Best Practices for Missing Data Management in Counseling Psychology

Gabriel L. Schlomer, Sheri Bauman, and Noel A. Card
University of Arizona

This article urges counseling psychology researchers to recognize and report how missing data are handled, because consumers of research cannot accurately interpret findings without knowing the amount and pattern of missing data or the strategies that were used to handle those data. Patterns of missing data are reviewed, and some of the common strategies for dealing with them are described. The authors provide an illustration in which data were simulated and evaluate 3 methods of handling missing data: mean substitution, multiple imputation, and full information maximum likelihood. Results suggest that mean substitution is a poor method for handling missing data, whereas both multiple imputation and full information maximum likelihood are recommended alternatives to this approach. The authors suggest that researchers fully consider and report the amount and pattern of missing data and the strategy for handling those data in counseling psychology research and that editors advise researchers of this expectation.

Keywords: missing data, best practices, counseling psychology, multiple imputation, full information maximum likelihood

Supplemental materials: http://dx.doi.org/10.1037/a0018082.supp

Missing data occur in most studies in the behavioral sciences (Acock, 2005; Allison, 2001; Orme & Reis, 1991; Pigott, 2001; Stevens, 1996; Streiner, 2002), and the APA Task Force on Statistical Inference (Wilkinson & Task Force on Statistical Inference, 1999) recommended that researchers report patterns of missing data and the statistical techniques used to address the problems such data create. Although adequate reporting and handling of missing data are essential for understanding results, this element of data analysis is often omitted from reports of research (Peng, Harwell, Liou, & Ehman, 2006; Saunders et al., 2006). With increased computing memory and processing speed, sophisticated analyses of missing data can now be accomplished by researchers without costly specialized software programs. However, many researchers are unaware of the importance of reporting and managing missing data, and editors have generally not insisted that authors provide this essential information.

Best practices related to missing data in research call for two items of essential information that should be reported in every research study: (a) the extent and nature of missing data and (b) the procedures used to manage the missing data, including the rationale for using the method selected. In this article, we first examine the most recent complete volume of the Journal of Counseling Psychology to determine how authors in this journal have addressed those two items. Then we present suggestions for improving the reporting of missing data in journal articles. Finally, we provide a review of several common approaches to managing missing data and present guidelines for aligning the approach to handling missing data with the specific situation of the study.

Current Practice in the Journal of Counseling Psychology

We examined all articles in the latest complete volume (Vol. 55 in 2008) of the Journal of Counseling Psychology to locate articles reporting quantitative data analysis. We found 37 such articles (out of 46 articles in that volume) and examined each to determine whether the percentage of missing data was reported, whether the method for handling these data was specified, and whether a rationale for the method was provided. Only 14 articles reported the percentage of missing data (one of those noted no missing data), although for three other articles the amount of missing data could be inferred from other information provided. In 11 articles, the method (either stated or inferred) was listwise deletion; cases with missing data were dropped from the analysis. In one article, cases with more than 10% missing data were deleted listwise, but no mention was made of how item nonresponse was handled in the remaining cases. In another, the author(s) justified use of listwise deletion (9% of cases) by noting that because no significant relations had been detected between the exclusion variable and other variables of interest, dropping cases did not bias results. Listwise deletion was used in another study with the explanation that the amount of missing data was not substantial, although the amount was not specified. In another study, mean substitution was used when the missing data were less than a predetermined cutoff.
rate, and cases were dropped if the missing data were on the outcome measure. In three articles no mention was made of missing data, but tables suggested that all participants were included in all analyses. Tables in another article did not include Ns, and no mention was made of missing data in the text; this research involved the administration of multiple measures, and it is difficult to imagine that every item was completed by every participant. In yet another article, no mention of missing data was made in the text, but inspection of the tables revealed that missing data existed. Median replacement was used in one study, except when a large number of missing data points were found on one variable, in which case regression imputation was the method used; a justification for those decisions was not provided. An additional article noted that missing values were deleted pairwise, with no explanation for that decision. Most striking is that in only one article was the percentage of missing data reported, the method for handling those data described, and a rationale for that method included.

This survey suggests that, despite the prevalence of missing data and recommendations for thorough consideration of missing data, counseling psychology has not yet adopted these recommendations. We recommend that researchers report the amount of missing data in a study, consider the potential sources and patterns of missing data, and use and report appropriate methods for handling missing data in their analyses. These recommendations are summarized in the Appendix, and we elaborate on each of these recommendations next.

**Best Practices in Reporting Missing Data**

**Amount of Missing Data**

At a minimum, researchers should always report the proportion of missing data (see Appendix). Just as the reporting of response rates in survey research is important in considering potential generalizability, the reporting of the amount of data missing from among participants is important. To frame reporting of the amount of missing data, researchers should consider two common sources of missingness.

In many studies, missing data are due to item nonresponse. Here, participants complete a survey, test, or other measure but do not give responses for every item. In this situation, it is ideal to report the percentage of missing responses for each item of the measure, which can be included in tables along with basic descriptive data such as means and standard deviations. However, if space limitations and the number of items make this option impractical, at a minimum, the range of missing data should be reported (see Appendix for an example). When the item-level percentage of missing data is not reported in an article, we encourage authors to report missing data at the item level in an online supplementary table.

A second source of missing data is participant attrition. For example, in longitudinal studies, in which data are collected multiple times from the same participants, some may be unavailable for at least one wave of data collection. This attrition can occur in either longitudinal naturalistic research (e.g., in a school-based survey, a child present in the first year may have left the school by the second year) or experimental/quasi-experimental designs employing multiple measures (e.g., some participants drop out of a prevention program between pre- and posttest). Participant attrition can also occur in questionnaire-based cross-sectional designs. In this case, participants often fail to complete the entire questionnaire due to fatigue or boredom. In any design, it is critical to report the percentage of attrition at each wave and within the entire study (see Appendix).

Experts have not reached a consensus regarding the percentage of missing data that becomes problematic. Schafer (1999) recommended 5% as the cutoff. However, Bennett (2001) suggested that when more than 10% of data is missing, statistical analyses are likely to be biased; and others have used 20% (e.g., Peng et al., 2006). In contrast to those who suggest a particular cutoff, we believe that two considerations determine whether a certain amount of missingness is problematic. The first is whether the resultant data set has adequate statistical power to detect the effects of interest. As we describe later in the Nonstochastic Imputation Methods section and the Stochastic Imputation Methods section, modern imputation procedures retain the maximum amount of possible statistical power and are therefore preferable to deletion methods. The second consideration is the pattern of missingness. As we discuss next, the pattern of missingness speaks to the potential biasing impact on the data.

**Pattern of Missing Data**

Researchers should consider, in addition to the amount and source of missing data, the pattern of missing data. This consideration involves the following question: Are the data that are missing random, or are they nonrandom and potentially biasing? Quantitative researchers have expanded this basic question to describe three patterns of missingness: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR).

**Missing completely at random.** With MCAR data, there are no patterns in the missing data and the missing values are not related to any variables under study (Acock, 2005; Bennett, 2001; Roth, 1994). If one examines a large table of all the data in the data set, the missing data points would be randomly distributed throughout the table. The implication of completely random missingness is that the cases with missing data would be equivalent to a random subset of the entire sample. In practice, it is difficult to determine whether data are MCAR; however, Little (1988) developed an omnibus statistical test of MCAR (see Appendix). Schafer and Graham (2002) consider MCAR to be a special case of MAR, described next.

**Missing at random.** The term missing at random is misleading in that data fitting this pattern are not randomly missing, at least not completely. With MAR data the probability of having a missing data point is related to another variable in the data set but is not related to the variable of interest (Allison, 2001). Put differently, under MAR, missing data are related to observed data (another variable in the data set) but not to missing data (Graham, Cumsille, & Elek-Fisk, 2003; Roth, 1994; Schafer & Graham, 2002). When this is the case, the researcher must include the observed variable in the analysis to avoid bias. For example, in a questionnaire completed by counseling psychologists, assume an item asks respondents to indicate their level of interest in conducting parent education programs. Responses on this item are missing for some respondents who indicate elsewhere on the survey (another variable) whether they have had specific training in deliver-
ing educational programs. If the responses to the item asking about interest in parent education are MAR within each of the two groups (had training vs. had no training), then they are missing at random, even if the data are not missing at the same rate for each group (Roth, 1994). To put it another way, if the missing data on the parent education item are related to the participants’ response on the training experiences variable (observed) but not related to the level of interest in parent education variable (missing), then the data are MAR (Acock, 2005). It may be that more of those without the training omitted this item, but the pattern is random within both groups. It is possible to distinguish between MCAR and MAR by computing a dummy variable representing whether data are missing on a variable of interest and then examining whether this dummy variable is associated with other variables in the model (see Appendix). If this dummy variable (missingness) is unrelated to other variables, then the pattern is not considered MAR but rather MCAR or NMAR (see next section). However, if the dummy variable is indeed associated with other variables, then we conclude MAR rather than MCAR, though we still cannot entirely rule out NMAR. This possibility of NMAR means that the researcher cannot determine unequivocally that data are MAR or MCAR. But, as we describe next, researchers often assume MAR or MCAR when there are no indications to the contrary.

Not missing at random. When there is a pattern to missing data such that the likelihood of missingness is related to the score on that same variable had the participant responded, the data are NMAR, also known as nonignorable nonresponse. The obvious difficulty in determining NMAR is that the association between missingness and how the participant would have responded cannot be evaluated, because we do not have the missing values. So, the possibility of NMAR becomes a conceptual consideration: Is it likely that participants who are high (or low) on the variable are more likely to have missing data (e.g., skip the item or leave the study)? Although the inability to empirically evaluate NMAR is unsatisfying, this does not mean that we should ignore this possibility. For example, we might observe that there is a high rate of missing data on an item inquiring about participants’ annual income. It may be the case that participants with high rates of income are more likely to omit this item because they are uncomfortable with others knowing their income. We encourage researchers to always consider the plausibility of NMAR and acknowledge this possibility when this pattern is plausible.

Methods of Handling Missing Data

There are a number of strategies for handling missing data, and the most common will be described here. These methods can be accomplished with standard statistical software packages (e.g., SAS, SPSS), available freeware (e.g., Amelia, Norm), or packages specifically designed for particular types of analyses (e.g., Mplus for structural equation modeling). There is not one best strategy; the strategy will depend on the data and the analyses. We describe these methods with continuous variables in mind. Special issues arise when dealing with missing data that are categorical, and treatment of these issues is beyond the scope of this article. We invite readers to refer to other existing literature that specifically deal with categorical missing data (e.g., Allison, 2001; Chen & Astebro, 2003; Graham, 2009).

Deletion Methods

Deletion methods are not as much a strategy for handling missing data as they are approaches to ignoring missing data. These methods are generally not recommended, so we review them only briefly.

Listwise deletion. In this method, cases with any missing values are deleted from an analysis. This method is sometimes called complete case analysis (Pigott, 2001), because only cases with complete data are retained. This is the default procedure for many statistical programs (e.g., SPSS), but it is generally not an advisable method. One problem with this strategy is that if the cases with missing values differ in some way from those with no missing values (i.e., they are not MCAR), then the remaining cases will be a biased subsample of the total sample and the analysis will therefore yield biased results (Bennett, 2001). On the one hand, as noted earlier, when missing data are MCAR, the observed data are essentially a random subset of the complete data. As such, parameters derived from MCAR data under listwise deletion are equivalent to those derived from complete data. On the other hand, listwise deletion can result in a loss of statistical power. Sheri Bauman collected data on a 58-item survey from 302 respondents with the intention of conducting a principal components analysis of these items. Although most cases with missing data were missing only one item, using the listwise option in SPSS resulted in only 154 cases being used in the analysis. In other words, nearly half of the data collected were dropped—a considerable loss of data and resources used to collect this otherwise large data set.

Pairwise deletion. In this method, the maximum amount of available data is retained, and so this method is sometimes referred to as available case analysis (Pigott, 2001). Cases are excluded from only operations in which data are missing on a variable that is required (Bennett, 2001; Roth, 1994). In a correlation matrix, for example, a case that was missing data on one variable would not be used to calculate the correlation coefficient between that variable and another but would be included in all other correlations. This means that different cases are used to calculate the different bivariate correlations. In the data set referred to in the previous section, the number of cases for each variable ranged from 271 to 302. The problems with pairwise deletion come from the use of different cases for each correlation, which results in difficulty in comparing correlations and oftentimes the inability to use these correlations in multivariate analyses (the resulting correlation matrix can be inadmissible for the underlying matrix algebra).

Nonstochastic Imputation Methods

The following methods are imputation strategies (our use of the adjective nonstochastic will become clear when contrasted with the stochastic methods later in the Stochastic Imputation Methods section). Imputation involves substituting a plausible value for data that are missing. Imputation based on either mean substitution or regression substitution are easily done with standard statistical software. These makes them easy to use, but they each have problems that must be considered.

Mean substitution. In this method, missing values are imputed with the mean value of that variable on the basis of the nonmissing values for that variable. This method assumes that data are MCAR and results in biased means when this assumption is
false. Furthermore, imputing the mean value into cases tends to reduce the variance of the variable, which also attenuates covariances that the variable has with other variables. This method produces biased means with data that are MAR or NMAR and underestimates variance and covariances (and resultant correlations) in all cases. Experts strongly advise against this method (Allison, 2001; Bennett, 2001; Graham et al., 2003; Pallant, 2007).

Regression substitution. Regression methods involve a regression equation based on the nonmissing data to predict expected values for the missing data. In other words, the missing values are the outcome variable, with the other variables in the data set serving as predictors. This approach represents the best “guess” as to what the participant would have scored on the missing variable. This approach also produces unbiased means under MCAR or MAR. However, this method is problematic in that it produces biases in the variances and covariances, so experts (e.g., Graham et al., 2003) advise against using this method.

Pattern-matching imputation. There are two other methods described in the literature that impute values based on matching the case with missing data with similar cases without missing data: hot-deck and cold-deck imputation. These do not require specialized programs and have been used with survey data (Roth, 1994). However, these methods suffer from the same disadvantages as the other nonstochastic imputation methods just described, in that they involve a single value that inevitably reduces the amount of variation in the data. In hot-deck imputation, values are imputed by finding participants who match the case with missing data on other variables. Bennett (2001) observed that this method results in less bias than does listwise deletion or mean imputation. However, strong evidence of the accuracy of this method has not been produced (Roth, 1994). Further, in most applications of this method, continuous data (scores on a measure) are collapsed into categories (e.g., high, medium, and low), which sacrifices information (e.g., MacCallum, Zhang, Preacher, & Rucker, 2002). Cold-deck imputation is a variation on this strategy in which information from external sources is used to determine the matching variables. The method suffers from the problems of hot-deck imputation and is also dependent on the availability of previous research and external information.

Stochastic Imputation Methods

It is useful to point out that here there is a distinct difference between the imputation model and the analysis model. The imputation model is the model used to impute missing values and may include variables that are not included in the analysis model (e.g., regression, analysis of variance, t test; see Collins, Schafer, & Kam, 2001, for a discussion). These variables that are included in the imputation model but not the analysis model are referred to as auxiliary variables. Auxiliary variables are useful in imputation techniques because they improve the precision of the imputation model by (a) including variables that account for the pattern of missing data and (b) improving the prediction of missing values by including variables that are correlated with the variable(s) that have missing data. In each of the methods described in the next four subsections (as well as regression imputation described previously), the imputation model and the analysis model may differ from one another, depending on whether the researcher has included auxiliary variables in the imputation model. Collins et al. (2001) suggested that including auxiliary variables can be quite beneficial and risks little in terms of cost.

Stochastic regression. This method is a variant of the regression-based approach in which a stochastic, or random, value is added to the imputed predicted value. These stochastic values are centered at zero, so they do not systematically change the mean; therefore, they provide the same unbiased means as does the regression imputation. However, these stochastic values introduce variance into the imputed data that results in unbiased variance estimates, thereby overcoming this limitation of the nonstochastic regression imputation.1

Expectation maximization (EM). This method is one of several maximum likelihood (ML) approaches. In all ML strategies, observed data are used to estimate parameters, which are then used to estimate the missing scores. These ML strategies have demonstrated superiority to deletion, nonstochastic imputation, and stochastic regression imputation methods (Roth, 1994) for multivariate normal distributions.

The EM strategy is based on a recursive process: The missing data have information that is useful in estimating various parameters, and the estimated parameter has information that is useful in finding the most likely value of the missing data (Bennett, 2001). Thus, the EM method is an iterative procedure with two steps in each iteration: In the expectation step, the process is similar to the regression-based imputation. First, starting values for the parameters (e.g., means, covariances) are obtained with available data. Regression methods are used to impute, on the basis of these initial values, the values for the missing data. When this step is completed, in the maximization step new values for the parameters are calculated with the newly imputed data along with the original observed data. Then the process starts over with the expectation step and continues until the estimates change very little from one iteration to the next (i.e., until the estimates converge; Allison, 2001).

The EM method provides “unbiased and efficient” (Graham et al., 2003, p. 94) parameters and is particularly useful for procedures such as exploratory factor analysis and internal consistency calculations, which do not require hypothesis testing. Because exploratory factor analysis requires relatively large sample sizes, the ability to impute missing data that are unbiased and retain all participants is an enormous advantage and is highly recommended. The disadvantage of EM is that the standard errors and confidence intervals are not provided, so obtaining those statistics requires an additional step. For inferential analyses for which those are essential, EM may not suffice.

Multiple imputation (MI). This is the most complex of the procedures described so far. It is an improvement over the EM approach because it involves the degree of similarity or difference between several imputed data sets as additional information for the standard errors of parameter estimates, and thus the solution is less biased than is a data set singly imputed with EM (Acock, 2005).

1 Specifically, the stochastic elements are randomly sampled from a normal distribution with a mean of zero and variance equal to the residual, or unexplained, variance of the regression imputation equations.
The first step in MI is to create several imputed data sets. Three to five imputations are usually adequate (Schafer, 1997), though with current computer speeds and programs to automate the process, there is little drawback to selecting a larger number. Then, the analyses are carried out on each data set, with the parameter estimates (e.g., factor loadings, group mean differences, correlations, regression coefficients) and their standard errors saved for each data set. Final results are obtained by averaging the parameter estimates across these multiple analyses, which results in an unbiased parameter estimate. The advantage of MI is that the final standard errors of these parameter estimates are based on both (a) the standard errors of the analysis of each data set and (b) the dispersion of parameter estimates across data sets. These combined standard errors from the multiply imputed data sets are used for significance testing and/or construction of confidence intervals around these parameter estimates. By accounting for the random fluctuations that occur between each imputation run, the MI procedure provides accurate standard errors and therefore accurate inferential conclusions.

The precision of parameter estimates and accuracy of standard errors make MI one of the best options for handling missing data. MI also has advantages over model-based approaches (e.g., full information maximum likelihood described in the next section) in that it is easy to include all variables in the imputation algorithms and then select a subset of variables for analysis. On the other hand, MI is computer-intensive, and it is difficult to combine data and then select a subset of variables for analysis. On the other hand, MI is computer-intensive, and it is difficult to combine data sets for analysis after the multiple data sets have been generated. In SAS, however, this can be done in PROC MIANALYZE (explained later in the Software section under SAS and in the online supplemental material).

**Full information maximum likelihood (FIML).** FIML is a direct model-based method for estimating parameters in the presence of missing data (Olinsky, Chen, & Harlow, 2003). The FIML approach computes a casewise likelihood function with observed variables for each case (see Arbuckle, 1996, for technical details). Unlike MI, FIML does not impute missing values into newly created data sets (and thus is not technically an imputation method) but rather estimates parameters on the basis of the available complete data as well as the implied values of the missing data given the observed data. For example, consider two variables, \(x\) and \(y\), where data are missing on the \(y\) variable and \(x\) and \(y\) are correlated to some degree. From a conceptual standpoint, FIML essentially “borrows” information about the probable values of \(y\) on the basis of the conditional expectation of \(y\) given \(x\) (Enders & Bandalos, 2001). This procedure is conceptually similar to regression imputation and produces results similar to EM and MI (Graham, Hofer, & MacKinnon, 1996; Olinsky et al., 2003). FIML has two primary advantages over imputation techniques that make this procedure attractive to researchers: (a) the imputation procedure and the analysis are conducted within the same step and (b) unlike EM, FIML produces accurate standard errors by retaining the sample size.

Simulation studies comparing FIML with other imputation techniques demonstrate that the FIML procedure produces approximately unbiased results across a variety of parameter estimates, particularly at small sample sizes (\(N = 100\)), and produces results similar to those for both EM and MI (Enders & Bandalos, 2001; Graham et al., 1996; Olinsky et al., 2003). Given these properties, FIML is one of the preferred methods for handling missing data. The ability to manage missing data and conduct analyses in one step makes this approach much simpler than multiple imputation. In addition, the ability to estimate accurate standard errors and confidence intervals by retaining the size of the sample is a distinct advantage over EM. On the other hand, it is important to ensure that variables that predict missing values are included in the analysis model; if they are not included in the model of interest, then they must be added for FIML to perform as well as MI or EM does (Graham, 2003).

**Software**

As one might expect from these descriptions, the best methods of imputing missing values are computation-intensive, and one would therefore certainly not attempt such an endeavor without specialized software. Fortunately, there exists a range of software for missing data imputation. Here we briefly review the capabilities of two popular commercial packages (SPSS and SAS), one specialized package (Mplus), and two freely downloadable programs specifically designed to handle missing data (Amelia and Norm).

**SPSS**

The base version of SPSS (we evaluated Version 16.0; note that SPSS is now PASW) can be used to compute percentage missing while computing basic descriptive information (the output reports the number of valid, nonmissing cases, which can be used to compute the percentage of total cases). To evaluate patterns of missingness—specifically to empirically evaluate between MCAR and MAR (recall that NMAR must be evaluated conceptually)—one can create a dummy variable by selecting the “recode into different variable” option from the drop-down menu, transforming missing values to 1 and all other values to 0, and then comparing these values on other variables in the data set. Once the researcher has determined the patterns of missing data, however, the base version of SPSS provides poor options for handling missing data, including only listwise and pairwise deletion or mean substitution. SPSS does offer an additional package, the Missing Values Analysis (MVA) module, at an additional cost. This module offers simplified steps to determine the amount of missingness and to explore patterns of missingness (using Little’s, 1988, test of missingness). It also offers some more acceptable imputation methods, including stochastic regression and EM imputation. The MVA module for Version 17 now includes an MI option. At the time of this writing, we had not evaluated that option and therefore cannot comment on the usability of this module.

**SAS**

The SAS-Stat system (we evaluated Version 9.2) allows for the same calculations of percentage missing (i.e., requesting descriptive statistics of variables of interest) and evaluation of missing data patterns (i.e., creating dummy codes for differentiating MCAR and MAR) as does the SPSS base system. However, SAS contains two procedures (i.e., PROCs) that offer more options than does SPSS, even with the additional MVA option. First, PROC MI (see SAS Institute, 2008, pp. 3738–3831) offers more algorithms for data imputation, including stochastic regression, EM, and a
more complex approach known as the Markov Chain Monte Carlo algorithm. PROC MI also allows the user to specify multiple imputations (MIs). The creation of multiple imputed data sets can then be analyzed with the PROC MIANALYZE (see SAS Institute, 2008, pp. 3833–3884), which performs many common analyses with these multiple imputed data sets and combines these results so as to provide unbiased parameter estimates with accurate standard errors.

**Mplus**

Mplus (we evaluated Version 5.21) is a statistical modeling program used primarily for estimating structural equation models, though it is flexible enough to perform the most basic (e.g., regression) and complex (e.g., categorical data, latent class analysis) analyses (Muthén & Muthén, 1998–2007). Mplus has the capability to conduct MI and also FIML. Starting with Version 5.0, the command TYPE = MISSING became the default procedure when analyzing data with missing values. Under this default and using ESTIMATOR = ML (maximum likelihood), missing data are handled with FIML, which makes it easy for new users of the program to implement. It should be noted that Mplus is not the only program that has FIML capabilities; other specialized software programs such as AMOS and LISREL (for structural equation modeling) and HLM (for multilevel modeling) can also estimate parameters with FIML.

**Freely Downloadable Software**

Both Amelia II (Honaker, King, & Blackwell, 2009) and Norm (Version 2.02; the site included in the Reference list also contains similar packages for other types of data in which missingness may be present) are freely downloadable programs that perform MI. Although there are some small differences in the usability and underlying imputation algorithms of these two programs, they are quite similar for most users’ purposes. Both programs can be used to create multiple imputed data sets and to combine parameter estimates from these multiple data sets, though one must perform most analyses of these imputed data sets (to obtain the parameter estimates within each imputed data set) within separate programs. Amelia interfaces with R (free), and Norm interfaces with S-Plus (must be purchased), making these multiple analyses reasonably simple for users of these programs.

**Summary**

The five software packages we have briefly reviewed do not capture the full range of possibilities. Nearly every statistical analysis package has some capabilities to summarize the amount of missing data, evaluate patterns of missingness, and perform some type of missing data imputation. Likely the most variability among software packages comes from the imputation procedures. We recommend using software that at least performs stochastic regression imputation, with EM being even better. The best methods are the MI and FIML procedures, though these can be more difficult to use and are not included in some packages (e.g., SAS has these capabilities; SPSS has added an MI option to the MVA module for Version 17). Fortunately, freely downloadable software can perform MI even if one’s usual software of choice does not.

**An Illustration**

**Method**

To illustrate the differences between the methods described above, we simulated a data set of 60 participants. The relatively small sample size was chosen to reflect the modest sample sizes of much of counseling psychology research. To provide context for this illustration, we might imagine that these data come from 60 clients under age 21 years at a large university counseling center who were referred for counseling by the dean of students due to underage drinking violations. The fictitious counseling center randomly assigned the students to one of two treatment programs, one of which uses the harm reduction approach, and the other of which is based on a 12-step model. The outcome (i.e., dependent variable) might be scores on a measure of self-efficacy for sobriety. In this fictitious example, we also have a measure of participants’ attitude toward authority, on which higher scores represent greater respect for authority.

These fictitious data were generated with SPSS such that characteristics of the data were known. We specified an effect size typically considered large (r = .50) for the impact of group (harm reduction vs. 12-step) on the outcome (self-efficacy for sobriety). However, we also created a covariate (attitude toward authority) with a large association with group (r = .50) and a very large correlation with outcome (r = .70). These parameter estimates can be confirmed in the data set available in an online supplement to this article, “Sample Syntax of Analyses for Illustration” (see the online supplemental material). We subjected this fictitious data set to various patterns and rates of missingness to illustrate different results obtained. Specifically, data were (a) randomly deleted (MCAR) at rates of 10%, 20%, and 50% or (b) deleted probabilistically on the basis of the value of the covariate (MAR), with the likelihood of missing values on the outcome being strongly negatively correlated (r = -.70) with the covariate, with rates of MAR missingness of 10%, 20%, and 50% considered. This MAR situation represents a plausible scenario in our illustration; it is likely that students with less respect for authority and who are mandated to attend counseling are less likely to complete a measure for a study of these programs than are those who have more respect for authority.

After data had been deleted, missing data were handled with three strategies: mean substitution (a poor method of handling missing data), MI, and FIML (with MI and FIML being recommended methods). SPSS was used to impute means for the mean substitution analysis. For MI, 10 imputations were used in SAS, and the mechanism that accounted for missingness, covariate, was included in the imputation model. Because FIML is model-based, we used a saturated correlates modeling technique3 (see Graham, 2003) to include the covariate (attitude toward authority) as well as the predictor (condition) and the outcome variable in the model. Mplus was used to estimate parameters using FIML.

---

2 The Markov Chain Monte Carlo (MCMC) algorithm is mathematically complex, and full description is beyond the scope of this article (see SAS Institute, 2008, pp. 3774–3784). MCMC has advantages over EM for missing data patterns that are arbitrary as opposed to monotone (see SAS Institute, 2008, p. 3766).

3 The covariate was included in the model by allowing it to covary with the predictor and the residual of the outcome.
For each data set, we regressed outcome (self-efficacy for sobriety) onto group to evaluate the magnitude of difference between participants who received the 12-step treatment (coded as 0) and those who received the harm reduction treatment (coded as 1), with 30 participants per group. Note that this regression analysis is equivalent to performing an independent samples t test in which the two groups are compared on the outcome (see e.g., Cohen, 1968); unstandardized regression coefficients are reported here and are equivalent to mean differences in t tests. A comparison of the relative effectiveness of each method across different amounts (10%, 20%, 50%) and types of missingness (MCAR, MAR) is summarized in Table 1. In Table 1 and in subsequent discussions, bias in parameters is defined as the percentage difference between the complete data results (0% missing) and the results using different methods of handling missing data with the six different data sets with missing data. Differences were converted into percentages to aid readers when comparing across different methods.

### Results

**Mean substitution.** As can be seen in Table 1, mean substitution created considerable bias in both regression coefficients and standard errors, even at low levels of missingness, and systematically underestimated both mean differences and standard errors. Surprisingly, mean substitution tended to perform worse, on average, under MCAR (regression coefficient = 27.55%; standard error = 13.42%) compared with MAR (regression coefficient = 25.00%; standard error = 9.59%). This can be attributed to the exceptionally poor performance at 50% MCAR. Under conditions of extreme missing data (e.g., 50%), mean substitution performed extremely poorly for both the MCAR and MAR situations and overall resulted in a 45.31% bias in calculating regression coefficients and 23.23% bias in standard errors. Furthermore, mean substitution at more common amounts of missing data (e.g., excluding 50% missing data) continued to result in considerable bias in regression coefficients (16.76%) and standard errors (5.64%).

**Multiple imputation.** As expected, MI performed well at common amounts of missing data and reasonably well at severe amounts of missing data. At common amounts of missing data (e.g., 10% and 20%), MI resulted in an average bias of 11.42% for regression coefficients and 7.3% for standard errors. When missing data were extreme, regression coefficients were biased on average by 12.05% and standard errors by 22.79%. A comparison of bias under MCAR versus MAR revealed that MI resulted in more bias in estimating regression coefficients under MAR (15.22%) compared with MCAR (8.03%), and the same was true for estimating standard errors (MAR: 15.34%; MCAR: 9.59%). These differences, however, are much improved over those for mean substitution. Overall, MI led to an 11.63% bias in regression coefficients and a 12.46% bias in standard errors. The largest bias in regression coefficients was 21.67% at 20% MAR, whereas the

### Table 1

**Regression Coefficients and Standard Errors for Complete and Missing Data: A Comparison of Different Methods for Handling Missing Data**

<table>
<thead>
<tr>
<th>Method and amount/ type of missing data</th>
<th>Regression coefficient</th>
<th>Standard error</th>
<th>t(58)</th>
<th>p</th>
<th>% bias for regression coefficient</th>
<th>% bias for standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>No missing data</td>
<td>0.992</td>
<td>0.226</td>
<td>4.40</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean substitution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10% MCAR</td>
<td>0.831</td>
<td>0.218</td>
<td>3.80</td>
<td>&lt;.001</td>
<td>-16.23</td>
<td>-3.54</td>
</tr>
<tr>
<td>10% MAR</td>
<td>0.856</td>
<td>0.222</td>
<td>3.86</td>
<td>&lt;.001</td>
<td>-13.71</td>
<td>-1.77</td>
</tr>
<tr>
<td>20% MCAR</td>
<td>0.807</td>
<td>0.200</td>
<td>4.04</td>
<td>&lt;.001</td>
<td>-18.65</td>
<td>-11.50</td>
</tr>
<tr>
<td>20% MAR</td>
<td>0.809</td>
<td>0.213</td>
<td>3.80</td>
<td>&lt;.001</td>
<td>-18.45</td>
<td>-5.75</td>
</tr>
<tr>
<td>50% MCAR</td>
<td>0.518</td>
<td>0.169</td>
<td>3.07</td>
<td>.003</td>
<td>-47.78</td>
<td>-25.22</td>
</tr>
<tr>
<td>50% MAR</td>
<td>0.567</td>
<td>0.178</td>
<td>3.19</td>
<td>.002</td>
<td>-42.84</td>
<td>-21.24</td>
</tr>
<tr>
<td>Multiple imputation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10% MCAR</td>
<td>0.880</td>
<td>0.240</td>
<td>3.66</td>
<td>&lt;.001</td>
<td>-11.29</td>
<td>6.19</td>
</tr>
<tr>
<td>10% MAR</td>
<td>1.06</td>
<td>0.247</td>
<td>4.28</td>
<td>&lt;.001</td>
<td>6.85</td>
<td>9.29</td>
</tr>
<tr>
<td>20% MCAR</td>
<td>0.934</td>
<td>0.223</td>
<td>4.19</td>
<td>&lt;.001</td>
<td>-5.85</td>
<td>-1.33</td>
</tr>
<tr>
<td>20% MAR</td>
<td>1.207</td>
<td>0.254</td>
<td>4.75</td>
<td>&lt;.001</td>
<td>21.67</td>
<td>12.39</td>
</tr>
<tr>
<td>50% MCAR</td>
<td>0.923</td>
<td>0.274</td>
<td>3.37</td>
<td>.001</td>
<td>6.96</td>
<td>21.24</td>
</tr>
<tr>
<td>50% MAR</td>
<td>1.162</td>
<td>0.281</td>
<td>4.14</td>
<td>&lt;.001</td>
<td>17.14</td>
<td>24.34</td>
</tr>
<tr>
<td>FIML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10% MCAR</td>
<td>0.867</td>
<td>0.230</td>
<td>3.77</td>
<td>&lt;.001</td>
<td>-12.60</td>
<td>1.77</td>
</tr>
<tr>
<td>10% MAR</td>
<td>1.045</td>
<td>0.237</td>
<td>4.41</td>
<td>&lt;.001</td>
<td>5.34</td>
<td>4.87</td>
</tr>
<tr>
<td>20% MCAR</td>
<td>0.924</td>
<td>0.226</td>
<td>4.08</td>
<td>&lt;.001</td>
<td>-6.85</td>
<td>0.00</td>
</tr>
<tr>
<td>20% MAR</td>
<td>1.195</td>
<td>0.258</td>
<td>4.64</td>
<td>&lt;.001</td>
<td>20.46</td>
<td>14.16</td>
</tr>
<tr>
<td>50% MCAR</td>
<td>0.893</td>
<td>0.283</td>
<td>3.16</td>
<td>.002</td>
<td>-9.98</td>
<td>25.22</td>
</tr>
<tr>
<td>50% MAR</td>
<td>1.133</td>
<td>0.313</td>
<td>3.61</td>
<td>&lt;.001</td>
<td>14.21</td>
<td>38.50</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean substitution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple imputation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIML</td>
<td>11.63</td>
<td>12.46</td>
<td></td>
<td></td>
<td>11.57</td>
<td>14.09</td>
</tr>
</tbody>
</table>

*Note.* Results are based on simulated data (N = 60). Parameter effect size $r_{xy} = .50$ (group on outcome); association of covariate used for imputation with independent variable $r_{yz} = .50$ (covariate) and dependent variable $r_{yx} = .70$. Overall results are based on absolute values averaged across type and amount of missing data. MCAR = missing completely at random; MAR = missing at random; FIML = full information maximum likelihood.
largest bias in standard errors was 24.34% at 50% MAR. MI did not systematically under- or overestimate regression coefficients or standard errors.

**Full information maximum likelihood.** The FIML procedure performed well and was quite comparable with MI, with the largest bias in regression coefficient also at 20% MAR (20.46%) and largest bias in standard error at 50% MAR (38.50%). At common amounts of missing data, FIML resulted in an average bias of 11.31% in estimating regression coefficients and 5.2% in estimating standard errors. At extreme missingness, FIML produced regression coefficients that were biased on average by 12.10% and standard errors that were biased by 31.86%. In a comparison of bias between MCAR and MAR, FIML tended to be more biased under MAR compared with MCAR for both the regression coefficients (13.34% vs. 9.81%, respectively) and their standard errors (19.18% vs. 13.50%). FIML did not systematically over- or underestimate regression coefficients but did systematically overestimate standard errors.

**Discussion**

In this article we examined three different methods for handling missing data: mean substitution, MI, and FIML. Results of independent samples t tests showed that, consistent with much previous research, mean substitution is a poor method of choice for dealing with missing data (Allison, 2001; Bennett, 2001; Graham et al., 2003; Pallant, 2007). In our illustration, the mean substitution method of imputation resulted in appreciable bias in both regression coefficients and standard errors even when the amount of missing data was typical (e.g., 10% and 20%). Furthermore, mean substitution resulted in considerable bias when missing data was severe (50%). The bias observed when mean substitution was used should be of considerable concern to counseling psychologists, because standard errors were systematically overestimated, which leads to a reduction in power to detect effects that are otherwise present. We therefore strongly recommend against using mean substitution as a method for imputing missing data. In our fictitious example, mean substitution might lead us to conclude that there was no difference (or only a small difference) between the two counseling approaches on students’ self-efficacy for sobriety, when in fact there actually was a moderate or large difference. In practice, that might lead to a faulty decision to offer the treatment that had the smaller effect.

So if mean substitution is a poor option, what other options does a researcher have? We have shown here that, consistent with much theoretical and simulation work that has preceded the present article, both MI and FIML methods are viable strategies for handling missing data. In our illustration, both MI and FIML provided acceptable estimations of regression coefficients and standard errors across both types of missing data (MCAR and MAR) as well as across 10% and 20% missing data. Although not analyzed in this study, both MI and FIML have also shown to perform reasonably well when data are NMAR and the amount of missing data is moderate (e.g., 25%; Buhi, Goodson, & Neilands, 2008). Although MI and FIML outperformed mean substitution, because they introduced enough bias to be of concern when data were missing at 50%, we suggest that when severe amounts of missing data are present, counseling psychologists make use of additional auxiliary variables that might predict missing values in order to improve the missing data model and subsequent analyses. Both of these methods (MI, FIML) would allow for more informed decisions about which treatment method is more effective.

**Conclusions Regarding Missing Data**

We strongly urge authors to follow best practices in reporting and handling missing data, and we urge editors to establish policies that insist that missing data be attended to in quantitative articles. Failure to do so means that findings, and interpretations of those findings, may be quite biased. Drawing conclusions (or determining policy) on the basis of biased results does not advance science and in fact may take it in an inappropriate direction.

In the field of counseling psychology, the scientist–practitioner model is the foundation upon which training and practice are constructed. If science is to accurately inform practice, it must not ignore such an important element of good science as accounting for the missing data that is so commonly found in research. Just as researchers routinely screen and clean their data prior to analysis and gather descriptive data about their samples, they should make missing data analysis part of the systematic first steps in data analysis. Researchers know that they must report response rates for survey data and include effect sizes for statistically significant results. The same expectation for accurate reporting should apply to missing data management. Ignoring this step is poor science, and results reported without attention to missing data can misinform our scientific understanding and misguide policy and practice. We believe that the field of counseling psychology endeavors to disseminate high-quality research and therefore must make reporting missing data a best practice for the field.

**References**


Appendix

Suggested Steps in Reporting and Managing Missing Data in Quantitative Analyses

A. Report the amount of missing data as a percentage of complete data.
   1. In the text, include a statement giving the range of missing data. For example, one could say, “Missing data ranged from a low of 4% for Attachment Anxiety to a high of 12% for Depression.”
   2. If a table with means, standard deviations, and/or internal consistency estimates of variables is included, add a column that shows the percentage of missing data for that variable.
   3. Include both item nonresponse and participant attrition in longitudinal studies.
   4. In longitudinal designs, report the percentage of attrition at each wave and over the entire study.
   5. When missing data are found on individual items that are used to calculate a total or scale score, impute the values for the items first, and then recalculate scale or total scores as needed.

B. Consider the pattern of missingness.
   1. To distinguish between missing completely at random (MCAR) and missing at random (MAR), select either of the following methods:
      a) Empirically evaluate relations between observed variables and missing values.
         (1) Create a dummy variable with two values: missing and nonmissing.
         (2) Use standard statistical procedures to test the relation between this variable and the other variables of interest in the data set.
      a) If the dummy variable is not related to any other variables, then the data are either MCAR or not missing at random (NMAR).
      (b) If the dummy variable is associated with other variables, then the data are MAR or NMAR

(Appendix continues)
   (1) Little's test can be computed in the Missing Values Analysis add-on module in SPSS. An SAS macro for this test can be downloaded at http://psychology.clasu.edu/files/LittlesTestofMCARMissingData.sas
   (2) If the p value for this test is not significant, then this indicates the data are MCAR.
   (3) Note that this is an omnibus test; it is looking at the data set as a whole, not the individual variables.
2. If the missing data are not MCAR or MAR, and NMAR is suspected, then researchers should consider that the missing data might well follow the NMAR pattern.
3. Include a conclusion about the pattern of missingness in the results after reporting the amount of missingness.

C. Determine the most appropriate method for handling missing data.
1. Examine published evidence regarding the performance of different missing data strategies to find similar situations.
2. Ensure that the method chosen is appropriate for the pattern of missingness determined in the previous step.
3. Consider expectation maximization, multiple imputation, and full information maximum likelihood.
4. After providing the amount and pattern of missing data, report the method chosen to handle the data, and give a brief rationale for that selection.

Received June 15, 2009
Revision received October 16, 2009
Accepted October 19, 2009

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the American Psychologist and APA Monitor. High School Teacher and International Affiliates receive subscriptions to the APA Monitor, and they may subscribe to the American Psychologist at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the Publication Manual of the American Psychological Association, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.