Chapter 7
Missing Data in Meta-analysis: Strategies and Approaches

Abstract This chapter provides an overview of missing data issues that can occur in a meta-analysis. Common approaches to missing data in meta-analysis are discussed. The chapter focuses on the problem of missing data in moderators of effect size. The examples demonstrate the use of maximum likelihood methods and multiple imputation, the only two methods that produce unbiased estimates under the assumption that data are missing at random. The methods discussed in this chapter are most useful in testing the sensitivity of results to missing data.

7.1 Background

All data analysts face the problem of missing data. Survey researchers often find respondents may refuse to answer a question, or may skip an item on a questionnaire. Experimental studies are also subject to drop-outs in both the treatment and control group. In meta-analysis, there are three major sources of missing data: missing studies from the review, missing effect sizes or outcomes for the analysis, and missing predictors for models of effect size variation. This chapter will provide strategies for testing the sensitivity of results to problems with missing data. As discussed throughout this chapter, current methods for missing data require strong assumptions about the reasons why data are missing, and about the distribution of the hypothetically complete data that cannot be verified empirically. Instead, reanalyzing the data under a number of differing assumptions provides the reviewer with evidence of the robustness of the results.

Over the past 20 years, statisticians have conducted an extensive amount of research into methods for dealing with missing data. Schafer and Graham (2002) point out that the main goal of statistical methods for missing data is not to recover or estimate the missing values but to make valid inferences about a population of interest. Thus, Schafer and Graham note that appropriate missing data techniques are embedded in the particular model or testing procedure used in the analysis. This chapter will take Schafer and Graham's perspective and provide missing data

T.D. Pigott, *Advances in Meta-Analysis*, Statistics for Social and Behavioral Sciences,
DOI 10.1007/978-1-4614-2278-5_7,

methods adapted from the statistical literature (Little and Rubin 1987; Schafer 1997) for use in meta-analysis. This chapter will focus on the sensitivity of results to missing data rather than on providing an alternative set of estimates that compensate for the missing data. For many missing data methods, the strategy involves recognizing the greater amount of uncertainty in the data caused by the missing information. Thus, many missing data methods result in a larger variance around the model estimates. This chapter will focus on methods that formally incorporate a larger amount of variance when missing data occurs.

The two most common strategies suggested for missing effect sizes in meta-analysis do not take into account the true level of uncertainty caused by missing data. These two strategies involve filling in either the observed mean using studies that report that missing variable, and filling in a zero for studies missing an effect size. Filling in the same value for missing observations in any data set will reduce the variance in the resulting estimate, making the estimates seem to contain more information than is truly available. Later in the chapter, we will discuss imputation strategies that incorporate a larger degree of uncertainty in the estimates reflecting the missing information in the data, and we will use these estimates to judge the sensitivity of results to assumptions about missing data.

7.2 Missing Studies in a Meta-analysis

One common form of missing data in a meta-analysis is missing studies. The most common cause of missing studies is publication bias. As many researchers have shown (Begg and Berlin 1988; Hemminki 1980; Rosenthal 1979; Smith 1980), there is a bias in the published literature toward statistically significant results. If a search strategy for a meta-analysis focuses only on published studies, then there is a tendency across many disciplines for the overall effect size to be biased toward statistically significant effects, thus over-estimating the true difference between the treatment and control group or the strength of the association between two measures. One strategy for addressing publication bias is the use of thorough search strategies that focus on published, unpublished and fugitive literatures. This section will provide an overview of strategies detecting and examining the potential for publication bias; more detailed information can be found in Rothstein et al. (2005).

7.2.1 Identification of Publication Bias

Even when a search strategy aims for a wide range of published and unpublished studies, the resulting sample of studies may still suffer from publication bias. One set of strategies associated with missing studies focuses on the identification of publication bias. The simplest and most widely known of these strategies is the funnel plot (Sterne et al. 2005). Funnel plots are scatterplots of the effect size on the

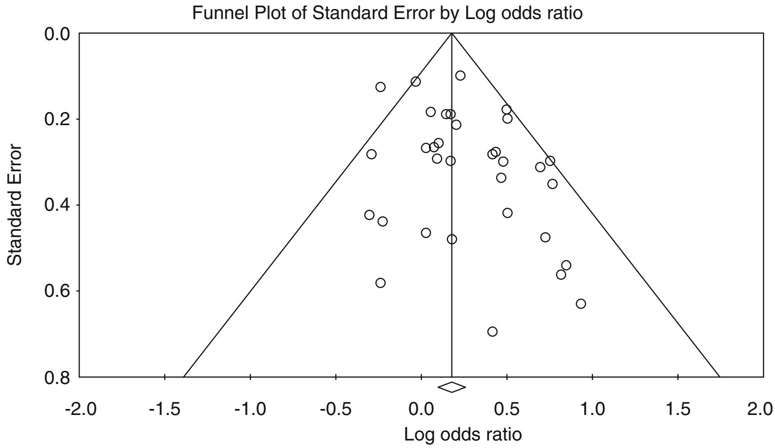

Fig. 7.1 Funnel plot for passive smoking data

x-axis and the sample size, variance, or study level weight of the effect size on the y-axis. (Recall that the study level weight is the inverse of the variance of the effect size). With no publication bias, the plot should resemble a funnel with the wider end of the funnel associated with studies of small sample sizes and large variances. The smaller end of the funnel should have effect sizes that have larger sample sizes and smaller variances, centered around the mean of the effect size distribution. If publication bias exists, then the plot will appear asymmetric. If small studies with small sample sizes are missing, then the funnel plot will appear to have a piece missing at its widest point. The quadrant with small effect sizes and small sample sizes would be those most likely to be censored in the published literature. If small effect sizes are, in general, more unlikely to appear in the literature, then the funnel will have fewer studies in the area of the graph corresponding to effect sizes close to zero, despite the sample size.

7.2.1.1 Example of Funnel Plot

Figure 7.1 is a funnel plot of data taken from Hackshaw et al. (1997) study of the relationship between passive smoking and lung cancer in women. The 33 studies in the meta-analysis compare the number of cases of lung cancer diagnosed in individuals whose spouses smoke with the number of cases of lung cancer in individuals whose spouses are non-smokers. The data used to construct this plot are given in the Data Appendix. The x-axis is the log-odds ratio, and the y-axis is the standard error of the log-odds ratio. There is a gap in the lower left-hand corner of the funnel plot, indicating that some studies with large standard errors and negative log-odds ratios could be missing. Thus, we see some evidence of publication bias here (Fig. 7.1).

While funnel plots are easily constructed, they can be difficult to interpret. Many conditions can lead to asymmetric plots even when the sample of studies is not affected by publication bias. A more formal test of publication bias was proposed by Egger et al. (1997) using regression techniques. Egger et al.'s method provides a test of whether a funnel plot shows evidence of asymmetry. The method involves standardizing the effect size into a standard normal deviate and regressing this transformed effect size on the precision of the effect size, defined as the inverse of the standard error of the effect size. The regression equation can be expressed as

$$\frac{T_i}{\sqrt{v_i}} = \hat{\beta}_0 + \hat{\beta}_1 \frac{1}{\sqrt{v_i}} \tag{7.1}$$

where T_i is the effect size for study i, and v_i is the standard error for the effect size in study i. When the funnel plot is symmetric, that is, when there is no evidence of publication bias, then $\hat{\beta}_0$ is close to zero. Symmetric funnel plots will produce an estimated regression equation that goes through the origin. Standardizing small effect sizes using the standard error should create a small standard normal deviate. In contrast, larger studies will produce larger standard normal deviates since their standard errors will be small. When publication bias is present, we may have large studies with normal deviates that are smaller than studies with small sample sizes – indicating that the small studies differ from large studies in their estimates of effect size.

Figure 7.2 provides the scatterplot of the standardized effect size by the inverse of the standard error of the effect size in the passive smoking data. The dotted line in the graph is the regression line given in (7.1). Table 7.1 provides the regression coefficients for (7.1) fit to the passive smoking data. The value for the intercept, β_0, is statistically different from zero, thus indicating that there is evidence of publication bias in the passive smoking data.

7.2.2 *Assessing the Sensitivity of Results to Publication Bias*

If a reviewer suspects publication bias in the meta-analytic data, there are two general classes of methods for exploring the sensitivity of results to publication bias. The first method, trim-and-fill (Duval and Tweedie 2000) is fairly easy to implement, but relies on strong assumptions about the nature of the missing studies. As Vevea and Woods (2005) point out, the trim-and-fill method assumes that the missing studies are one-to-one reflections of the largest effect sizes in the data set, in essence, that the missing studies have effect sizes values that are the negative of the largest effect sizes observed in the data set. In addition, the trim-and-fill method may lead to biased results if the effect sizes are in fact heterogeneous. Vevea and Woods present a second method that produces estimates for models of effect size under a series of possible censoring mechanisms, addressing the shortcomings they

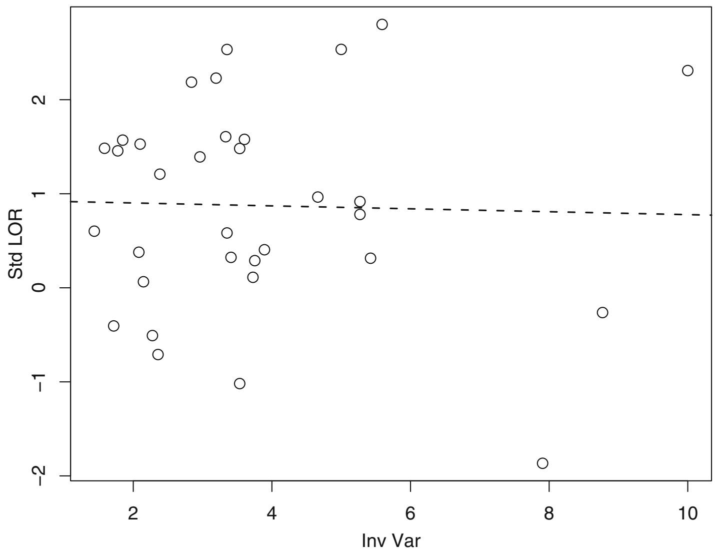

Fig. 7.2 Egger's test for publication bias for passive smoking data

Table 7.1 Egger's test for the passive smoking data set

Coefficient	Estimate	SE	t-Value	p
β_0	0.933	0.417	2.236	0.033
β_1	−0.0155	0.098	−0.158	0.875

find with the trim-and-fill method This second method provides more flexibility than trim-and-fill since it allows the examination of sensitivity for a range of models of effect size.

The trim-and-fill method (Duval and Tweedie 2000) is based on asymmetry in the funnel plot. This method builds on the funnel plot by "filling in" effect sizes that are missing from the funnel and then estimating the overall mean effect size including these hypothetical values. The assumption in this method is that there are a set of missing effect sizes that are "mirror images" of the largest effect sizes in the data set. In other words, the method inserts missing effect sizes that are of the opposite sign, and are mirror reflections of the largest observed effect sizes with similar sample sizes. The theoretical basis for the method is beyond the scope of this text, but the analysis itself is fairly simple. The idea is to first estimate how many effect sizes are "missing" in order to create a symmetric funnel plot. This computation may require a few iterations to obtain the estimate. Once the researcher computes the number of missing effect sizes, hypothetical values for these missing observations are used to re-estimate the mean effect size. This new mean effect size incorporates the possible effects of publication bias.

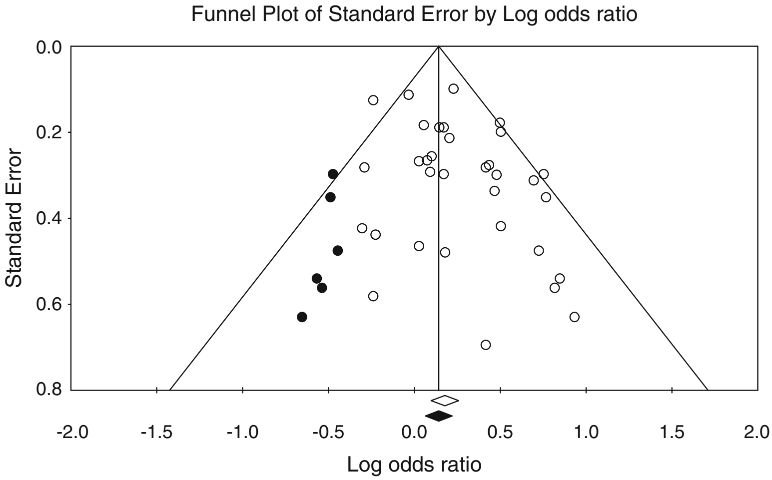

Fig. 7.3 Funnel plot for passive smoking data with Trim and Fill results

In practice, knowing how much bias is possible allows a test of the sensitivity of results. The reviewer should decide if the difference in these two values is of substantive importance, and should report these values to readers. While space does not permit a full example illustrating trim-and-fill, Duval (2005) provides a step-by-step outline for completing the method. Figure 7.3 is a funnel plot of the smoking data with the "missing" effect sizes represented by solid circles. The bottom of the plot shows the mean effect size computed with only the observed studies, and the mean effect size when the "missing" studies are included. As seen in Fig. 7.3, the mean effect size does not change significantly if we assume some publication bias.

A second strategy involves modeling publication bias using a censoring mechanism as described by Vevea and Woods (2005). The model illustrated by Vevea and Woods proposes a number of censoring mechanisms that could be operating in the literature. These censoring mechanisms are based on the type of censoring and the severity of the problem. In general, Vevea and Woods examine the impact of censoring where the smallest studies with the smallest effect sizes are missing (one-tailed censoring) and where the studies with non-significant effect sizes (two-tailed censoring) are missing. Vevea and Woods point out that the trim-and-fill method assumes that the missing studies are one-to-one reflections of statistically significant effect sizes, and that the method can only examine the sensitivity of estimates of the mean effect size. Vevea and Woods' method can examine the sensitivity of fixed, random and mixed effects models of effect size to publication bias. These methods do require the use of more flexible computing environments, and reviewers may find them more difficult. Readers interested in methods for publication bias will find more details in Rothstein et al. (2005).

7.3 Missing Effect Sizes in a Meta-analysis

When effect sizes are missing from a study, there are few missing data strategies to analyze the data. The most common method used by reviewers is to drop these studies from the analysis. The problem is similar to missing data in primary studies. If an individual patient does not have a measure for the target outcome, then that patient cannot provide any information about the efficacy of the treatment.

One reason for missing effect sizes is that reviewers either cannot compute an effect size from the information given in a study or do not know how to compute an effect size from the results of the study. For example, studies may fail to report summary statistics needed to compute an effect size such as means and standard deviations for standardized mean differences, or frequencies for the odds ratio. Other studies may report only the summary test statistic such as a t-test or an F-test, or only the p-value for the test. These difficulties often occur with older studies since professional organizations such as the American Psychological Association and the American Medical Association now have standards for reporting that assist reviewers in extracting information from a meta-analysis.

Another reason for missing effect sizes in more recently published studies arises when a reviewer does not know how to translate complex statistical results derived from techniques such as factorial ANOVA, regression models, or hierarchical linear models into an effect size for the review. A related problem occurs when the studies in the review report a wide variety of statistics. For example, one study in the review may report a t-test from a quasi-experimental study, while another study may report a correlation coefficient from an observational study. The question is whether these differing measures of effect size can and should be combined in the meta-analysis, or whether the reviewer should consider certain types of effect sizes as missing. Lipsey and Wilson (2001) and Shadish et al. (1999) provide a number of tools for computing effect sizes from commonly provided information in a study. Wilson (2010) provides a free, web-based, effect size calculator that obtains effect sizes from a large array of reported statistics. In practice, the reviewer should try multiple methods for computing an effect size from the data given in the study, and to contact primary authors for available information. If effect sizes are still not available, then the reviewer should explore the sensitivity of results to publication bias. One suggestion often used by reviewers is to impute a value of 0 for all missing effect sizes. While this method seemingly provides a conservative estimate of the mean effect size, the standard errors for the mean effect size will be underestimated. As will be discussed in the next sections, imputing a single value for any missing observation will not reflect accurately the uncertainty in the data.

A second reason for missing outcomes in a study is selective reporting. A number of researchers in medicine have documented primary studies where researchers have not reported on an outcome that was gathered. In many cases, these outcomes are also ones that are not statistically significant, or may be outcomes reporting on an adverse outcome. In the Cochrane Collaboration, reviewers are required to report on whether selective reporting of outcomes has occurred in a study in the Risk of

Bias table. Reviewers who suspect outcome reporting bias have few strategies for dealing with this problem aside from contacting the primary authors for the missing information. Researchers (Chan et al. 2004; Williamson et al. 2005) have also developed methods for assessing the potential bias in results when some outcomes are selectively missing. In the long term, reporting standards and registries of primary studies may be the most effective strategies for ensuring complete reporting of all outcomes gathered in a study.

7.4 Missing Moderators in Effect Size Models

Another form of missing data in a meta-analysis is missing moderators or predictors in effect size models. This form of missing data occurs when the reviewer wants to code particular aspects of the primary study, but this information is not provided in the study report. A reviewer examining a body of research on a topic has, in one sense, fewer constraints than a primary researcher. When a primary researcher plans a study, they must make decisions about the number and type of measures used, the optimal design and methods, and in the final report, about what information is most relevant. In contrast, a reviewer can examine questions about how the different decisions made by the primary researcher about study design, measures, and reporting relates to variation in study results.

Disciplinary practices in a given area of research may constrain how primary authors collect, measure and report information in a study. Orwin and Cordray (1985) use the term macrolevel reporting to refer to practices in a given research area that influence how constructs are defined and reported. In a recent meta-analysis, Sirin (2005) found multiple measures of socioeconomic status in studies examining the relationship between income and academic achievement. These measures included ratings of parents' occupational status, parental income, parental education, free lunch eligibility, as well as composites of these measures. Researchers in some fields may be more inclined to use ratings of parents' occupation status whereas other educational researchers may only have a measure of free lunch eligibility available. Thus, parents' occupation status may be missing in some studies since the researchers in one field may rely on free lunch eligibility, for example, as the primary measure of socioeconomic status. Differences in reporting among primary studies could be related to disciplinary practices in the primary author's field. Primary authors may also be constrained by a particular journal's publication practices, and thus do not report on information a reviewer may consider important.

Researchers also differ in their writing styles and thoroughness of reporting. Orwin and Cordary (1985) use the term microlevel reporting quality to refer to individual differences among researchers in their reporting practices. In some cases, a moderator that is not reported among all studies in a review could be missing in a random way due to the individual differences among researchers.

Another reason study descriptors may be missing from primary study reports relates to bias for reporting only statistically significant results. For example, a primary researcher may be interested in the relationship between the percentage of low-income children in a classroom and classroom achievement. If the primary researcher finds that the percentage of low-income children in a classroom is not related to classroom achievement, then they may report only that the statistic did not reach statistical significance. Williamson et al. (2005) provide a discussion of this problem raising the possibility that the likelihood of a study reporting a given descriptor variable is related to the statistical significance of the relationship between the variable and the outcome. In this case, we have selective predictor reporting that operates in a similar manner to selective outcome reporting as described above.

Researchers may also avoid reporting study descriptors when those values are not generally acceptable. For example, if a researcher obtains a racially homogeneous sample, the researcher may hesitate to report fully the actual distribution of ethnicity in the sample. This type of selective reporting may also occur when reporting in a study is influenced by a desire to "clean up" the results. For example, attrition information may not be reported in a published study in order to present a more positive picture. In both of these cases, the values of the study descriptor influence whether the researcher reports those values.

Missing data in a meta-analysis occur in the form of missing studies from the review, missing effect sizes and missing predictors. The cases of missing studies and missing effect sizes correspond to missing an outcome in a primary study. The only approach to missing outcomes is checking the sensitivity of results to publication bias or censoring. The major class of missing data methods in the statistical literature applies most directly to missing data on predictors in models of effect size. The remainder of this chapter discusses the assumptions, techniques and interpretations of missing data methods applied to meta-analysis with an emphasis on handling missing predictors in effect size models.

7.5 Theoretical Basis for Missing Data Methods

The general approach used in current missing data methods involves using the data at hand to draw valid conclusions, and not to recover all the missing information to create a complete data set. This approach is especially applicable to meta-analysis since missing data frequently occur because a variable was not measured, and is not recoverable. Thus, the approaches taken in this chapter do not attempt to replace individual missing observations, but instead either estimate the value for summary statistics in the presence of missing data or sample several possible values for the missing observations from a hypothetical distribution based on the data we do observe.

The methods used in this chapter make strong assumptions about the distribution of the data, and about the mechanism that causes the missing observations.

Generally, the methods here require the assumption that the joint distribution of the effect size and moderator variables is multivariate normal. A second assumption is that the reasons for the missing data do not depend on the values of the missing observations. For example, if our data are missing measures of income for studies with a large proportion of affluent participants, then the methods we discuss here could lead to biased estimates. One major difficulty in applying missing data methods is that assumptions about the nature of the missing data mechanism cannot be tested empirically. These assumptions can only be subjected to the "is it possible" test, i.e., is it possible that the reasons for missing observations on a particular variable do not depend directly on the values of that variable? Missing observations on income usually fail the test, since it is a well-known result in survey sampling that respondents with higher incomes tend not to report their earnings. The following section examines these assumptions in the context of meta-analysis.

7.5.1 Multivariate Normality in Meta-analysis

The missing data methods used in this chapter rely on the assumption that the joint distribution of the data is multivariate normal. Thus, meta-analysts must assume that the joint distribution of the effect sizes and the variables coded from the studies in the review follow a normal distribution. One problematic issue in meta-analysis concerns the common incidence of categorical moderators in effect size models. Codes for characteristics of studies often take on values that indicate whether a primary author used a particular method (e.g., random assignment or not) or a certain assessment for the outcome (e.g., standardized protocol or test, researcher developed rating scale, etc.). Schafer (1997) indicates that in the case of categorical predictors, the normal model will still prove useful if the categorical variables are completely observed, and the variables with missing observations can be assumed multivariate normal conditional on the variables with complete data. The example later in the chapter examines a meta-analysis on gender differences in transformational leadership. Say we have missing data from some studies on the percent of men in the sample of participants who are surveyed in each study. We can still fulfill the multivariate normality condition if we can assume that the variable with missing observations, the percent of males in the sample, is normally distributed conditional on a fully observed categorical variable such as whether or not the first author is a woman. If the histogram of the percent of male subjects in the sample is normally distributed for the set of studies with male first authors and for the set of studies with female first authors, then we have not violated the normality assumption. Some ordered categorical predictors can also be transformed to allow the normal assumption to apply. If key moderators of interest are non-ordered categorical variables, and these variables are missing observations, then missing data methods based on the multinomial model may apply. There is currently no research on how to handle missing categorical predictors in meta-analysis.

7.5.2 Missing Data Mechanisms or Reasons for Missing Data

In addition to assuming that the joint and/or conditional distribution of the data is multivariate normal, the methods discussed in this chapter also require the assumption that the missing data mechanism is ignorable. There are two conditions that meet the conditions of ignorability, missing completely at random (MCAR) data, and missing at random (MAR) data (Rubin 1976). Missing data are missing completely at random when the cases with completely observed data are a random sample of the original data. When there are small amounts of missing data on an important variable, often analysts assume that the completely observed cases are as representative of the target population as the original sample. Thus, when data are missing completely at random, the data analyst does not need to know the exact reasons or mechanisms that caused the missing data; analyzing only the cases with complete observations will yield results that provide unbiased estimates of population parameters.

As applied to meta-analysis, individual differences among primary authors in their reporting practices may result in missing predictors that could be considered missing completely at random. The difficulty lies in gathering evidence that the missing predictors are missing completely at random. One strategy suggested is using logistic regression models to examine the relationships between whether a given predictor is observed or not and values of other variables in the data set. The difficulty arises when these models do not adequately explain the probability of missing a predictor. The relationships between observed variables and missing variables could be more complex than represented in the logistic regression model, or the probability of observing a value could be dependent on other unknown information. Schafer (1997) suggests that a more practical solution is to use as much information in the data set to estimate models in the presence of missing data, a point that will be elaborated later in the chapter.

A second condition that meets the conditions of ignorability is data missing at random. Unlike MCAR data, the cases with completely observed data are not a random sample of the original data. When data are MAR, the probability of missing an observation depends on the values of completely observed variables. Data are MAR if the conditional distribution of the missing values depends only on completely observed variables and not on variables with missing observations. This assumption is less stringent than MCAR, and is plausible in many missing data situations in meta-analysis. For example, some studies may report the income level of subjects as a function of the percent of students who qualify for free lunch, while others report income level as the average income reported by parents. The differences between these studies could be due to the discipline of the primary author – studies in education tend to use the percent of students with low income in a school while larger-scale studies may have the resources to conduct a survey of parents to obtain a more direct measure of income. A missing value for a particular measure of income in a particular study is not necessarily related to the value of income itself but to the choices of the primary author and constraints on the published version of the study. Thus, if we can posit a plausible mechanism for the missing observations, say the disciplinary background of the primary author, and this variable is completely observed in the data, then we can consider income in this instance MAR.

One set of methods that cannot be addressed fully in this chapter are methods for nonignorable missing data. Nonignorable missing data occur when the reason for a missing observation is the value of that observation. For example, self-reports of income are more frequently missing for those respondents with high income, a classic example of nonignorable missing data. In the case of nonignorable missing data, the analysis must incorporate a model for the missing data mechanism instead of ignoring it as in the case of MCAR and MAR data. A special case of nonignorable missing data is publication bias. As discussed earlier in the chapter, these methods usually require specialized computing environments, and provide information about the sensitivity of results to assumptions about the missing data.

7.6 Commonly Used Methods for Missing Data in Meta-analysis

Prior to Little and Rubin's (1987) work, most researchers employed one of three strategies to handle missing data: using only cases with all variables completely observed (listwise deletion), using available cases that have particular pairs of variables observed (pairwise deletion), or replacing missing values on a given variable with a single value such as the mean for the complete cases (single value imputation). The performance of these methods depends on the validity of the assumptions about why the data are missing. In general, these methods will produce unbiased estimates only when data are missing completely at random. The main problem with the use of these methods is that the standard errors do not accurately reflect the fact that variables have missing observations.

7.6.1 Complete-Case Analysis

In complete-case analysis, the researcher uses only those cases with all variables fully observed. This procedure, listwise deletion, is usually the default procedure for many statistical computer packages. When the missing data are missing completely at random, then complete-case analysis will produce unbiased results since the complete cases can be considered a random sample from the originally identified set of cases. Thus, if a synthesist can make the assumption that values are missing completely at random, using only complete cases will produce unbiased results.

Table 7.2 Complete case analysis using the gender and leadership data

Variable	Coefficient	SE	Z	p
Intercept	58.033	22.650	2.562	0.005
Publication year	−0.028	0.011	−2.446	0.007
Average age of sample	−0.040	0.005	−7.946	0.000
Percent of male leaders	0.001	0.002	4.472	0.000
First author female	−0.372	0.087	−4.296	0.000
Size of organization	−0.308	0.103	−2.989	0.001
Random selection used	0.110	0.035	3.129	0.000

In meta-analysis as in other analyses, using only complete cases can seriously limit the number of observations available for the analysis. Losing cases decreases the power of the analysis, and also ignores the information contained in the incomplete cases (Kim and Curry 1977; Little and Rubin 1987).

When data are missing because of a nonignorable response mechanism or are MAR, complete case analysis yields biased estimates since the complete cases cannot be considered representative of the original sample. With nonignorable missing data, the complete cases observe only part of the distribution of a particular variable. With MAR data, the complete cases are also not a random sample of the original sample. But with MAR data, the incomplete cases still provide information since the variables that are completely observed across the data set are related to the probability of missing a particular variable.

7.6.1.1 Example: Complete-Case Analysis

The data for the examples that follow are adapted from a meta-analysis by Eagly et al. (2003) that examines gender differences in transformational, transactional, and laissez-faire leadership styles. Six moderators are used in this example: Year the study was published (Year), if the first author is female, the average age of the participants, the size of the organization where the study was conducted, whether random methods were used to select the participants from the organization, and the percent of males in the leadership roles in the organization. Table 7.2 provides the complete case results for a meta-regression of the transformational leadership data. Only 22 or 50% of the cases include all five predictor variables. Positive effect sizes indicate that males were found to score higher on transformational leadership scales, while negative effect sizes favor females. For this sample of studies, gender differences in favor of females are associated with more recently published studies, with studies that have older samples of participants, when the study's first author is female, and in studies conducted in larger organizations. Gender differences in favor of males are found in studies with a higher percentage of men in leadership roles in the sample, and when random methods were used to select the sample from the target population. The complete case results can be obtained from any weighted regression program usually by default. Most statistical computing packages automatically delete any cases missing at least one variable in the model.

Table 7.3 Missing data patterns in leadership data

Pattern	Effect size	% male leaders	Average age	N
0	O	O	O	22 (50%)
1	O	O	–	15 (34%)
2	O	–	O	3 (7%)
3	O	–	–	4 (9%)
N	44 (100%)	37 (84%)	25 (57%)	44 (100%)

O indicates observed in data

7.6.2 Available Case Analysis or Pairwise Deletion

An available-case analysis or pairwise deletion estimates parameters using as much data as possible. In other words, if three variables in a data set are missing 0%, 10% and 40% of their values, respectively, then the correlation between the first and third variable would be estimated using the 60% of cases that observe both variables, and the correlation between the first and second variable would use the 90% of cases that observe these two variables. A different set of cases would provide the estimate of the correlation of the second and third variables since there is the potential of having between 40% and 50% of these cases missing both variables. For example, Table 7.3 shows the pairs of cases that would be used to estimate parameters using available case analysis in the leadership data. The letter O indicates that the variable was observed in that missing data pattern. Assuming the effect sizes are completely observed, the estimate of the correlation between the effect size and the percentage of men in leadership roles would be based on 37 studies or 84% of the studies. The correlation between effect size and average age of the sample would use 25 studies or 57% of the sample. The estimated correlation of percentage of men in leadership roles and average age of the sample would use only 22 studies or 50% of the sample.

This simple example illustrates the drawback of using available case analysis – each correlation in the variance-covariance matrix estimated using available cases could be based on different subsets of the original data set. If data are MCAR, then each of these subsets are representative of the original data, and available case analysis provides estimates that are unbiased. If data are MAR, however, then each of these subsets is not representative of the original data and will produce biased estimates.

Much of the early research on methods for missing data focuses on the performance of available case analysis versus complete case analysis (Glasser 1964; Haitovsky 1968; Kim and Curry 1977). Fahrbach (2001) examines the research on available case analysis and concludes that available case methods provide more efficient estimators than complete case analysis when correlations between two independent variables are moderate (around 0.6). This view, however, is not shared by all who have examined this literature (Allison 2002).

One statistical problem that could arise from the use of available cases under any form of missing data is a non-positive definite variance-covariance matrix, i.e., a

Table 7.4 Available case analysis

Variable	Coefficient	SE	Z	p
Intercept	−46.283	19.71	−2.348	0.009
Publication year	0.024	0.010	2.383	0.009
Average age of sample	−0.047	0.006	−7.625	0.000
Percent of male leaders	0.013	0.002	6.166	0.000
First author female	−0.260	0.058	−4.455	0.000
Size of organization	0.185	0.062	3.008	0.001
Random selection used	0.037	0.041	0.902	0.184

variance-covariance matrix that cannot be inverted to obtain the estimates of slopes for a regression model. One reason for this problem is that different subsets of studies are used to compute the elements of the variance-covariance matrix. Further, Allison (2002) points out that a more difficult problem in the application of available case analysis concerns the computation of standard errors of available case estimates. At issue is the correct sample size when computing standard errors since each parameter could be estimated with a different subset of data. Some of the standard errors could be based on the whole data set, while others may be based on the subset of studies that observe a particular variable or pair of variables. Though many statistical computing packages implement available case analysis, how standard errors are computed differs widely.

7.6.2.1 Example: Available Case Analysis

Table 7.4 provides the results from SPSS estimating a meta-regression for the leadership studies using pairwise deletion. In this analysis, gender differences favoring females are associated with an older sample and with studies whose first author is female. Gender differences favoring males are associated with a higher percentage of men in leadership roles, larger organizations in the sample, and with more recent studies. The last two findings, related to larger organizations and more recent findings, contradicts the findings from the complete case analysis.

While available case analysis is easy to understand and implement, there is little consensus in the literature about the conditions where available case analysis outperforms complete case analysis when data are MCAR. As described above, the performance of available case analysis may relate to the size of the correlations between variables in the data, but there is no consensus about the optimal size of these correlations needed to produce unbiased estimates.

7.6.3 Single Value Imputation with the Complete Case Mean

When values are missing in a meta-analysis (or in any statistical analysis), many researchers have replaced the missing value with a “reasonable” value such as the

Table 7.5 Comparison of complete-case and mean imputation values

Variable	N	Mean	SD
Average age of sample	22	44.88	6.629
Average age, mean imputed	*44*	*44.88*	*4.952*
Percent of male leaders	22	64.97	17.557
Percent of male leaders, mean imputed	*44*	*64.97*	*16.060*

Table 7.6 Linear model of effect size for leadership data using mean imputation

Variable	Coefficient	SE	Z	p
Intercept	21.165	11.468	1.846	0.032
Publication year	−0.01	0.006	−1.783	0.037
Average age of sample	−0.024	0.004	−6.537	0.000
Percent of male leaders	0.006	0.001	4.753	0.000
First author female	−0.076	0.035	−2.204	0.014
Size of organization	−0.064	0.034	−1.884	0.030
Random selection used	−0.013	0.028	−0.469	0.319

mean for the cases that observed the variable. Little and Rubin (1987) refer to this strategy as single-value imputation. Researchers commonly use two different strategies to fill in missing values. One method fills in the complete case mean, and the other uses regression with the complete cases to estimate predicted values for missing observations given the observed values in a particular case.

Replacing the missing values in a variable with the complete case mean of the variable is also referred to as unconditional mean imputation. When we substitute a single value for all the missing values, the estimate of the variance of that variable is decreased. The estimated variance thus does not reflect the true uncertainty in the variable – instead the smaller variance wrongly indicates more certainty about the value. These biases get compounded when using the biased variances to estimate models of effect size. Imputation of the complete case mean never leads to unbiased variances of variables with missing data.

7.6.3.1 Example: Mean Imputation

In Table 7.5, the missing values of average age and percent of men in leadership roles were imputed with the complete case mean. These values are given under the complete case means and standard deviations. While the means for the variables remains the same, the standard deviations are smaller for the variables when missing values are replaced by the complete case mean. The problem is compounded in the regression analysis in Table 7.6. These results would lead us to different conclusions from those based on either the complete-case or pairwise deletion analyses.

Using mean imputation in this example would lead us to conclude that men score higher on transformational leadership scales only in studies that have a larger percentage of male leaders in the sample. The use of random selection is not related to variability in the effect size across studies. All the other predictors favor women's scores on transformational leadership.

7.6.4 Single Value Imputation Using Regression Techniques

A single-value imputation method that provides less biased results with missing data was first suggested by Buck (1960). Instead of replacing each missing value with the complete case mean, each missing value is replaced with the predicted value from a regression model using the variables observed in that particular case as predictors and the missing variable as the outcome. This method is also referred to as conditional mean imputation or as regression imputation. For each pattern of missing data, the cases with complete data on the variables in that pattern are used to estimate regressions using the observed variables to predict the missing values. The end result is that each missing value is replaced by a predicted value from a regression using the values of the observed variables in that case. When data are MCAR, then each of the subsets used to estimate prediction equations are representative of the original sample of studies. This method results in more variation than in unconditional mean imputation since the missing values are replaced with values that depend on the regression equation. However, the standard errors using Buck's method are still too small. This underestimation occurs since Buck's method replaces the missing values with predicted values that lie directly on the regression line used to impute the values. In other words, Buck's method results in imputing values that are predicted exactly by the regression equation without error.

Little and Rubin (1987) present the form of the bias for Buck's method and suggest corrections to the estimated variances to account for the bias. If we have two variables, Y_1 and Y_2, and Y_2 has missing observations, then the form of the bias using Buck's method to fill in values for Y_2 is given by

$$(n - n^{(2)})(n - 1)^{-1}\sigma_{22.1},$$

where n is the sample size, $n^{(2)}$ is the number of cases that observe Y_2, and $\sigma_{22.1}$ is the residual variance from the regression of Y_2 on Y_1. Little and Rubin also provide the more general form of the bias with more than two variables. Table 7.7 compares the complete case means and standard deviations of average age of the sample and percent of male leaders with those obtained using regression imputation (Buck's method) and Little and Rubin (1987) correction.

The standard deviations for the corrected regression imputation results are larger than for both complete cases and for the uncorrected regression imputation. The correction reflects the increased uncertainty in the estimates due to missing data.

Note that correcting the bias in Buck's method involves adjusting the variance of the variable with missing observations. Using Little and Rubin (1987) correction

Table 7.7 Comparison of methods for imputing missing data

	Complete cases N = 22		Regression imputation (uncorrected) N = 44		Regression imputation (corrected) N = 44	
Variable	Mean	SD	Mean	SD	Mean	SD
Average age of sample	45.23	6.70	44.52	5.52	44.52	6.91
Percent of male leaders	64.23	17.44	65.44	16.22	65.44	17.52

results in a corrected covariance matrix, and not individual estimates for each missing observation. Thus, estimating the linear model of effect size in our example will require estimating the coefficients using only the variance-covariance matrix.

To date, there has not been extensive research on the performance of Buck's method to other more complex methods for missing data in meta-analysis. As seen above, one advantage of using Buck's method with the corrections suggested by Little and Rubin is that the standard errors of estimates reflect the uncertainty in the data and lead to more conservative and less biased estimates than complete case and available case methods. While using Buck's method is fairly simple, the adjustments of the variances and covariances of variables with missing observations adds another step to the analysis. In addition, it is not clear how to utilize the corrected variances and covariances when estimating weighted regression models of effect sizes in meta-analysis. While it is possible to estimate a linear model using a covariance matrix in the major statistical packages, it is not clear how to incorporate the weights into the corrected covariance matrix.

When missing predictors are MCAR, then complete case analysis, available case analysis and conditional mean imputation have the potential for producing unbiased results. The cost of using these methods lies in the estimation of standard errors. For complete case analysis, the standard errors will be larger than those from the hypothetically complete data. In available case and conditional mean imputation, the standard errors will be too small, though those from conditional mean imputation can be adjusted. When data are MAR or are missing due to a nonignorable response mechanism, none of the simpler methods produce unbiased results.

7.6.4.1 Example: Regression Imputation

Table 7.8 provides the estimates of the linear model of effect size when using regression imputation. The results in the table were produced using SPSS Missing Values Analysis, saving a data set where missing values are imputed using regression.

These results differ from the mean imputation results in that both when the first author is female and whether random selection was used are not related to effect size magnitude. As in the mean imputation results, only percent of male leaders is related to high scores on transformational leadership for men. These results are again not consistent with the available case or complete case results.

Table 7.8 Linear model of effect size for leadership data using regression imputation

Variable	Coefficient	SE	Z	p
Intercept	38.436	11.181	3.438	0.001
Publication year	−0.019	0.005	−3.515	<0.001
Average age of sample	−0.024	0.003	−7.105	<0.001
Percent of male leaders	0.008	0.001	5.921	<0.001
First author female	−0.044	0.035	−1.252	0.105
Size of organization	−0.079	0.034	−2.339	0.010
Random selection used	−0.018	0.028	−0.65	0.258

7.7 Model-Based Methods for Missing Data in Meta-analysis

The simple methods for missing data discussed above do not provide unbiased estimates under all circumstances. The general problem with these ad hoc methods is that they do not take into account the distribution of the hypothetically complete data. For example, filling in a zero for a missing effect size may be a reasonable assumption, but it is not based on a plausible distribution for the effect sizes in a review. The missing data methods outlined in this section begin with a model for the observed data. Maximum-likelihood methods using the EM algorithm are based on the observed data likelihood while multiple imputation techniques are based on the observed data posterior distribution. Given the assumptions of ignorable missing data and multivariate normal data, the observed data likelihood and the observed data posterior distribution will provide the information needed to estimate important data parameters. The next sections outline both methods, providing an example of its application.

7.7.1 Maximum-Likelihood Methods for Missing Data Using the EM Algorithm

In statistical inference, we are interested in obtaining an estimate of a parameter from our data that has optimal properties such as minimum variance and bias. The method most often used to obtain parameter estimates is maximum likelihood. Maximum likelihood methods are based on a joint density distribution of the data. For example, if we assume that our data consist of a sample of n observations from the normal distribution, then we can write down our joint density as a product of a series of n normal densities. The maximum value for this density function is a parameter estimate that has optimal properties. For example, the arithmetic average of a set of observations from a normal distribution is the maximum likelihood estimate of the mean of the population.

When the data include missing observations, our data likelihood becomes complicated. Since our goal is to make inferences about the population, the relevant likelihood for our problem is the hypothetically complete data likelihood. This complete data likelihood includes the density of the observed data given the unknown parameters of the population distribution and the density of the missing data given the observed data and the unknown population parameters. Since we do not know the density of the missing data, it would seem impossible to compute the maximum likelihood estimates of the complete data. However, Dempster et al. (1977) developed an algorithm called the Expectation-Maximization algorithm, or EM algorithm. As its name denotes, obtaining maximum likelihood estimates requires an iterative process. In the first step, the Expectation or E-step, the algorithm uses an estimate of the data parameters (such as the mean and covariance matrix) to estimate plausible values for the missing observations. These missing observations are then "filled in" for the Maximization or M-step where the algorithm re-estimates the population parameters. Thus, in the E-step, we assume that we have estimates of the population parameters, and we use these to obtain values for the missing observations. In the M-step, we assume that the missing observations are "real" and use them to re-compute the population parameters. The iterations continue until the estimates of the population parameters do not change, i.e., when the algorithm converges.

When we can assume the data is multivariate normal, the distribution for the missing values given the observed values is also multivariate normal. It is important to note that the algorithm provides the maximum likelihood estimates of the sufficient statistics, the means and covariance matrix, and not the maximum likelihood of any particular missing observation. Thus, we have maximum likelihood estimates of the means, variances and covariances of our data that we can use to obtain estimates of other parameters such as regression coefficients for a linear model. We must also assume that the response mechanism is ignorable, i.e., that our data are either MCAR or MAR so that the distribution of our data does not need to include a specification of the response mechanism.

7.7.1.1 Example Using the EM Algorithm

There are a number of options available in commercially-available statistical packages and in freeware for use in computing the EM estimates. Using Schafer's NORM (1999) program, we obtain the maximum likelihood estimates of the means and standard deviations for our two variables with missing data, average of the sample and percent of male leaders as seen in Table 7.9. These estimates are compared with the estimates from the complete case and corrected regression imputation analyses. The standard deviation from the EM algorithm falls between the complete case and regression imputation estimates. The standard deviations are of the same magnitude as the complete case estimates, generally reflecting the same amount of "information" as in complete cases.

Table 7.9 Comparison of estimates from the EM algorithm, complete-case analysis and regression imputation

	Complete cases N = 22		Regression imputation (corrected) N = 44		EM algorithm N = 44	
Variable	Mean	SD	Mean	SD	Mean	SD
Average age of sample	45.23	6.70	44.52	6.91	44.44	6.67
Percent of male leaders	64.23	17.44	65.44	17.52	65.65	17.21

The difficulty with using the EM algorithm in the context of meta-analysis is similar to that of the corrected regression imputation analysis – it is not clear how to estimate the weighted regression coefficients using the sufficient statistics matrix (the matrix of means, variances and covariances). Thus, the EM algorithm has limited applicability to meta-analysis since we do not yet have a method for computing the weighted regression coefficients from the means, variances and covariances of the variables in the data.

7.7.2 Multiple Imputation for Multivariate Normal Data

Multiple imputation has become the method of choice in many contexts of missing data. The main advantage of multiple imputation is that the analyst uses the same statistical procedures in the analysis phase that were planned for completely observed data. In other words, in the analysis phase of multiple imputation, the researcher does not need to adjust standard errors as in Buck's method, and does not need to estimate a regression from the covariance matrix as in maximum likelihood with the EM algorithm. Multiple imputation, as its name implies, is a technique that generates multiple possible values for each missing observation in the data. Each of these values is used in turn to create a complete data set. The analyst uses standard statistical procedures to analyze each of these multiply imputed data sets, and then combines the results of these analyses for statistical inference.

Thus, multiple imputation consists of three phases. The first phase consists of generating the possible values for each missing observation. The second phase then analyzes each completed data set using standard statistical procedures. The third phase involves combining the estimates from the analyses of the second phase to obtain results to use for statistical inference. Each of these phases is discussed conceptually below. Readers interested in more details should consult the following works (Enders 2010; Schafer 1997).

7.7.2.1 Generating Multiple Imputations

Multiple imputation, like maximum likelihood methods for missing data, relies on a model for the distribution of missing data given the observed data under the condition of MAR data. As in maximum likelihood, this distribution is complex. The previous section discussed the use of the EM algorithm to estimate sufficient statistics from this distribution assuming that the hypothetically complete data is multivariate normal. Multiple imputation uses Bayesian methods to obtain random draws from the posterior predictive distribution of the missing observations given the observed observations. These random draws are completed in an iterative process much like the EM algorithm. Given the means and covariance matrix of our hypothetically complete multivariate normal data, we can then obtain the form of the distribution of the missing observations given the observed data, and draw a random observation from that distribution. That observation would be one plausible value for a missing value for a given case. Once we have drawn plausible values for all our missing observations, we obtain a new estimate of our means and covariance matrix, and repeat the process. Note that again we are assuming that our response mechanism is ignorable so that the posterior distribution also does not include a specification of the response mechanism.

In order to generate these random draws, however, we need to use simulation techniques such as Markov Chain Monte Carlo. These methods allow the use of simulation to obtain random draws from a complex distribution. While this phase is the most complex statistically, there are many commercial software packages and freeware available to generate these imputations, especially in the case where we can assume the complete data is multivariate normal. The Appendix provides details about these computer packages.

7.7.2.2 Analyzing the Completed Data Sets

Multiple imputation was first developed in large-scale survey research to assist researchers who wanted to use public data sets. The idea was to provide researchers with a way to handle missing data that did not require specialized computer programming skills or statistical expertise when using these publicly-available data sets. In this second step, the researcher will obtain a series of completed data sets, with each missing observation filled in using the methods in the prior section. Once the imputations are generated, the analyst uses whatever methods were originally planned for the data. These analyses are repeated for each completed data set. As Schafer (1997) argues, for most applications of multiple imputation, five imputations is sufficient to obtain estimates for statistical inference. In this phase, the analyst takes each completed data set and obtains estimates for the originally planned model. Table 7.10 provides the estimates for the linear model of effect size for each of five imputations generated as discussed in the Appendix.

7.7.2.3 Combining the Estimates

Rubin (1987) provides the formulas for combining the multiply-imputed estimates to obtain overall estimates and their standard errors. Let us denote the mean of our target estimate for the ith parameter across all j imputations as

Table 7.10 Imputations for the leadership data

Variable	Imputation 1	Imputation 2	Imputation 3	Imputation 4	Imputation 5
Intercept	17.219	3.276	31.380	0.429	54.661
Publication year	−0.008	−0.001	−0.016	0.000	−0.027
Average age of sample	−0.018	−0.026	−0.010	−0.021	−0.031
Percent of male leaders	0.005	0.004	0.003	0.007	0.009
First author female	−0.174	−0.186	−0.004	−0.147	−0.181
Size of organization	−0.016	0.015	−0.046	−0.021	0.063
Random selection used	0.047	0.068	−0.043	−0.018	0.112

Table 7.11 Multiply-imputed regression coefficients

Variable	Coefficient	SE	Z	p
Intercept	21.393	27.726	0.77	0.29
Publication year	−0.010	0.014	−0.75	0.30
Average age of sample	−0.021	0.009	−2.34	0.13
Percent of male leaders	0.006	0.003	2.06	0.14
First author female	−0.138	0.094	−1.47	0.19
Size of organization	−0.001	0.059	−0.02	0.49
Random selection used	0.033	0.076	0.43	0.37

$$\bar{q}_i = \frac{1}{m}\sum_{j=1}^{m} q_{ij}$$

where q_{ij} is the estimate of the ith parameter from the jth completed-data sets. To obtain the standard errors of the q_{ij}, we need two estimates of variance. Denote the variance of the estimate, q_{ij}, from the jth completed sets as $se^2(q_i)$. The variance across the j data sets of the estimate q_i is given by

$$s_{q_i}^2 = \frac{1}{m-1}\sum_{j=1}^{m} (q_{ij} - \bar{q}_i)^2.$$

The standard error of the points estimates of the q_{ij} is then given by

$$SE(q_i) = \sqrt{\frac{1}{m}\sum_{j=1}^{m} se^2(q_i) + s_{q_i}^2\left[1 + \frac{1}{m}\right]}.$$

Table 7.11 presents the multiply-imputed results for the leadership data. In this analysis, none of the coefficients are significantly different from zero with $p = 0.05$.

Multiple imputation is now more widely implemented in statistical computing packages. SAS (Yuan 2000) implements multiple imputation procedures with multivariate normal data. The examples in this chapter were computed with Schafer (1999) NORM program, a freeware program for conducting multiple imputation with multivariate normal data. The Appendix provides information about options for obtaining multiple imputation estimates and for combining those estimates.

In general, multiple imputation is the recommended method for handling missing data in any statistical analysis, including meta-analysis. The methods illustrated in this chapter produce divergent results, indicating that the results of this analysis are sensitive to missing data. A potential difficulty in this data could be power since there are just slightly over 40 studies available for analysis. More research is needed to understand the conditions where meta-analysts should use multiple imputation.

Appendix

Computing Packages for Computation of the Multiple Imputation Results

There are a number of options for obtaining multiple imputation results in a meta-analysis model. Two freeware programs are available. The first is the program Norm by Schafer and available at http://www.stat.psu.edu/~jls/misoftwa.html. The Norm program runs as a stand alone program on Windows 95/98/NT. The second is a program available in R by Honaker et al. called Amelia II and available at http://gking.harvard.edu/amelia/. Schafer's norm program was used for the example given earlier.

The program SAS includes two procedures, one for generating the multiple imputations, PROC MI, and a second for analyzing the completed data sets, PROC MIANALYZE. For obtaining the weighted regression results for meta-analysis, the SAS procedure PROC MIANALYZE will have limited utility since the standard errors of the weighted regression coefficients will need to be adjusted as detailed by Lipsey and Wilson (2001). Below is an illustration of the use of PROC MI for the leadership data.

R Programs

One program available in R for generating multiple imputations is Amelia II (Honaker et al. 2011). Directions for using the program are available at http://gking.harvard.edu/amelia/. Once the program is loaded into R, the following command was used to generate $m = 5$ imputed data sets.

```
> a.out <- amelia(leadimp, m = 5, idvars = "ID")
```

Table 7.12 Regression estimates from each imputation generated using Amelia

Variable	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5
Intercept	51.05[a] (11.28)[b]*	68.56 (12.03)*	45.79 (11.061)*	41.854 (10.984)*	−5.974 (13.665)
Year	−0.025 (0.006)*	−0.034 (0.006)*	−0.023 (0.006)*	−0.020 (0.005)*	0.003 (0.007)
Average age	−0.033 (0.004)*	−0.019 (0.003)*	−0.022 (0.003)*	−0.037 (0.004)*	−0.019 (0.003)*
Percent of male leaders	0.009 (0.001)*	0.006 (0.001)*	0.006 (0.001)*	0.007 (0.001)*	0.002 (0.001)*
First author female	−0.284 (0.045)*	−0.127 (0.040)*	−0.060 (0.033)*	−0.338 (0.048)*	−0.154 (0.041)*
Size of organization	−0.113 (0.039)*	−0.056 (0.037)	−0.140 (0.042)*	−0.237 (0.047)*	−0.001 (0.034)*
Random selection used	0.049 (0.019)*	0.101 (0.023)*	0.059 (0.019)*	0.075 (0.020)*	0.068 (0.021)*

[a]Coefficient estimate
[b]Standard error of coefficient in parentheses
*Coefficient is significantly different from zero

Table 7.13 Multiply-imputed estimates from Amelia

Variable	Coefficient	SE	Z	p
Intercept	40.255	32.660	1.232	0.217
Year	−0.0198	0.016	−1.207	0.210
Average age	−0.026	0.010	−2.617	0.116
Percent of men	0.006	0.003	1.989	0.148
First author female	−0.193	0.133	−1.451	0.192
Size of organization	−0.110	0.106	−1.039	0.127
Random selection used	0.070	0.030	2.377	0.244

The imputed data sets can be saved for export into another program to complete the analyses using the command,

>write.amelia(obj = a.out, file.stem = "outdata").

where “obj” refers to the name given to the object with the imputed data sets (the result of using the command **Amelia**), and “file.stem” provides the name of the data sets that will be written from the program.

Table 7.12 are the weighted regression estimates for the effect size model from each imputation obtained in Amelia. The two variables missing observations are average age of subjects and percent of male leaders. There is variation among the five data sets in their estimates of the regression coefficients. This variation signals that there is some uncertainty in the data set due to missing observations.

Table 7.13 provides the multiply-imputed estimates for the linear model of effect size. These estimates were combined in Excel, and are fairly consistent with the earlier multiple imputation analysis using Schafer's Norm program. None of the coefficients are significantly different from zero.

SAS Proc MI

The SAS procedure PROC MI provides a number of options for analyzing data with missing data. For the example illustrated in this chapter, we use the Monte Carlo Markov Chain with a single chain for the multiple imputations. We also use the EM estimates as the initial starting values for the MCMC analysis. The commands below were used with the leadership data to produce the five imputed data sets:

```
proc mi data = work.leader out = work.leaderimp seed = 101897;
var year ageave perlead gen2 sizeorg2 rndm2 effsize;
mcmc;
```

The first line of the command gives the name of the data set to use, the name of the created SAS data set with the imputations, and the seed number for the pseudo-random number generator. The second command line provides the variables to use in the imputations. Note that the effect size is included in this analysis. The third line specifies the use of Markov Chain Monte Carlo to obtain the estimates of the joint posterior distribution as described by Rubin (1987). Note that the number of imputations are not specified; the default number of imputed data sets generated is five, the number recommended by Schafer (1997).

SAS Proc MI provides a number of useful tables, including one outlining the missing data patterns and the group means for each variable within each missing data pattern. Once the imputations are generated, the procedure gives the estimates for the mean and standard error of the variables with missing data as illustrated below.

Multiple Imputation Parameter Estimates

Variable	Mean	SE	95% confidence limits		DF
Average age of sample	44.109	1.619	40.341	47.877	7.596
Percent of male leaders	65.691	2.898	59.743	71.640	26.869

Variable	Minimum	Maximum	Mu0	t for Mean = Mu0	Pr > |t|
Average age of sample	42.659	45.481	0	27.25	<.0001
Percent of male leaders	64.586	67.390	0	22.67	<.0001

To obtain the weighted regression results for each imputation, we use Proc Reg with weights. The command lines are shown below.

Table 7.14 Multiple imputations generated using SAS Proc MI

Variable	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5
Intercept	29.36[a] (14.38)[b]*	23.70 (11.49)*	−7.962 (14.16)	71.548 (12.560)*	14.935 (11.71)
Year	−0.012 (0.006)*	−0.012 (0.006)*	0.004 (0.007)	−0.036 (0.006)*	−0.007 (0.006)
Average age	−0.026 (0.004)*	−0.026 (0.004)*	−0.019 (0.004)*	−0.013 (0.003)*	−0.024 (0.003)*
Percent of male leaders	0.009 (0.001)*	0.009 (0.001)*	0.004 (0.001)*	0.005 (0.001)*	0.004 (0.001)*
First author female	−0.198 (0.042)*	−0.198 (0.042)*	−0.171 (0.042)*	−0.053 (0.037)	−0.202 (0.040)*
Size of organization	−0.038 (0.034)	−0.038 (0.034)	−0.072 (0.034)*	−0.084 (0.034)*	0.072 (0.041)*
Random selection used	0.006 (0.031)*	0.062 (0.031)*	0.066 (0.033)*	−0.021 (0.028)*	0.034 (0.029)

[a]Coefficient estimate
[b]Standard error of coefficient in parentheses
*Coefficient is significantly different from zero

Table 7.15 Multiply-imputed estimates generated by SAS

Variable	Coefficient	SE	t	p
Intercept	26.315	34.305	0.767	0.292
Year	−0.013	0.017	−0.749	0.295
Average age	−0.020	0.007	−2.744	0.111
Percent of men	0.005	0.003	1.706	0.169
First author female	−0.143	0.084	−1.705	0.169
Size of organization	−0.026	0.078	−0.336	0.397
Random selection used	0.030	0.050	0.591	0.330

```
proc reg data = work.leaderimp outest = work.regout covout;
model effsize = year ageave perlead gen2 sizeorg2 rndm2;
weight wt;
by _Imputation_;
run;
```

The lines given above use the SAS data set generated by Proc MI, and estimate the coefficients for the effect size model using weighted regression. The results are computed for each imputation as indicated in the **by** statement. Table 7.14 provides the weighted regression results for each imputation.

Table 7.15 gives the multiply-imputed estimates for the weighted regression results. As in the prior analyses, none of the regression coefficients were significantly different from zero.

References

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

Begg, C.B., and J.A. Berlin. 1988. Publication bias: A problem in interpreting medical data (with discussion). *Journal of the Royal Statistical Society Series A* 151(2): 419–463.

Buck, S.F. 1960. A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society Series B* 22(2): 302–303.

Chan, A.-W., A. Hrobjartsson, M.T. Haahr, P.C. Gotzsche, and D.G. Altman. 2004. Empirical evidence for selective reporting of outcomes in randomized trials. *Journal of the American Medical Association* 291(20): 2457–2465.

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* 39(1): 1–38.

Duval, S. 2005. The Trim and Fill method. In *Publication bias in meta-analysis: Prevention, assessment and adjustments*, ed. H.R. Rothstein, A.J. Sutton, and M. Borenstein. West Sussex: Wiley.

Duval, S., and R. Tweedie. 2000. Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56(2): 455–463.

Eagly, A.H., M.C. Johannesen-Schmidt, and M.L. van Engen. 2003. Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin* 129(4): 569–592.

Egger, M., G.D. Smith, M. Schneider, and C. Minder. 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 315(7109): 629–634.

Enders, C.K. 2010. *Applied missing data analysis. Methodology in the Social Sciences*. New York: Guilford.

Fahrbach, K.R. 2001. *An investigation of methods for mixed-model meta-analysis in the presence of missing data*. Lansing: Michigan State University.

Glasser, M. 1964. Linear regression analysis with missing observations among the independent variables. *Journal of the American Statistical Association* 59(307): 834–844.

Hackshaw, A.K., M.R. Law, and N.J. Wald. 1997. The accumulated evidence on lung cancer and environmentaly tobacco smoke. *British Medical Journal* 315(7114): 980–988.

Haitovsky, Y. 1968. Missing data in regression analysis. *Journal of the royal Statistical Society Series B* 30(1): 67–82.

Hemminki, E. 1980. Study of information submitted by drug companies to licensing authorities. *British Medical Journal* 280(6217): 833–836.

Honaker, J., G. King, and M. Blackwell (2011) Amelia II: A program for missing data. http://r.iq.harvard.edu/src/contrib/

Kim, J.-O., and J. Curry. 1977. The treatment of missing data in multivariate analysis. *Sociological Methods and Research* 6(2): 215–240.

Lipsey, M.W., and D.B. Wilson. 2001. *Practical meta-analysis*. Thousand Oaks: Sage Publications.

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

Orwin, R.G., and D.S. Cordray. 1985. Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. *Psychological Bulletin* 97(1): 134–147.

Rosenthal, R. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86(3): 638–641.

Rothstein, H.R., A.J. Sutton, and M. Borenstein. 2005. *Publication bias in meta-analysis: Prevention, Assessment and Adjustments*. West Sussex: Wiley.

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

Rubin, D.B. 1987. Multiple imputation for nonresponse in surveys. Wiley, New York, NY

Schafer, J.L. 1997. *Analysis of incomplete multivariate data*. London: Chapman Hall.

Schafer, J.L. 1999. *NORM: Multiple imputation of incomplete multivariate data under a normal model. Software for Windows*. University Park: Department of Statistics, Penn State University.

Schafer, J.L., and J.W. Graham. 2002. Missing data: Our view of the state of the art. *Psychological Methods* 7(2): 147–177.

Shadish, W.R., L. Robinson, and C. Lu. 1999. *ES: A computer program and manual for effect size calculation*. St. Paul: Assessment Systems Corporation.

Sirin, S.R. 2005. Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research* 75(3): 417–453. doi:10.3102/00346543075003417.

Smith, M.L. 1980. Publication bias and meta-analysis. *Evaluation in Education* 4: 22–24.

Sterne, J.A.C., B.J. Becker, and M. Egger. 2005. The funnel plot. In *Publication bias in meta-analysis: Prevention, assessment and adjustment*, ed. H.R. Rothstein, A.J. Sutton, and M. Borenstein. West Sussex: Wiley.

Vevea, J.L., and C.M. Woods. 2005. Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods* 10(4): 428–443.

Williamson, P.R., C. Gamble, D.G. Altman, and J.L. Hutton. 2005. Outcome selection biase in meta-analysis. *Statistical Methods in Medical Research* 14(5): 515–524.

Wilson, D.B. 2010. Practical meta-analysis effect size calculator. Campbell Collaboration. http://www.campbellcollaboration.org/resources/effect_size_input.php. Accessed 16 July 2011.

Yuan, Y.C. 2000. Multiple imputation for missing data: Concepts and new developments. http://support.sas.com/rnd/app/papers/multipleimputation.pdf. Accessed 2 April 2011.