EVALUATION TECHNICAL ASSISTANCE BRIEF
for OAH & ACYF Teenage Pregnancy Prevention Grantees

May 2013 • Brief 3

Coping with Missing Data in Randomized Controlled Trials

Missing data in randomized controlled trials (RCTs) can lead to biased impact estimates and a loss of statistical power. In this brief we describe strategies for coping with missing data in RCTs.[1] Specifically, we discuss strategies to:

- ***Describe the extent and nature of missing data.*** We suggest ways to provide a clear and accurate accounting of the extent and nature of missing outcome data. This will increase the transparency and credibility of the study, help with interpreting findings, and also help to select an approach for addressing missing data in impact analyses.
- ***Address missing data in impact analyses.*** Missing outcome or baseline data pose mechanical and substantive problems when conducting impact analyses. The mechanical problem is that mathematical operations cannot be performed on missing values. The substantive problem is that missing data can lead to biased impact estimates, especially if outcome data are missing or if missing data are handled inappropriately. We discuss methods that address both of these problems.

I. Describing the Extent and Nature of Missing Data

We suggest an approach to describing the extent and nature of missing outcome data that is primarily focused on assessing the potential for missing data to lead to biased study impacts. Missing outcome data can lead to two types of bias in impact estimates:

- ***Generalizability Bias.*** Sometimes a study is interested in the impacts of an intervention for a specific population of youth. If the youth who fail to respond to a follow-up survey are systematically different from the youth who do respond (for example, if nonrespondents have higher school absentee rates), then it is possible that the impact of the intervention might be different for survey respondents than it is for nonrespondents (for example, impacts might be smaller for youth with higher absentee rates). The difference between the impact calculated for respondents and the true impact for all youth (both respondents and nonrespondents) is "generalizability bias."[2]
- ***Causal Validity Bias.*** While youth who respond to follow-up surveys may be different from youth who do not, it may still be possible to calculate a causally valid impact for respondents[3] if there are no systematic differences between respondents in the treatment group and respondents in the control group. However, if respondents in the treatment group are systematically different from respondents in the control group, then the impact we calculate for respondents lacks causal validity. We call the difference between the calculated impact for respondents and the true impact for respondents "causal validity bias." This type of bias is the focus of the evidence review of teen pregnancy prevention programs (Mathematica Policy Research and Child Trends 2012).

We suggest conducting descriptive analyses designed to assess the potential for both types of bias in an RCT.

With regard to generalizability, we suggest conducting two analyses. First, we recommend creating either a CONSORT diagram (Campbell et al. 2012) or tables that clearly describe the size of the samples of schools and students that were randomly assigned, the number who consented to participate in the study, and the number who responded to surveys. For example, consider the case of a study that randomly assigned schools to treatment and control groups and then sought consent and surveyed students in those schools at baseline and follow-up. In Example Table 1, we show sample size information that can be presented at the school level, and in Example Table 2, we show sample size information that can be presented at the student level. Second, we recommend creating tables that describe differences in measured baseline characteristics between follow-up survey respondents and nonrespondents (that is, those with and without outcome data). The first two columns of Example Table 3 illustrate how this can be presented.

Example Table 1. School Sample Size by Treatment Status

	All Schools	Treatment	Control
(1) Number of schools randomized			
(2) Number of schools contributing follow-up data			
Study Retention Rate [(2)/(1)]			

Note: Additional rows should be added to account for multiple follow-up periods.

Office of Adolescent Health

Example Table 2. Youth Response Rates by Treatment Status

	All Students	Treatment	Control	T/C Difference	*p*-value
Number of Youth					
(1) Ineligible randomized sample				NA	NA
(2) who consented				NA	NA
(3) Completed a baseline survey				NA	NA
(4) Completed a follow-up survey				NA	NA
(5) Completed both baseline and follow-up surveys				NA	NA
Response Rates Among Youth					
Baseline Survey					
In study schools at random assignment [(3)/(1)]					
who consented [(3)/(2)]					
who completed a follow-up survey [(5)/(4)]					
Follow-Up Survey					
In study schools at random assignment [(4)/(1)]					
who consented [(4)/(2)]					
who completed a baseline survey [(5)/(3)]					

Note: Additional rows should be added to account for multiple follow-up periods.

Example Table 3. Baseline Characteristics of Youth by Response and Treatment Status, for Youth Who Completed a Baseline Survey

	All Youth Who Completed a Baseline Survey		Youth Not Missing Outcome Data (Analytic Sample)	
Characteristic	**Not Missing Outcome Data**	**Missing Outcome Data**	**Control**	**Treatment**
Female (proportion)				
Age (mean)				
Sexually Active (proportion)				

Note: In final reports, tables that are used to demonstrate whether two groups are statistically different from each other typically include the following table note: *Statistically significant differences in characteristics between treatment and control youth in the analytic sample are indicated by a * on values in the "Treatment" column. Statistical significance is defined as a p-value less than or equal to 0.05.*

With regard to causal inference, we suggest creating tables that describe baseline differences between the treatment and control groups for the analytic sample (that is, the sample of youth with nonmissing follow-up data who are included in impact estimation). The characteristics included in the second set of tables should be identified prior to analysis and selected based on some systematic criteria. For example, it is often advisable to examine baseline covariates that are highly correlated with outcomes of interest (such as baseline measures of the outcomes). The last two columns of Example Table 3 illustrate how this type of information can be presented. Note that one limitation of Table 3 is that it necessarily excludes youth who did not respond to the baseline survey (for example, because they did not consent to the study).

II. Addressing Missing Data in Experimental Impact Analyses

Note: Using the approaches described in this brief to address missing outcome data does *not* affect how a study will be assessed with respect to attrition by the teen pregnancy evidence review under current evidence review standards. That is, attrition is defined as the proportion of randomly assigned units that are missing outcome data prior to any adjustments such as imputation.

Missing data can occur in different contexts and a variety of methods exist for addressing missing data. In this section, we:

A. Describe contexts that applied researchers conducting RCTs commonly face

B. Provide an overview of several common methods for handling missing outcome and baseline data

C. Make specific recommendations for which methods to use in each context

A. Description of Contexts

The contexts described below are all defined in terms of whether outcome data are missing and the type of bias of interest in a study. Whether baseline data are missing does not affect the definition of the contexts—instead, we will make recommendations for how to address missing baseline data in each of these contexts.

Context 1—Little or No Missing Outcome Data

In some studies, descriptive analyses may show that missing data simply are not an issue. Specifically, we may find that the overall missing data rate is very low (for example, so low that the study's statistical power is unaffected by missing data), that the rate is not significantly different between the treatment and control groups, and that the baseline characteristics of the analytic sample do not differ significantly between the treatment and control groups.

Context 2—Outcome Data Are Missing, but Only Causal Validity Bias Matters

This might be the most common context for evaluators of teen pregnancy prevention programs funded by the Office of Adolescent Health (OAH). As described above, the attrition standard used by the teen pregnancy prevention evidence review is only focused on causal validity bias. Furthermore, all RCTs funded by OAH are being conducted using samples of convenience and are not intended to generate impacts that generalize to a larger population.

Context 3—Outcome Data Are Missing, Causal Validity Bias and Generalizability Bias Matter

Generalizability bias matters when impacts are being compared across multiple follow-up periods. In these cases, the proportion of youth missing outcome data—and the characteristics of these youth—can vary over time. One option in this situation is to analyze only the youth with data at every point in time, but that sample of youth could be substantially different from the sample that was randomly assigned (leading to generalizability bias), which could severely reduce the sample size and study power.

For example, at the first follow-up, we may find that 5 percent of the treatment and control groups are missing outcome data, and that those missing outcome data are no different from those with outcome data. But at the second follow-up we may find that 10 percent of the treatment group is missing outcome data and 15 percent of the control group is missing outcome data. Also, we may find that those missing outcome data were more likely to be sexually active at baseline than those not missing outcome data. If we find different impacts at those two points in time, we cannot know if the difference is "real" or if it is only due to the compositional differences in the sample.

B. Description of Methods

Missing Outcome Data

Below we describe five methods for coping with missing outcome data. These methods vary in terms of the type of bias they address, implementation pitfalls, and the ease of implementation using standard statistical software. Table 4 presents the advantages and disadvantages of each method.

1. ***Complete Case Analysis with No Regression Adjustment.*** The most straightforward approach to handling missing outcome data is to drop observations with missing outcomes from the analysis.

This method should only be considered when (1) descriptive analyses show that missing data rates are low in both the treatment and control groups, and (2) descriptive analyses show that the baseline characteristics of the analytic sample do not differ between the treatment and control groups. In that ideal context, the simplicity of this method can outweigh its disadvantages. However, even in the ideal context, other methods may be preferable. For example, complete case analysis with regression adjustment for baseline covariates is likely to yield more precise impact estimates even when no outcome data are missing.

2. ***Complete Case Analysis with Regression Adjustment for Baseline Covariates.*** One approach to account for pre-intervention differences between the treatment and control groups that may arise from attrition is to conduct a complete case analysis with statistical adjustment for pre-intervention differences (for example, through regression or analysis of covariance [ANCOVA]). Examples of baseline covariates may include pre-intervention measures of the outcome and demographic characteristics.

   This is an appealing method in many contexts. It is very simple to implement and has been found to perform as well as more complex methods with respect to reducing causal validity bias in the context of RCTs in which the outcome is an academic test score (Puma et al. 2009).

3. ***Multiple Imputation.*** Multiple imputation (Rubin 1987) involves creating multiple imputed data sets through the repeated application of an imputation algorithm (such as imputation by chained equations; Raghunathan et al. 2001). There are a variety of valid approaches to multiple imputation in the literature and many software packages allow easy implementation of this method (for example, Su et al. 2011). However, the method is more complex than regression adjustment and, if not implemented correctly, has the potential to do more harm than good. Some key implementation requirements are: (1) variables used in the imputation model must include at least all covariates used for statistical adjustment in the impact estimation; (2) in the context of an RCT, imputation should be conducted separately for the treatment and control groups (Puma et al. 2009); and (3) appropriate imputation methods should be used for different data types—in particular, methods to impute continuous variables should generally not be used to impute binary variables.

   This approach is the most flexible of all the approaches considered here. It can be used in nearly any context, combined with any other estimation method, and can address both generalizability and causal validity bias. In contexts with multiple outcomes that are highly correlated but differ in which observations are missing, this method can perform particularly well since nonmissing values on one outcome can inform the imputation of missing values on another outcome. Its primary drawback is that it has more implementation pitfalls than a method such as case deletion with regression adjustment. Also, the computational cost (in terms of the amount of computer processing time and memory needed) can be very high for data with a large number of covariates.

4. ***Maximum Likelihood.*** Many statistical packages use maximum likelihood methods (for example, the EM algorithm) to account for missing data, often in the context of estimating a hierarchical linear model (HLM). This approach is essentially another form of imputation. In general terms, the EM algorithm uses nonmissing data to estimate the distributional characteristics of, and relationships between, variables in a data set. The algorithm then uses those estimated relationships to impute missing values.

   For researchers using HLM (or similar methods) to estimate impacts, this can be a natural choice since software packages implementing HLM often use this approach to "automatically" handle missing data. Other methods are better suited in cases in which HLM is not already being used. Also, note that it is not necessary to use this method when estimating impacts using HLM—all other methods described in this brief can be used instead, if desired.

5. ***Nonresponse Weights.*** With nonresponse weights, individuals with nonmissing outcome data who "look like" individuals with missing outcome data are given greater weight in the impact analysis. Specifically, weights are constructed to be proportional to the inverse of the predicted probability of having nonmissing outcome data. The predicted probabilities are typically constructed either by calculating the rate of nonmissing data within groups of study subjects with similar pre-intervention covariate values, or by estimating a logit or probit model of the probability of having nonmissing outcome data conditional on covariates.

   Constructing these weights is a complex process typically requiring an experienced statistician. If not implemented correctly, this method has the potential to do more harm than good. Some key implementation requirements are: (1) the probabilities of having nonmissing outcome data should be predicted conditional on treatment status (for example, treatment status could be included as a covariate in the logit or probit); (2) diagnostic analyses should be conducted to ensure that the constructed weight has desirable distributional properties (for example, a small number of observations typically should not receive an extremely large

weight); and (3) the design effect for the weight must be properly accounted for when calculating standard errors in impact estimation. Standard statistical packages that permit analysis of complex survey data provide commands that account for the design effect of weighting, but creation of the weights themselves is typically not a built-in feature. See Scheaffer et al. 2005 for more information about nonresponse weights.

This is a valid method that can address both causal validity and generalizability bias. However, we do not recommend it for most researchers because of the implementation complexity and pitfalls. It is best used only by statisticians with experience implementing this approach.

Missing Baseline Data

Below we describe four methods for addressing missing baseline data in an RCT. Table 4 presents the advantages and disadvantages of each method.

1. ***Drop Variable from Analysis.*** The most straightforward approach to handling missing baseline data is to drop baseline variables with missing observations from the analysis.

   We only recommend dropping a baseline variable when it contributes nothing to the statistical precision of impact estimates and when there is no observed difference between the treatment and control groups and/or between respondents and nonrespondents with respect to the baseline variable.

2. ***Missing Value Dummy and Impute to Constant Value.*** Under this approach, every missing value of a baseline variable is imputed to the same value (for example, 0). Also, a missing value dummy variable is included in the impact analysis. The missing value dummy variable equals 1 for every observation where the baseline variable of interest was missing (before imputing) and 0 otherwise. Using this approach, the relationship between the baseline variable and the outcome is estimated using only nonmissing values of the baseline variable.

   This approach is often appealing because it is easy to implement and it is valid when treatment status is uncorrelated with baseline variables (which is not generally true in nonexperimental studies, but is often true in RCTs).

3. ***Multiple Imputation.*** This is the same approach described for missing outcome data.

   When multiple imputations are being used to address missing outcome data, missing baseline data are handled automatically, and there is no reason to use any other approach for addressing missing baseline data.

Summary of a Study that Compared Methods for Coping with Missing Data

Puma et al. (2009) used simulations to compare methods for coping with missing outcome or baseline data in the context of education studies where schools are randomly assigned to treatment and control groups, impacts are calculated on academic test scores, and where a pre-intervention measure of the outcome is available. Their simulations were conducted under three different scenarios: (1) missing data is related to treatment status only; (2) missing data is related to treatment status and a pre-intervention measure of the outcome; and (3) missing data is related to treatment status, a pre-intervention measure of the outcome, and an unobserved variable that is correlated with the outcome. They considered missing baseline data and missing outcome data separately. They assessed the performance (in terms of causal validity bias) of all methods described in this brief except one—they did not assess the performance of "Drop Variable From Analysis" as a means for coping with missing baseline data. This summary focuses on the methods they did assess, for the case where 40 percent of student data was missing (Exhibit 3, Puma et al. 2009).

Under their first scenario, for either missing outcome or missing baseline variables, all assessed methods described in this brief resulted in "low bias" (which they defined as less than 0.05 standard deviations).

Under their second scenario, for either missing outcome or missing baseline variables, all assessed methods described in this brief except "case deletion with no regression adjustment" resulted in "low bias."

Under their third scenario, for missing baseline variables, all assessed methods described in this brief resulted in "low bias." For missing outcome variables, all assessed methods described in this brief resulted in "high bias."

Bottom line for missing baseline data: The methods "Missing Value Dummy and Impute to Constant Value," "Multiple Imputation," and "Maximum Likelihood" all resulted in "low bias" regardless of scenario. This is because random assignment ensures balance with respect to baseline covariates, which means that any strategy for coping with missing baseline data in an RCT need only "do no harm."

Bottom line for missing outcome data: The methods "Complete Case Analysis with Regression Adjustment for Baseline Covariates," "Multiple Imputation," "Maximum Likelihood," and "Nonresponse Weights" all had the same results—either "low bias" or "high bias," depending on the scenario. This is because with respect to causal validity bias (which was the focus of Puma et al. 2009), all of these methods are able to adjust for *observed* differences between the treatment and control groups, but none of them can adjust for *unobserved* differences between the treatment and control groups.

Table 4. Advantages and Disadvantages of Methods for Handling Missing Data

Method for Handling Missing Data	Advantage	Disadvantage
Outcome Data		
Complete Case Analysis with No Regression Adjustment	Simplest to implement.	Does not provide adjustment for known differences between the treatment and control groups in the characteristics of students in the analytic sample. Does not adjust for differences between respondents and nonrespondents in general. Provides no protection against either type of bias arising from unobserved differences between the treatment and control groups or between respondents and nonrespondents.
Complete Case Analysis with Regression Adjustment for Baseline Covariates	Second simplest to implement. Reduces causal validity bias arising from observed differences in pre-intervention characteristics between the treatment and control groups.	Does not provide protection against generalizability bias. Does not make use of information available from nonmissing values of other outcome variables. Does not reduce either kind of bias arising from unobserved differences between the treatment and control groups.
Multiple Imputation	Convenient to implement in some software packages. Reduces causal validity bias arising from observed differences between the treatment and control groups. Reduces generalizability bias arising from observed differences between the respondents and nonrespondents. Makes use of information available from nonmissing values of other outcome variables.	Though some software packages have built-in routines for implementing multiple imputation, these routines are often a "black box" that is difficult for researchers and readers to understand and explain. Difficult to implement in software packages that lack built-in imputation commands (we suggest consulting the manual for your software). Does not reduce either type of bias arising from unobserved differences between the treatment and control groups or between respondents and nonrespondents.
Maximum Likelihood	Convenient to implement in some software packages. Reduces causal validity bias arising from observed differences between the treatment and control groups. Reduces generalizability bias arising from observed differences between the respondents and nonrespondents.	Though some software packages have built-in routines for implementing maximum likelihood, these routines are often a "black box" that is difficult for researchers and readers to understand and explain. Difficult to implement in software packages that lack built-in routines for implementing the method. Does not make use of information available from nonmissing values of other outcome variables. Does not reduce either type of bias arising from unobserved differences between the treatment and control groups or between respondents and nonrespondents.
Nonresponse Weights	Reduces causal validity bias arising from observed differences between the treatment and control groups. Reduces generalizability bias arising from observed differences between the respondents and nonrespondents.	Difficult to implement with many implementation pitfalls. Cumbersome to use in situations with varying missing rates across multiple outcomes and follow-up periods. Does not make use of information available from nonmissing values of other outcome variables. Does not reduce either type of bias arising from unobserved differences between the treatment and control groups or between respondents and nonrespondents.

Method for Handling Missing Data	Advantage	Disadvantage
Baseline Data		
Drop Variable from Analysis	Simplest to implement.	Impact estimates may be less precise. When outcome data are missing, baseline variables may be needed to adjust for observed differences between the treatment and control groups and/or between respondents and nonrespondents. In that situation, dropping baseline variables could exacerbate bias.[a]
Missing Value Dummy and Impute to Constant Value	Simple to implement. Retains precision advantage of covariate adjustment.	In cases in which missing outcome data lead to a correlation between a baseline variable and treatment status, multiple imputation may lead to less bias than this approach.
Multiple Imputation	Easy to implement when multiple imputation is being used to address missing outcome data.	Though some software packages have built-in routines for implementing multiple imputation, these routines are often a "black box" that is difficult for researchers and readers to understand and explain. Difficult to implement in software packages that lack built-in imputation commands (consult the technical documentation for your preferred statistical software). Does not reduce either type of bias arising from unobserved differences between the treatment and control groups or between respondents and nonrespondents.
Maximum Likelihood	Easy to implement when maximum likelihood is being used to address missing outcome data.	Though some software packages have built-in routines for implementing maximum likelihood, these routines are often a "black box" that is difficult for researchers and readers to understand and explain. Difficult to implement in software packages that lack built-in routines for implementing the method. Does not make use of information available from nonmissing values of other outcome variables. Does not reduce either type of bias arising from unobserved differences between the treatment and control groups or between respondents and nonrespondents.

[a] Puma et al. (2009) found that adjusting for baseline variables that are highly correlated with the outcome can dramatically reduce bias.

4. ***Maximum Likelihood.*** This is the same approach described for missing outcome data.

   When maximum likelihood methods are being used to address missing outcome data, missing baseline data are handled automatically, and there is no reason to use any other approach for addressing missing baseline data.

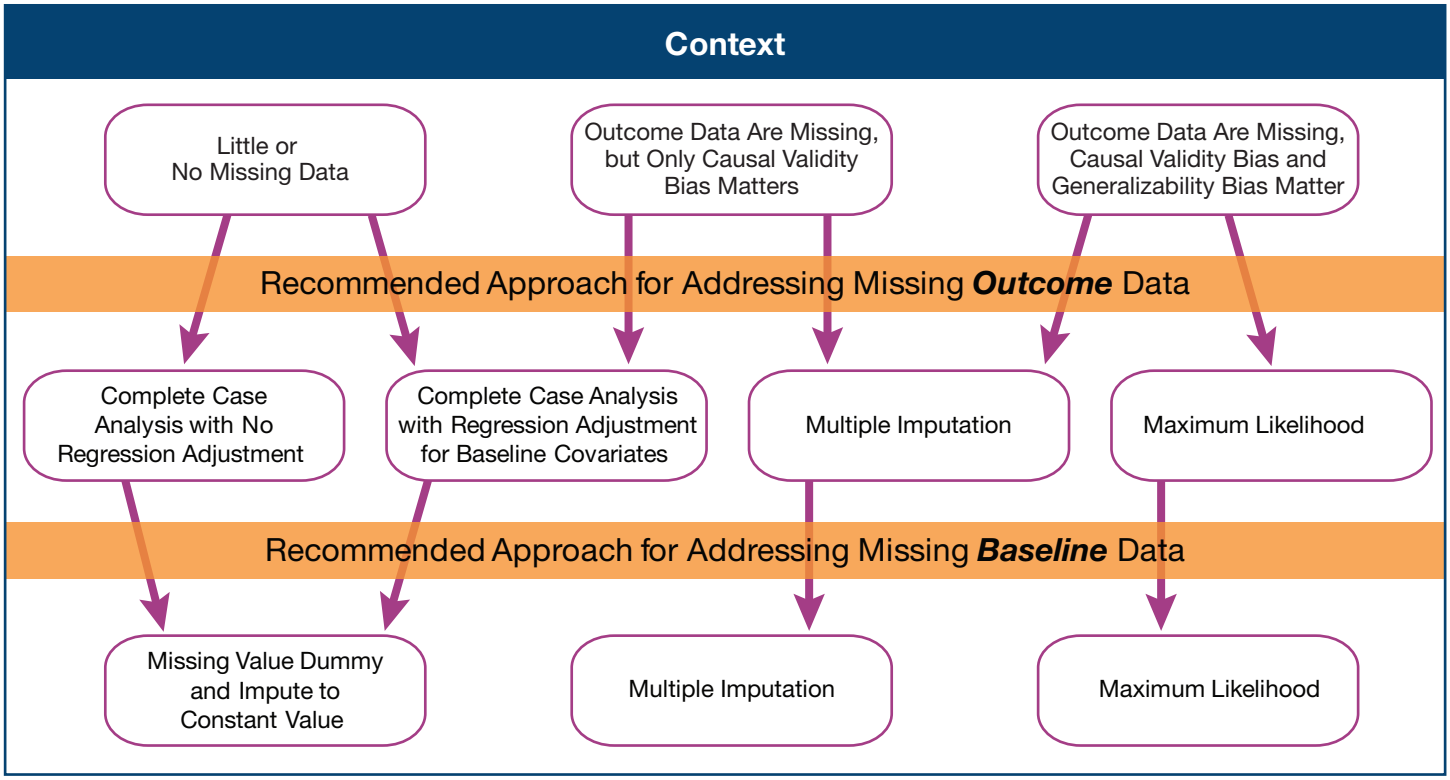
C. Recommendations

No single method described above is clearly preferable to all the others in every possible context. Furthermore, every method described above is valid under the right conditions. We cannot say that any method should never be used, nor can we say that any method should always be used.

However, we can recommend the use of certain methods in particular contexts that applied researchers conducting RCTs commonly face. The choice of methods for addressing missing baseline data should take into account the method used to address missing outcome data. Once a method has been selected for addressing missing outcome data, a compatible method can be selected to address missing baseline data.

Figure 1 presents the contexts and the recommended method for handling missing outcome and baseline data.

Figure 1. Contexts and Recommended Methods for Handling Missing Outcome and Baseline Data



Context 1—Little or No Missing Data

In this context, complete case analysis (with or without regression adjustment) is the recommended approach for handling missing outcome data. When using complete case analysis to address missing outcome data, we recommend addressing missing baseline data using a missing value dummy and imputing missing values to a constant value so long as the rates of missing outcome data in the treatment and control groups are low enough to pass the attrition standard used by the evidence review.[4]

Context 2—Outcome Data Are Missing, but Only Causal Validity Bias Matters

In this context we recommend either complete case analysis with regression adjustment or multiple imputation. If multiple highly correlated outcomes are being analyzed, multiple imputation may be the preferred approach if various youth are missing different outcomes so that one outcome can inform the imputation of another outcome.[5] Otherwise, complete case analysis with regression adjustment is preferred because it is easy to implement and—like more complex methods—can reduce causal validity bias (Puma et al. 2009).[6]

When using multiple imputation to address missing outcome data, we recommend also using multiple imputation to address missing baseline data (note that in statistical packages that support multiple imputations, this will happen "automatically").

Similarly to Context 1, if using complete case analysis with regression adjustment to address missing outcome data, we recommend addressing missing baseline data using a missing value dummy and imputing missing values to a constant value.

Context 3—Outcome Data Are Missing, Causal Validity Bias and Generalizability Bias Matter

We recommend using multiple imputation or maximum likelihood to address missing data in this context. These methods will adjust for observed differences between youth in the treatment and control groups, and for observed differences between respondents and nonrespondents at different points in time. These adjustments increase the chance that differences in impacts across time are "real" rather than a result of compositional shifts. We note, however, that multiple imputation only adjusts for observed characteristics—if the unobserved characteristics of youth differ between the treatment and control groups and/or between respondents and nonrespondents at different points in time, then variation in impacts across time could be due to unobserved compositional differences.

When using multiple imputation or maximum likelihood to address missing outcome data, we recommend using the same method to address missing baseline data (note that in statistical packages that support multiple imputation or maximum likelihood, this will happen "automatically").

III. CONCLUSION

Missing data—particularly missing outcome data—can pose a serious threat to the validity of experimental impacts. The best solution to missing data is to avoid the problem in the first place by making every reasonable effort to collect data—particularly outcome data—for everyone in the study sample. Yet despite all best efforts, some missing data may be inevitable. When that happens, the strategies described in this brief can help contain the damage from missing data, first by clearly describing the problem and second by using valid statistical methods to adjust for it. It is important to note, however, that while the use of these statistical adjustments may improve statistical power and the face validity of a study's findings, it will not improve a study's rating with respect to the attrition standard under the teen pregnancy prevention evidence review's current evidence standards.

Endnotes

[1] This brief does not examine missing data in the context of quasi-experimental designs (QEDs). (By QEDs, we mean designs in which impacts are calculated by comparing the outcomes of participants to a group of matched nonparticipants; these designs do not involve randomization.) QEDs are based on the stronger assumption that baseline equivalence of treatment and comparison sample members *in the analytic sample* (that is, the sample actually used to calculate the impact) is sufficient to establish the causal validity of the impact estimates. Therefore, observations that are missing outcomes or key baseline variables can simply be dropped from the analysis prior to establishing baseline equivalence for the analytic sample.

[2] This type of bias was the focus of a recent *American Journal of Public Health* "Statistically Speaking" column (Harel and Boyko 2013) and the subject of the study referenced by that column (Harel et al. 2012).

[3] That is, the impact we expect to observe for respondents really is the impact for respondents.

[4] With high rates of missing outcome data treatment status may become correlated with baseline variables in the analytic sample, undermining the validity of the missing value dummy approach for missing values of baseline variables. Evaluators who encounter this scenario should contact their TA liaison for further guidance.

[5] For example, the outcomes "ever had sex" and "ever had unprotected sex" are, by construction, highly correlated and can have different patterns of missing values when item nonresponse, rather than survey nonresponse, is the main cause of missing outcome data. Note that if outcome data are missing primarily due to survey nonresponse (meaning that an individual is missing *all* outcomes), then this advantage of imputation is lost.

[6] Puma et al. (2009) compared methods for addressing missing data in the context of clustered RCTs where schools are the unit of assignment and the outcomes of interest are test scores.

References

Campbell, M.K., G. Piaggio, D.R. Elbourne, and D.G. Altman. "Consort 2010 Statement: Extension to Cluster Randomised Trials." *BMJ*, vol. 345, 2012, p. e5661.

Harel, Ofer, and Jennifer Boyko. "Mi??ing Data: Should We C?re?" *American Journal of Public Health*, vol. 103, no. 2, February 2013.

Harel, Ofer, Jennifer Pellowski, and Seth Kalichman. "Are We Missing the Importance of Missing Values in HIV Prevention Randomized Clinical Trials?" *AIDS and Behavior*, vol. 16, no. 6, 2012, pp. 1382–1393.

Mathematica Policy Research and Child Trends. "Identifying Programs That Impact Teen Pregnancy, Sexually Transmitted Infections, and Associated Sexual Risk Behaviors. Review Protocol. Version 2.0." Princeton, NJ: Mathematica Policy Research, 2012. Available at: [http://www.hhs.gov/ash/oah/oah-initiatives/tpp/eb-programs-review-v2.pdf].

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. *What to Do When Data Are Missing in Group Randomized Controlled Trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009.

Raghunathan, T.E., J.M. Lepkowski, J. Van Hoewyk, and P. Solenberger. "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models." *Survey Methodology*, vol. 27, no. 1, 2001, pp. 85–95.

Rubin, D.B. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley, 1987.

Scheaffer, R.L., W. Mendenhall III, and R. L. Ott. *Elementary Survey Sampling*. 6th ed. Boston: Duxbury, 2005.

Su, Yu-Sung, A. Gelman, J. Hill, and M. Yajimsa. "Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box." *Journal of Statistical Software*, vol. 45, no 2, December 2011.

This brief was written by John Deke from Mathematica Policy Research and Mike Puma from Chesapeake Research Associates for the HHS Office of Adolescent Health under contract #HHSP23320095642WC, Task Order No. HHSP23337017T.