core aspects of missing data analysis here, however, the current work can be thought of as a companion piece to any of the previously cited reviews.

The current article is organized into three sections. First, I describe three *missing data levels* (item-level, construct-level, and person-level missingness), three *missing data mechanisms* (missing completely at random [MCAR], missing at random [MAR], and missing not at random [MNAR]; Rubin, 1976), two major *missing data problems* (parameter bias and inferential error), and five widely available *missing data treatments*: listwise deletion, pairwise deletion, single imputation/ad hoc approaches, maximum likelihood (ML) approaches (full information maximum likelihood [FIML] and the expectation-maximization [EM] algorithm), and multiple imputation. Second, I enumerate several missing data considerations that must precede data analysis (e.g., the partial avoidability of missing data, and the basic fact that incomplete data analysis always requires a choice from among several imperfect alternatives—abstinence is not an option). Third and most important, I describe five practical guidelines for handling missing data. These guidelines are:

(1) Use all the available data (e.g., do not use listwise deletion).
(2) Do not use single imputation.
(3) For construct-level missingness that exceeds 10% of the sample, ML and multiple imputation (MI) techniques should be used under a strategy that includes auxiliary variables and any hypothesized interaction terms as part of the imputation/estimation model.
(4) For item-level missingness, one item is enough to represent a construct (i.e., do not discard a participant's responses simply because he or she failed to complete some of the items from a multi-item scale).
(5) For person-level missingness that yields a response rate below 30%, simple missing data sensitivity analyses should be conducted (also see Figure 1).

Following these five practical guidelines should curtail a large portion of the avoidable missing data bias and error in the fields of psychology and management. An appendix is also presented that gives syntax (in R, SAS, and LISREL) to aid in implementing state-of-the-art missing data routines (ML and MI). For readers who are in a hurry, I advise skipping down to the section titled "Five Practical Guidelines." For those who want to understand more of the bases and terminology underlying the guidelines, I offer the intervening sections. In the next section, I begin by defining *missing data*.

What Are Missing Data, and How Much Should We Care?

The term *missing data* is defined here as a statistical problem characterized by an incomplete data matrix that results when one or more individuals in a sampling frame do not respond to one or more survey items (Newman, 2009). Most missing data are due to survey nonresponse, which can vary from an intentional decision (discarding a survey or skipping sensitive items) to a rather

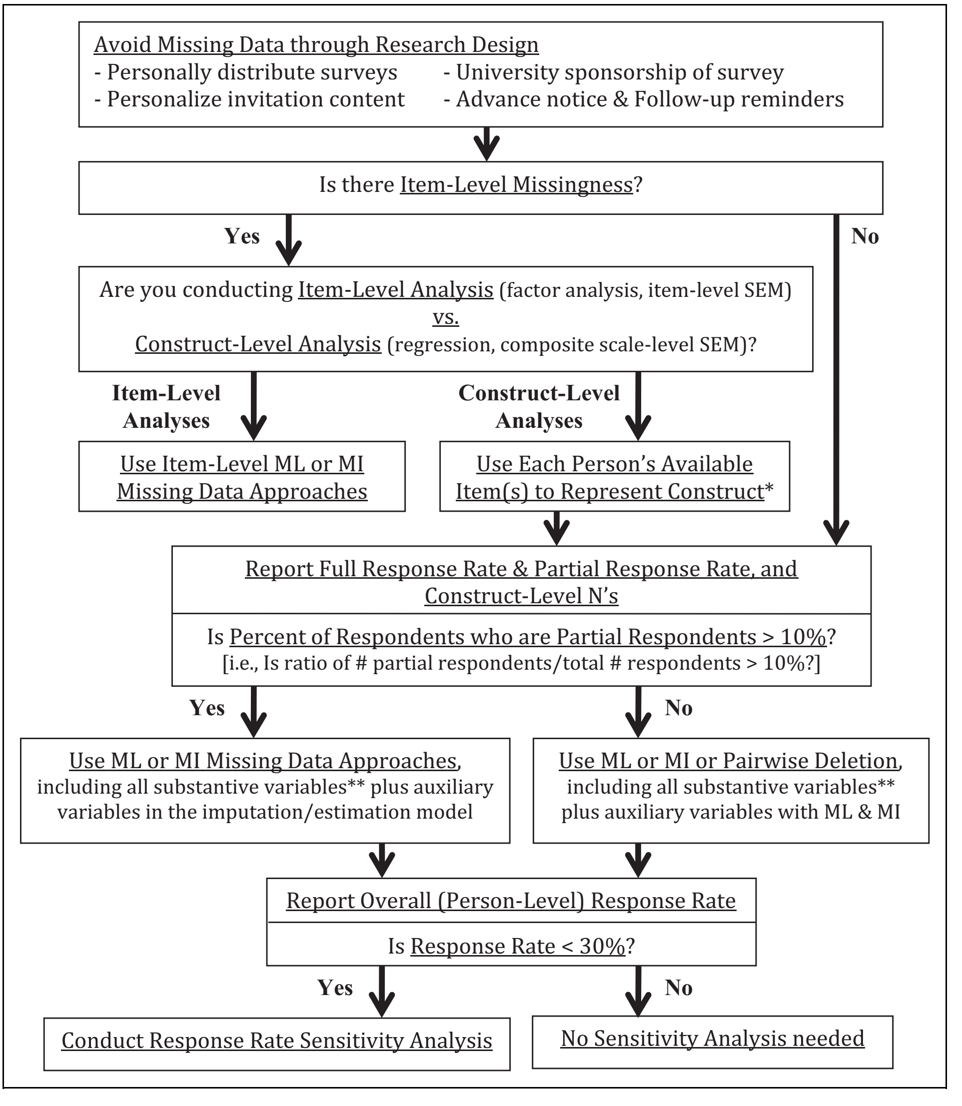


Figure 1. Decision tree for choosing missing data treatments.
Note: Listwise deletion and single imputation are never recommended. ML = maximum likelihood missing data routines (i.e., full information maximum likelihood [FIML] or expectation-maximization [EM] algorithm). MI = multiple imputation.
*This rule applies even if the respondent only answers one item from a multi-item scale (also see Appendix B).
**Includes hypothesized interaction terms.

unintentional act (forgetting a survey or being too busy to attend to a survey; Rogelberg et al., 2003); but missing data can also arise from technical errors on the part of the researcher or equipment (online survey programming errors or computer malfunction).

Three Levels of Missing Data: Item-, Construct-, and Person-Levels

It is helpful to think of missing data as corresponding to three levels of analysis: item-level missingness, construct-level missingness, and person-level missingness (see Figure 2). *Item-*

Complete Data					Incomplete Data				
	X_1	X_2	X_3	Y		X_1	X_2	X_3	Y
*person*1	3	2	2	1	*person*1	3	.	2	1
*person*2	2	2	2	3	*person*2	.	.	.	3
*person*3	4	3	4	4	*person*3	4	3	4	4
*person*4	3	3	3	3	*person*4	.	.	.	.
*person*5	2	3	2	3	*person*5	2	3	2	3
*person*6	4	4	4	3	*person*6	.	.	.	.
*person*7	4	4	3	5	*person*7	4	4	3	5
*person*8	3	2	3	5	*person*8	3	2	.	5
*person*9	5	5	4	5	*person*9	5	5	4	.
*person*10	2	3	2	3	*person*10	2	3	2	3

Three Levels of Missingness

- Item-level missingness
- Construct-level missingness
- Person-level missingness

Figure 2. Three levels of missing data: Example (10-person sampling frame, three-item measure of construct *X*, single-item measure of construct *Y*).

level missingness occurs when the respondent leaves a few items blank on a multi-item scale (i.e., the respondent answers only j out of J possible items, where $1 \leq j < J$). Items can be skipped for a variety of reasons (e.g., items deal with sensitive information such as drug use or employee theft, items are at the end of a survey and respondents quit before getting to these items, items have unusual wording or are otherwise confusing, or respondents are skipping items quasi-randomly). *Construct-level missingness* occurs when the respondent answers zero items from a scale (i.e., omitting an entire scale or an entire construct). *Person-level missingness* involves failure by an individual to respond to any part of the survey.

I note that the levels of missingness are nested, such that item-level missingness can aggregate into construct-level missingness (i.e., when an individual fails to respond to all of the items on a multi-item scale), and construct-level missingness can aggregate into person-level missingness (i.e., when a person fails to respond to all of the constructs on a survey). One advantage of distinguishing the three levels of missingness (item-, construct-, and person-level) is that the choice of a missing data treatment can depend on which level of missing data you have, as discussed in the following sections (e.g., see Table 1). Generally speaking, person-level missingness is far more problematic (i.e., more difficult to address) than either item-level or construct-level missingness, because with person-level missingness the researcher often possesses no relevant information about the nonrespondent that could be used to improve estimation and reduce missing data bias and error.

At this point, I also note that the notion of construct-level missingness can be used to sort the individuals in the sampling frame[1] into three mutually exhaustive categories: full respondents, partial respondents, and nonrespondents.

Full respondents – individuals who responded to every single construct on the survey.
Partial respondents – individuals who responded to part of the survey (i.e., more than zero but fewer than all constructs on the survey),
Nonrespondents – individuals who did not respond to any constructs on the survey.

To restate, partial respondents are individuals with construct-level missingness, whereas full respondents are individuals with no construct-level missingness.[2] I also point out that person-level missingness determines the *nonresponse rate*, which is equal to 1.0 minus the response rate.

Table 1. Three Levels of Missing Data and their Corresponding Missing Data Techniques.

Level of Missing Data	Recommended Missing Data Technique	Favorable Condition for Technique
Item level	Use each person's $\text{mean}_{(\text{across available items})}$ to represent the construct.	Parallel items[a]
Construct level	Use maximum likelihood (ML) or multiple imputation (MI), with auxiliary variables.	Missing at random (MAR) mechanism (probability of missingness is correlated with observed variables) or missing completely at random (MCAR) mechanism (completely random missingness)
Person level (i.e., as reflected in response rate)	Use sensitivity analysis.	Data are available from previous studies that compare respondents to nonrespondents on the constructs of interest (e.g., $r_{miss,x}$ can be estimated)

Note: Table adapted from Newman (2009).
[a]See Appendix B.

Three Mechanisms of Missing Data: Random Missingness (MCAR) and Systematic Missingness (MAR and MNAR)

Data can be missing randomly or systematically. According to Rubin's (1976) typology, there are three missing data mechanisms (Little and Rubin, 1987; Schafer & Graham, 2002):

MCAR (*missing completely at random*) – the probability that a variable value is missing does not depend on the observed data values nor on the missing data values [i.e., p(missing|complete data) = p(missing)]. The missingness pattern results from a process completely unrelated to the variables in one's analyses, or from a completely random process (similar to flipping a coin or rolling a die).

MAR (*"missing at random"*) – the probability that a variable value is missing partly depends on other data that are observed in the dataset, but does not depend on any of the values that are missing [i.e., p(missing|complete data) = p(missing|observed data)].

MNAR (*missing not at random*) – the probability that a variable value is missing depends on the missing data values themselves [i.e., p(missing|complete data) ≠ p(missing|observed data)].

Of the aforementioned missing data mechanisms, one is random (i.e., the MCAR mechanism), and the other two are systematic (i.e., the MAR mechanism and the MNAR mechanism). I highlight the seemingly odd labeling of the MAR mechanism. Despite being referred to as *missing at random*, MAR is actually a *systematic* missing data mechanism (the MAR label is confusing and stems from the unintuitive way statisticians [versus social scientists] use the word *random*).

To better understand the three missing data mechanisms, it is useful to borrow an example from Schafer and Graham (2002; see Little & Rubin, 1987). Imagine two variables X and Y, where some of the data on Y are missing. Now imagine a dummy variable $miss_{(y)}$, which is coded as 0 when Y is observed and coded as 1 when Y is missing. Under MCAR, $miss_{(y)}$ is not related to Y or to X. Under MAR, $miss_{(y)}$ is related to X (i.e., one can predict whether Y is missing based on observed values of X), but $miss_{(y)}$ is not related to Y after X is controlled. Under MNAR, $miss_{(y)}$ is related to Y itself (i.e., related to the missing values of Y), even after X is controlled (see Figure 3).

It is often impossible in practice to determine whether data are MNAR, because doing so would require comparing observed Y values to missing Y values, and the researcher does not have access to the missing Y values. Generally speaking, the point of delineating the three missing data mechanisms

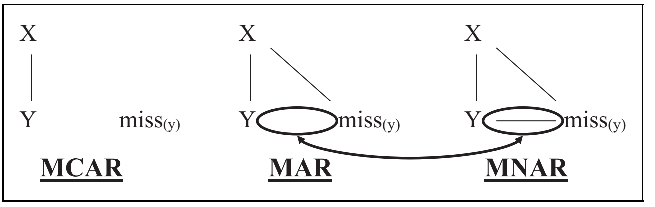


Figure 3. Three missing data mechanisms (MCAR, MAR, MNAR) and the continuum between MAR and MNAR.

Note: Adapted from Schafer and Graham (2002, p. 152). Each line represents the relationship between two variables. *Y* is an incomplete variable (partly missing), and X is an observed variable. $Miss_{(y)}$ is a dummy variable that captures whether data are missing on variable *Y*. Notice that the difference between MAR and MNAR is simply the extent to which $miss_{(y)}$ is related to *Y* itself after *X* has been controlled. MCAR = missing completely at random; MAR = missing at random; MNAR = missing not at random.

is not to determine which missing data mechanism is at work in a particular data set. Instead, the point of describing MCAR, MAR, and MNAR mechanisms is to illustrate the assumptions underlying different missing data treatments (i.e., *listwise and pairwise deletion are unbiased under MCAR, whereas ML and MI missing data treatments are unbiased under both MCAR and MAR missingness mechanisms*).

I agree with Graham's (2009) cogent observation that,

> These three kinds of missingness should not be thought of as mutually exclusive categories of missingness, despite the fact that they are often misperceived as such. In particular, MCAR, pure MAR, and pure MNAR really never exist because the pure form of any of these requires almost universally untenable assumptions. *The best way to think of all missing data is as a continuum between MAR and MNAR* [italics added]. Because all missingness is MNAR (i.e., not purely MAR), then whether it is MNAR or not should never be the issue. (p. 567)

In other words, missing data are almost never missing completely randomly (MCAR).[3] As such, most missing data fall on a continuum between one extreme—where the systematic missingness pattern depends entirely on the observed data (pure MAR), and the other extreme—where the systematic missingness pattern depends entirely on the missing data (pure MNAR). In typical scenarios, systematic missingness depends in part on the observed data (MAR) and in part on the missing data (MNAR), to varying degrees. A corollary of this view is that even though the strict MAR assumption might not be fully met in practice, missing data techniques based on this assumption (e.g., ML and MI missing data techniques) can still provide less biased, more powerful estimates than any of the other available missing data techniques.

Two Missing Data Problems: Bias and Inaccurate Standard Errors/Hypothesis Tests

Generically speaking, the purpose of data analysis is to give unbiased estimates of population parameters, as well as to provide accurate (error-free) hypothesis testing. Relatedly, the two chief problems caused by missing data are bias and error. *Bias* refers to the systematic over- or underestimation of a parameter (e.g., underestimated mean, correlation, or regression coefficient). Parameter estimation bias can be thought of as an external validity problem, because the biased estimates reflect a different population from the target population the researcher intends to understand. Missing data bias typically occurs when the missingness mechanism is systematic/nonrandom (i.e., under MAR or MNAR missingness; see Table 2).

Table 2. Missing Data Bias and Error Problems of Common Missing Data Techniques.

	Missingness Mechanism		
Missing Data Technique	MCAR	MAR	MNAR
Listwise Deletion	Unbiased; Large Std. Errors (Low Power)	Biased; Large Std. Errors (Low Power)	Biased; Large Std. Errors (Low Power)
Pairwise Deletion	Unbiased; Inaccurate Std. Errors	Biased; Inaccurate Std. Errors	Biased; Inaccurate Std. Errors
Single Imputation	Often Biased; Inaccurate Std. Errors	Often Biased; Inaccurate Std. Errors	Biased; Inaccurate Std. Errors
Maximum Likelihood (ML)	**Unbiased; Accurate Std. Errors**	**Unbiased; Accurate Std. Errors**	Biased; Accurate Std. Errors
Multiple Imputation (MI)	**Unbiased; Accurate Std. Errors**	**Unbiased; Accurate Std. Errors**	Biased; Accurate Std. Errors

Note. Recommended techniques are in boldface. Adapted from Newman (2009).

Error refers to hypothesis testing errors of inference, such as Type I error (a.k.a., false positive or "mirage"—errantly concluding a false hypothesis is supported) and Type II error (low power; a.k.a., false negative or "blindness"—errantly concluding a true hypothesis is unsupported). Hypothesis testing errors can be caused by inaccurate standard errors (*SE*s), which come about when a particular parameter being significance tested is associated with a sample size that is either too low or too high. Note that statistical significance testing typically involves calculating a p value for a t distribution using the equation $t = estimate/SE$, where the numerator is the *parameter estimate* being evaluated (e.g., correlation, regression coefficient), and the denominator (*SE*) is the degree of uncertainty associated with that parameter estimate (the *SE* term is proportional to $1/\sqrt{n}$). Thus, if a researcher's choice to use a missing data treatment like listwise deletion causes n to decrease by a factor of 4, then the t value will decrease by a factor of $\sqrt{4} = 2$, making her or him much less likely to obtain $p < .05$. This is why the choice of a missing data treatment (e.g., listwise deletion) can decrease statistical power to detect true effects, even in the absence of parameter bias.[4]

Before moving on, I note that the amount of missing data bias is a multiplicative function of the amount of missing data (response rate), the strength of the missingness mechanism (from completely random missingness [MCAR] to strongly systematic missingness [MAR or MNAR]), and the missing data treatment (see Table 2). As an example of this, Newman and Cottrell (in press) showed that the amount of missing data bias in the correlation can be estimated as a special case of Thorndike's (1949) formula for indirect range restriction:

$$r_{xy(comp)} = \frac{r_{xy(resp)} + r_{miss,\, x(resp)} r_{miss,\, y(resp)} (1/u^2 - 1)}{\sqrt{(1/u^2 - 1) r^2_{miss,\, x(resp)} + 1}\sqrt{(1/u^2 - 1) r^2_{miss,\, y(resp)} + 1}}, \tag{1}$$

where x and y are the two variables whose correlation we seek to estimate; $r_{xy(comp)}$ is the *unbiased* correlation between x and y (with complete data; i.e., no missing data bias); $r_{xy(resp)}$ is the missing data *biased* correlation between x and y (i.e., the pairwise-deleted correlation that was observed based on the subset of respondents whose data were available for both x and y); the variable labeled *miss* is a hypothetical *selection variable* that defines the missing data mechanism—*miss* has a continuous distribution and cut score below which all individuals are missing data on x and/or y—and above which x and y are both observed (not missing);[5] $r_{miss,x(resp)}$ is the range-restricted correlation between the variables *miss* and x; $r_{miss,y(resp)}$ is the range-restricted correlation between *miss* and y

(note that $r_{miss,x(resp)}$ and $r_{miss,y(resp)}$ are *systematic nonresponse parameters* that capture the extent to which the missing data on x and y are missing randomly versus systematically); and u^2 is the variance ratio of restricted (respondents-only) variance to unrestricted (complete-data) variance, for the selection variable *miss* (i.e., $u^2 = s^2_{miss}/S^2_{miss}$; note also that u^2 is a monotonic function of the amount of missing data [response rate], under the assumption of normality; see Newman & Cottrell, in press; Schmidt, Hunter, & Urry, 1976).[6]

To see a depiction of how missing data bias works, look at Figure 4. The most extreme form of missing data bias occurs under *direct range restriction* (e.g., when data on one variable [y] are missing due to truncation on another variable [x]—this is *selection on* x)[7] and can lead to substantial underestimation of the correlation (see scatterplot in Figure 4a). A much less extreme (and more realistic) form of missing data bias occurs when one variable (y) only has a probabilistic tendency to be selected on (x)—this has been labeled *stochastic direct range restriction* (*selection on* $x + e$; Newman & Cottrell, in press; see Figure 4b scatterplot)—and leads to much smaller negative missing data bias, compared to direct range restriction. A third category of missing data bias is *indirect range restriction*, where y and/or x is selected on a third variable called *miss*, while *miss* is correlated with y and/or x. When $r_{miss,x}$ and $r_{miss,y}$ have the same sign (e.g., both positive [or both negative]), then the missing data bias is negative (the observed correlation is biased in the negative direction [Table 3]; e.g., see Figure 4c scatterplot, where data are missing from the low end of x and from the low end of y). But, when $r_{miss,x}$ and $r_{miss,y}$ have opposite signs, the missing data bias can be substantial and positive (see Table 3 and Figure 4d scatterplot, where data are missing from the low end of x and the high end of y, which increases the observed positive correlation).

Figure 5 illustrates how the magnitude of missing data bias in the correlation under pairwise deletion is a function of three factors: (a) the amount of missing data (response rate ranges from 0% to 100%), (b) the strength of missingness mechanism (can be [i] completely random [i.e., MCAR; where $r_{miss,x} = 0$ and $r_{miss,y} = 0$] or [ii] systematic [i.e., MAR or MNAR; where $r_{miss,x} \neq 0$ and/or $r_{miss,y} \neq 0$]), and (c) whether $r_{miss,x}$ and $r_{miss,y}$ have the same sign. If $r_{miss,x}$ and $r_{miss,y}$ have the same sign, missing data bias is negative (leads to underestimation of a positive correlation or overestimation of a negative correlation). However, bias can become positive when the product term $(r_{miss,x})(r_{miss,y})$ is negative (see Equation 1)—which happens when $r_{miss,x}$ and $r_{miss,y}$ have opposite signs.

As an aside, I reiterate that MAR and MNAR are both systematic missingness mechanisms, and they can yield the same amount of missing data bias as each other (see Table 3; both MAR and MNAR correspond to $r_{miss,x} \neq 0$ and/or $r_{miss,y} \neq 0$). The key difference between MAR and MNAR is whether the nonrandom component of the selection variable (*miss*) has been observed in the dataset at hand (as I describe in the section below on auxiliary variables).

In sum, Figure 5 gives a sense of exactly how bad the missing data bias problem is, in the context of the bivariate correlation parameter when using pairwise deletion. When the response rate is close to 100%, there is no missing data bias. When the missingness mechanism is completely random (MCAR; $r_{miss,x}$ and $r_{miss,y} = 0$), there is no missing data bias. When the missingness mechanism is systematic and the systematic nonresponse parameters ($r_{miss,x}$ and $r_{miss,y}$) have the same sign, there is negative missing data bias. When $r_{miss,x}$ and $r_{miss,y}$ have opposite signs, there is positive missing data bias—see Figure 5.

How much should we care about missing data bias and error? The answer to this depends on our answers to two other questions: (a) How large is the expected degree of missing data bias? and (b) What can we reasonably do to reduce the amount of missing data bias and error? With regard to the former question, we note that Anseel, Lievens, Schollaert, and Choragwicka (2010) have reported the average response rate in the organizational sciences to be 52%—this amount of missing data can be compared to Figure 5 to estimate how much missing data bias might be expected in the correlation. With regard to the second question (What can we reasonably do to reduce missing data bias and error?), the easiest answer comes in the form of choosing the least biased and least error-prone missing data treatments from among the available options.

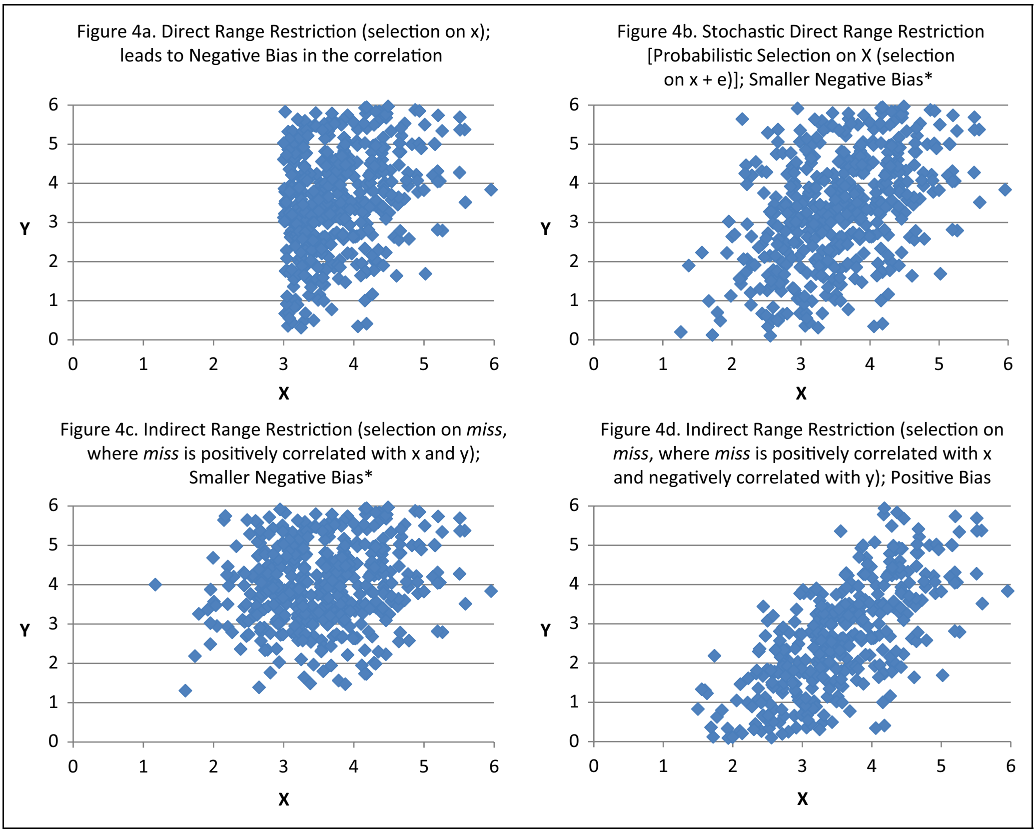


Figure 4. Missing Data Bias.
Note: Both stochastic direct range restriction (b) and indirect range restriction (c) typically yield smaller missing data bias than does direct range restriction (a). The precise amount of missing data bias can be understood by referring to Equation 1, because all the scenarios depicted in Figure 4 are special cases of Equation 1 (as described in Table 3, third column).

Five Missing Data Treatments: Listwise Deletion, Pairwise Deletion, Single Imputation, ML Routines, Multiple Imputation

Before moving to the next section, I briefly review the currently available missing data treatments. When data are missing, there are five major categories of missing data treatments, a researcher must choose among. The choice of missing data treatment has major implications for missing data bias and error (see Tables 2 and 3; as well as simulations by Enders, 2010; Newman, 2003; Schafer & Graham, 2002). Five missing data treatments are described in Table 4. Because this table is essential to what comes next, I recommend that the reader take a very careful look at Table 4. I will discuss aspects of the various missing data treatments in the following sections.

Three Key Considerations Prior to Data Analysis

Before continuing with the discussion of missing data analysis, however, there are three major considerations that should first be understood. These are presented briefly now.

Table 3. Missing Data Bias in the Correlation, under Pairwise Deletion versus Maximum Likelihood (ML) Estimation, for 11 Missing Data Selection Mechanisms.

Missing Data Selection Mechanism	Rubin's (1976) Mechanism	Selection Variable (*miss*)	Pairwise Deletion Bias	ML Estimation Bias
(1) Completely random missingness (*y* and/or *x* selected randomly)	MCAR	miss = e	Zero bias	Zero bias
(2a) Direct range restriction (*y* selected on *x*) [maximally systematic missingness]	MAR	miss = x	Negative bias	Zero bias
(2b) Direct range restriction (*x* selected on *y*)	MAR	miss = y	Negative bias	Zero bias
(3a) Stochastic direct range restriction (*y* probabilistically selected on *x*) [weaker systematic missingness]	MAR	miss = x + e	Smaller negative bias	Zero bias
(3b) Stochastic direct range restriction (*x* probabilistically selected on *y*)	MAR	miss = y + e	Smaller negative bias	Zero bias
(4) Indirect range restriction (*y* and/or *x* selected on *miss*; *miss* is observed)$r_{miss,\ x}$ and $r_{miss,\ y}$ have [same sign] {opposite signs}	MAR	miss = miss	[Smaller negative bias] {Positive bias}	Zero bias
(5a) Direct range restriction (*x* selected on *x*)	MNAR	miss = x	Negative bias (Same as MAR)	Same negative bias as pairwise
(5b) Direct range restriction (*y* selected on *y*)	MNAR	miss = y	Negative bias (Same as MAR)	Same negative bias as pairwise
(6a) Stochastic direct range restriction (*x* probabilistically selected on *x*)	MNAR	miss = x + e	Smaller negative bias (Same as MAR)	Same negative bias as pairwise
(6b) Stochastic direct range restriction (*y* probabilistically selected on *y*)	MNAR	miss = y + e	Smaller negative bias (Same as MAR)	Same negative bias as pairwise
(7) Indirect range restriction (*y* and/or *x* selected on *miss*; *miss* is unobserved)$r_{miss,\ x}$ and $r_{miss,\ y}$ have [same sign] {opposite signs}	MNAR	miss = miss	[Smaller negative bias] {Positive bias} (Same as MAR)	[Same negative bias as pairwise] {Same positive bias as pairwise}

Note. Pairwise deletion is unbiased under MCAR, while ML estimation is unbiased under MCAR and MAR. Adapted from Newman and Cottrell (in press). e = random error term; MCAR = missing completely at random; MAR = missing at random (i.e., a type of systematic missingness, with a confusing label; Rubin, 1976); MNAR = missing not at random.

Missing Data Are Partly Unavoidable, and Partly Avoidable

To some extent, missing data are a natural and *unavoidable* consequence of the ethical principle of *respect for persons* and its application in the requirement that research participation be voluntary (National Commission for the Protection of Human Subjects, 1979). So long as would-be participants who are sampled from the target population are allowed to autonomously opt out of the study (or to opt out of part of the study), missing data will be an ethically unavoidable data analysis problem. Pertaining to this, we should be wary of research reports that claim response rates near 100%, and should ask questions about how these extremely high response rates were secured (e.g., How was confidentiality maintained? Did the supervisors know which employees had responded? [i.e.,

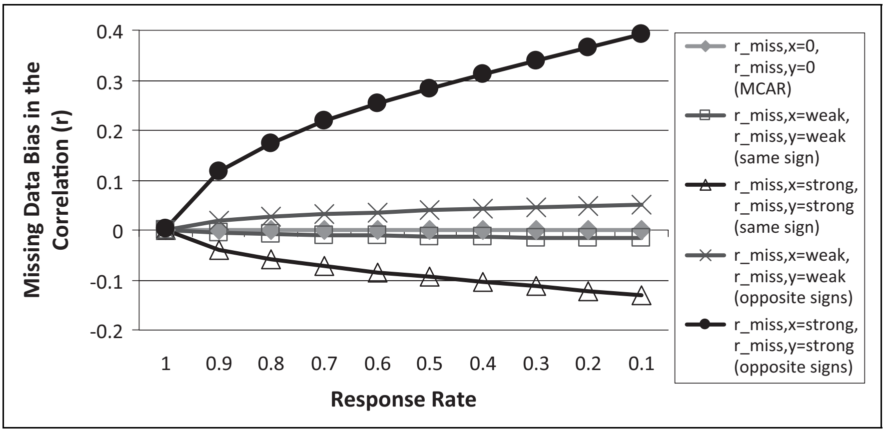


Figure 5. Missing data bias in the correlation, under pairwise deletion.
Note: Missing data bias = $r_{xy(resp)} - r_{xy(comp)}$. In the present example, $r_{xy(comp)} = .5$. For completely random missingness (MCAR), $r_{miss,x} = .0$ and $r_{miss,y} = .0$. For weak systematic missingness (MAR or MNAR), $r_{miss,x} = .2$ or –.2, and $r_{miss,y} = .2$ or –.2. For strong systematic missingness (MAR or MNAR), $r_{miss,\,x} = .5$ or –.5, and $r_{miss,y} = .5$ or –.5. "Same sign" indicates that $r_{miss,x}$ and $r_{miss,y}$ were both positive or both negative; "opposite signs" indicates that one was positive and the other was negative. MCAR = missing completely at random; MAR = missing at random; MNAR = missing not at random.

they should not know]; Did the online survey administration require that all items be answered before proceeding to the next page? [i.e., proceeding to the next page should not be contingent on completing all items]).

On the other hand, much missing data are *avoidable*. Anseel et al. (2010) have conducted a major meta-analysis of response rates in the organizational sciences (see also Cycyota & Harrison, 2006; Dillman, 1978; Roth & BeVier, 1998; Yammarino, Skinner, & Childers, 1991), and found that the major predictors of high response rates are: (a) personally distributing surveys ($r = .38$); (b) using identification numbers ($r = .18$), which I note can appear to threaten respondent confidentiality in some cases and should therefore *not* be used as an explicit response-enhancing technique, although identification methods of some sort are a requirement for longitudinal studies, multisource studies to reduce common method bias, and social network studies; (c) personalization of the survey invitation ($r = .14$); (d) university sponsorship of the survey ($r = .11$); and (e) giving advance notice ($r = .08$). Interestingly, incentives appeared to have no positive average effect on response rates ($r =$ –.04). Also, another way to prevent avoidable missing data is that researchers conducting longitudinal data collections should not give up on initial nonrespondents; each individual in the sampling frame should be contacted at every wave of data collection, regardless whether she or he has responded to past waves of data collection.

Define the Target Population of Interest

In the organizational and psychological sciences, researchers rarely attempt to explicitly define the target population to which a study should generalize. Rather, we often draw inferences to vaguely or implicitly defined populations, such as "all working adults" or "working adults in customer service jobs." For most single-organization studies, it would probably be more appropriate for researchers to limit their generalizations to the original sampling frame (e.g., to "working adults in this particular company" and not to "all working adults").

Table 4. Missing Data Treatments.

Missing Data Treatment	Definition	Major Issues
Listwise Deletion	Delete all cases (persons) for whom any data are missing, then proceed with the analysis.	Discards real data from partial respondents. Smallest *n*, lowest power. Biased under MAR and MNAR.
Pairwise Deletion	Calculate summary estimates (means, SDs, correlations) using all available cases (persons) who provide data relevant to each estimate, then proceed with analysis based on these estimates.	Different correlations represent different subpopulation mixtures. Sometimes covariance matrix is not positive definite. Biased under MAR and MNAR. No single *n* makes sense for whole correlation matrix (SEs inaccurate).
Single Imputation (ad hoc techniques)	Fill in each missing value [e.g., using mean (across persons) imputation, regression imputation, hot deck imputation, etc.], then proceed with analysis based on partially-imputed 'complete' dataset.	Mean (across persons) imputation and regression imputation are both biased under MCAR! No single *n* makes sense for whole correlation matrix (SEs inaccurate). SEs underestimated if you treat dataset as complete.
Maximum Likelihood	Directly estimate parameters of interest from incomplete data matrix (e.g., FIML); or Compute summary estimates [means, SDs, correlations] (e.g., EM algorithm), then proceed with analysis based on these summary estimates.	Unbiased under MCAR and MAR. Improves as you add more variables to the imputation model. Number of variables should be < 100. Accurate SEs for FIML. For EM algorithm, no single *n* makes sense for whole correlation matrix (SEs inaccurate).
Multiple Imputation	Impute missing values multiple times, to create 40, partially-imputed datasets. Run the analysis on each imputed dataset. Combine the 40 results to get parameter estimates and standard errors.	Unbiased under MCAR and MAR. Improves as you add more variables to the imputation model. Number of variables should be < 100. Accurate SEs. Gives slightly different estimates each time. When used with SEM, suffers more nonconvergences.

Note: See Allison (2002), Enders (2001b, 2010), Graham (2009), Marsh (1998), Newman (2003, 2009), Schafer and Graham (2002). MAR = missing at random; MNAR = missing not at random; MCAR = missing completely at random.

When considering missing data, the problem of limited target populations gets worse. By conducting all of our analyses on survey respondents, we can now only generalize our study results to "working adults who fill out surveys." And perhaps worst of all, a listwise deletion missing data strategy only makes sense if one's target population is restricted to "working adults who fill out surveys *completely*"—such a target population is rarely theoretically defensible.

Guideline 0: Abstinence Is Not an Option

There is one last consideration that must be made prior to data analysis, but that is essential for every researcher to understand prior to using the five practical guidelines below—as such, I refer to this principle as "Guideline 0." After data collection, every researcher must understand that choosing a missing data treatment involves choosing the lesser of evils. Avoiding missing data treatments is not an option. The data analyst must choose listwise deletion, pairwise deletion, a single imputation/ad hoc technique, a maximum likelihood technique, or multiple imputation—and then defend that choice. Missing data problems cannot be avoided by simply ignoring them. If you are using a default approach of listwise or pairwise deletion, then you are—in reality—*choosing* listwise or pairwise deletion. Given the widespread availability of software that implements ML and MI missing data routines (e.g., see Appendix A), it is no longer defensible to simply say, ''We are not going to bother with the fancy missing data routines.'' Each researcher must now be in the position to defend why his or her chosen missing data technique is equal or superior to its available alternatives in terms of missing data bias and error. Such arguments are increasingly hard to make in defense of listwise and pairwise deletion (at least for traditional correlation, regression/ANOVA, factor analysis, and SEM analyses—which are all based on a covariance matrix and vector of means, and for which ML and MI routines are now widely available; see Appendix A). With that said, the decision tree in Figure 1 does provide a rule of thumb to help designate when pairwise deletion might be similarly as accurate as a state-of-the-art (ML or MI) technique.

Five Practical Guidelines for Missing Data Analysis

Guideline 1: Use All the Available Data

(Do Not Use Listwise Deletion). The dictum that researchers should use all the available data is the ''fundamental principle of missing data analysis'' (according to Newman, 2009, p. 11). As sensible as this principle may sound, it is still routinely ignored by researchers in management and psychology who regularly opt for listwise deletion (cf. Peugh & Enders, 2004). *Listwise deletion* involves deleting all cases (persons) for whom any data are missing, then proceeding with the analysis. In other words, the analysis is based on the full respondents only, and the data from partial respondents get discarded by the researcher. Listwise deletion has the ill effect of converting item-level and construct-level missingness into person-level missingness.

The prevalence of listwise deletion is attested by phrases of the following sort, which are regularly found in the Method sections of our top journals: ''Out of 542 surveys returned, 378 provided usable data and were included in the analysis.'' The problem with this sort of statement is that it is inherently false. *All* of the respondents who provided data provided ''usable data,'' but the researcher chose to throw away some of this precious information. Indeed, listwise deletion compounds the problem of sample nonresponse, by adding to it the extra problem that the researcher herselft or himself creates additional missing data by discarding the partial respondents.

There are several problems with listwise deletion. First, regardless whether missingness is systematic versus random, listwise deletion often greatly reduces sample size and statistical power (i.e., it increases *SE*s and Type II error). Second, even when statistical power seems adequate, listwise deletion yields biased parameter estimates under systematic (MAR and MNAR) missingness. Third, listwise deletion only supports inferences to a target population of ''individuals who fill out surveys completely,'' and restricting the target population in this way is almost never theoretically defensible. Fourth, because listwise deletion involves discarding data that cost the partial respondents' valuable time and energy to provide, failure to use all the available data in one's analyses may violate an ethical imperative (Rosenthal, 1994). Even throwing out data from a small number of partial respondents (2%-3%) would be suboptimal, potentially unethical, and totally unnecessary.

In sum, because listwise deletion often leads to extreme levels of inferential error (low power) and missing data bias (over- or underestimation of effect sizes), and because it is based on theoretically and ethically indefensible rationales, it should be avoided outright. This prohibition should also extend to the intuitive yet misguided practice of using listwise deletion to ''double check'' the accuracy of more robust missing data approaches like ML estimation and multiple imputation. There is no good reason for this. When listwise deletion yields discrepant results from ML or MI techniques, this does not in any way cast doubt onto the ML and MI results; rather, it only suggests that the missingness mechanism is in part MAR (which is quite often the case).[8] Further, there is no logical basis for using listwise deletion in this way. If the listwise result *agrees* with the ML and MI result, then we will accept the ML and MI result; and if the listwise result *disagrees* with the ML and MI result, then we will still accept the ML and MI result (because ML and MI provide accurate *SE*s and are unbiased under both MAR and MCAR mechanisms, whereas listwise deletion provides highly inflated *SE*s and is only unbiased under MCAR)—the information value of the listwise deletion result is nil either way.

Guidelines 3 and 4 (reviewed below) follow directly from the current principle. Once we are using all of the available data, the question arises of *how* we should use the available data. Guidelines 3 and 4 involve the cases of construct-level missingness and item-level missingness. But first, I address the dangers of single imputation.

Guideline 2: Do Not Use Single Imputation

(Do Not Simply Impute Data Once and Then Proceed as Though You Have Complete Data). Single imputation techniques involve filling in each missing datum with a "good guess" as to what the missing datum should be. Fortunately, single imputation techniques are much less popular now than they once were. Common examples of single imputation are: (a) mean (across persons) imputation—replacing each missing datum with the group mean for the corresponding variable, (b) hot deck imputation—replacing each missing datum with a value from a "donor" who has similar scores on other variables (which can be more error prone than listwise deletion; see Switzer, Roth, & Switzer, 1998), and (c) regression imputation—replacing each missing datum with a predicted value based on a multiple regression equation derived from observed cases.

Single imputation suffers two major drawbacks. First, most single imputation techniques are biased under MCAR. For example, because mean imputation imputes a constant mean for each missing value (see Figure 6a), the resulting sample estimates of the variance and the correlation will be downwardly biased—even if the missingness mechanism is completely random (MCAR). As another example, regression imputation leads to underestimation of the variance and overestimation of the correlation (because imputed values fall exactly on the regression line; see Figure 6b)—even if the missingness mechanism is MCAR! It is possible to improve regression imputation methods, however, by adding a random error term to the imputed values (i.e., stochastic regression imputation; see Figure 6c). Stochastic regression imputation works to remove the missing data bias in regression imputation (described below) that previously underestimated the variance and overestimated the correlation (i.e., stochastic regression imputation is unbiased under both the MCAR and MAR missingness mechanisms). Nonetheless, even when considering stochastic regression imputation (which is unbiased under MAR), I still do not recommend single imputation, for the following reason.

The second major drawback is that single imputation suffers the inability to calculate accurate *SE*s for hypothesis testing (i.e., there is usually no single value of n that corresponds well to all the parameter estimates). This problem is coupled with the real and common danger that many researchers tend to use the maximum n (treating the partially imputed data set as though it were a complete data set), which naturally leads to deflated *SE*s and creates Type I errors of inference (a.k.a., mirages, where incorrect hypotheses are falsely supported). As described in the following, multiple imputation solves this problem.

Overall, the main reason to place a moratorium on single imputation is because multiple imputation has all of the advantages of single imputation, but none of its major drawbacks.[9] Thus, for typical data-analytic applications (e.g., involving correlation, regression/ANOVA, factor analysis, and SEM), single imputation should be forbidden.

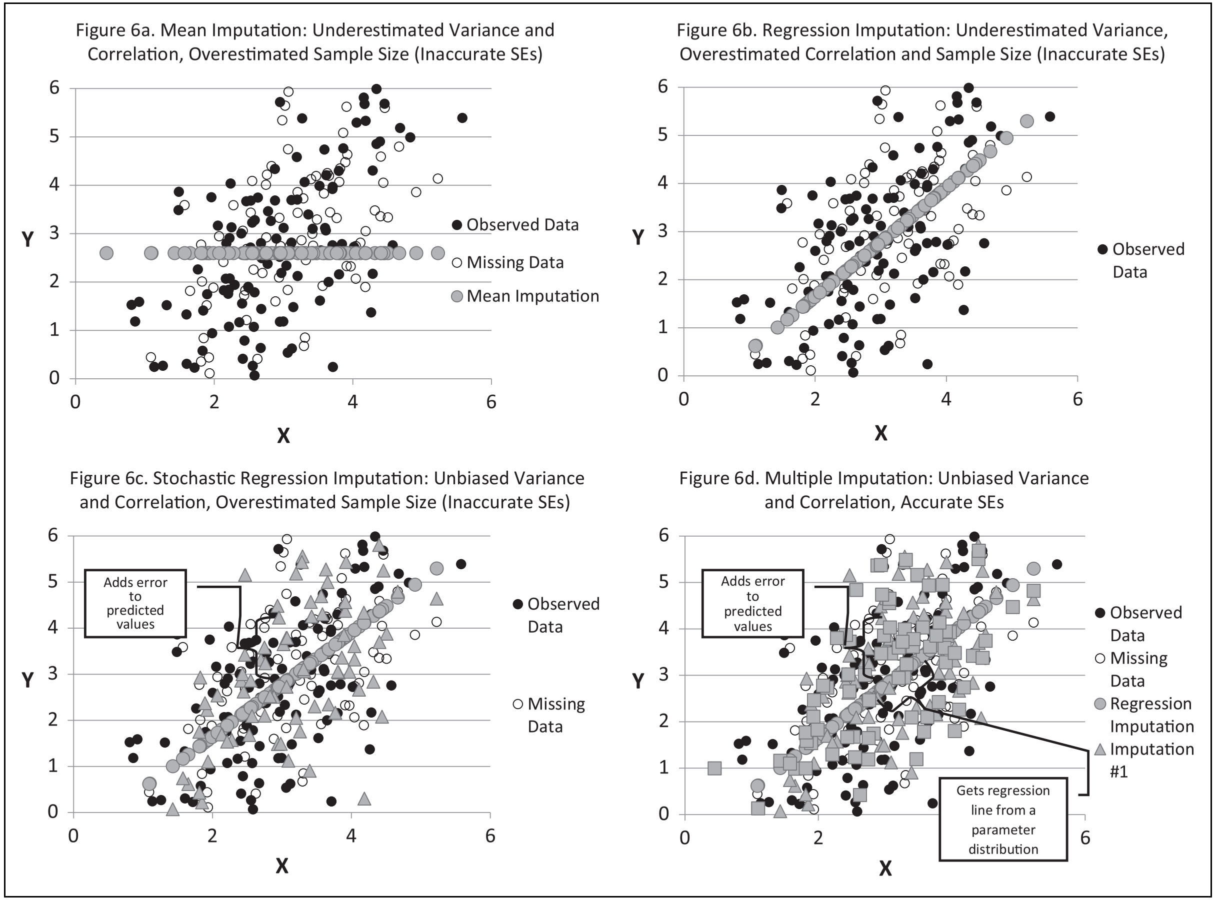


Figure 6. Missing Data Bias.
Note: For mean imputation, the mean of observed *Y* scores is used in place of each missing *Y* score. For regression imputation, the predicted *Y* score for each individual (based on the individual's observed *X* score and a regression equation of *Y* onto *X* estimated from the observed X-Y pairs) is used in place of each missing *Y* score. For stochastic regression imputation, the predicted *Y* score plus a random error term proportional to $\sqrt{(1 - r^2)\sigma_Y^2}$ is used in place of each missing *Y* score, for the purpose of removing bias in the variance and correlation for the imputed dataset. For multiple imputation, the predicted *Y* score plus a random error term is used in place of each missing *Y* score, except that the missing data are imputed $\sim$40 different times and the regression equation varies slightly across imputations, because it is drawn from a parameter distribution. The parameter estimates are still unbiased under multiple imputation, but the 40 data sets are also combined in such a way as to render accurate *SE*s.

Figure 6a. Mean Imputation: Underestimated Variance and Correlation, Overestimated Sample Size (Inaccurate SEs)

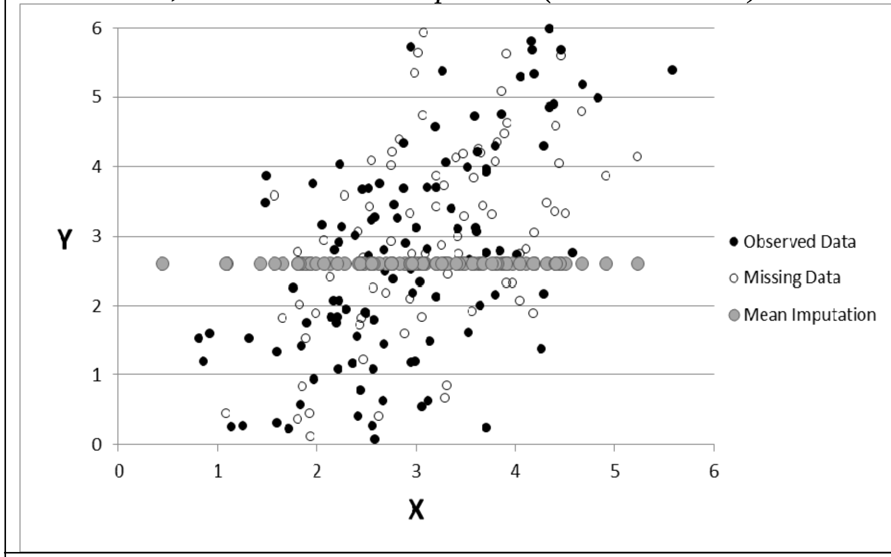


Figure 6b. Regression Imputation: Underestimated Variance, Overestimated Correlation and Sample Size (Inaccurate SEs)

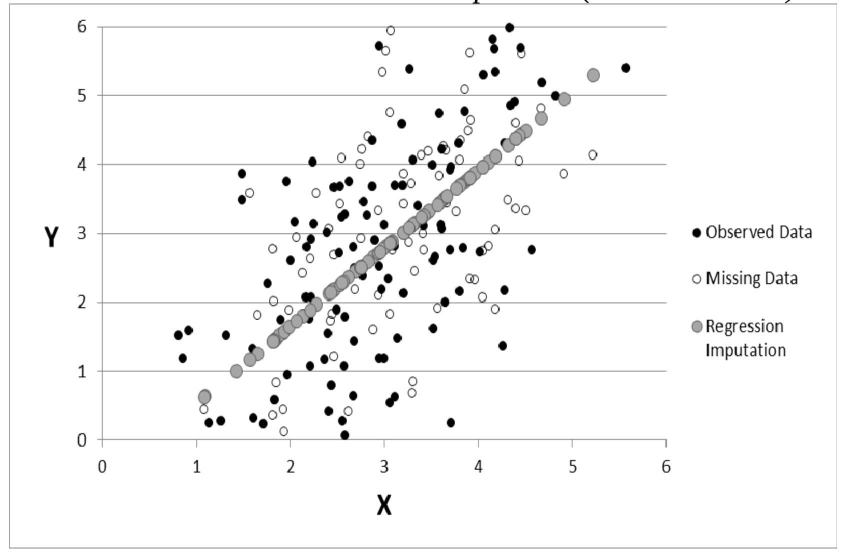


Figure 6c. Stochastic Regression Imputation: Unbiased Variance and Correlation, Overestimated Sample Size (Inaccurate SEs)

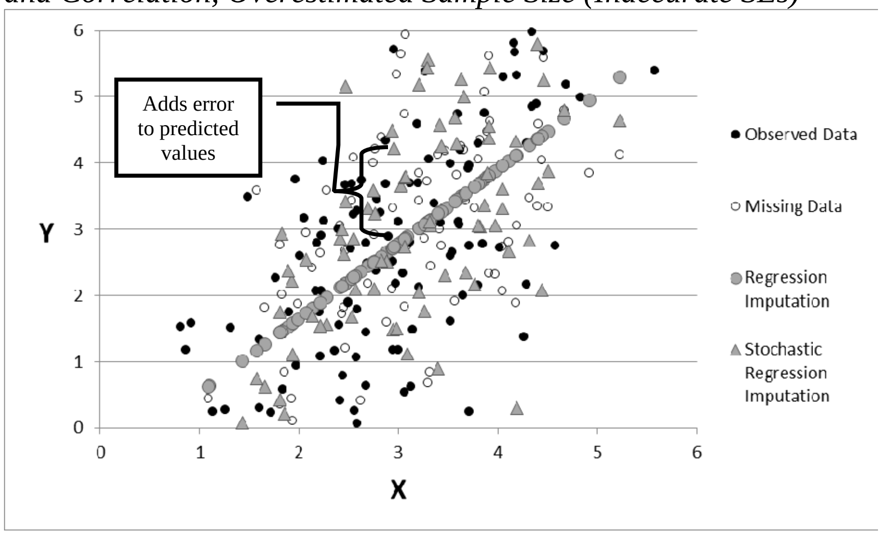
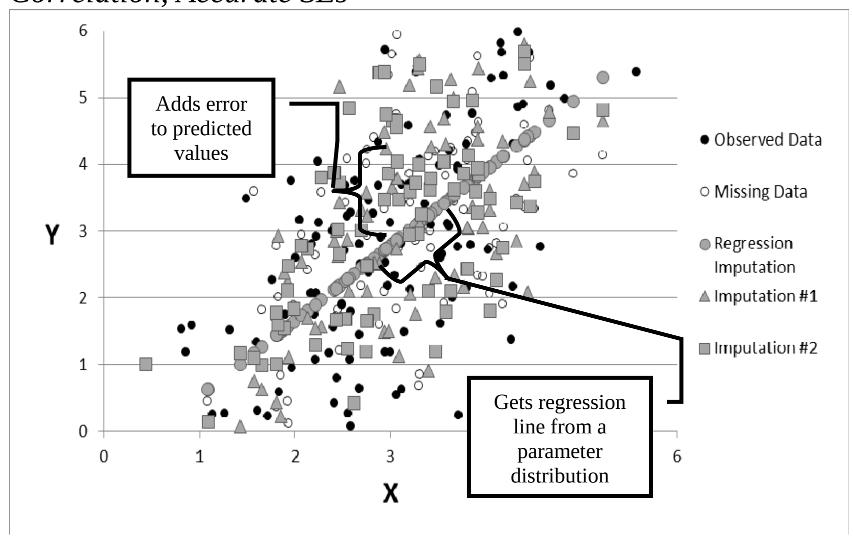


Figure 6d. Multiple Imputation: Unbiased Variance and Correlation, Accurate SEs



Guideline 3: Construct-Level Missingness: Use Maximum Likelihood or Multiple Imputation Missing Data Treatments Whenever 10% or More of The Respondent Sample Is Made Up of Construct-Level Partial Respondents (i.e., Respondents Who Reported on at Least One Construct but Who Omitted at Least One Construct)

1. Report the response rate, full response rate, and partial response rate; as well as the number of respondents for each construct (or for sets of constructs).

The *response rate* can be formally defined using the equation:

$$\text{Response rate} = (n \text{ partial respondents } + n \text{ full respondents})/n \text{ contacted.} \tag{2}$$

For the purpose of Equation 2, we can treat *n partial respondents* as the number of respondents with construct-level missingness; whereas *n full respondents* is the number of respondents with no construct-level missingness. *n contacted* is the number of individuals contacted with the survey invitation. Equation 2 can be distinguished from three other, related expressions:

$$\text{Full response rate} = n \text{ full respondents}/n \text{ contacted,} \tag{3}$$

$$\text{Partial response rate} = n \text{ partial respondents}/n \text{ contacted,} \tag{4}$$

$$\text{Nonresponse rate} = n \text{ nonrespondents } /n \text{ contacted.} \tag{5}$$

Note that: (a) Full response rate + Partial response rate + Nonresponse rate = 1.0, and (b) Response rate = Full response rate + Partial response rate = 1.0 – Nonresponse rate.

In order for the consumers of our research to be able to understand likely missing data bias, it is essential for researchers to report the response rate, partial response rate, and full response rate. For example, a researcher could report:

> In the current study, surveys were distributed to 500 employees, 300 of whom provided responses (response rate = 60%). Two hundred and fifty of these were full respondents who answered every scale (full response rate = 50%), whereas 50 of these were partial respondents who answered some but not all of the scales (partial response rate = 10%);

or more succinctly:

> Surveys were returned by 300 out of 500 employees (response rate = 60%; full response = 50%; partial response = 10%).

Additionally, the presence of partial respondents implies that the response rate varies systematically across constructs. This information should be reported in the footnote of a paper's correlation matrix. For example:

> $N = 246$ to 276 for variables A, B, and G to J; and $N = 172$ to 189 for variables C to F.

Some researchers already follow this practice of concisely reporting variable-wise response rate information, which I applaud.

2. If 10% or more of the respondent sample is made up of partial respondents (i.e., if partial response rate / [partial response rate + full response rate] > .10), then maximum likelihood

(EM or FIML) or multiple imputation approaches should be used, instead of pairwise deletion.

In this section, I will first provide some advice, and then give the rationale behind it. Generally speaking, the advice is to use ML and MI missing data routines when there is construct-level missingness (i.e., when there is a sizeable portion of partial respondents). When there is a sizeable amount of construct-level missingness, then ML and MI routines typically outperform listwise and pairwise deletion substantially in terms of reduction in missing data bias and error (Allison, 2002; Enders, 2010; Newman, 2003; Schafer & Graham, 2002). On the other hand, when there is no construct-level missingness, then ML and MI routines perform no better than listwise and pairwise deletion. To introduce the current practical guideline, I begin by defining the ratio, percentage of respondents who are partial respondents (PRPR):

$$\text{PRPR} = n \text{ partial respondents}/(n \text{ partial respondents} + n \text{ full respondents}), \tag{6}$$

which is also equal to the partial response rate divided by the response rate (see Equations 2 and 4). This ratio indexes the extent to which the respondents are partial respondents (as opposed to full respondents). If this percentage of respondents who are partial respondents falls below 10%, then it usually doesn't make much difference whether the researcher is using pairwise deletion versus state-of-the-art ML and MI missing data techniques. (To be fair, I acknowledge that statisticians would recommend ML and MI techniques over pairwise deletion even in this case, because ML and MI are robust to MAR missingness; but the point I am proposing here is that it will not make much practical difference in this particular case [i.e., when PRPR < 10%]).

The choice of a 10% cutoff is arbitrary, but it attempts to reflect a consistent standard that appreciates the fact that—when there is little construct-level missingness—then the choice of using ML and MI techniques versus using pairwise deletion will make little difference. One example of a research design that nearly always exhibits >10% PRPR (i.e., a high portion of construct-level missingness) is a longitudinal design, where many of the respondents at Time 1 drop out before Time 2. In order to understand why this PRPR <10% guideline works, I briefly explain ML and MI missing data routines (Table 4).

Multiple Imputation. *Multiple imputation* is a procedure that operates by performing an unbiased single imputation routine (like stochastic regression imputation) over and over again (i.e., it makes multiple, different guesses at what the missing data might have been). It then takes advantage of the variation between those different guesses/imputations when indexing the degree of uncertainty (*SE*) associated with each parameter estimate. As such, significance tests/hypothesis tests based on MI are more accurate (i.e., fewer errors of inference).

The MI missing data routine (Rubin, 1987; Schafer, 1997) operates in three phases. In Phase 1 (imputation phase), the available data are used to impute multiple data sets (Graham, Olchowski, & Gilreath, 2007, recommend imputing at least $m = 40$ different data sets to approach optimal statistical power). Data sets are imputed using a routine similar to stochastic regression imputation, which is unbiased under MAR and for which the regression parameters are drawn from a Bayesian parameter distribution (see Figure 6d). In Phase 2 (analysis phase), the researcher analyzes each of the (e.g., $m =$ 40) data sets using whichever analysis she or he would have ordinarily used on complete data (as if there had been no missing data), and she or he then saves the parameter estimates (e.g., correlations, regression coefficients, factor loadings, SEM path coefficients) and their corresponding *SE*s for all (e.g., $m = 40$) data sets. Finally, in Phase 3 (pooling phase), the parameter estimates and their *SE*s from the multiple, partly imputed data sets are combined. The parameter estimates are simply averaged across the m imputed data sets to get the final parameter estimates. The standard errors are combined across m imputed data sets using Rubin's (1987) formula:

$$S.E. = \sqrt{\frac{1}{M}\sum_{m=1}^{M} S.E._m^2 + \left(1+\frac{1}{M}\right)\left(\frac{1}{M-1}\right)\sum_{m=1}^{M}\left(b_m - \bar{b}\right)^2}, \tag{7}$$

where $\frac{1}{M}\sum_{m=1}^{M} S.E._m^2$ is the average squared *SE* across imputations, $\left(\frac{1}{M-1}\right)\sum_{m=1}^{M}\left(b_m - \bar{b}\right)^2$ is the variance of the parameter estimates (e.g., b's) across imputations, and $\left(1+\frac{1}{M}\right)$ is a correction factor that converges to 1 as the number of imputations increases.

The two important things to remember about MI are that: (a) the pooled MI parameter estimates are unbiased under both MAR and MCAR missing data mechanisms, and (b) the pooled MI *SE*s are accurate (i.e., the standard errors, upon which hypothesis tests are based). The parameter estimates are unbiased under MAR because they are based on stochastic regression imputation. The *SE*s are accurate because of the second term in Equation 7, $\left(\frac{1}{M-1}\right)\sum_{m=1}^{M}\left(b_m - \bar{b}\right)^2$, which is the variance of the parameter estimates between imputations. As mentioned previously, multiple imputation works by performing an unbiased single imputation routine over and over again (i.e., by making multiple, different guesses at what the data might have been), and then takes advantage of the variance between those guesses/imputations when indexing the degree of uncertainty (*SE*) associated with each parameter estimate. As such, the operative word in multiple imputation is *multiple*, not *imputation*—the whole point is that each single imputation contains some inaccuracy, so the imputations are performed multiple times and then aggregated in a way that accounts for the uncertainty of each imputation. This way, significance tests/hypothesis tests based on MI have the appropriate level of uncertainty.

This advantage of MI (i.e., the accurate *SE*s/hypothesis tests) can perhaps be most easily understood by comparison to other missing data techniques. Under listwise deletion, the *SE*s are too large (because the sample size is too small due to discarding real data from partial respondents); under single imputation, the *SE*s are too small (because the sample size is too large due to pretending one has a complete data set when in fact one does not); but under multiple imputation, the *SE*s are just right. This is because the multiple imputation *SE*s are essentially single-imputation *SE*s that have been adjusted upward using the between-imputations variance in parameter estimates. To restate, listwise deletion overestimates the uncertainty of one's results by discarding partial respondents (increasing Type II error), single imputation underestimates the uncertainty of one's results by treating partial respondents as though they were full respondents (increasing Type I error), and multiple imputation is in between—it appropriately treats partial respondents as partial respondents, and thereby provides the accurate level of uncertainty corresponding to each parameter and hypothesis test. This is why Guideline 3 makes sense: When there are no (or very few) partial respondents, then it makes almost no difference whether one uses MI versus a less robust missing data routine (i.e., vs. pairwise deletion or stochastic regression imputation).

Maximum Likelihood. ML missing data routines are mathematically complex, although some of the most user-friendly descriptions of them have been provided by Enders (2001b, 2010). I refer the reader to those excellent summaries to understand the mechanics of the approach. In brief, ML routines operate by choosing parameter estimates that maximize the probability of the observed data. Stated differently, ML routines use a *likelihood function* (e.g., see Finkbeiner, 1979, for a FIML likelihood function) that describes the relationship between a likelihood (i.e., a probability based on the observed data) and different values of the parameter estimates. ML techniques then select the parameter estimates that maximize the likelihood function based on the available data. For the current article, we emphasize the following points with regard to ML techniques:

- ML missing data routines provide results that are essentially identical to results from MI routines (Collins, Schafer, & Kam, 2001, p. 33). This is because both ML and MI are designed to provide unbiased parameter estimates under MAR and MCAR missingness mechanisms.[10]
- ML missing data techniques are not overtly imputation techniques, and so they might be perceived as more palatable by naïve readers and reviewers who are philosophically opposed to multiple imputation because they fear that multiple imputation routines are "making up data." (This fear is unfounded, because the point of MI is not to make up data but rather to render unbiased parameter estimates and accurate *SE*s; however, the philosophical opposition that lives in the minds of some reviewers can be very real.) This point is essentially cosmetic.
- There are two common ML missing data routines: FIML and the EM algorithm.
- FIML is a direct estimation technique and operates by directly analyzing the incomplete data set to yield unbiased parameter estimates and accurate *SE*s.
- The EM algorithm is not a direct estimation technique, but instead operates by providing summary statistics (a covariance matrix and vector of means), which can then be used as input to another analysis routine (e.g., one can perform multiple regression and SEM on a covariance matrix). The chief problem with the EM algorithm is that, even though the parameter estimates will be unbiased under MAR, there is typically not one single sample size that appropriately corresponds to the entire covariance matrix. As such, the EM algorithm is not recommended for use with hypothesis testing (with the possible exception of tests that conservatively use the minimum observed sample size to correspond to the EM covariance matrix—such tests provide adequate Type I error protection, but are still vulnerable to Type II error).

3. When using ML or MI missing data treatments, the researcher should report the ML correlation matrix, standard deviations, and means (estimated via the EM algorithm), instead of the listwise- or pairwise-deleted correlation matrix, standard deviations, and means. The reasoning here is that the ML correlation matrix, *SD*s, and means are unbiased under both MAR and MCAR missingness, whereas the listwise- and pairwise-deleted parameter estimates are biased whenever the data are not MCAR.
4. When using ML or MI missing data treatments, the missing data imputation or estimation model should include all of the variables in the theoretical model under consideration (including product terms when testing interaction effects).

When implementing ML or MI routines, the researcher must specify which variables will be used as part of the missing data routine. It is important to include all variables in the imputation model that will appear in the substantive theoretical model being tested, including any interaction terms that will be used to assess moderator hypotheses. If interaction terms are left out of the missing data imputation model, then the estimated interaction effect will be biased toward zero (Graham, 2009). For an excellent summary of missing data treatments for interaction effects, see Enders, Baraldi, and Cham (2014).

5. When using ML or MI missing data treatments, the missing data imputation or estimation model should include extra, *auxiliary variables* that are not part of the theoretical model under consideration, when possible.

In addition to using all the substantive variables from one's theoretical model (including interaction terms) as part of the missing data imputation/estimation model, some researchers have helpfully advised that researchers should also use auxiliary variables. *Auxiliary variables* are variables included in the missing data imputation/estimation model that are not part of one's theoretical model, nor do they have any particular substantive interest in the study at hand (Collins et al., 2001; Graham, 2003). That is, auxiliary variables are variables that the researcher includes in the imputation model for the express purpose of reducing missing data bias and error. The rationale behind auxiliary variables is summarized in the following.

Auxiliary Variables Can Convert MNAR Missingness Into MAR Missingness. When looking at Tables 2 and 3, one of the big problems in missing data analysis that becomes painfully apparent is that there are no widely available missing data treatments that are especially good at treating the MNAR missingness mechanism (i.e., the best available missing data techniques, ML and MI, are both still biased under MNAR). Enders (2010) has summarized that MNAR missing data problems have often been treated using either selection models (Heckman, 1979; Puhani, 2000; Winship & Mare, 1992) or pattern mixture models (Glynn, Laird, & Rubin, 1986; Little, 1993; Rubin, 1987). Unfortunately, both selection models and pattern mixture models are necessarily based on assumptions about the missing data mechanism that are potentially wrong and essentially untestable, and as such these alternatives often perform worse than ML or MI techniques, even under MNAR (see Enders, 2010).

One especially good piece of advice for dealing with MNAR missingness is to use auxiliary variables as part of the imputation model (Collins et al., 2001), for the reason that including auxiliary variables in the imputation model can convert an MNAR missingness mechanism into an MAR missingness mechanism. To understand why, look at Figure 3. In Figure 3, note that the one factor that distinguishes MNAR missingness from MAR missingness is the extent to which there still exists a relationship between the incomplete variable (Y) and the missingness pattern on Y ($miss_{(y)}$), after the other observed variables (X variables) have been controlled. So in order to convert an (untreatable) MNAR missingness mechanism into an (easily treatable) MAR missingness mechanism, one needs to simply choose the right observed (X) variables to include in the imputation model. This is where auxiliary variables come in, because they can play the role of observed (e.g., X) variables, which help to erase the leftover relationship between Y and $miss_{(y)}$.

Now, when looking at Table 3, we also see that one way to distinguish MNAR from MAR missingness is to notice whether the selection variable (which I have labeled *miss*) has been observed. In the most common scenario, the selection variable (*miss*) is only a hypothetical variable and has *not* been directly observed (i.e., when the missing data are not due to personnel selection or some other intentional missingness procedure, then the selection variable *miss* [which I am using to describe the missingness mechanism] has not been directly observed). In such cases, one purpose of auxiliary variables is to serve as an approximate surrogate for the unobserved selection variable, *miss*. As such, if one chooses auxiliary variables that are: (a) correlated with the hypothetical selection variable *miss* (i.e., auxiliary variables that are correlated with the probability of missingness on the substantive variables of interest) and also (b) correlated with the substantive variables of interest themselves (e.g., X and Y), then such auxiliary variables will go a long way toward helping convert an MNAR missingness mechanism into an MAR missingness mechanism.[11] This auxiliary variables procedure thus helps to remove missing data bias, because ML and MI approaches are unbiased under the MAR missingness mechanism.

One final issue with using auxiliary variables is that under the FIML approach, the auxiliary variables must be included in the estimation model (e.g., in the SEM model). On the one hand, if useful missing data "auxiliary" variables are the variables that tend to be correlated with X and Y (as well as with the hypothetical selection variable *miss*), then these so-called auxiliary variables might well make sense as control variables or as mediator variables in one's substantive regression or SEM model. On the other hand, if useful missing data auxiliary variables are truly *auxiliary* in the sense that they cannot be incorporated into the substantive theoretical model at hand, then procedures exist for including auxiliary variables in a FIML analysis without disturbing one's substantive model. Two such approaches were recommended by Graham (2003; i.e., the "extra dependent variables (extra DV)" approach, and what has come to be known as the "saturated correlates" approach), although the two approaches yield essentially identical results. In Appendix A, I provide LISREL syntax and R syntax for conducting multiple regression while implementing Graham's (2003) FIML procedure that involves specifying the auxiliary variables as "extra dependent variables" in one's analytic model (i.e., the "extra DV" procedure). This extra DV approach for incorporating auxiliary variables into FIML analyses is *highly recommended*, because it combines the advantages of FIML (FIML reduces missing data bias and gives accurate standard errors/more accurate hypothesis tests) with the advantages of auxiliary variables (auxiliary variables reduce missing data bias and increase statistical power).

Guideline 4: Item-Level Missingness—One Item Is Enough!

1. When conducting an item-level analysis (e.g., item-level factor analysis or item-level SEM), the analysis should be based on ML or MI missing data techniques.
2. When conducting a construct-level analysis, if a participant responds to any items (even a single item) from a multi-item scale, then the participant's average response across the item(s) answered should be used to represent the participant's scale/construct score.

Under ideal conditions, it would be nice if researchers could treat item-level missingness using the same practices that are recommended for construct-level missingness (see Guideline 3). That is, ideally one could use ML or MI missing data techniques to treat item-level missingness. I recommend that whenever possible, ML (i.e., FIML or EM algorithm) or MI techniques should be used when conducting item-level analyses such as item-level factor analysis, item-level SEM, and computing Cronbach's alpha. When such analyses involve hypothesis testing/significance testing (i.e., item-level SEM), then I recommend using FIML or MI when available; otherwise one should analyze the EM algorithm covariance matrix but should base the *SE*s on the minimum observed sample size in order to be conservative about hypothesis testing with the EM algorithm (i.e., emphasizing Type I error protection when using an EM covariance matrix with item-level analyses; see Enders & Peugh, 2004; cf. Savalei & Bentler, 2009).

In practice, however, using ML and MI techniques on item-level data is often difficult to do. One major problem is that ML and MI techniques can encounter difficulties converging when the number of variables exceeds 100—an issue that led Graham (2009) to conclude that the number of variables used with ML and MI missing data techniques should be kept to fewer than 100 when the sample size is large (over $N = 1{,}000$), and the number of variables should be kept even smaller when the sample sizes are smaller. Because of this issue, it is often much easier for the researcher to use a two-step procedure: (Step 1) First, combine (e.g., average) sets of items to form their respective composite scores representing each theoretical construct being studied, (which reduces the total number of variables to under 100), and then (Step 2) conduct ML or MI analyses on the construct-level scores (see Guideline 3). Step 2 (using ML or MI on the construct-level data set) is fairly straightforward (see Appendix A), but Step 1 (combining items into composite scale scores in the presence of item-level missing data) is less straightforward, as discussed below.

The problem is that, because ML and MI techniques do not always work for item-level missingness (i.e., because the number of items is large), then when forming scale composite scores from items the researcher must choose between two missing data treatments that are not state of the art: (a) listwise deletion cutoffs, versus (b) using the mean across available items. After describing these two approaches, I will then recommend using the mean across available items.

Listwise Deletion Cutoffs. When calculating scale composite scores for multi-item survey scales, it is relatively common practice to drop respondents from the analysis for a particular construct if they fail to respond to (approximately) half (or more) of the construct's scale items. This practice is widely taught in research methods graduate seminars, and has even been advocated by missing data experts (e.g., Graham, 2009, said, "forming a scale score based on partial data will be acceptable [a] if a relatively high portion of variables are used to form the scale score [and never fewer than half of the variables], p. 565)."

This commonly recommended practice—dropping construct scores if an individual fails to respond to at least half of the items for the construct—is nonetheless arbitrary, and it has the damaging effect of converting item-level missingness into construct-level missingness, by deleting actual data from respondents who do not finish at least half of the items for a particular construct. In other words, the practice of dropping respondents' construct scores when they do not complete most of the scale items violates the principle to *use all the available data*, and as such this practice is a particular form of item-level listwise deletion. I label this approach *listwise deletion cutoff* because a cutoff point (usually half of the items on the scale) is used to decide whether to delete the respondent's construct score.

Mean Across Available Item(s). A second approach is to calculate an individual's scale score for a multi-item scale by simply using the items that are available for that individual. This is like the practice that Roth, Switzer, and Switzer (1999) recommend, which they referred to as "mean substitution across items (and within an individual)" or "$\text{mean}_{\text{person}}$ imputation" (pp. 214, 222; also see Downey & King, 1998), although using the technique as I am describing it here (averaging across the subset of scale items with available responses for each person to calculate each person's scale score) does not technically involve any imputation (i.e., at no point am I replacing any missing values with a "good guess").

Choosing Between Listwise Deletion Cutoffs versus Using the Mean Across Available Items. When making a choice between the aforementented two strategies for addressing item-level missing data, one must attempt to choose the lesser of evils (neither approach is ideal). Both techniques work better when the items on the scale are parallel (Newman, 2009; that is, when scale items are approximately interchangeable and do not have grossly differing means or factor loadings), as well as when the available items are good representations of the content domain and when Cronbach's α is relatively high (Graham, 2009). Also, when an individual has responded to most of the scale items, then the two techniques (using mean across available item[s] and listwise deletion cutoffs) are identical—the difference between the two approaches only affects the rarer cases, for whom the number of available items (for an individual) falls below the listwise deletion cutoff. In other words, the listwise deletion cutoff method is a special case of the mean across available items method—the available items are being used in both methods, except that the former method opts for listwise deletion when the item-level response rate is low (i.e., it has a cutoff).

Strictly speaking, when using the mean of available items method, then an individual's answers to item(s) should be used to represent that individual's construct score, even if the individual responds to only a single item from the scale. This is what I mean by the phrase *one item is enough* for calculating a scale composite score for individuals who have item-level missing data. The alternative is to discard this person's data from the analysis altogether (i.e., the listwise deletion cutoff method), which is less defensible on theoretical and ethical grounds and—as I discuss next—is typically less defensible on statistical grounds as well.

Importantly, neither one of these two techniques for dealing with item-level missingness is unbiased under MAR or MNAR, and it is not clear that one technique is more biased than the other. Thus, the choice between the two approaches to item-level missing data must be made using another criterion. In particular, I recommend distinguishing between these two options based on statistical power (i.e., avoidance of Type II error).

Item-level missing data harms statistical power under both alternative methods, but in different ways and to differing degrees. For the listwise deletion cutoff method, the researcher is discarding individuals from the analysis, which impairs power by reducing the sample size. For the mean of available items method, the fact that individuals with fewer item responses are still included in the analysis means that those individuals' scale scores are less reliable on average due to their use of fewer items (see Spearman-Brown prophecy formula—having fewer items leads to lower reliability of the scale composite score). The inclusion of individuals who are using a smaller number of items (and thus who have less reliable measures) then attenuates the observed effect size (e.g., correlation), which in turn also reduces statistical power. (Recall that listwise deletion cutoff methods will also bias the observed effect size whenever the item-level missingness is not MCAR.) So item-level missingness harms statistical power, regardless which technique is used (listwise deletion cutoffs vs. mean across available items). Because a thorough treatment of this issue is unfortunately beyond the scope of the current review, I will suffice to say that the statistical power compromise caused by dropping respondents (listwise deletion cutoff method) is, under typical conditions, worse than the statistical power compromise caused by including partial respondents who only answered a subset of the items (mean of available items method)—even under the extreme case when only one item has been answered. As a result, I recommend the use of the mean of available item(s) method, and I discourage the commonly used listwise deletion cutoff method. Both methods tend to suffer bias under MAR and MNAR missingness mechanisms, but the mean of available items method typically offers greater expected statistical power.

Extreme Items. One final issue with item-level missingness involves the possibility that some items with missing data are *extreme items*—namely, items with especially high or low endorsement rates, compared to the other items on the multi-item scale. For the most part, these items are a rarity on validated scales, and differential missing data on these items is an even greater rarity; so extreme items will be unlikely to make a practical difference in the vast majority of data analyses. For those rare scales on which extreme items do exist, I provide advice for dealing with missing data on extreme items in Appendix B.

Guideline 5: Person-Level Missingness: If the Response Rate Is Below 30%, Report Systematic Nonresponse Parameters and Consider Conducting Sensitivity Analyses

As shown in Tables 2 and 3, there are available missing data treatments (ML and MI techniques) that are unbiased under MAR (systematic) missingness mechanisms. However, person-level missingness cannot be MAR. This is because the MAR mechanism requires that the probability of missing data be predictable by observed variables, and for individuals with person-level missingness, there are no observed variables. This is a big problem because state-of-the-art missing data treatments (ML and MI) were designed to be unbiased under MAR missingness, and therefore the advantages of these techniques over listwise and pairwise deletion disappear in the case of person-level missingness. As mentioned previously, this situation makes person-level missingness quite difficult to handle, because we have no available missing data techniques that can yield unbiased parameter estimates.

For facing person-level missingness (i.e., high nonresponse rates), there are no great guidelines I can offer that are able to root out missing data bias and error. The best option for researchers facing person-level missingness at this point is to attempt to report all the relevant information that might be useful for aiding future readers in understanding the degree of nonresponse bias likely present in a particular set of results.

If the response rate is especially low (below 30%), then the researcher should attempt to provide information that can be used to gauge the likely amount of missing data bias in the parameter estimates. Because missing data bias is a function of the response rate and the systematic nonresponse parameters ($r_{miss,x}$; see Newman & Cottrell, in press, and Equation 1), researchers with especially low response rates should provide three pieces of information:

1. Report the overall response rate (i.e., [n full respondents + n partial respondents] / n contacted; see Equation 2). (This advice was also given as part of Guideline 3.)
2. Report the systematic nonresponse parameters (e.g., $r_{miss,x}$, $r_{miss,y}$) pertaining to each substantive variable in the study, if possible.

According to Newman (2009), *systematic nonresponse parameters* (SNPs) capture the difference between respondents and nonrespondents on the variables of interest in a particular study. For example, Newman and Sin (2009) provided an expression for an SNP called d_{miss}:

$$d_{miss} = (\bar{X}_{nonrespondents} - \bar{X}_{respondents}) / s_{pooled}, \tag{9}$$

which can be equivalently expressed as:

$$r_{miss,\,x} = d_{miss} \Big/ \sqrt{d_{miss}^2 + 1/p(1-p)}, \tag{10}$$

where p is the response rate.

These SNPs can be used to index whether the person-level missingness mechanism is random (MCAR; where $r_{miss,x} = 0$) versus systematic (MAR or MNAR; where $r_{miss,\,x} \neq 0$). The closer the missingness mechanism is to completely random (MCAR), the less concern there is about missing data bias in one's results.

As an example, Newman (2009) has quantitatively summarized a few studies of survey respondent-nonrespondent differences conducted by Rogelberg and colleagues (Rogelberg et al., 2003; Spitzmuller, Glenn, Barr, Rogelberg, & Daniel, 2006), to estimate that the average systematic nonresponse parameter is approximately $d_{miss} = -.40$ ($r_{miss,x} \approx -.2$) for variables like perceived fairness, conscientiousness, and agreeableness personality traits; and $d_{miss} = -.14$ ($r_{miss,x} \approx -.07$) for variables like job satisfaction/attitudes, perceived organizational support, and turnover intentions (although for turnover intentions, d_{miss} has a positive sign). That is, this initial evidence suggests that survey respondents tend to be more satisfied, feel more supported, and have lower turnover intentions than do survey nonrespondents (suggesting weak systematic missingness for job attitudes and behavioral intentions); but respondents also tend to have much higher fairness perceptions, conscientiousness, and agreeableness than nonrespondents (suggesting stronger systematic missingness for personality traits and justice perceptions).

Comparing these tentative estimates of $r_{miss,x}$ (which are relatively small and primarily have the same [negative] sign) against Figure 5, the conclusion can be drawn that personality-attitude correlations and personality-behavioral intentions correlations often suffer fairly small amounts of missing data bias due to person-level missingness (i.e., due to low response rates), whereas attitude-attitude correlations and attitude-behavioral intentions correlations suffer even smaller and near-zero missing data bias due to person-level missingness. Altogether, these early results for systematic nonresponse parameters ($r_{miss,x}$) support the approximate viewpoint that missing data bias due to person-level missingness (i.e., low response rates) is typically small or negligible, and tends to produce underestimation of the relationships among constructs. To restate the previous guideline, because the expected amount of nonresponse bias due to person-level missingness will depend on which substantive constructs are being studied (e.g., nonresponse bias will be higher for studies of conscientiousness and justice perceptions), researchers should attempt to report data on respondent-nonrespondent differences ($r_{miss,x}$) for each substantive construct being studied. These estimates will usually need to come from other, nonlocal studies that have compared respondents against nonrespondents (e.g., Rogelberg et al., 2003; Spitzmuller et al., 2006; see Newman, 2009).

3. Where possible, conduct response rate sensitivity analyses by estimating the response rate–corrected correlations using Equation 1.

When the key inference from one's study relies on a particular correlation or a particular set of relationships among three variables (e.g., a mediation test), then it would be useful to calculate response rate–corrected versions of these two or three important correlations (using Equation 1; Newman & Cottrell, in press). That is, the $r_{miss,x}$ values collected in response to the aforementioned recommendation can then be plugged into Equation 1 to yield response rate–corrected correlation estimates. (Note that for the common case where $r_{miss,x}$ is unknown, one can simply try a realistic range of $r_{miss,x}$ values—I recommend using $r_{miss,x}$ values between 0.0 and –0.2, consistent with Newman's [2009] small-scale review, described previously. Also note that Equation 1 requires the response rate for a given study to be transformed into a u^2 estimate; Newman & Cottrell, in press.)[12] These corrected correlation estimates from Equation 1 can then be used to perform a simple response rate sensitivity analysis to demonstrate that the study's key result (e.g., a bivariate correlation or a mediation parameter[13]) still obtains even after making rough corrections for the low response rate.

Such simple sensitivity analyses are primarily useful because they help to indicate the direction of the missing data bias due to person-level missingness (i.e., Is the parameter of interest likely to be underestimated vs. overestimated due to the low response rate?). I surmise that a large portion of effect sizes in the psychological literature are likely to be *underestimated*—not overestimated—due to low response rates (see Newman & Cottrell, in press). This surmise is based on the fact that many of the known $r_{miss,x}$ estimates in the psychological and organizational sciences are negative (Newman, 2009; i.e., respondents have more positive attitudes and personalities and lower turnover intentions compared to nonrespondents) and thus have the same sign as each other, which would suggest that missing data bias typically leads to small negative bias (usually underestimation) of one's theorized parameters (see Figure 5; Newman & Cottrell, in press).

Finally, I note that Guideline 5 is the most tentative of the five guidelines I have presented in the current article. This is because, to a realistic extent, our science still does not have very good solutions to offer that can address person-level missingness. Guideline 5 is an early attempt to do something to acknowledge the issue of response rate bias—rather than simply ignoring the problem or simply rejecting all manuscripts that are based on low response rates. My choice of a 30% response rate cutoff for Guideline 5 is arbitrary (indeed, nonresponse bias can matter at much higher response rates too), but it roughly corresponds to the 20th percentile of response rates found in organizational research (Anseel et al., 2010). The idea here is to present missing data guidelines that are practical (cf. requiring everyone with less than perfect response rates to conduct nonresponse bias sensitivity analyses seems impractical, given the nascent state of the current science for precisely estimating and using the systematic nonresponse parameters [e.g., $r_{miss,x}$], which are a necessary part of the sensitivity analyses). As such, Guideline 5 only applies to the most egregious instances of person-level missingness (when the response rate falls below 30%).

Conclusion

The five practical guidelines offered in the current article are built upon statistical theory (see reviews by Allison, 2002; Enders, 2001b, 2010; Dempster, Laird, & Rubin, 1977; Little & Rubin, 2002; Newman, 2003; Rubin, 1976, 1987; Schafer, 1997; Schafer & Graham, 2002), but the guidelines themselves are *practical* guidelines and not intended to be statistically exact. That is, I am offering a set of compromised standards that are midway between current research practice (e.g., in which listwise and pairwise deletion are routinely implemented) and statistical best practice (e.g., in which one could likely insist that all data analyses ought to be based on FIML). In an attempt to propose a set of missing data standards on which most researchers can generally agree, the compromise is that I am only recommending state-of-the-art missing data routines (ML and MI) be used in those instances when they are likely to make the biggest difference (e.g., when the percentage of respondents who are partial respondents >10%).

If the five practical guidelines were followed, it would represent a big step forward in the accuracy with which results are reported in the social sciences (both in terms of less biased effect size estimates and more accurate hypothesis tests). The decision rules involved in using the five practical guidelines articulated here are designed for the purpose of assisting researchers who want to choose the lesser of evils among missing data treatments, under the types of missing data conditions typically found in the social and organizational sciences. Because the guidelines are a decision aid, they are forced to somewhat arbitrarily convert a set of continuous phenomena into a binary decision tree (Figure 1). More research would still be useful on a wide variety of imaginable boundary conditions under which the various missing data techniques might have different degrees of relative performance (e.g., under violations of normality [Enders, 2001a; Gold & Bentler, 2000; Gold, Bentler, & Kim, 2003], small sample size conditions [Graham & Schafer, 1999], nonlinear missing data patterns [Collins et al., 2001; Roth et al., 1999], or in multilevel models [Mistler, 2013; van Buuren, 2011]). Under the current state of scientific knowledge (Enders, 2010; Graham, 2009; Schafer & Graham, 2002), though, following the five guidelines would produce immediate and palpable improvements in the accuracy and believability of research results. This is because research results would no longer narrowly apply only to individuals who respond completely to surveys—results would instead generalize to a target population including both full survey respondents and partial survey respondents, without introducing unnecessary bias and error that can be caused by listwise deletion, pairwise deletion, and single imputation.

Appendix A

Annotated Syntax (in SAS, LISREL, and R) for Maximum Likelihood (expectation-maximization [EM] algorithm, full information maximum likelihood [FIML]) and Multiple Imputation

For most research projects involving correlation and multiple regression, the following code labeled "R syntax for FIML and EM algorithm" will directly and easily provide all the estimates the researcher needs (i.e., ML [EM] correlation matrix, ML [EM] means, ML [EM] standard deviations, ML [FIML] regression coefficients, and ML [FIML] accurate standard errors for significance tests).

Brief description of annotated syntax:

SAS syntax:
(A) Listwise Deletion (Multiple Regression) ** This is the SAS default missing data routine for multiple regression, but it is NOT RECOMMENDED (see Guideline 1).
(B) Pairwise Deletion (Correlation and Multiple Regression) ** This is the SAS default missing data routine for correlation, and it can also be used for multiple regression; but it is NOT RECOMMENDED UNLESS THE PORTION OF CONSTRUCT-LEVEL MISSINGNESS IS <10% (see Guideline 3).
(C) EM Algorithm (ML missing data routine) (Correlation and Multiple Regression) The EM algorithm calculates the covariance/correlation matrix and vector of means. ** This is the RECOMMENDED PROCEDURE for CALCULATING A CORRELATION MATRIX, MEANS, AND STANDARD DEVIATIONS. One can also conduct multiple regression using the EM covariance/correlation matrix. ** This provides least biased regression coefficients, but SEs are still inaccurate (no single sample size makes sense for the entire correlation matrix). So if this technique is used for hypothesis testing, conservative minimum-*N* procedures are recommended to control Type I error (Enders & Peugh, 2004).
(D) Multiple Imputation (MI) (Multiple regression) ** This is a RECOMMENDED PROCEDURE for conducting multiple regression and structural equation modeling (SEM; use auxiliary variables and relevant interaction terms in the imputation model).
LISREL syntax and R syntax (lavaan R package):
(E) FIML: Full Information ML (ML missing data routine) (Correlation and Multiple Regression) One can conduct multiple regression using FIML by treating multiple regression as a special case of SEM (e.g., in LISREL or in R [lavaan package]). Both the LISREL and R syntax provided below use Graham's (2003) "Extra DV" auxiliary variable method.The ML covariance/correlation matrix and the ML means are also output by both the LISREL syntax and the R syntax below. These are the exact same as the EM algorithm ML covariance/correlation matrix and means. ** This is a RECOMMENDED PROCEDURE for conducting multiple regression and SEM (use auxiliary variables in the estimation model, via the extra DV method [Graham, 2003]). ** This is also the RECOMMENDED PROCEDURE for calculating the CORRELATION MATRIX, MEANS, AND STANDARD DEVIATIONS (i.e., the ML covariance/correlation matrix and means are the exact same as the EM covariance/correlation matrix and means).

This appendix provides syntax intended for use with analyses based on covariance/correlation matrices (i.e., multiple regression, factor analysis, and SEM). The specific examples involve correlation and multiple regression.

Also, when implementing analyses based on an EM algorithm correlation matrix, I recommend recording the correlations to at least five decimals (i.e., to limit error due to rounding).

SAS, LISREL, and R Syntax for Missing Data Analysis (Multiple Regression and Correlation)

SAS Syntax:

Enter the dataset, using a dot '.' to represent missing data.

```
*INPUT RAW INCOMPLETE DATA;

data INCOMP;                    *Label the dataset 'INCOMP';
input y x z aux1 aux2 @@;       *5 variables—y, x, z, plus 2 auxiliary vars.;
cards;
 3.04 3.31 4.32 2.42 3.20     .  2.81 2.41 2.59 2.77    2.32 2.78 2.93 3.28 2.39
  .   1.05  .   2.23 0.85     .  0.31 5.10 0.83 0.37     .   3.51 3.69 2.68  .
  .    .   4.26 1.29 2.34     .  3.07 4.11 3.53 1.27     .   2.27 4.19  .   1.54
  .   4.36 3.41 2.98 2.93     .  2.59 3.74  .   1.45    4.22 3.93 3.49  .   3.10
  .   2.27 3.54 2.68 3.47     .  2.26 4.38 2.00 1.65    4.33 3.78 1.75  .   4.47
  .   2.90 3.29 2.49 2.36     .  1.33 4.47 0.75 2.38    4.98 3.97 2.27 3.56  .
  .   3.61 1.38  .   2.92     .   .   2.00 4.20 2.39     .    .   2.47 3.16 1.78
 4.05 4.84 1.61 5.58 2.56     .   .    .    .   3.76     .   2.81  .   3.81 2.88
  .   3.23 2.69  .   2.88   4.94 2.19 3.62 3.24 3.25    3.16 3.27 2.54  .   3.33
 3.55 3.92 4.62 2.92 3.43   2.69 3.48 1.42 2.51 3.25     .   3.04 5.17 2.11  .
 6.38 3.78 2.87 4.13  .       .   .    .    .   1.73    2.16 3.25 3.28 2.60 2.78
 3.41 2.30 3.45 2.38 3.15     .  3.32  .    .   2.45     .   3.03 4.30  .   1.91
  .   2.57 2.75 2.36 3.03   4.64 5.07 1.10 5.43 4.04     .   3.14 4.24 2.18 2.97
 2.75 3.87 3.01 3.53 2.73   3.32 3.55 2.93  .   2.46     .   3.16 2.34 2.57 2.74
  .    .   2.32 3.66 3.06   5.08 4.02 3.27 4.68 3.26    2.95 2.59 3.53  .   2.76
  .   1.22 3.82 3.59 1.19     .  2.54  .   2.26 1.71     .   2.37 3.15 2.60 2.01
 3.78 3.67 1.15  .   3.32     .  1.80 3.52 2.00 1.71    3.37 2.55 2.56  .   3.97
  .   2.41 2.99 1.48  .     3.18 2.75 2.16 2.74 4.31     .    .   2.57 3.67  .
 3.37 4.21 3.78 2.92 3.72   3.36 2.92 2.98 4.02 2.28     .   0.98 2.68 1.20 3.17
 3.13 3.04 1.25  .   3.02     .  0.10 4.29 0.84 0.20     .   2.92 2.58 3.04 2.37
  .    .   2.14 2.45  .       .   .   4.39 2.50  .      3.01 2.10 1.78 2.32  .
  .    .   3.54 2.63 3.64     .   .    .   1.89  .       .   1.08 3.35 2.09 1.69
 3.79 3.12 3.79 2.78 4.65     .  2.24 1.29 4.25 2.65    2.65 4.19 3.72 2.86  .
 4.67 5.57 1.78  .   3.93   2.65 4.56 1.40 4.25  .      3.60 2.32 2.65  .    .
 2.62 2.76 2.73 2.94 2.69   3.06  .   2.43 4.00 3.61    2.96 3.17 3.48 3.42 2.45
 4.47 3.28 3.16 3.73  .       .  1.24 4.03 2.50 1.06     .    .   3.56 3.19 2.78
  .   3.66 3.85 3.75  .     3.19 3.13 4.35 2.69 2.50     .   3.72  .   1.17 3.93
 2.33 5.21 2.45 4.46 3.59   3.62 4.33 2.18 3.58 4.41     .   2.85 4.14 3.05  .
  .   3.10 2.88 2.79  .     3.00 2.93 3.58 2.33 3.32     .   3.18 3.03 3.15 1.34
  .   2.58 2.42  .    .     3.73 1.88 2.07 3.21 4.10     .   1.00 2.67 2.96 1.51
 3.64 4.50 1.76  .    .       .  2.58 5.27  .    .       .   2.71 3.51 2.11 1.48
 4.27 3.88  .   4.12 3.96   2.45 1.33 4.10  .   2.99     .   4.03  .   5.28  .
 3.20 4.42 2.18 2.61 5.20     .   .   4.06 3.89 2.81    4.55 4.17 2.73 4.91  .
 3.61 4.18 3.85  .   4.02   3.06 3.97 3.13  .   3.39     .   2.28 6.34 0.73 2.11
  .   2.11  .   2.70 2.57     .  3.95  .   1.89 2.86    1.93 3.93 2.73 3.03 3.25
 3.07 2.39 3.40 3.54 2.17     .  1.77 1.69 2.77 3.26    3.46 3.19 3.03 3.70 2.32
 2.62 3.48  .   2.53 3.03     .  1.15  .   1.31  .       .    .   6.13 0.34  .
 3.36  .   2.52  .   2.93   3.19 4.31 2.49  .   4.56    3.09 4.40 2.34 4.39 2.94
 2.36 3.42 2.61 2.95 3.40   4.43  .   1.41 2.46 6.53    3.84 3.26 1.85 3.21 4.10
 3.23 2.35  .   4.60 2.87     .  4.70 3.90 2.80 2.52     .   3.23 3.87 3.51 2.24
  .   3.37  .   2.06 2.76     .  2.01 3.85 2.14 1.77    3.68 3.95 2.90 3.18 3.89
  .   3.65 3.34 1.95 3.43   3.47 3.08 2.63 5.07 2.29     .   2.99 2.25 2.84 3.31
 3.00 5.53 2.35 2.68 5.22   1.98 3.71 2.91 3.86 2.38    3.10 2.80 3.56  .   3.73
 3.00 3.05 3.57 4.28 2.01     .  2.17 3.49 3.06 0.99     .   1.65 3.81 2.40 1.42
  .   2.33 3.35 2.20 2.01   3.14 5.40 3.88 3.59 3.61     .   2.08 3.44 2.96 2.29
  .   2.82 4.15  .   3.15   3.04 3.04 3.31 3.53 2.24     .   3.84 3.40 2.99 2.23
 4.43 3.47 1.10 4.45 3.54     .  2.33 3.59 3.22  .       .    .   3.38 2.52 3.04
  .    .   3.26 3.22 1.85   3.80 2.59 3.70  .   3.45    3.77 3.73 2.38 2.68 4.63
 3.56 3.24 3.09 3.56 3.43     .  2.79 3.59 3.17 2.55    3.88 3.73 1.52 4.68 3.10
  .   2.04 3.48 1.95 2.06     .  2.78 4.91 2.63  .      4.18 3.68 2.86 3.25 4.10
 4.31 3.41 2.65 3.58 3.53     .   .    .    .   1.81     .   3.48 2.16 3.49 2.76
 3.49 5.13 2.80 3.44 4.27   3.60 0.91 1.58 3.01 2.84     .   3.25 2.92  .   2.06
```

```
   .  1.78 1.34 2.07 3.09   2.04 3.15 3.68 1.93 3.38     .  2.22 5.38   .  1.19
 3.47 2.38 2.21 3.09 3.49     .  2.40 2.83 2.51 1.75     .  2.29 2.11 3.13 3.79
   .  3.45   .  3.20 3.27     .    .  2.53   .  3.29   4.51 3.78 1.99   .  3.07
 2.40 1.64 1.74   .    .      .  1.52 3.61 3.77 0.75     .  2.30 2.04 2.27 4.18
   .  3.19 3.41   .  2.28   3.67 3.21 3.55 1.99 4.18   3.18 4.88 1.37 4.27 3.84
 4.82 4.93   .  4.64 5.31   5.07 3.05 2.30 4.30 4.14   2.95 3.65 4.61 2.15 3.28
 2.83 3.31 5.37   .    .      .  1.04 3.09 2.28 2.33   3.38 0.84 3.10   .  1.50
   .  1.61 1.70 2.80 1.71   4.20 2.56 2.48 3.91 3.44   3.41 3.03 2.43 3.51 3.13
 3.35 4.13 1.99   .  3.50   1.87 3.18 2.98 1.81 3.71   3.21 3.06 1.86 2.93 4.20
 3.15 3.60 1.52 4.54 3.03   3.49 3.30 3.90 3.71   .      .  2.60 3.63 1.94   .
 2.75 4.33 2.49 2.92 3.66   3.25 2.99   .  3.51   .    3.11 3.45 2.46 5.38 1.92
 3.62 4.37   .    .  3.31   3.09 3.00 3.42 3.46 2.20   3.92 3.36 4.32   .  3.17
   .  3.19 1.94   .  3.15     .    .  3.25 1.39 3.14     .  2.61 4.49 1.17 1.09
 3.03 2.93 2.24 2.53 2.81     .  2.41 2.82 1.51 3.81     .  2.75 2.56 3.16 2.56
   .  2.26 4.38 2.50 2.52     .    .  2.12 2.59 1.96   3.02 1.61 3.12 1.39   .
   .  1.31 3.61 1.12   .      .  3.23   .  1.68 2.81     .  3.00   .  2.26 4.70
   .  2.44   .  2.78   .      .  3.38 4.94 2.77 1.82     .  2.20 3.10 1.42 2.93
 2.51 3.13 3.13 2.55 3.54   2.53 3.11 1.24 1.97 4.05   2.92 2.85 2.74   .  2.53
 3.62 2.54 2.85 3.57 2.71     .  1.89 2.87   .    .      .  2.28 4.71 1.95 1.80
 5.07 3.91 2.25 3.88 4.63   2.90 3.19 3.58 2.68 2.94   4.24 3.63   .  2.94 4.72
 3.81 1.62 2.58 3.50   .      .  4.03 3.17 2.14 3.53   2.54 2.60 2.95 3.03   .
 3.85 2.44 2.02 3.96 3.09     .  1.90 3.17   .  2.12   3.11 1.93 2.99 2.29 2.66
 3.01 2.40 3.80 3.59 2.24     .  2.49 3.45 2.45 2.51     .  2.12   .    .  2.57
 3.43 4.48 1.11 3.96 3.92     .  1.31 2.96 2.25 2.42     .    .  3.89   .  2.58
   .  2.40 4.08 1.47 2.82     .  3.14 3.05 3.20 2.46     .  4.42   .    .  3.35
   .  3.10 3.57 2.44 3.00     .  3.24 4.18   .  2.38   4.42 2.75 4.05 3.19 2.67
   .  2.42 2.93 0.90 2.82   5.18 5.48 2.09 5.25 4.64     .  1.80 4.58 2.09 1.68
   .    .  2.28 2.75 3.86   4.27 1.10 2.98 2.03 4.20     .    .    .  4.74 3.54
 3.57 5.21 3.06 3.42 4.34   3.00 0.80 2.90 2.06 2.99     .  3.64   .  2.74   .
 3.95 3.26 2.65 3.76 3.60     .    .    .  3.67 3.90   3.32   .  3.69 3.65 2.82
 3.40 4.60 3.08   .    .    4.75 3.00 2.89 3.73   .    2.79 2.73 1.60 2.97 3.18
   .  2.20 2.44 2.65   .      .  2.53 3.34 2.28 2.53   5.15 2.83   .  5.73 3.63
   .  2.41 3.36 2.93 3.12     .    .    .  2.08 1.98   3.94 4.39 1.99 4.10 3.47
   .    .  2.69 2.58 2.12     .  2.77 2.48 1.93 2.15     .  2.51 3.33 2.75 2.53
 4.42 2.31 1.69 3.30 3.66     .  2.25 4.53 2.51 1.96     .    .  4.48 1.39 2.86
 3.26 3.39 3.47 2.59 4.07   3.18 4.08 0.98   .  3.91     .  2.48 4.43 2.51 2.54
   .  2.10   .  1.87 3.07     .  3.92 4.31 2.25 3.25   3.32 3.33 3.24 3.75 1.86
   .  2.74 3.67 2.43 2.08     .  3.65 4.56 2.19 2.73   4.61 3.91 2.10 4.14 4.55
   .  1.69   .  2.43 3.31     .  2.81 2.78 3.77 1.91   3.51 3.11 2.46 3.28 4.08
 4.51 2.01 2.36 4.88   .      .  3.98   .  2.97 2.12     .  2.93 3.54 2.58   .
   .    .  2.86 2.52 3.24   3.00 3.35 1.68 3.93 3.49   3.93 4.58 2.61 3.58 4.29
 3.87 2.60 3.24 4.10   .      .  2.33   .  1.57 2.19   2.76   .  2.19 3.63 2.60
 2.92 4.10 1.90   .    .    5.42   .  2.93 4.12 3.13     .  1.44 3.29 2.14 2.86
   .  1.34 2.63   .  3.46     .    .  5.05 1.54 1.98   2.49 3.26 2.95 4.22   .
 3.09 2.10 3.60 1.65   .    4.60 4.17 1.04 5.45 4.00     .    .  2.53 3.71 2.40
 4.46 3.69 0.79 3.29 5.20   1.66 2.85 2.07 2.83 2.64   3.72 2.52 2.11 2.88 4.71
   .  2.78 2.78 3.16 2.87     .  1.34 6.48 0.61 1.40   5.06 3.62 4.48 3.45 3.95
 3.53 3.19 1.83 3.50 3.77     .  3.92 4.38 3.19 2.40     .  1.76   .  0.59 3.20
;

proc print data=INCOMP; run;  *Check that the dataset was entered correctly;
```

(A) Listwise Deletion

```
*LISTWISE DELETED MULTIPLE REGRESSION;

Proc reg data=INCOMP;          *Use incomplete dataset 'INCOMP' entered above;
model y = x z/STB;             *Estimate regression model with 2 predictors;
run;
```

(B) Pairwise Deletion

```
*PAIRWISE DELETED MULTIPLE REGRESSION;

proc corr data=INCOMP outp=PAIRWISE;  *Create Pairwise Deleted corr. matrix;
var   y x z aux1 aux2;
run;

proc print data=PAIRWISE; run; *Display Pairwise Deleted correlation matrix
and variable means, SDs, and pairwise sample sizes;

Proc reg data=PAIRWISE (type=CORR); *Input Pairwise Deleted corr. matrix;
model y = x z/STB;                  *Estimate regression model on pairwise-
deleted corr. matrix;
run;
```

(C) EM Algorithm
(gives ML estimates, but conservatively uses minimum N for hypothesis testing)

```
* EM ALGORITHM MULTIPLE REGRESSION, USE SUBSTANTIVE MODEL VARIABLES (Y, X, &
Z) AND AUXILIARY VARIABLES (Aux1 & Aux2);

proc mi data=INCOMP nimpute = 0 seed = 51075 simple;
var   y x z aux1 aux2;
em outem = EMCOVS;            *Create EM Algorithm covariance matrix;
run;

proc print data=EMCOVS; run; *Display EM Algorithm covariance matrix and
vector of variable means. To get EM correlation matrix, look at Proc Reg
EMCOVS statement below;

*Use minimum N with the EM covariance matrix, for Type I Error protection
(Enders & Peugh, 2004). This involves setting N for the EM cov. matrix to
minimum pairwise N (in the current example, minimum N = 146);
Data N_for_EMCOVS;
input _TYPE_ $ y x z aux1 aux2;
cards;
N 146 146 146 146 146
;

Data EMCOVS_N; set EMCOVS N_for_EMCOVS; run;

Proc reg corr simple data=EMCOVS_N (type=COV); *Input EM covariance matrix,
and also display EM correlation matrix, means, and SDs;
model y = x z/STB;                  *Estimate regression model on EM
algortithm covariance matrix;
run;
```

(D) Multiple Imputation

```
* MULTIPLE IMPUTATION, USE SUBSTANTIVE MODEL VARIABLES (Y, X, & Z) AND
AUXILIARY VARIABLES (Aux1 & Aux2);

proc mi data=INCOMP nimpute = 40 seed = 51075 out = IMPUTED noprint; *Creates
40 imputed datasets, and saves them to a file called 'IMPUTED';
var   y x z aux1 aux2;
em outem = EMCOVS;      *Proc MI routine automatically provides EM algorithm
```

```
covariance matrix;
mcmc nbiter = 100 niter = 100; *Specify number of burn-in iterations (nbiter)
and number of iterations between imputations (niter, see Enders, 2010);
run;

proc sort data = IMPUTED;      * Sort data by imputation, from 1 to 40;
by _Imputation_;
run;

* ESTIMATE REGRESSION MODELS;

proc reg data = IMPUTED outest = regparms covout noprint; *Input 40 imputed
datasets, and run regression on each of the 40;
model y = x z/STB;
by _Imputation_;
run;

proc print data=regparms; run;  * Display regression results from each of the
40 imputations;

proc mianalyze data = regparms; * Combine regression coefficients and SEs;
modeleffects intercept x z;
run;

* MULTIPLE IMPUTATION PROCEDURES PROVIDE UNSTANDARDIZED REGRESSION
COEFFICIENTS. TO OBTAIN STANDARDIZED REGRESSION COEFFICIENTS, MULTIPLY EACH
UNSTANDARDIZED COEFFICIENT BY SDX/SDY, WHERE THE SD'S ARE TAKEN FROM THE EM
ALGORITHM COVARIANCE MATRIX (I.E., SD = SQUARE ROOT OF VARIANCE TERM IN THE
DIAGONAL OF THE EM COVARIANCE MATRIX);

*TO GET CHANGE IN R-SQUARED WITH MULTIPLE IMPUTATION;
*If change in (pseudo)R-squared estimate is negative, report as zero;

proc reg data = IMPUTED outest = regparms covout noprint;
model y = x z; *Estimate regression model with full set of predictor
variables for which you want the R-squared change;
by _Imputation_;
run;

proc print data=regparms; run; *Display 40 regression results;

data Rsq; set regparms;
if _Type_ = 'PARMS';
MSE=_RMSE_**2;    *Calculate and save the full-model mean squared error;
keep _Imputation_ MSE;

proc reg data = IMPUTED outest = regparms covout noprint;
model y = ; *Drop the predictor variables for which you want the R-squared
change;
by _Imputation_;
run;

data Rsq_without; set regparms;
if _Type_ = 'PARMS';
MSE_without=_RMSE_**2; *Calculate and save the submodel mean squared error;
keep _Imputation_ MSE_without;
```

```
data RsqCHANGE; merge Rsq Rsq_without; *Combine the 2 MSE files;
by _Imputation_;
RsqCHNG=1-(MSE/MSE_without); *Compute R-squared change for each of the 40
imputed datasets;

proc print data=RsqCHANGE; run; *Display R-squared change for each dataset;

proc means data=RsqCHANGE; *Display mean R-squared change across imputed
datasets;
var RsqCHNG;
run;
```

(E) FIML (gives ML estimates)

LISREL Syntax for FIML and EM algorithm:	Annotation
FIML Multiple Regression Example, Simulated data	!Label the dataset
da ni=5 no=300 ma=cm mi='9999'	!ni=# of variables, no=sample size N, cm=analyze covariance matrix !mi designates the missing data code in raw data file (must be numeric; '9999')
la Y X Z Aux1 Aux2	!Label the variables
ra fu fi 3.04 3.31 4.32 2.42 3.20 9999 2.81 2.41 2.59 2.77 2.32 2.78 2.93 3.28 2.39 9999 1.05 9999 2.23 0.85 … 9999 1.34 6.48 0.61 1.40 5.06 3.62 4.48 3.45 3.95 3.53 3.19 1.83 3.50 3.77 9999 3.92 4.38 3.19 2.40 9999 1.76 9999 0.59 3.20	!Enter raw data. This is the same dataset used with the SAS example above, except that the missing data code in LISREL is now '9999', instead of the '.' used in SAS !Only a portion of the dataset is shown here, to save space
se 1 4 5 2 3/	!Select the [5] variables to be used in regression analysis, in order [DVs first: Y, Aux1, Aux2; followed by X, Z]
mo ny=3 nx=2 ne=3 nk=2 ly=di,fi lx=di,fi ps=sy,fr ph=sy,fr ga=fu,fr te=ze td=ze al=fu,fr ka=fu,fr	!Model statement specifies multiple regression with 3 outcome variables (ny and ne) and 2 predictor variables (nx and nk). Multiple regression is a saturated (df=zero) manifest variable model [i.e., diagonal factor loading matrix with factor loadings (ly and lx) fixed at 1.0, uniquenesses (te and td) fixed at zero, all covariances (ps and ph) freely estimated, and all regression coefficients (ga) freely estimated]. Alpha parameters are estimated to give regression model intercept.
le Y Aux1 Aux2	!Label the outcome variables. Note that the auxiliary variables are modeled as Extra DVs (Graham, 2003).
lk X Z	!Label the predictor variables.

```
st 1.0 ly 1 1 ly 2 2 ly 3 3   !For manifest variable model, factor
st 1.0 lx 1 1 lx 2 2          loadings are fixed at 1.0.
pd                            !Display path diagram.
ou sc nd=3 AD=OFF             !Display standardized output too, and
                              use 3 digits after the decimal.
```

R Syntax for FIML and EM algorithm:

R is a free software that can be downloaded at http://www.r-project.org/ .
Select Packages > Install package(s) > lavaan . (If prompted, allow R to create a library.)
(The lavaan package in R performs 'latent variable analysis' (i.e., SEM), similar to LISREL, Mplus, or EQS.)

After opening R, in the R Console window, type in `library(lavaan)` .
Copy the following syntax, and paste it into the R Console window:

```
INCOMP<-matrix(c(             #Enter raw data as a matrix, and use a
3.04,3.31,4.32,2.42,3.20,     #comma-delimited format(commas between numbers)
NA,2.81,2.41,2.59,2.77,       #similar to a csv file (comma-separated values)
2.32,2.78,2.93,3.28,2.39,     #Label the datafile 'INCOMP'
NA,1.05,NA,2.23,0.85,         #Use 'NA' for missing data
…
NA,1.34,6.48,0.61,1.40,       #Only a portion of the data is shown here,
5.06,3.62,4.48,3.45,3.95,     #to save space
3.53,3.19,1.83,3.50,3.77,
NA,3.92,4.38,3.19,2.40,
NA,1.76,NA,0.59,3.20),
nrow=300,ncol=5,byrow=TRUE)   #Enter the # rows and # columns in raw datafile


print(INCOMP)                 #Check that the data were entered correctly

INCOMP<-data.frame(INCOMP)    #Treats the data like a spreadsheet table
names(INCOMP)<-c("y","x","z","aux1","aux2") #Name the variables, in order

MODEL <- '                    #In lavaan, label the SEM model 'MODEL'
  # measurement model         #This part is a factor analysis.
    AUX1 =~ aux1              #For multiple regression, make the measurement
    AUX2 =~ aux2              #model a manifest variable model, which is a
    X =~ x                    #single-indicator model with factor loadings
    Y =~ y                    #set to 1.0 and uniquenesses set to zero.
    Z =~ z                    #I use upper-case for latent variables.
  # regressions         #This part is the structural model (i.e., regression)
    Y ~ B1*X + B2*Z     #Label the regression coefficients 'B1' and 'B2'
    AUX1 ~ X + Z        #Use auxiliary variables as Extra DVs (Graham, 2003)
    AUX2 ~ X + Z
  # residual correlations
    Y ~~ AUX1                 #Allow all DV residual terms to correlate
    Y ~~ AUX2
    AUX1 ~~ AUX2
  # intercepts          #This part gives the regression intercept
    y ~ 0               #Set all measurement model intercepts to zero
    x ~ 0
    z ~ 0
```

```
    aux1 ~ 0
    aux2 ~ 0
    Y ~ B0*1                  #Label the regression intercept 'B0'
    X ~ 1
    Z ~ 1
    AUX1 ~ 1
    AUX2 ~ 1
'

RESULTS <- sem(MODEL, data = INCOMP, verbose=TRUE, missing = "FIML")
                        #Label the results file 'RESULTS'
                        #Use FIML for missing data.
                        #The 'verbose' command gives the EM algorithm
                        #covariance matrix 'Sigma' and the EM means 'Mu'.

summary(RESULTS, standardized = TRUE) #Get standardized results too

fitted(RESULTS)   #Display the ML (EM) covariance matrix and ML (EM) means
                  #(this only works for saturated models like the current
                  #multiple regression SEM, where df = zero and therefore the
                  #fitted covariance matrix is the EM covariance matrix)

cov2cor(fitted(RESULTS)$cov)  #Convert ML (EM) covariance matrix into ML (EM)
                              #correlation matrix

diag(fitted(RESULTS)$cov)^.5  #Convert ML (EM) cov. matrix into ML (EM)
                              #standard deviations (use diagonal of cov.
                              #matrix)

#If desired, you can change the order in which the variables in the
#EM correlation matrix are displayed, and round the corr.s to 2 decimals:

round(cov2cor(fitted(RESULTS)$cov[c(4,3,5,1,2),c(4,3,5,1,2)]),2)
```

***RESULTS FROM SIMULATED EXAMPLE;**

Regression Results

	Complete Data: *b* (*SE*), β (*p* value)	Listwise Deleted: *b* (*SE*), β (*p* value)	Pairwise Deleted: *b* (*SE*), β (*p* value)	EM Algorithm: *b* (*SE*), β (*p* value)	FIML: *b* (*SE*), β (*p* value)	Multiple Imputation: *b* (*SE*), β (*p* value)
Intercept	3.10 (.26), 0 (.000)	3.42 (.35), 0 (.000)	3.45 (.33), 0 (.000)	3.25 (.35), 0 (.000)	3.25 (.30), 0 (.000)	3.25 (.30), (.000)
X	.26 (.05), .26* (.000)	.10 (.07), .12 (.180)	.10 (.07), .13 (.143)	.22 (.07), .23* (.003)	.22 (.06), .23* (.001)	.21 (.06), .22* (.001)
Z	–.31 (.05), –.31* (.000)	–.10 (.08), –.11 (.197)	–.08 (.07), –.10 (.239)	–.28 (.07), –.30* (.000)	–.28 (.06), –.30* (.000)	–.27 (.06), –.29* (.000)
R^2	.20	.03	.03	.18	.18	.16
N for analysis	300	132 (listwise *N*)	146 (minimum pairwise *N*)	146 (minimum pairwise *N*)	Varies across variables (from 146 to 264)	Varies across variables (from 146 to 264)

Note: b = unstandardized regression coefficient, β = standardized regression coefficient, *SE* = standard error. For FIML, intercept b_0 = alpha parameter from the SEM FIML output, and $R^2 = 1 -$ standardized ψ (for *Y*). For multiple imputation, $\beta_I = b_1(SD_X/SD_Y)$, where SD_X and SD_Y are ML estimates from the EM algorithm. Notice how listwise and pairwise deletion give strongly biased parameter estimates and significance test results in this example. Also notice how the EM algorithm, FIML, and multiple imputation yield very similar (essentially identical) parameter estimates (with far less bias). FIML and MI also yield nearly identical *SE*s and significance test results (and the EM algorithm approach conservatively uses larger *SE*s, providing Type I error protection at least as well as FIML and MI do). Although this one example is not intended to prove the generality of ML and MI missing data techniques, it does show the expected result under conditions where the missingness mechanism is (at least partly) MAR—namely, ML and MI techniques outperform listwise and pairwise deletion. Alternatively, under MCAR missingness, listwise deletion, pairwise deletion, ML (EM and FIML), and MI techniques would all be equally unbiased. For a more complete set of simulation examples, see Collins, Schafer, and Kam (2001); Enders (2010); Graham (2003); Newman (2003); Newman and Cottrell (in press); and Schafer and Graham (2002).

Appendix B

If items on a scale have widely differing means (e.g., if an item mean [across persons] differs from the overall composite mean [across items and persons] by more than two standard deviations), then—for each partial respondent with item-level missingness on an extreme item—use an extreme item adjustment (i.e., Equation B1).

Dealing With Item-Level Missingness for Scales That Contain Extreme Items

As mentioned previously, item-level missingness is a worse problem if the items on a multi-item scale are not interchangeable. The key consideration here is whether the set of available items (i.e., if there is item-level missing data) represents the complete set of items from the whole multi-item survey instrument (i.e., if there were no item-level missing data). For example, if administering a survey of counterproductive work behavior (CWB), the mean for the item, "Falsified a receipt to get reimbursed for more money than you spent on business expenses," is lower than the means for other items on this scale (Bennett & Robinson, 2000, p. 354). The likely potential reasons for this low item mean are that (a) the "falsified receipts" item represents a more extreme form of counterproductive work behavior (i.e., stealing money) than do many of the other items on the multi-item CWB scale (e.g., lateness, break-taking, and neglecting to follow the boss's instructions), so it is only enacted by a small number of individuals who possess a high standing on the underlying CWB trait, or (b) some respondents do not file receipts as part of their jobs (i.e., the item is not relevant to them). Indeed, both of these might be reasons that the item is missing—because it is a more extreme manifestation of CWB, respondents will be more reticent to answer the question (the item divulges sensitive information), and because the item is irrelevant to some people's jobs, they might leave it blank to indicate that the behavior is not applicable to their jobs.

Because this "falsified receipts" item has a lower mean than the other items on the CWB scale, the likely consequence of excluding this item from the scale composite score would be to increase the individual's scale composite mean CWB score by a small amount. Whether this small amount of bias due to omitting a low-mean item is negligible depends on: (a) the portion of respondents who omitted this item, (b) the total number of items on the multi-item scale, and (c) whether the small positive bias due to omitting a low-mean item was offset by a countervailing small negative bias due to omitting different, high-mean items. In most practical scenarios (i.e., real data sets with validated multi-item survey instruments and steps taken to ensure participant confidentiality), the small bias due to actual item-wise missingness patterns will be negligible for all practical purposes.

However, for extreme cases where the items on a multi-item scale have highly discrepant means (i.e., if a scale contains extreme items), I offer the following recommendation. If an item from a multi-item scale has an observed mean that is two standard deviations away from the composite score mean, then individuals who are missing that item should have their composite construct scores adjusted to account for the fact that they are missing an extreme item. This procedure should take place in three steps. First, tabulate the item means for each item on the multi-item scale and then calculate the mean and *SD* of these item means (across items). Second, screen the item means to identify any "extreme items," which are items with a mean that is more than two standard deviations away from the mean of item means. If no extreme items are identified, then no composite scale score adjustments are needed. Third, for any individuals who are missing a response for an extreme item, adjust those individuals' scale composite scores using the formula:

$$\text{Individual's adjusted composite (mean) score} = \qquad \text{(Equation B1)}$$

$$\text{Individual's observed composite (mean) score (i.e., with extreme item missing)} \\ + [\text{item grand mean (acrosspersons)} - \text{overall composite score grand mean (across persons)}]/ \\ \text{total } n \text{ items on the full-length multi-item scale.}$$

For example, if there is a 10-item scale that contains 1 extreme item, then for any individual who failed to respond to that 1 item, her or his construct score should be adjusted using Equation B1. This would involve taking the individual's mean composite score (across available items) without the missing extreme item, and adding an adjustment term equal to the missing item's grand mean (across persons) minus the scale composite score grand mean (across persons), divided by 10 (i.e., the number of items on the full-length scale). This adjustment formula is based on a technique that Bernaards and Sijtsma (2000) called "two-way imputation," which they recommended for addressing item-level missingness (although, unlike Bernaards and Sijtsma, 2000, I am not recommending that this approach be used for imputation or for item-level analyses [e.g., item-level factor analysis]; I am only recommending the approach to adjust a few individuals' construct scores as a precursor to construct-level analyses). In the vast majority of cases, items will not be extreme enough to require the aforementioned adjustment. The whole point of this particular ad hoc adjustment is that it helps to address item-level missingness only in those extreme cases where items differ enough for item-level missingness to practically affect the construct scale score.

Acknowledgments

The author would like to thank R. Chris Fraley for his assistance with the R code for running multiple regression using the FIML missing data routine, as well as James LeBreton, Adam Meade, Scott Tonidandel, and Glenn Roisman for helpful comments on earlier drafts.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. For the current article, I define a *population* as a group from which a sample is drawn and to which inferences will be made (e.g., all working adults); a *sampling frame* is the list of all individuals from the population who were contacted with a survey invitation (i.e., in organizational research, it is typical to send surveys to everyone in the sampling frame); and a *sample* is the group of individuals who responded to at least part of the survey (i.e., full respondents and partial respondents).
2. As an aside, although one might reasonably define *partial respondents* with regard to item-level missingness in addition to construct-level missingness, in the current article and for the sake of developing consistent response rate reporting standards (discussed in the following sections), I am choosing to confine the *partial respondent* terminology to individuals with construct-level missingness. Again, construct-level missingness is a special case of item-level missingness where an individual fails to respond to all of the items on a multi-item scale.
3. One notable exception is the rare case where the researcher intentionally creates an MCAR planned missingness mechanism by flipping a coin to determine which individuals will receive different versions of a survey (Graham, Taylor, Olchowski, & Cumsille, 2006). Such *planned missing data* designs are sometimes used when the researcher wants to study relationships among a larger set of questions/variables than the average respondent wants to answer.

4. Although missing data bias and inaccurate standard errors (*SE*s) are two distinct issues, both can affect Type I and Type II errors of inference. For instance, an underestimation missing data bias in the observed effect size can lead to low statistical power, just as much as a large *SE* (e.g., small sample size) can. Alternatively, an overestimation missing data bias in the observed effect size can offset a large *SE* (e.g., small sample size) by increasing power. The ideal scenario for minimizing Type I and Type II errors is to have zero missing data bias, accurate *SE*s, and a large sample size.
5. Note that the selection variable *miss* is the more general and continuously-distributed version of the binary dummy variable $miss_{(y)}$ that I previously defined in reference to Figure 3.
6. My purpose in providing Equation 1 is merely to illustrate the factors that determine the magnitude of missing data bias. I do not intend to suggest that Equation 1 should be used to correct for missing data bias, because for most applications the local r_{miss} parameters are not known with adequate certainty to permit such corrections.
7. Technically, *direct range restriction* means that data on *X* and/or *Y* are missing on the basis of truncation on the observed values of either *X* or *Y* (Thorndike, 1949). *Indirect range restriction* means that data on *X* and/or *Y* are missing on the basis of truncation on a third variable, *Z*, which is correlated with *X* and/or *Y* (e.g., see Equation 1, where *miss* is the third variable).
8. As seen in Table 2, missing at random (MAR) missing data conditions naturally lead to biased parameter estimation under listwise deletion, but to unbiased parameter estimation under maximum likelihood (ML) and multiple imputation (MI) techniques. As such, MAR missingness is a common reason why listwise deletion results might differ from ML and MI results.
9. The only scenarios where single imputation might be defensible would be for unusual data structures (like social network data), for which no multiple imputation model nor ML missing data routine is available. For social network data, for example, it is sometimes defensible to use symmetry imputation (e.g., imputing a peer's nomination of a dyadic friendship in place of one's own missing self-report of the friendship, under the assumption of reciprocity). This can be a preferable alternative to listwise deletion.
10. For those who speak Bayesian language (see Brannick, 2001; Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin, 2013; Newman, Jacobs, & Bartram, 2007), multiple imputation approximates a Bayesian posterior estimate (which is a weighted average of the prior and the likelihood), whereas ML estimation provides the likelihood. So in the common case of a relatively uninformative prior, MI and ML techniques yield essentially the same results.
11. Collins, Schafer, and Kam (2001) further showed that auxiliary variables can improve missing data estimation even when the auxiliary variables only meet the second condition previously described—being correlated with the partially missing substantive variables of interest—regardless whether the auxiliary variables are correlated with the cause of missingness.
12. Newman and Cottrell (in press) showed that the variance ratio u^2 can be approximated as a function of the response rate only, under normality assumptions. That is, $u^2 = 1 + c_{x_z}/p_c\sqrt{2\pi e^{c^2}} - (1/p_c^2 2\pi e^{c^2})$, where p_c is the response rate, and c_{x_z} is the selection cut-score in standard score (*z*-score) form, which can be looked up in a *z* table in the back of any statistics textbook or approximated using "= - NORMSINV('response rate')" in Microsoft EXCEL.
13. For assessing missing data bias in parameters of a simple mediation model with three variables ($X \to M \to Y$), one can use equations for the regression coefficient as a function of the missing data–corrected correlations. For example, $\beta_x = (r_{XY} - r_{MY}r_{XM})/(1 - r_{XM}^2)$.

References

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response rates in organizational science, 1995-2008: A meta-analytic review and guidelines for survey researchers. *Journal of Business and Psychology*, *25*, 335-349.

Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology*, *85*(3), 349-360.

Bernaards, C. A., & Sijtsma, K. (2000). Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. *Multivariate Behavioral Research*, *35*(3), 321-364.

Brannick, M. T. (2001). Implications of empirical Bayes meta-analysis for test validation. *Journal of Applied Psychology*, *86*, 468-480.

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.

Cycyota, C. S., & Harrison, D. A. (2006). What (not) to expect when surveying executives a meta-analysis of top manager response rates and techniques over time. *Organizational Research Methods*, *9*(2), 133-160.

Dempster, A. P., Laird, N. H., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, *B39*, 1-38.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York, NY: Wiley.

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, *125*(2), 175-191.

Enders, C. K. (2001a). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*, 352-370.

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

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

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

Enders, C. K., & Peugh, J. L. (2004). Using an EM covariance matrix to estimate structural equation models with missing data: Choosing an adjusted sample size to improve the accuracy of inferences. *Structural Equation Modeling*, *11*, 1-19.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*(4), 409-420.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Taylor & Francis.

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In H. Wainer (Ed.), *Drawing inferences from self-selected samples* (pp. 115-142). New York, NY: Springer-Verlag.

Gold, M. S., & Bentler, P. M. (2000). Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Structural Equation Modeling*, *7*, 319-355.

Gold, M. S., Bentler, P. M., & Kim, K. H. (2003). A comparison of maximum-likelihood and asymptotically distribution-free methods of treating incomplete nonnormal data. *Structural Equation Modeling*, *10*, 47-79.

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling*, *10*, 80-100.

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

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206-213.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In *Statistical Strategies for Small Sample Research*, ed. R. Hoyle, 1:1-29. Thousand Oaks, CA: Sage.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323-343.

Heckman, J. T. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153-161.

Little, R. J. A. (1993). Pattern mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, *88*, 125-134.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Marsh, H. W. (1998). Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling*, *5*, 22-36.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: Guilford Press.

Mistler, S. A. (2013). A SAS macro for applying multiple imputation to multilevel data. In *Proceedings of the SAS Global Forum.*

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Bethesda, MD. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Washington, DC: ERIC Clearinghouse.

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*, 328-362.

Newman, D. A. (2009). Missing data techniques and low response rates: The role of systematic nonresponse parameters. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in the organizational and social sciences* (pp. 7-36). New York, NY: Routledge.

Newman, D. A., & Cottrell, J. M. (in press). Missing data bias: Exactly how bad is pairwise deletion? In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends*. New York, NY: Routledge.

Newman, D. A., Jacobs, R. R., & Bartram, D. (2007). Choosing the best method for local validity estimation: Relative accuracy of meta-analysis versus a local study versus Bayes-analysis. *Journal of Applied Psychology*, *92*, 1394-1413.

Newman, D. A., & Sin, H. P. (2009). How do missing data bias estimates of within-group agreement? Sensitivity of SD_{WG}, CV_{WG}, rWG(J), rWG(J)*, and ICC to systematic nonresponse. *Organizational Research Methods*, *12*, 113-147.

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*, 525-556.

Puhani, P. A. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, *14*, 53-67.

Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmuller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*, *88*, 1104-1114.

Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, *5*(3), 127-134.

Roth, P. L., & BeVier, C. A. (1998). Response rates in HRM/OB survey research: Norms and correlates, 1990-1994. *Journal of Management*, *24*, 97-117.

Roth, P. L., Switzer, F. S., & Switzer, D. M. (1999). Missing data in multiple item scales: A Monte Carlo analysis of missing data techniques. *Organizational Research Methods*, *2*, 211-232.

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

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.

Savalei, V., & Bentler, P. M. (2009). A two-stage approach to missing data: Theory and application to auxiliary variables. *Structural Equation Modeling*, *16*(3), 477-497.

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.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, *61*, 473-485.

Spitzmuller, C., Glenn, D. M., Barr, C. D., Rogelberg, S. G., & Daniel, P. (2006). "If you treat me right, I reciprocate": Examining the role of exchange in survey response. *Journal of Organizational Behavior*, *27*, 19-35.

Switzer, F. S., Roth, P. L., & Switzer, D. M. (1998). Systematic data loss in HRM settings: A Monte Carlo analysis. *Journal of Management*, *24*, 763-779.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York, NY: Wiley.

Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology*, *18*, 327-350.

van Buuren, S. (2011). Multiple imputation of multilevel data. In J. K. Roberts & J. J. Hox (Eds.), *The Handbook of Advanced Multilevel Analysis*, (pp. 173-196). New York: Routledge.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior. *Public Opinion Quarterly*, *55*, 613-629.

Author Biography

Daniel A. Newman (Ph.D., Pennsylvania State University) is an associate professor in the Department of Psychology, and in the School of Labor & Employment Relations, at the University of Illinois at Urbana-Champaign. His research deals with emotional intelligence, adverse impact/diversity, social networks, narcissism, and research methods.