Published in final edited form as:

Exp Clin Psychopharmacol. 2012 June ; 20(3): 243–250. doi:10.1037/a0027146.

Missing Data in Substance Abuse Treatment Research: Current Methods and Modern Approaches

Sterling McPherson, Celestina Barbosa-Leiker, G. Leonard Burns, Donelle Howell, and John Roll

Sterling McPherson, Celestina Barbosa-Leiker, and John Roll, College of Nursing, Department of Psychology, Program of Excellence in the Addictions, and Program for Rural Mental Health and Substance Abuse Treatment, Washington State University; G. Leonard Burns, Department of Psychology, Washington State University; Donelle Howell, College of Nursing, Program of Excellence in the Addictions, and Program for Rural Mental Health and Substance Abuse Treatment, Washington State University

Abstract

Two common procedures for the treatment of missing information, listwise deletion and positive urine analysis (UA) imputation (e.g., if the participant fails to provide urine for analysis, then score the UA positive), may result in significant biases during the interpretation of treatment effects. To compare these approaches and to offer a possible alternative, these two procedures were compared to the multiple imputation (MI) procedure with publicly available data from a recent clinical trial. Listwise deletion, single imputation (i.e., positive UA imputation), and MI missing data procedures were used to comparatively examine the effect of two different buprenorphine/naloxone tapering schedules (7- or 28-days) for opioid addiction on the likelihood of a positive UA (Clinical Trial Network 0003; Ling et al., 2009). The listwise deletion of missing data resulted in a nonsignificant effect for the taper while the positive UA imputation procedure resulted in a significant effect, replicating the original findings by Ling et al. (2009). Although the MI procedure also resulted in a significant effect, the effect size was meaningfully smaller and the standard errors meaningfully larger when compared to the positive UA procedure. This study demonstrates that the researcher can obtain markedly different results depending on how the missing data are handled. Missing data theory suggests that listwise deletion and single imputation procedures should not be used to account for missing information, and that MI has advantages with respect to internal and external validity when the assumption of missing at random can be reasonably supported.

Keywords

substance abuse treatment; missing data; positive drug test imputation; multiple imputation

Missing data in the social and behavioral sciences are often treated in a suboptimal manner (Graham, 2009; Hedden, Woolson, & Malcolm, 2008; Wilkinson & Task Force on Statistical Inference, 1999; Yang & Shoptaw, 2005). Under most circumstances, two common procedures (i.e., listwise deletion, single imputation) represent poor options due to

© 2012 American Psychological Association

Correspondence concerning this article should be addressed to Sterling McPherson, Washington State University, College of Nursing, P.O. Box 1495, SNRS 414E, Spokane, WA 99210-1295. smcpherson05@wsu.edu.

All authors contributed in a significant way to this article and all authors have read and approved the final manuscript. None of the authors have any financial, personal, or other type of relationship that would cause a conflict of interest that would inappropriately impact or influence the research and interpretation of the findings.

the loss of power as well as inaccurate measures of effect size and standard errors (Allison, 2001; Enders, 2010; Schafer & Graham, 2002), thus potentially resulting in incorrect conclusions. Although researchers should try everything to reduce the likelihood of missing data in substance abuse research, it is likely that given the nature of the research there will continue to be a significant amount of missing information. When evaluating different strategies of missing data treatment, the researchers should keep in mind three important goals (Acock, 2012): maximize the information used in the analysis, minimize the bias in estimating model parameters, and minimize the bias in estimating standard errors by correctly reflecting the degree of uncertainty associated with parameter estimates. With these goals in mind, what then are the best approaches to handle missing information?

Substance abuse treatment research is often characterized by having a nontrivial amount of missing data, especially in longitudinal, randomized clinical trials (e.g., Hedden et al., 2008; Wood, White, & Thompson, 2004; Graham, Hofer, & Piccinin, 1994; Yang & Shoptaw, 2005). In addition, a review of substance abuse clinical trials indicates that listwise deletion and single imputation are the most commonly used procedures to deal with missing information (Arndt, 2009; Wood et al., 2004). For example, Arndt (2009) reported that 94% of clinical trials filled in the missing value by assuming that the UA would have been positive or with the level of use that was observed at baseline. The remaining 6% used the observation prior to the missing value to fill in the missing value. All three of these choices represent a type of single imputation.

Many of these substance abuse studies assume that the participants' failure to show up was due to the participants' anticipation that the UA would be positive, or that the participant was intoxicated and unable to attend. Although this may be the case for many participants, it is also likely that the failure to appear is due to a host of other reasons (e.g., the participant decides the treatment was successful, thus no further need to attend; the participant decides the treatment was not helpful, thus no further reason to attend; the participant was placed in a controlled environment such as a jail or hospital, thus not able to attend; and so on). Due to the prevalence and complex nature of missing data in substance abuse research, the best missing data procedures need be employed and tailored to each unique missing data situation so as to maximize the likelihood of arriving at the most accurate conclusions (e.g., Arndt, 2009; Enders, 2010; Graham et al., 1994; Hedden et al., 2008).

This paper will first describe three missing data mechanisms to provide the foundation for the discussion of different procedures for the treatment of missing information and then review the potential problems associated with the use of the listwise and single imputation procedures. The paper will then describe a modern procedure for the treatment of missing data—multiple imputation (MI). Lastly, the use of listwise deletion, single imputation, and MI to treat missing information with a recent substance abuse clinical trial (Clinical Trial Network 0003; Ling et al., 2009) will be demonstrated. Our purpose is to compare the outcomes from the listwise and single imputation procedures with the outcomes from the MI procedure in order to demonstrate that interpretations of treatment effectiveness can change as a function of how the missing information is handled.

Causes of Missing Data (Why Are My Data Missing?)

There are three different ways to characterize missing data—missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR; Rubin, 1976). Missing completely at random means the data are missing due to a completely random process that cannot be explained by any variable inside or outside of the dataset. For example, if a researcher slips in the rain and drops a handful of participant surveys, these

survey responses would now be missing due to an entirely random process that has nothing to do with any variables inside or outside of the dataset.

The second mechanism, MAR, means that variables within the dataset can be used in the analysis process that will account for the missing information even though the data are not missing completely at random. For example, if some participants leave a question blank on a predictor variable due solely to their age and gender (e.g., older participants and men tend to leave the question blank), then age and gender can be used to account for (e.g., assist with estimating the missing values and recover important characteristics of the data) the missingness by utilizing them during the missing data analysis. If this procedure is done properly and all of the variables responsible for the propensity for missingness are included in the analysis model, it results in the data effectively becoming MCAR.

The third mechanism, MNAR, is the most problematic. Here the missingness is not completely random or explainable by other predictor variables. The most common MNAR situation is when the missingness is due to the outcome itself (Enders, 2010). For example, a participant involved in a cocaine treatment program does not show up for a UA due to current cocaine use.

The MAR and MNAR mechanisms are probably not mutually exclusive with the missing data in most substance abuse research being due to a combination of both (Graham, 2009). Unfortunately, there is currently no test to determine if the missing information is due to the MAR or the MNAR mechanism, and justification for assuming either MAR or MNAR often relies on a well-reasoned, logical argument (Enders, 2010; Graham, 2009; Little & Rubin, 2002). This issue can be even further complicated when considering longitudinal data that contain both intermittent missing data and monotonic (i.e., permanent dropout) missing data that could indicate different mechanisms of missing data being at work simultaneously. It is important that any investigate team address the missing data in the presentation of results, and understand that one of the three assumptions described above is assumed no matter what method of missing data treatment is used in conjunction with the chosen analytic strategy.

While most researchers' missing data are likely somewhere between MAR and MNAR, the MI procedures are not commonly implemented in substance abuse research. However, it should be noted that in most instances of missing data (i.e., MNAR or MAR), it is impossible to know for sure why one's data are missing, and thus, impossible to know what the missing values would have been. The researcher's decision for what mechanism to assume becomes one of deciding how well s/he can predict why the values in question are missing based on the variables at his or her disposal. The better s/he can account for the missing values given relevant variables in the dataset, the more tenable the MAR assumption becomes.

Common Methods for the Treatment of Missing Data in Substance Abuse Research

As noted earlier, listwise deletion and single imputation are the most commonly used procedures to treat missing information in substance abuse research. Listwise deletion (also referred to as *complete case* analysis) discards the participant if the participant is missing any data on variables needed to perform the analysis. This method assumes that the data are MCAR. This procedure can lead to significant bias as well as reduce the power of the analysis to detect a true effect (Enders, 2010; Graham, 2009). This method is not recommended unless the data are MCAR with the amount of missing data being very small (see Enders, 2010). Under these limited circumstances, the listwise deletion procedure and the modern procedure of MI will likely yield unbiased estimates (Graham, 2009; Schafer &

Graham, 2002). Unless there is a dramatic power difference between the analyses, the final results will also be similar.

Single imputation involves a set of procedures (e.g., replacing the missing value with the mean for the variable; replacing the missing value with an earlier value in the longitudinal study; replacing the missing value with a positive score on the UA). Two of these procedures are especially common in substance abuse research. One is referred to as the *last value (or observation) carried forward* procedure (e.g., the missing value at time 2 is replaced by an earlier value from time 1). This first procedure has been used for both intermittent missingness or for permanent dropout missingness. The second procedure we refer to as *positive UA imputation* (i.e., the participant who fails to show up for a UA receives a positive UA). Here the missing UA data are scored as a positive UA. Although both of these single imputation procedures have been commonly considered “conservative approaches” to the treatment of missing information in substance abuse research, each procedure can result in biased parameter estimates as well as inappropriately small standard errors, which artificially increases the likelihood of a statistically significant outcome (e.g., an incorrect conclusion that treatment was effective; see Arndt, 2009; Cook, Zeng, & Yi, 2004; Hedden et al., 2008; Shao & Zhang, 2004). Potential alternatives exist for dealing with missing data in substance abuse research, the focus of the next section.

Recommended Methods for the Treatment of Missing Data in Substance Abuse Research

Maximum likelihood (ML) estimation and MI are currently considered two newer procedures for the treatment of missing information (Enders, 2010; Graham, 2009; Schafer & Graham, 2002). Both procedures assume the missing data mechanism is MAR and the data are multivariate normal. When performed appropriately, ML and MI yield nearly identical results (Enders, 2006; Enders, 2010). Some authors recommend one procedure over the other in limited circumstances, but for the most part, ML and MI will perform equally well in nearly all missing data situations, assuming MAR is tenable. Because the analysis in the current paper uses MI, the paper will only describe this procedure (see Enders, 2001 for a description of the ML approach to missing data). We chose to use MI in the current situation because MI more easily allows for missingness in explanatory variables (e.g., UA result during the first week of treatment), which is a common situation for substance abuse researchers (see Enders, 2010 for a thorough comparison of MI and ML).

MI is a Bayesian approach that uses a regression equation (specified by the researcher) where the complete variables predict the incomplete variables. This process then generates predicted values for the missing data from these regression equations, but adds a random residual to each predicted value. For each coefficient in the imputation regression equations, MI will add a random residual to the coefficient itself. Adding random error to each term in the regression equation produces an alternate equation that is different, yet equivalent, to the one that generated the original imputed values. It is this newly constructed regression equation that is used to start the process over again and produce an additional dataset. MI carries out this process multiple times in order to generate multiple complete datasets.

This first phase of MI can (should) include the use of auxiliary variables as predictor variables in the initial imputation regression equation to increase the accuracy of the imputed values in each data set (Collins, Schafer, & Kam, 2001; Graham, 2009). Auxiliary variables are variables that are potential correlates of missingness or correlates of the analysis variables. Although the auxiliary variables do not need to be of substantive interest (i.e., are not predictors in the model), such variables may be related to the missing data mechanism or to variables that are responsible for the missingness. For example, a

researcher may not be interested in how age affects treatment effectiveness, but if older participants tend to be missing, age would be used as an auxiliary variable. This inclusion of auxiliary variables will refine the missing data estimation process while also increasing the likelihood of satisfying the MAR assumption (Enders, 2010; Graham, 2009). Acock (2012) has suggested that an ideal set of auxiliary variables is fairly small in number and can serve at least one (but ideally both) of the following functions: predict what the missing value would have been if it were observed, predict the propensity for producing a missing value. Also, gender and race have been shown to be commonly associated with missing values (Collins et al., 2001), and their inclusion alone can make the MAR assumption much more tenable.

The second MI phase involves the analysis of the multiple, newly created datasets, and the third phase aggregates the parameter estimates generated for each dataset into one final set of estimates. The sampling variance (i.e., the squared standard error) aggregation does not involve a simple average, rather the process takes an adjusted average that accounts for the variance between the imputed estimates and the within variance (i.e., variance generated during the iterative optimization process for each data value estimate) for the missing values (Enders, 2006; Schafer & Graham, 2002). The MI procedure has demonstrated exceptional performance in comparison to other methods of treating missing data with the exception of ML (e.g., Graham, 2009; Enders, 2010) when the assumption of MAR is tenable.

Clinical Trial Network Dataset 0003

The paper will now demonstrate the use of listwise deletion, single imputation, and MI procedures for the treatment of missing information with the Clinical Trial Network dataset 0003 (Ling et al., 2009). This dataset was selected for two reasons. First, the primary hypothesis and subsequent analysis (i.e., chi-square test) reported by Ling et al. (2009) were straightforward and easily lent themselves to a pedagogical example. Our analyses attempted to replicate and extend those analyses using a slightly more complex analytic strategy (i.e., logistic regression), but one that did not require more sophisticated, longitudinal techniques that could be easily understood by a wide audience. While our rationale for selecting this dataset was in part due to ease of understanding the different missing data strategies, it is also important to note that all of these procedures can be used in conjunction with common longitudinal statistical analyses (i.e., generalized estimation equations) as well. Second, Ling et al. (2009) noted that the treatment of missing data represented a limitation of the study. One of our goals was to build on Ling et al.'s (2009) work and address this limitation.

The primary purpose of this trial was to compare a 7-day versus 28-day taper (i.e., stepwise decrease in the amount of administered buprenorphine/naloxone) on the likelihood of a positive UA for opioid use at the end of the taper. The results showed that those enrolled in the 7-day taper had significantly less chance of testing positive for opioid use compared to the 28-day taper (Ling et al., 2009). This study used positive UA imputation to treat the missing information (i.e., if a participant failed to show at the end of the taper then the UA was scored positive). Our purpose was to demonstrate how to use the MI procedure in conjunction with auxiliary variables to provide an alternative way to treat missing data in substance abuse research. Given the results of multiple simulation studies on missing data procedures (see Enders, 2010; Graham, 2009 for an overview), we predicted that the listwise procedure would result in the smallest estimated effect size while the positive UA imputation would result in the largest estimated effect size. The MI procedure was expected to result in an estimated effect size in the middle.

Method

Participants and Procedures

Participants were from National Drug Abuse Treatment Clinical Trials Network No. 0003—a publicly available dataset (<http://www.ctndatashare.org/>). This was a randomized, parallel-group, open-label study design with the procedures for both arms of the study being identical until the start of two taper periods. A total of 11 sites in 10 different U.S. cities were used for the trial. Participants were a sample of opioid-dependent individuals. A total of 990 individuals agreed to participate in the study with 894 being eligible and 748 eventually receiving the treatment medication, buprenorphine/naloxone. A total of 232 individuals were terminated from the trial for various reasons during the induction and stabilization phases. This resulted in a final total intention-to-treat sample of 516 participants who were potentially available for data collection at the end of the 7- or 28-day taper.

After the completion of the baseline assessments, the participants were stabilized on buprenorphine/naloxone across a 4-week stabilization period. After this stabilization phase, the patients were stratified across the maintenance dose of buprenorphine/naloxone and then randomized to the 7-day or 28-day taper groups. The taper groups did not differ statistically on age, gender, ethnicity, drug use, life-time drug use, past 30 days drug use, withdrawal symptoms (either clinically assessed or self-reported), number of concomitant medications taken during withdrawal symptoms, or opioid use during stabilization (Ling et al., 2009). There were also no documented differences in demographics or drug use characteristics between those who dropped out and those who completed treatment (Ling et al., 2009). The primary outcome of interest in the Ling et al. (2009) study was whether there was a significant difference between the 7-day and 28-day taper groups for the percentage of opioid-free urine specimens at the completion of the taper. A total of 28% (i.e., 144 participants) of the sample of 516 participants were not available for urine data collection at the completion of the taper. As noted earlier, these missing values were assigned a positive UA score. There was no missingness in any other variables, except for previous UAs collected that were used as auxiliary variables in the MI analysis. The data for the current analyses involved the same 516 participants.

Analytic Strategy

Mplus 6.1 (Muthén & Muthén, 1998–2010) was used to perform three separate logistic regression analyses on the data. When conducting MI using logistic regression via chained equations, the details of the estimation algorithm are somewhat different from the more general description given above, but the process is conceptually very similar. The first analysis used listwise deletion and the second analysis positive UA imputation to treat the missing information. The positive UA imputation is the same missing data treatment procedure used in the Ling et al. (2009) study. The third analysis used the MI procedure to deal with the missing information. Each logistic analysis regressed the probability of obtaining a positive UA score on the taper condition (7 days vs. 28 days).

The MI procedure also included multiple auxiliary variables (i.e., Weeks 1 through 4 opioid UA results, Wk1Topi to Wk4Topi in the Appendix) in the first phase of the procedure. These auxiliary variables were not of primary interest for the research question, but were included in the first MI phase to demonstrate their use as well as increase the likelihood of the MAR assumption being met. We also included the covariates of gender and race not only to replicate analyses performed by Ling et al. (2009), but also to help justify the MAR assumption (see earlier discussion of auxiliary variables).

One potential reason for a modern approach such as MI not being used more commonly is the lack of readily available computer syntax to start from and adapt for one's own analyses.

Thus, we have included the Mplus code for the first and second phase of the MI procedure, shown in the Appendix.

Results

Listwise Logistic Regression

Table 1 shows the results from the logistic regression where the listwise deletion procedure was used to handle the missing information ($n = 373$). We regressed UA results at the end of the each arm's respective taper onto age, sex, race, and trial arm. Neither the taper condition (7-day vs. 28-day taper) nor any of the covariates were significantly related to the probability of a positive UA. The listwise deletion procedure thus indicates that the treatment condition failed to have a significant effect—a different outcome than the significant effect reported by Ling et al. (2009)—likely due to an increase in bias associated with the listwise procedure, but there is no way to tell for sure that this is the case.

Positive UA Imputation Logistic Regression

Table 1 also shows the results of logistic regression where the positive UA imputation procedure was used to handle the missing information. We regressed UA results at the end of the each arm's respective taper onto age, sex, race, and trial arm. Here the main effect for the taper condition was statistically significant ($\beta = 0.568, p = .003$). The taper condition resulted in an odds ratio of 1.766 (95% C. I.: 1.216–2.564), such that those in the 28-day taper group were approximately 77% more likely to submit a positive UA at the end of the taper. Ling et al. (2009) reached the same conclusion, which is undoubtedly correct.

It is important to note that the standard errors for the positive UA imputation method are markedly smaller than the standard errors for the listwise deletion procedure (i.e., standard error difference = 0.017). This outcome is consistent with the earlier discussion that a weakness of the positive UA imputation procedure is that this procedure produces inappropriately small standard errors because it assumes absolute certainty in the imputed values. This weakness inflates the Type I error rate (i.e., an incorrect conclusion that the treatment was effective). In addition, there is a significant power difference between the two analyses that is likely a contributing factor as to why the standard errors are markedly different.

Multiple Imputation Logistic Regression

Table 1 also shows the results from the logistic regression where the MI procedure was used to handle the missing information. The first phase of imputation (i.e., creating multiple datasets with imputed values) used weeks 1–4 opioid UA results of the clinical trial, in addition to age, race, and sex in order to impute the newly created datasets (20 imputed data sets were created, see Appendix). In this example, weeks 1 through 4 opioid UA results were treated as auxiliary variables. Since these variables were only part of the imputation phase, but not the analysis phase, their primary purpose was to help ensure that the mechanism of missingness was closer to MAR than MNAR, and assisted in accurate imputation of the datasets. The Mplus code that performed the aggregated logistic regression on the 20 datasets (recommended minimum is five; see Graham, Olchowski, & Gilreath, 2007) is shown in the second section of the Appendix. We imputed 20 datasets rather than the commonly cited minimum of five because Acock (2012) has shown that the relative efficiency (i.e., minimization of standard errors compared to complete dataset) increases from 90.9% with five datasets to 97.6% with 20 datasets and 50% missing data. Acock (2012) cites 20 imputed datasets as a reasonable minimum in order to achieve a higher relative efficiency. The only limitation associated with imputing additional datasets is the

computational time and resources required, which is less of an issue with modern computing capability.

We regressed UA results at the end of the each arm's respective taper onto age, sex, race, and trial arm. The results indicated the impact of the taper condition was statistically significant ($\beta = 0.411$, $p = .037$). This treatment effect resulted in an odds ratio of 1.508 (95% C. I.: 1.025–2.219), such that those in the 28-day taper group were approximately 51% more likely to submit a positive UA at the end of the taper. However, in comparison to the positive UA imputation results, all the standard errors for the MI procedure are larger with one exception (see Table 1). The standard errors are larger because the MI procedure accounts for the uncertainty and imprecision that is inherent whenever missing values are imputed. In addition, all the parameter estimates for the MI procedure are smaller than parameter estimates from the positive UA imputation procedure with one exception. Assuming that the data are MAR, this means that when the missing data are treated using the positive UA imputation approach, the parameter estimates could be inflated and the standard errors suppressed, which are two biases that lead to Type I errors (i.e., concluding that the effect of treatment was effective when in actuality, it was not).

Figure 1 shows how the positive UA imputation method, although generally considered a conservative procedure to treat missing information, can exaggerate group differences that may not reflect the most likely UA for some individuals. Observe that the percent of individuals with a positive UA is meaningfully higher when using the positive UA imputation approach compared to the MI approach. While it could be argued from a clinical standpoint that this is the most likely outcome for individuals who do not show up for treatment, this assumption is tested within the context of MI in that we were able to utilize other variables (i.e., previous UAs and other auxiliary variables) that can help estimate the most likely value.

Discussion

Missing data are a ubiquitous problem in substance abuse treatment research, especially in longitudinal clinical trials. Listwise deletion and single imputation of missing information, two common procedures used to deal with missing information, are potentially problematic. Listwise deletion is based on the MCAR assumption and, even if this assumption is met, nearly always results in a significant erosion of power and can increase overall bias if the MCAR assumption is not met (which is most often the case), thus decreasing the likelihood of the identification of a real treatment effect. The MCAR assumption is made with respect to missing information in studies that analyze substance abuse treatment clinical trial data utilizing the general estimation equation procedures. Because of the MCAR assumption being made, this common strategy essentially ignores the missing information and could lead to severely biased estimates (Enders, 2010). Single imputation procedures (e.g., replacing the missing value with the mean for the variable; replacing the missing value with an earlier value in the longitudinal study; replacing the missing value with a positive score on the UA) is the most favored approach but will often result in biased treatment effects and decreased standard errors, thus increasing the likelihood of incorrectly concluding the treatment was effective. MI, a modern procedure to the treatment of missing information, does not entail the weaknesses of listwise or the single imputation procedures when the MAR assumption can be justified (Enders, 2010).

It cannot be stressed enough that these different approaches to missing data represent the researchers' best estimate of the reason for missing values because, in most cases, it is impossible to recover missing values and therefore impossible to know exactly what those values would have been and, hence, why the researcher is faced with missing data. For

example, it is possible in our analysis that the true mechanism of missing data is MNAR, but our strong set of auxiliary variables (e.g., previous UAs) helps to justify the MAR mechanism. To the best of our knowledge, the current study represents the first comparison of the listwise, single imputation (i.e., positive UA imputation), and MI procedures with data from a substance abuse clinical trial.

Listwise deletion, as expected due to the inability of this procedure to make full use of the data and the injection of bias into the results, indicated that the effect for the taper condition was not significant. Although the positive UA imputation approach did indicate that the effect of the taper condition was significant, the same result as reported by Ling et al. (2009) in the original study where positive UA imputation was used to deal with the missing information, the positive UA imputation procedure resulted in an inflated treatment effect relative to the more appropriate MI procedure, assuming MAR for the current set of analyses. For example, the positive UA imputation procedure produced a 0.25 larger odds ratio (25% odds ratio inflation for positive UA in 28-day taper arm compared to 7-day taper arm) compared to the MI procedure. Because the positive UA imputation procedures assume absolute certainty when filling in the missing values with positive UA scores, this procedure produces biased estimates and decreased standard errors (i.e., greater likelihood of concluding an effect exists when it does not; Arndt, 2009). Moreover, our analysis suggests that contrary to conventional wisdom that positive UA imputation represents a conservative approach to missing data, this method can make treatment differences appear larger than they are in reality. While the difference of 0.25 between the two odds ratios (positive UA imputation result vs. multiple imputation result) may seem modest, we believe that the most accurate effect size estimate of treatment in all settings is critical, especially in light of effect sizes being the primary component of systematic reviews and meta-analyses.

The missing data situation described in this investigation generalizes to many other clinical trial missing data situations wherein there is missing data on the outcome of interest only. Thus, we consider our analyses to be a prototype to guide future sensitivity analyses conducted by investigating teams interested in evaluating how the interpretation of treatment effectiveness varies as a function of how the missing data are handled. It is also important to note that this same set of analyses could be conducted in the context of longitudinal analytic techniques (i.e., generalized estimation equations, multilevel modeling).

It may be argued that since we do not know what the missing values were, they could have all been positive values and therefore the positive UA imputation would represent a good approach to handle the missing values. While it is true that the missing values are impossible to recover, the MI approach builds on the positive UA imputation approach by allowing the researcher to use clinical judgment (i.e., the appropriate selection of variables used to impute the missing values) and statistical inference to arrive at the *most likely* estimate, given the data and assuming MAR. If the missing values were most likely all positive UAs, then the data should bare this out by producing positive UAs as the most likely estimates based on the data. In addition, the MI approach allows the researcher the flexibility of using different variables that are unique to the missing data situation to account for the missing values. Thus, MI is a tool that builds on the positive UA imputation approach by allowing the researcher to exercise his or her clinical expertise more precisely by way of choosing variables to assist in the estimation of the missing values.

Based on previous research and theoretical work done on missing data theory, we conclude that the MI procedure produced estimates that have the greatest likelihood of reflecting reality assuming the data is MAR. Given the problems associated with the listwise and single imputation procedures, researchers should consider other analytic strategies, especially given the ease of use of the MI procedure within the Mplus statistical software

(see the Appendix). By including relevant auxiliary variables (e.g., previous 4 weeks of UAs), we likely accounted for much of the reason for missingness, which makes the MAR assumption more tenable. As a result, this makes the MI procedure more valid than procedures that assume MCAR or otherwise do not account for the inherent uncertainty of imputing missing values.

One weakness of the current study was that the demonstration was only in the context of one substance abuse clinical trial data set. While simulation studies suggest that the same results would occur with the application of the three procedures to other clinical trial data sets (e.g., Allison, 2001; Enders, 2010), this remains an empirical question. However, missing data theory and simulation studies (Enders, 2010; Graham, 2009) along with the current results caution against the continued use of the listwise and single imputation procedures.

A more central weakness of the current study is that the MI procedure is based on the MAR assumption. This assumption could prove difficult to justify in different applications. However, it is also important to note that undertaking an analysis that assumes MNAR also requires a set of assumptions that are just as difficult to justify (Enders, 2010). Thus, the MI procedure should not be viewed as a “fix-all” technique that will work in every missing data situation. Even if an assumption of MAR is tenable (and hence, the MI procedure allowable), the researcher may be specifically interested in the nature of the missingness and have hypotheses about why dropout occurs. For example, some researchers may be interested in simultaneously estimating the model of interest and a discrete time-survival model for “time to dropout” (i.e., Diggle-Kenward Selection Model; Diggle & Kenward, 1994; Enders, 2010; Muthén & Muthén, 1998–2010). Another example would be the estimation of a latent class for each group of participants with the same missing data pattern. This would allow the researcher to investigate invariance (e.g., significant covariates, treatment effectiveness) across the missing data pattern classes (i.e., Pattern-Mixture Modeling; Enders, 2010; Little, 1993; Muthén & Muthén, 1998–2010).

Consistent with previous investigations of missing data in substance abuse treatment (Arndt, 2009; Hedden et al., 2008; Yang & Shoptaw, 2005), we encourage researchers to understand and report the missing data mechanism as well as use newer procedures for the treatment of missing information (i.e., MI or direct maximum likelihood procedures) that are based on a researcher-specified “best estimate” of the missing values. It is our hope that this brief introduction and example can serve as the starting point for researchers dealing with missing data.

Acknowledgments

This research was supported by grants from the Department of Justice to the Program of Excellence in the Addictions, and a grant from the Life Science Discovery Fund to the Rural Mental Health and Substance Abuse Treatments Program. Dr. John Roll is the principal investigator on both of these grants, and Dr. Donelle Howell is the research director of the Program of Excellence in the Addictions. In addition, this project was supported by funds from a grant to the Clinical Trials network Pacific Northwest Node (award number 5 U10 DA013714-10) from the National Institute on Drug Abuse (NIDA). Dr. Dennis Donovan and Dr. John Roll are co principal investigators on this grant. These funding sources had no role than financial support. This work would not have been possible if it weren't for the original clinical trial conducted by Dr. Walter Ling and colleagues, and the National Drug Abuse Treatment Clinical Trials Network making the data available to the public. We would also like to thank Suzette Evans for her constructive and helpful feedback on previous versions of this article.

Appendix

Appendix

Example Syntax of Multiple Imputation in the Mplus Software

<p>TITLE: Program 1 Multiple Imputation: Imputation Phase</p> <p>DATA: File is CTN 0003 Missing Data Project Mplus Dataset N= 516.dat; Format is 12F10.2;</p> <p>VARIABLE: Missing = All (999); ! Comment: This indicates the missing value “flag” that you have used. Names are Age Sex Race TriArm EdTpOpi Wk1Topi Wk2Topi Wk3Topi Wk4Topi Wk5Topi Wk6Topi Wk7Topi; Usevariables are Age Sex Race TriArm EdTpOpi Wk1Topi Wk2Topi Wk3Topi Wk4Topi; ! Comment: All variables in this list are used to help impute the values (i.e., auxiliary variables).</p> <p>DATA IMPUTATION: Impute = Race (c) EdTpOpi (c); ! Comment: The (c) indicates that these are categorical variables. Ndatasets = 20; ! Comment: The default number of datasets is 5. Save = CTN003MissingProjectImputed*.dat;</p> <p>ANALYSIS: Type = Basic;</p> <p>OUTPUT: Tech8;</p>	<p>TITLE: Program 2 Multiple Imputation: Analysis Phase</p> <p>DATA: File is CTN003MissingProjectImputedlist.dat; ! Comment: This is the file created from Program 1 above, and it tells Mplus where your imputed data files are so it can perform the aggregated analysis. Type = Imputation;</p> <p>VARIABLE: Missing = *; ! Comment: This command indicates how missing values are represented in the raw dataset. When datasets are created in Mplus, the default “flag” for missing values is an asterisk. Names are Age Sex Race TriArm EdTpOpi Wk1Topi Wk2Topi Wk3Topi Wk4Topi; Usevariables are Age Sex Race TriArm EdTpOpi; ! Comment: Notice that we are not using the auxiliary variables for the analysis phase. Categorical = EdTpOpi; ! Comment: It is necessary to alert Mplus of an outcome that is not continuous.</p> <p>MODEL: EdTpOpi on Age Sex Race TriArm;</p> <p>ANALYSIS: Type = General; Estimator = Mir;</p> <p>OUTPUT: Tech4 Crosstabs;</p>
---	---

References

- Acock, A. What to do about missing values. In: Cooper, H., editor. APA handbook of research methods in psychology. American Psychological Association; Washington, DC: 2012.
- Allison, PD. Missing data. Sage; Newbury Park, CA: 2001.
- Arndt S. Stereotyping and the treatment of missing data for drug and alcohol clinical trials. Substance Abuse Treatment, Prevention, and Policy. 2009; 4:2–3. doi:10.1186/1747-597X-4-2.
- Collins LM, Schafer JL, Kam C-M. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychological Methods. 2001; 6:330–351. doi:10.1037/1082-989X.6.4.330. [PubMed: 11778676]
- Cook RJ, Zeng L, Yi GY. Marginal analysis of incomplete longitudinal binary data: A cautionary note on LOCF imputation. Biometrics. 2004; 60:820–828. doi:10.1111/j.0006-341X.2004.00234.x. [PubMed: 15339307]
- Diggle PJ, Kenward MG. Informative dropout in longitudinal data analysis (with discussion). Applied Statistics. 1994; 43:49–73. doi:10.2307/2986113.
- Enders CK. A primer on maximum likelihood algorithms available for use with missing data. Structural Equation Modeling. 2001; 8:128–141. doi:10.1207/S15328007SEM0801_7.

- Enders CK. A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosomatic Medicine*. 2006; 68:427–436. doi:10.1097/01.psy.0000221275.75056.d8. [PubMed: 16738075]
- Enders, CK. *Applied missing data analysis*. Guilford Press; New York, NY: 2010.
- Graham JW. Missing data analysis: Making it work in the real world. *Annual Review of Psychology*. 2009; 60:549–576. doi:10.1146/annurev.psych.58.110405.085530.
- Graham, JW.; Hofer, SM.; Piccinin, AM. Analysis with missing data in drug prevention research. In: Collins, LM.; Seitz, L., editors. *Advances in data analysis for prevention intervention research*. National Institute on Drug Abuse; Washington, DC: 1994. p. 13-63.
- Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*. 2007; 8:206–213. doi:10.1007/s11121-007-0070-9. [PubMed: 17549635]
- Hedden SL, Woolson RF, Malcolm RJ. A comparison of missing data methods for hypothesis tests of the treatment effect in substance abuse clinical trials: A Monte-Carlo simulation study. *Substance Abuse Treatment, Prevention, and Policy*. 2008; 3:13–21. doi:10.1186/1747-597X-3-13.
- Ling W, Hillhouse M, Domier C, Doraimani G, Hunter J, Thomas C, Bilangi R. Buprenorphine tapering schedule and illicit opioid use. *Addiction*. 2009; 104:256–265. doi:10.1111/j.1360-0443.2008.02455.x. [PubMed: 19149822]
- Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*. 1993; 88:125–134. doi:10.2307/2290705.
- Little, RJA.; Rubin, DB. *Statistical analysis with missing data*. 2nd ed.. Wiley; Hoboken, NJ: 2002.
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. 6th Edition. Muthén & Muthén; Los Angeles, CA: 1998–2010.
- Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–592. doi:10.1093/biomet/63.3.581.
- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2002; 7:147–177. doi:10.1037/1082-989X.7.2.147. [PubMed: 12090408]
- Shao J, Zhang B. Last observation carry-forward and last observation analysis. *Statistics in Medicine*. 2004; 22:2429–2441. doi:10.1002/sim.1519. [PubMed: 12872300]
- Wilkinson L, Task Force on Statistical Inference. Statistical methods used in psychology journals: Guidelines and explanations. *American Psychologist*. 1999; 54:594–604. doi:10.1037/0003-066X.54.8.594.
- Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials*. 2004; 1:368–376. doi:10.1191/1740774504cn0320a. [PubMed: 16279275]
- Yang X, Shoptaw S. Assessing missing data assumptions in longitudinal studies: An example using a smoking cessation trial. *Drug and Alcohol Dependence*. 2005; 77:213–225. doi:10.1016/j.drugalcdep.2004.08.018. [PubMed: 15734221]

Figure 1.
Estimated percent of participants with positive or negative UA result at the end of taper by type of missing data treatment and Trial Arm.

Table 1

Likelihood of a Positive Urine Analysis for a 7-Day and 28-Day Drug Taper: Impact of Missing Data Treatment Choice on Outcomes

Parameters	(SE)	z	Odds Ratio (95% CI)
<i>Listwise Deletion of Missing Data (n = 373)</i>			
Age	−0.004 (0.010)	−0.359 ^{ns}	0.996 (0.98–1.02)
Sex	0.148 (0.219)	0.675 ^{ns}	1.159 (0.76–1.78)
Race	−0.101 (0.093)	−1.082 ^{ns}	0.904 (0.75–1.09)
Trial Arm	0.335 (0.207)	1.618 ^{ns}	1.398 (0.93–2.10)
<i>Positive (Single) Urine Analysis Imputation of Missing Data (n = 516)</i>			
Age	−0.011 (0.009)	−1.213 ^{ns}	0.989 (0.97–1.01)
Sex	0.059 (0.203)	0.290 ^{ns}	1.061 (0.71–1.58)
Race	−0.126 (0.085)	−1.488 ^{ns}	0.881 (0.75–1.04)
Trial Arm	0.568 (0.190)	2.986 [*]	1.766 (1.22–2.56)
<i>Multiple Imputation of Missing Data (n = 516)</i>			
Age	−0.002 (0.009)	−0.246 ^{ns}	0.998 (0.98–1.02)
Sex	0.218 (0.205)	0.205 ^{ns}	1.244 (0.83–1.86)
Race	−0.098 (0.093)	−1.060 ^{ns}	0.907 (0.76–1.09)
Trial Arm	0.411 (0.197)	2.091 [*]	1.508 (1.03–2.22)

Note. The dependent measure was a negative or positive UA at the end of the taper (0 = negative UA; 1 = positive UA). Trial Arm represents the 7- and 28-day taper groups (0 = 7-day; 1 = 28-day).

ns = non-significant.

^{*}
p < .05.