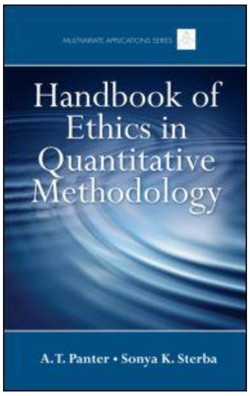

Handbook of Ethics in Quantitative Methodology

A.T. Panter, Sonya K. Sterba

The Impact of Missing Data on the Ethical Quality of a Research Study

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9780203840023.ch14
Craig K. Enders, Amanda C. Gottschall
Published online on: 20 Jan 2011

14

The Impact of Missing Data on the Ethical Quality of a Research Study

Craig K. Enders
Arizona State University
Amanda C. Gottschall
Arizona State University

The purpose of this chapter is to explore the impact of missing data on the ethical quality of a research study. In doing so, we borrow heavily from the work of Rosenthal (1994) and Rosenthal and Rosnow (1984). The overarching principle of Rosenthal's (1994) work is that ethics is closely linked with the quality of a research study, such that high-quality studies are more ethically defensible than low-quality studies. Missing data pose an obvious threat to quality at the analysis stage of a study (e.g., when a researcher uses a missing data handling technique that is prone to bias), but ethical issues arise throughout the entire research process. Accordingly, we explore the linkage between quality and ethics at the design and data collection phase, the analysis phase, and the reporting phase. In doing so, we also apply Rosenthal and Rosnow's (1984) *cost–utility model* to certain missing data issues (see also Rosnow & Rosenthal, Chapter 3, this volume). In this framework, the costs associated with a study (e.g., potential harm to participants, time, money, resources) are weighed against its utility (e.g., potential benefits to participants, science, or society). As it relates to ethics, a study is more defensible when its benefits exceed its costs.

Missing Data Mechanisms

To fully appreciate the impact that missing data can have on the quality (and thus the ethics) of a research study, it is necessary to understand missing data theory. Rubin and colleagues (Little & Rubin, 2002; Rubin, 1976) developed a classification system for missing data problems that is firmly entrenched in the methodological literature. These so-called missing data mechanisms describe how the propensity for missing data is related to measured variables, if at all. From a practical perspective, missing data mechanisms serve as assumptions that dictate the performance of different analytic approaches. This section gives a brief conceptual description of Rubin's missing data mechanisms, and more detailed accounts are available elsewhere in the literature (Enders, 2010; Little & Rubin, 2002; Rubin, 1976; Schafer & Graham, 2002).

To begin, a *missing completely at random* (MCAR) mechanism occurs when the probability of missing data on a variable is unrelated to other measured variables and to the values of the variable itself. When these conditions hold, the observed data are a simple random sample of the hypothetically complete data set. To illustrate, suppose that an industrial organizational psychologist is studying psychological well-being in the workplace and finds that some of the well-being scores are missing for purely haphazard reasons (e.g., an employee left the study because she went on maternity leave, an employee quit because his spouse accepted a job in another state, or an employee was on vacation when the surveys were administered). An MCAR mechanism would describe this scenario if the reasons for missingness are uncorrelated with well-being scores and with other measured variables.

Under a *missing at random* (MAR) mechanism, the probability of missing data on a variable is related to the observed values of other variables in the analysis model, but not to the unobserved values of the variable itself. As an example, consider a health study where researchers restrict the administration of a sensitive sexual behavior questionnaire to participants that are above the age of 15. Provided that age is included in any analysis that involves the sexual behavior variable, this example satisfies the MAR mechanism because missingness is unrelated to sexual behavior. Said differently, there is no residual relationship between the propensity for missing data and sexual behavior after controlling for age.

Finally, a *missing not at random* (MNAR) mechanism occurs when the probability of missingness is directly related to the scores on the incomplete variable itself. For example, suppose that a psychologist is studying quality of life in a clinical trial for a new cancer medication and finds that a number of patients become so ill (i.e., their quality of life becomes so poor) that they can no longer participate in the study. This example is consistent with an MNAR mechanism because the probability of missing data on the quality of life measure is directly related to a participant's quality of life. As an aside, an MNAR mechanism can also result when a cause or correlate of missingness is omitted from an analysis (e.g., the health researchers from the MAR example analyze the sexual behavior data without incorporating age into the analysis).

From a practical perspective, Rubin's mechanisms are vitally important because they serve as assumptions that dictate the performance of different *missing data handling techniques*. For example, an analysis method that assumes an MCAR mechanism will produce accurate parameter estimates under a more limited set of circumstances than a technique that requires MAR data because MCAR is a more restrictive condition than MAR. Based on theory alone, it is possible to reject MCAR-based methods in favor of approaches that assume MAR or MNAR. The inherent difficulty with missing data problems is that there is no way to determine which mechanism is at play in a given analysis. Although the observed data may provide evidence against an MCAR mechanism, there is no way to empirically differentiate MAR from MNAR (establishing that there is or is not a relationship between an incomplete variable and the probability of missingness on that variable requires knowledge of the missing values). Consequently, the statistical and ethical quality of a missing data analysis ultimately relies on the credibility of one or more untestable assumptions, and the onus is on the researcher to choose and defend a particular set of assumptions. We explore the ethical ramifications of different analytic choices in more detail later in the chapter.

Missing Data Techniques

A brief description of common analytic approaches is necessary before addressing ethical issues. Space limitations preclude a comprehensive overview of missing data handling options, so the subsequent sections describe techniques that are used with some regularity in the social and the behavioral sciences. Throughout the chapter, we make the argument that the ethical quality of a particular analysis is linked to the credibility of its assumptions, so this section organizes the techniques according to their assumptions about the missing data mechanism. The following descriptions are necessarily brief, but a number of resources are available to readers who want additional details (Allison, 2002; Enders, 2010; Graham, 2009; Little & Rubin, 2002; Schafer, 1997; Schafer & Graham, 2002).

Atheoretical Methods

The group of atheoretical missing data handling procedures includes methods that are known to produce biases under any missing data mechanism or do not have a theoretical foundation that dictates their expected performance. This category includes many of the ad hoc solutions that have appeared in the literature over the past several decades, at least three of which have enjoyed widespread use: mean imputation, last observation carried forward, and averaging the available items. Mean imputation replaces missing values on a variable with the arithmetic average of the complete observations. This method is among the worst approaches available because it severely distorts estimates of variation and association under any missing data mechanism. Last observation carried forward is an imputation procedure for longitudinal data that replaces missing repeated measures variables with the observation that immediately precedes dropout. This is one of the most widely used imputation techniques in medical studies and in clinical trials (Wood, White, & Thompson, 2004), despite the fact that the procedure is prone to bias, even under an MCAR mechanism (Molenberghs et al., 2004). Finally, in the context of item-level missing data on questionnaires, researchers in the social and behavioral sciences routinely compute scale scores by averaging the available item responses. For example, if a respondent answered 13 of 15 items on a one-dimensional scale, the scale score would be the average of the 13 complete items. We include this procedure in the atheoretical category because the methodological literature has yet to establish the conditions under which the procedure may or may not work. The lack of empirical support for this procedure is troubling, given that it is a widely used method for handling item-level missing data (Schafer & Graham, 2002).

MCAR-Based Methods

The category of MCAR-based analyses includes three common approaches: listwise deletion, pairwise deletion, and regression imputation. Listwise deletion removes cases with missing data from consideration, whereas pairwise deletion discards cases on an analysis-by-analysis basis. Regression imputation takes the different tack of filling in missing values with predicted scores from a regression equation (this method is unbiased under MCAR, but only after applying adjustments to variance and covariance terms). Listwise and pairwise deletion are perhaps the most widely used missing data handling techniques in the social and behavioral sciences (Peugh & Enders, 2004), most likely because of their widespread implementation in computer software packages. Considered as a whole, MCAR-based methods are virtually never better than MAR-based approaches, even when the MCAR mechanism is plausible (e.g., because they lack power). For this reason, we argue that these techniques detract from the ethical quality of a study.

MAR-Based Methods

Maximum likelihood estimation and *multiple imputation* are the principal MAR-based missing data handling procedures. Maximum likelihood estimation (also referred to as *direct maximum likelihood* and *full information maximum likelihood*) uses an iterative algorithm to audition different combinations of population parameter values until it identifies the particular set of values that produce the best fit to the data (i.e., the highest log likelihood value). The estimation process is largely the same with or without missing data, except that missing data estimation does not require individuals to contribute a full set of scores. Rather, the estimation algorithm uses all the available data to identify the most probable population parameters. Importantly, the estimation process does not impute missing values during this process. On the other hand, multiple imputation is a three-step process that consists of an imputation phase, an analysis phase, and a pooling phase. The purpose of the imputation phase is to create multiple copies of the data set, each of which contains different estimates of the missing values. The imputation phase is essentially a regression-based procedure where the complete variables predict the incomplete variables. In the analysis phase, the researcher performs each analysis m times, once for each imputed data set. The analysis phase yields m sets of parameter estimates and standard errors that are subsequently combined into a single set of results in the pooling phase. Relative to MCAR-based analysis methods, maximum likelihood and multiple imputation are desirable because they yield unbiased estimates under either an MCAR or MAR mechanism. Even when the MCAR mechanism is plausible, MAR-based analyses are still superior because they maximize statistical power.

MNAR-Based Methods

MNAR-based analyses simultaneously incorporate a model for the data (i.e., the analysis that would have been performed had the data been complete) and a model for the propensity for missing values. The *selection model* and the *pattern mixture model* are the two well-established frameworks for performing MNAR-based analyses. The selection model is a two-part model that combines the substantive analysis with an additional set of regression equations that predict the response probabilities for each incomplete outcome variable. For example, in a linear growth curve analysis, the selection part of the model consists of a set of logistic regressions that describe the probability of response at each measurement occasion. In the logistic model, each incomplete outcome variable has a binary missing data indicator, and the probability of response at wave t depends on the outcome variable at time t and the outcome variable from the previous data collection wave. Simultaneously estimating the two parts of the model adjusts the substantive model parameters to account for the MNAR mechanism.

The basic idea behind a pattern mixture model is to form subgroups of cases that share the same missing data pattern and to estimate the substantive model within each pattern. Doing so yields pattern-specific estimates of the model parameters, and computing the weighted average of these estimates yields a single set of results that appropriately adjusts for an MNAR mechanism. For example, to apply the pattern mixture model to a linear growth curve analysis, the growth model parameters are first estimated separately for each missing data pattern, and the pattern-specific parameter values are subsequently averaged into a single set of estimates. Conceptually, the pattern mixture model is akin to a multiple group model, where the missing data patterns define the subgroups (e.g., in a longitudinal study, cases with two waves of data form a group, cases with three waves of data form a group, etc.).

Despite their intuitive appeal, MNAR models require potentially tenuous assumptions that go beyond the missing data mechanism. Among other things, the selection model relies heavily on multivariate normality, and even modest departures from normality can severely distort the estimates from the substantive portion of the model. In the case of a pattern mixture model, certain pattern-specific estimates are usually inestimable. For example, in a linear growth curve model, it is impossible to estimate a growth trajectory for a pattern of cases with a single observation, and it is impossible to estimate certain variance components for a pattern with two observations. Consequently, the researcher must specify values for all inestimable parameters in the model. Again, the final estimates are prone to substantial bias if these user-specified values are incorrect. It is worth noting that MNAR-based analysis models have received considerable attention in the methodological literature. Methodologists have proposed methods for addressing the practical limitations of these models (e.g., approaches for generating values for the inestimable parameters in a pattern mixture model), but these methods have not been thoroughly studied. Until further research accumulates, MNAR-based analysis techniques should be viewed with some caution.

Ethical Issues Related to Design and Data Collection

Having established some necessary background information, this section explores ethical considerations that arise during the design and data collection phases of a study. When faced with the prospect of missing data, it may seem that a researcher's primary goal is to do damage control by minimizing the negative consequences to the study. This is largely true in situations where missingness is beyond the researcher's control, and attending to missing data issues before and during data collection can mitigate damage. Although it may seem counterintuitive to do so, researchers can also incorporate intentional missing data into the data collection design. These so-called planning missingness designs can bolster the ethical quality of a study by reducing its costs and maintaining its utility. The subsequent sections describe these issues in more detail.

An Ounce of Prevention

Obviously, the best approach to dealing with missing data is to avoid the problem altogether. For the purposes of this chapter, it is more interesting to explore the ethical issues that arise from missing data, so we chose not to focus on prevention strategies (a detailed discussion of this topic could easily fill an entire chapter by itself). Nevertheless, it is important to note that researchers have developed a variety of techniques for reducing attrition, and there is substantial literature available on this topic. Some of these retention strategies are specific to particular disciplines (e.g., Bernhard et al., 1998), whereas others are quite general. For a detailed review of retention and tracking strategies, we recommend that readers consult Ribisl et al. (1996) and the references contained therein. For a general discussion of design-based strategies for preventing missing data, interested readers can also refer to Chapter 4 in McKnight, McKnight, Sidani, and Figueredo (2007).

The Role of Auxiliary Variables

When performing missing data analyses, methodologists frequently recommend a so-called *inclusive analysis strategy* that incorporates a number of auxiliary variables into the analysis (Collins, Schafer, & Kam, 2001; Enders, 2010; Graham, 2003, 2009; Rubin, 1996; Schafer & Graham, 2002). An *auxiliary variable* is a variable that would not have appeared in the analysis model had the data been complete but is included in the analysis because it is a potential correlate of missingness or a correlate of an incomplete variable. Auxiliary variables are beneficial because they can reduce bias (e.g., by making the MAR assumption more plausible) and can improve power (e.g., by recapturing some of the lost information). For these reasons, incorporating auxiliary variables into an MAR-based analysis is a useful strategy for mitigating the negative impact of missing data.

To illustrate the idea behind an inclusive analysis strategy, consider a health study where researchers restrict the administration of a sensitive sexual behavior questionnaire to participants who are above the age of 15. An analysis of the sexual behavior variable would only satisfy the MAR assumption if age (the cause of missingness) is part of the statistical model. Omitting age from the model would likely distort the estimates, whereas incorporating age into the analysis as an auxiliary variable would completely eliminate nonresponse bias (assuming that age was the only determinant of missingness). Correlates of the incomplete analysis model variables are also useful auxiliary variables, regardless of whether they are also related to missingness. For example, a survey question that asks teenagers to report whether they have a steady boyfriend or girlfriend is a useful auxiliary variable because it is correlated with the incomplete sexual activity scores. Introducing correlates of the incomplete variables as auxiliary variables may or may not reduce bias, but doing so can improve power by recapturing some of the missing information.

The previous health study is straightforward because the researchers control the missing data mechanism and the cause of missingness is a measured variable. In most realistic situations, the missing data mechanism is unknown, and the true causes of missingness are unlikely to be measured variables. Consequently, implementing an effective inclusive analysis strategy requires proactive planning to ensure that the data include potential correlates of missingness and correlates of the variables that are likely to have missing data. Identifying correlates of incomplete variables is relatively straightforward (e.g., via a literature review), but selecting correlates of missingness usually involves educated guesswork (see Enders, 2010, for additional details). When all else fails, asking participants about their intentions to complete the study is also a possibility. For example, in the context of a longitudinal design, Schafer and Graham (2002) recommend using a survey question that asks respondents to rate their likelihood of dropping out of the study before the next wave of data collection. These authors argue that incorporating auxiliary variables such as this into the analysis "may effectively convert an MNAR situation to MAR" (p. 173). Given the potential pitfalls associated with MNAR models, taking proactive steps to satisfy the MAR assumption may be the best way to maximize the quality of the analysis.

Despite the potential benefits of doing so, collecting a set of auxiliary variables raises ethical concerns. In particular, adding variables to a study increases respondent burden, requires participants to devote more time to the study, and generally increases the potential for unintended negative consequences. The impact on participants is an important consideration in and of itself, but collecting additional variables may also affect the overall integrity of the data by inducing fatigue or reducing motivation. Of course, increasing respondent burden can increase the probability of attrition, which defeats the purpose of collecting auxiliary variables. The conventional advice from the missing data literature is to incorporate a comprehensive set of auxiliary variables (Rubin, 1996), but there is clearly a need to balance statistical issues with practical and ethical concerns. Establishing definitive guidelines for the size of the auxiliary variable set is difficult because the costs and benefits associated with these additional variables will likely vary across studies. Nevertheless, the most useful auxiliary variables are those that have strong correlations with the incomplete analysis variables. For example, incorporating a single pretest score as an auxiliary variable is often more beneficial than using several variables with weak to moderate correlations. Consequently, identifying a small set of auxiliary variables that are likely to maximize the squared multiple correlation with the incomplete variables is often a good strategy.

Documenting the Reasons for Missing Data

Researchers often view missing data as an analytic problem that they can address after the data are collected. However, documenting the reasons for attrition during the data collection phase is an important activity that can bolster the ethical quality of a study by making the subsequent analytic choices more defensible. Later in the chapter we propose an ethical continuum that differentiates missing data handling techniques according to the quality of the estimates that they produce. These classifications are inherently subjective because the data provide no mechanism for choosing between MAR and MNAR analysis models. Ultimately, researchers have to weigh the credibility of different untestable assumptions when choosing among missing data handling techniques. Defending analytic choices is difficult without knowing why the data are missing, so devoting resources to tracking the causes of attrition is important. Documenting and reporting the causes of missingness is also important for planning future studies because the information can help guide the selection of effective auxiliary variables (e.g., if a school-based study reports that student mobility is a common cause of attrition, then future studies might include a survey question that asks parents how likely they are to move during the course of the school year). Ultimately, this may improve the overall quality of scientific research by converting some MNAR analyses to MAR. Of course, it is usually impossible to fully document the real world causes of missing data, but this information is still highly valuable, even if it is incomplete or partially speculative. Interested readers can consult Enders, Dietz, Montague, and Dixon (2006) and Graham, Hofer, Donaldson, MacKinnon, and Schafer (1997) for examples of longitudinal studies that tracked and reported the sources of attrition.

Planned Missing Data Designs

Much of this chapter is concerned with ethical issues related to unintentional missing data. The development of MAR-based analysis techniques has made planned missing data research designs a possibility. The idea of intentional missing data may sound absurd, but planned missingness designs solve important practical problems, and they do so without compromising a study's integrity. In particular, these designs can cut research costs (e.g., money, time, resources) and can reduce respondent burden. Rosenthal and Rosnow (1984) and Rosenthal (1994) argue that research studies that minimize costs (e.g., those that require fewer resources and reduce respondent burden) are more ethically defensible than studies with high costs, so the use of intentional missing data can actually improve the ethical quality of a study (see also Rosnow & Rosenthal, Chapter 3, this volume). This section provides a brief description of planned missing data designs, and interested readers can consult Graham, Taylor, Olchowski, and Cumsille (2006) for additional details.

To illustrate the issue of reducing costs and resources, consider a schizophrenia study where a researcher wants to use magnetic resonance imaging (MRI) to collect neuroimaging data. In this scenario, data collection is both expensive and potentially inconvenient for study participants (e.g., because the researcher may only have access to the MRI facility during off-peak hours). To reduce costs, the researcher could obtain less expensive computed tomography (CT) scan data from all subjects and could restrict the MRI data to a random subsample of participants. As a second example, consider an obesity prevention study that uses body composition as an outcome variable. Using calipers to take skinfold measurements is a widely used and inexpensive approach for measuring body fat. However, caliper measurements are often unreliable, so a better approach is to use hydrostatic (i.e., underwater) weighing or air displacement. Like the MRI, hydrostatic weighing and air displacement measures are expensive, and the equipment is difficult to access. In a planned missing data design, the prevention researchers could use calipers to collect body composition data from the entire sample and could restrict the more expensive measures to a subset of participants. Importantly, MAR-based analysis methods allow the researchers from the previous examples to use the entire sample to estimate the associations between the expensive measures and other study variables, even though a subset of the sample has missing data. If the researchers simultaneously incorporate the inexpensive measures (e.g., the CT scan data and the caliper measurements) into the analysis as auxiliary variables, the reduction in power resulting from missing data may be minimal.

Next, consider the issue of respondent burden. In Rosenthal and Rosnow's (1984) cost–utility framework, respondent burden would be one of the costs associated with doing research. Consequently, studies that minimize respondent burden are more ethically defensible than studies that impose a high burden. The previous scenarios illustrate how planned missing data can reduce respondent burden (e.g., by reducing the number of subjects who need to undergo an MRI during off-peak hours), but there are other important examples. For instance, researchers in the social and behavioral sciences routinely use multiple-item questionnaires to measure constructs (e.g., psychologists use several questionnaire items to measure depression, each of which taps into a different depressive symptom). Using multiple-item scales to measure even a small number of constructs can introduce a substantial time burden. Obviously, reducing the number of questionnaires or reducing the number of items on each questionnaire can mitigate the problem, but these strategies may be undesirable because they can limit a study's scope and can reduce the content validity of the resulting scale scores. Planned missing data designs are an excellent alternative. In the context of questionnaire research, planned missingness designs distribute questionnaire items (or entire questionnaires) across different forms, such that any single participant responds to a subset of the items. Again, it is important to note that MAR-based methods allow researchers to perform their analyses as though the data were complete. Consequently, these designs reduce respondent burden without limiting the scope of the study or the content validity of the scale scores.

Respondent burden is also a serious problem in longitudinal studies. Graham, Taylor, and Cumsille (2001) describe a number of planned missingness designs for longitudinal studies. The basic idea behind these designs is to divide the sample into a number of random subgroups, each of which has a different missing data pattern. For example, in a study with six data collection waves, one subgroup may have intentional missing data at the third wave; another subgroup may be missing at the fifth wave; and yet another subgroup may have missing values at the second and the fourth waves. Interestingly, Graham et al. show that longitudinal planned missingness designs can achieve higher power than complete-data designs that use the same number of data points. This finding has important implications for maximizing the resources in a longitudinal study. For example, in a five-wave study with a budget that allows for 1,000 total assessments, collecting incomplete data from a sample of 230 respondents can produce higher power than collecting complete data from a sample of 200 respondents.

Researchers are sometimes skeptical of planned missing data designs, presumably because they hold the belief that missing data are harmful and something to avoid. It is important to emphasize that planned missingness designs produce MCAR data, so the intentional missing values that result from these designs are completely benign and are incapable of introducing bias. The primary downside of these designs is a reduction in statistical power. However, empirical studies suggest that the loss in power may be rather small (Enders, 2010; Graham et al., 2001, 2006), and researchers can mitigate this problem by carefully deciding which variables to make missing (preliminary computer simulations are particularly useful in this regard). Given their potential benefits, planned missing data designs may be an ethical imperative, particularly for high-cost studies.

Ethical Issues Related to Data Analysis

During the analysis phase, researchers have to make a number of important decisions, the most obvious being the choice of analytic technique. Later in this section, we explore quality differences among missing data handling techniques and propose an ethical continuum that ranks analytic methods according to the quality of the estimates that they produce. This section also explores a number of other analytic issues that can impact the quality of a research study.

How Much Missing Data Is Too Much?

One question that often arises with missing data is, "How much is too much?" A recent report by the American Psychological Association (APA) speculated that publishing missing data rates in journal articles will prompt researchers to "begin considering more concretely what acceptable levels of attrition are" (APA Publications and Communications Board Work Group on Journal Article Reporting Standards, 2008, p. 849). Establishing useful cutoffs for an acceptable level of attrition is difficult because it is the missing data mechanism that largely dictates the performance of an analytic method, not the percentage of missing data. In truth, the missing data rate may not be that important, provided that underlying assumptions are satisfied. As an example, some planned missing data designs (e.g., the three-form design) produce a 67% missing data rate for certain pairs of variables. This seemingly alarming amount of missing data causes no problems because the data are MCAR, by definition. Using MAR-based methods to analyze the data from such a design can produce unbiased parameter estimates with surprisingly little loss in power (Enders, 2010; Graham et al., 2006).

To be fair, high missing data rates (or even small to moderate missing data rates, for that matter) can be detrimental when the missing data mechanism is beyond the researcher's control, as it typically is. For example, most researchers would be uncomfortable with 67% attrition in a longitudinal clinical trial, and rightfully so. Here again, the missing data mechanism is the problem, not the attrition rate per se. If the reasons for missingness are largely unrelated to the outcome variable after controlling for other variables in the analysis (i.e., the mechanism is MAR), then the resulting parameter estimates should be accurate, albeit somewhat noisy. However, if missingness is systematically related to the outcome variable (i.e., the mechanism is MNAR), then the parameter estimates may be distorted. Unfortunately, when the reasons for missingness are beyond the researcher's control, it is impossible to use the observed data to differentiate these two scenarios, so there is usually no way to gauge the impact of the missing data rate on the validity of a study's results.

Given that the missing data mechanism is usually unknown, determining what is and is not an acceptable level of attrition becomes a bit of an arbitrary exercise. Nevertheless, journal editors and reviewers do impose their own subjective criteria when evaluating manuscripts. As an example, a former student recently contacted me for advice on dealing with a manuscript revision where 90% of the sample had dropped out by the third and final wave of data collection. Not surprisingly, the journal editor and the reviewers voiced legitimate concerns about missing data. In this situation, assuaging the reviewers that the missing data pose no problem is impossible because the missing data mechanism is unknown. This scenario raises an interesting ethical question: In light of extreme attrition, is it better to report a potentially flawed set of results, or is it better to discard the data altogether? The word "potentially" is important because a high missing data rate does not necessarily invalidate or bias the analysis results.

Rosenthal and Rosnow's (1984) cost–utility framework is useful for considering the ethical ramifications of abandoning analyses that suffer from serious attrition problems. The basic premise of the cost–utility model is that the decision to conduct a study depends on the cost–utility ratio of doing the research (e.g., conducting a study with high costs and low utility is indefensible) and the cost–utility ratio of *not* doing the research (e.g., failing to conduct a study that may produce positive outcomes may be unethical). When considering the ethics of conducting a study, Rosenthal and Rosnow argue that "The failure to conduct a study that could be conducted is as much an act to be evaluated on ethical grounds as is the conducting of a study" (p. 562). Applying this idea to missing data, it is reasonable to argue that the failure to report the results from a study with high attrition is as much an act to be evaluated on ethical grounds as is the reporting of such results. We suspect that editors and reviewers generally consider only one side of the ethical issue, the costs of reporting results that are potentially distorted by missing data. However, the costs of discarding the data are not necessarily trivial and are also important to consider. Among other things, these costs include (a) the loss of all potential benefits (e.g., new knowledge, positive outcomes) from the study, (b) the waste of time, resources, and money that accrued from conducting the study, and (c) the fact that any negative impact that the study might have had on participants was for naught.

As an aside, researchers sometimes believe that MAR-based analyses work as long as the missing data rate falls below some critical threshold. The thought is, if the proportion of missing data exceeds this threshold, MAR methods become untrustworthy and ad hoc missing data handling approaches (e.g., MCAR-based methods) provide more accurate results. To be clear, this view is not supported by the methodological literature.

Simulation studies have repeatedly shown that the advantages of using MAR-based approaches over MCAR-based methods increase as the missing data rate increases. With small amounts of missing data, the differences between competing methods tend to be relatively small, but the relative benefits of MAR methods increase as the proportion of missing data increases. Consequently, there is no support for the notion that high missing data rates are a prescription for avoiding MAR-based methods in favor of more "conservative" traditional approaches.

Imputation Is Not Just Making Up Data

Researchers sometimes object to imputation, presumably because they equate it to the unethical practice of fabricating data. For example, in the decision letter to my former student, the journal editor stated, "I have never been a fan of imputation." This type of cynicism is largely valid for single imputation (e.g., mean imputation, regression imputation, last observation carried forward, averaging the available items) techniques because filling in the data with a single set of values and treating those values as though they are real data produces standard errors that are inappropriately small. Of course, the other problem with most single imputation procedures is that they tend to produce biased parameter estimates, irrespective of their standard errors.

Importantly, MAR-based multiple imputation does not suffer from these problems because it (a) has a strong theoretical foundation, (b) produces accurate estimates under an MCAR and MAR mechanism, and (c) incorporates a correction factor that appropriately adjusts standard errors to compensate for the uncertainty associated with the imputed values. From a mathematical perspective, it is important to realize that multiple imputation and maximum likelihood estimation are asymptotically (i.e., in large samples) equivalent procedures. Maximum likelihood estimation produces estimates that effectively average over an infinite number of imputed data sets, although it does so without filling in the values. Multiple imputation uses a simulation-based approach that repeatedly fills in the missing data to accomplish the same goal. Some of the objections to imputation may stem from the fact that researchers place undue emphasis on the filled-in data values without considering the fact that the data set is just a means to a more important end, which is to estimate the population parameters. In truth, multiple imputation is nothing more than a mathematical tool for achieving this end goal. In that sense, it is the final parameter estimates that matter, not the imputed values themselves.

Revisiting an Inclusive Analysis Strategy

Earlier in the chapter we described an inclusive analysis strategy that incorporates auxiliary variables (correlates of missingness or correlates of the incomplete analysis variables) into the statistical model. In line with the premise that research quality and ethics are linked, we believe that MAR-based analyses that incorporate auxiliary variables are more ethically defensible than analyses that do not. For one, an inclusive analysis strategy is more likely to satisfy the MAR assumption, thereby reducing the potential for bias. Auxiliary variables can also mitigate the power loss resulting from missing data, thereby maximizing resources and reducing costs. As an example, Baraldi and Enders (2010) used data from the Longitudinal Study of American Youth to illustrate the impact of auxiliary variables. In their analysis, including three useful auxiliary variables in a regression model reduced standard errors by an amount that was commensurate with increasing the total sample size by 12% to 18% (the magnitude of the reduction in standard errors varied across regression coefficients). From a cost–utility perspective, there is no question that an inclusive analysis strategy is desirable because it maximizes existing resources, whereas collecting more data requires additional costs (e.g., time, money, risks to participants). The standard error reduction in the Baraldi and Enders (2010) study is probably close to the upper limit of what would be expected in practice, but even a modest improvement (e.g., a reduction in standard errors that is commensurate with a 5% increase in the sample size) supports our argument.

As an aside, it is possible for the same analysis to produce different estimates with and without the auxiliary variables. When this happens, there is no way of knowing which set of estimates is more accurate, but two points are worth remembering: The auxiliary variable analysis has the most defensible set of assumptions (i.e., it is more likely to satisfy MAR), and methodological studies have yet to identify detrimental effects of an inclusive strategy (e.g., including a large set of useless auxiliary variables does not appear to negatively impact the resulting estimates and standard errors; Collins et al., 2001). These two factors clearly favor the estimates from the auxiliary variable model, but ethical issues can arise if researchers are tempted to choose the analysis results that best align with their substantive hypotheses. To avoid this ethical pitfall, researchers should disclose the fact that the two analyses produced conflicting results, perhaps reporting the estimates from the alternate analysis in a footnote or in a supplementary appendix in the electronic version of the manuscript.

An Ethical Continuum of Analysis Options

Researchers have a variety of options for analyzing data sets with missing values. Earlier in the chapter, we described four categories of missing data handling procedures: atheoretical methods, MCAR-based methods, MAR-based methods, and MNAR-based methods. In this section, we propose an ethical continuum that differentiates missing data handling techniques according to the quality of the estimates that they produce. On one end, the continuum is anchored by low-quality approaches that are difficult to defend on ethical grounds, whereas the other end of the continuum is defined by defensible approaches that have a strong theoretical foundation. When comparing certain categories of methods, there are distinct and consistent quality differences that are difficult to dispute (e.g., there is little question that MAR-based procedures are more defensible than MCAR-based analyses). However, differentiating among techniques that rely on one or more untestable assumptions is a subjective exercise. Consequently, some readers will disagree with certain aspects of our proposed continuum, and rightfully so. In proposing the ethical continuum, it is not our intent to form rigid distinctions that cast a negative light on certain analytic choices. Quite the contrary, choosing among the theory-based approaches at the high end of the quality continuum requires researchers to judge the credibility of different sets of assumptions. The veracity of these assumptions will vary across situations, so the ordering of certain procedures is fluid.

Figure 14.1 shows a graphic of our proposed continuum. The low-quality end of the continuum is anchored by the collection of atheoretical analysis techniques. This group of procedures includes missing data handling procedures that (a) are known to produce biases under any missing data mechanism, (b) do not have a theoretical framework that dictates their expected performance, or (c) lack empirical evidence supporting their widespread use. It is worth noting that the low-quality endpoint includes at least three procedures that enjoy widespread use (mean imputation, last observation carried forward, and averaging the available items). As seen in the figure, MCAR-based approaches provide an improvement in quality. MCAR methods require a rather strict assumption about the cause of missing data (i.e., the propensity for missing data is unrelated to all study variables), but the situations where these techniques produce accurate parameter estimates are well established. However, even if the MCAR mechanism is plausible, MAR-based analyses generally increase

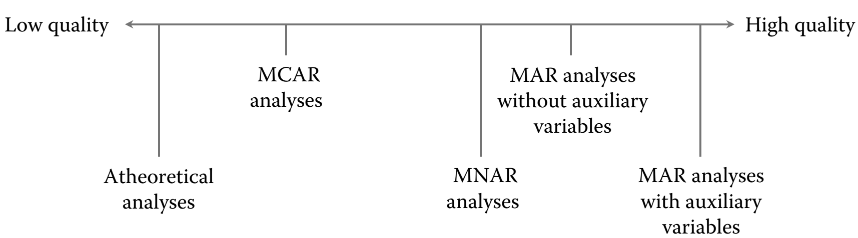

FIGURE 14.1
An ethical continuum of missing data handling techniques.

statistical power, thus making better use of available resources. This alone provides a strong ethical justification for abandoning MCAR-based missing data handling methods.

Assessing the relative quality of MAR- and MNAR-based analysis techniques is less clear-cut because the accuracy of these procedures relies on one or more untestable assumptions. To be clear, all the procedures at the high-quality end of the continuum are capable of producing unbiased parameter estimates when their requisite assumptions hold. Although some readers will likely disagree, we believe that the range of conditions that satisfies the assumptions of an MNAR-based analysis will generally be narrower than the range of conditions that satisfies the assumptions of an MAR-based analysis. Consequently, we assigned a slightly higher quality ranking to MAR-based analyses (i.e., multiple imputation and maximum likelihood estimation). In part, our rationale was based on the fact that MNAR models require assumptions that go beyond the missing data mechanism. For example, Enders (2010) gives an analysis example where a modest departure from normality causes a selection model to produce estimates that are less accurate than those of maximum likelihood estimation, despite the fact that the selection model perfectly explains the MNAR missing data mechanism. Pattern mixture models rely on equally tenuous assumptions because they require the analyst to specify values for one or more unknown parameters.

The ethical continuum assigns the highest-quality rating to MAR-based analyses that incorporate auxiliary variables. Although this choice will not be met with universal agreement, we believe that a well-executed MAR analysis generally has the most defensible assumptions, even when there is reason to believe that dropout is systematically related to the incomplete outcome variable. Other methodologists have voiced a similar opinion. For example, Schafer (2003, p. 30) discussed the tradeoffs between MAR and MNAR analysis models, stating that "Rather than rely heavily on poorly estimated MNAR models, I would prefer to examine auxiliary variables that may be related to missingness … and include them in a richer imputation model under assumption of MAR." Similarly, Demirtas and Schafer (2003, p. 2573) stated that "The best way to handle drop-out is to make it ignorable [i.e., consistent with an MAR mechanism]." They go on to recommend that researchers should collect data on variables that predict attrition and incorporate these variables into their analyses.

Again, it is important to reiterate that some of the classifications in Figure 14.1 are subjective and open to debate. Ultimately, the data provide no mechanism for choosing between MAR and MNAR analyses, so researchers have to weigh the credibility of different sets of untestable assumptions. Missing data techniques are only as good as the veracity of their assumptions, so adopting a defensible analysis that minimizes the risk of violating key assumptions will maximize the ethical quality of a research study. The need to defend analytic choices has important implications for data collection (e.g., documenting the reasons for missingness, planning for attrition by collecting data on auxiliary variables) and for reporting the results from a missing data analysis. We address the latter topic in a subsequent section.

Sensitivity Analyses

The purpose of a sensitivity analysis is to explore the variability of a parameter estimate across models that apply different assumptions. For example, in longitudinal studies, methodologists often recommend fitting MAR- and MNAR-based growth models to the same data. This strategy seems eminently sensible given the difficulty of defending a set of untestable assumptions. If the key parameter estimates are stable across different missing data models, then the choice of analytic procedure makes very little difference. Unfortunately, it is relatively common for *sensitivity analyses* to produce discrepant sets of estimates. For example, Enders (2010) used an artificial data set to illustrate a sensitivity analysis for a linear growth model with a binary treatment status variable as a predictor. None of the five analysis models accurately reproduced the true parameter estimates, and the estimates varied dramatically across models (e.g., the MAR-based growth model and the MNAR-based selection model underestimated the true effect size, whereas MNAR-based pattern mixture models overestimated the true effect). Other methodologists have reported similar discrepancies from sensitivity analyses (Demirtas & Schafer, 2003; Foster & Fang, 2004).

In many situations, sensitivity analyses add no clarity to analytic choices, and researchers are left to decide among potentially disparate sets of results. Unfortunately, the data provide no mechanism for choosing among competing analyses, and two models that produce different estimates can produce comparable fit. Ideally, researchers should choose the estimates from the model with the most defensible set of assumptions, but it may be tempting to adopt the estimates that best align with the substantive hypotheses. Sensitivity analyses with longitudinal data seem particularly prone to this ethical dilemma because estimates can vary so dramatically from one model to the next. To avoid this ethical dilemma, researchers should present an argument that supports their assumptions and should disclose the fact that estimates differed across analysis models. Ideally, the estimates from alternate analyses could appear in a footnote or in a supplementary appendix in the electronic version of the manuscript.

Ethical Issues Related to Reporting

A final set of ethical concerns arises when reporting the results from a missing data analysis. Rosenthal (1994) describes a number of ethical issues related to research reporting, most of which involve misrepresentation of research findings (e.g., inappropriately generalizing, making claims that are not supported by the data, failing to report findings that contradict expectations). In the context of a missing data analysis, two additional forms of misrepresentation are problematic: providing an insufficient level of detail about the missing data and the treatment of missing data, and overstating the benefits of a missing data handling technique. This section explores these two issues in detail.

Reporting Standards for Missing Data Analyses

A 1999 report by the APA Task Force on Statistical Inference encouraged authors to report unanticipated complications that arise during the course of a study, including "missing data, attrition, and nonresponse" (Wilkinson & Task Force on Statistical Inference, 1999, p. 597). At the time of the Task Force report, missing data reporting practices were abysmal, and many published research studies failed to report any information about missing data. In a comprehensive methodological review, Peugh and Enders (2004) examined hundreds of published articles in the 1999 and 2003 volumes of several education and psychology journals. In the 1999 volumes, approximately one third of the articles with detectable missing data (e.g., studies where the degrees of freedom values unexpectedly changed across a set of analyses of variance) explicitly acknowledged the problem. Whether it was a result of the Task Force report or a general increase in awareness of missing data issues, reporting practices improved in the 2003 volumes, such that three quarters of the studies with detectable missing data disclosed the problem. Obviously, failing to report any information about missing data is a gross misrepresentation, regardless of intent (to be fair, the researchers that authored the papers in the review were probably unaware that missing data pose a problem). Missing data reporting practices have arguably progressed since 1999, but there is still room for improvement.

Recently, several organizations have published detailed guidelines aimed at improving reporting practices in scientific journals. In the social sciences, the American Educational Research Association (2006) published the *Standards for Reporting on Empirical Social Science Research in AERA Publications*, and APA published *Reporting Standards for Research in Psychology: Why Do We Need Them? What Might They Be?* (APA Publications and Communications Board Work Group on Journal Article Reporting

Standards, 2008). Similar reports have appeared in the medical and clinical trials literature (Altman et al., 2001; Des Jarlais, Lyles, Crepaz, & the TREND Group, 2004; Moher, Schulz, & Altman, 2001). Although these reports have a general focus, they do provide specific recommendations for dealing with missing data. The APA Journal Article Reporting Standards (JARS) report provides the most comprehensive recommendations concerning missing data, so we briefly summarize its main points here.

The JARS report recommends that researchers describe (a) the percentage of missing data, (b) empirical evidence or theoretical arguments in support of a particular missing data mechanism, (c) the missing data handling technique that was used for the analyses, and (d) the number and characteristics of any cases that were deleted from the analyses. Following guidelines from the clinical trials literature (the Consolidated Standards of Reporting Trials, or CONSORT statement; Moher et al., 2001), the JARS report recommends a diagrammatic flowchart that, among other things, describes the amount of and the reasons for missing data at each stage of the research process (see p. 846). We believe that the JARS recommendations are adequate for most studies, but some analyses may require additional details (e.g., planned missing data designs).

Given the rather abysmal state of missing data reporting practices, it is hard to argue against the need for more detailed research reports. Nevertheless, devoting additional space to missing data issues decreases the amount of journal space that is available for reporting substantive results. In some situations, satisfying the JARS recommendations requires relatively little journal space, whereas other situations are more demanding. For example, a thorough description of a multiple imputation procedure may be somewhat lengthy because it involves many nuances and subjective choices. Similarly, planned missing data designs often require preliminary computer simulations to assess the power of different missing data patterns, and describing these preliminary analyses may require an excessive amount of journal space. As a compromise, researchers may want to rely more heavily on electronic resources to convey the procedural details of their missing data analyses. In situations where the missing data handling procedure is very involved, the printed version of the manuscript could include a brief description of the core analytic details, and the electronic version could include an appendix that documents the analytic steps in more elaborate detail.

Overstating the Benefits of an Analytic Technique

MAR and MNAR analyses are sometimes met with skepticism because they are relatively new to many disciplines. When faced with the prospect of "selling" an unfamiliar missing data technique, a natural inclination is to provide a detailed description of the analysis along with empirical evidence that supports its use (e.g., references to computer simulation studies that demonstrate the procedure's efficacy). However, many researchers are unfamiliar with Rubin's missing data mechanisms, so describing the benefits of an analysis without also describing its assumptions can mislead the reader into believing that the procedure is an analytic panacea for missing data. A similar type of misrepresentation can occur when a manuscript provides insufficient details about the missing data handling procedure. Consequently, it is important for authors to provide a thorough but balanced description of the missing data handling procedure that addresses the benefits and the assumptions of their analytic choices.

The recommendation to use honest and balanced reporting practices is unlikely to be met with objections. However, the pressure to publish in top-tier journals creates situations that are at odds with this practice. As an example, I previously described an interaction with a former student who contacted me for advice on dealing with a manuscript revision that involved a substantial amount of missing data. In the decision letter, the journal editor responded to the use of maximum likelihood estimation by saying that "I have never been a big fan of imputation." The editor's response is misguided because it incorrectly characterizes maximum likelihood estimation as an imputation technique and because it implies that imputation is an inherently flawed procedure (presumably, the editor's opinion stems from the misconception that imputation is "making up data"). Ignoring the problems associated with the high attrition rate, the editor's objection is easily addressed by describing the benefits of maximum likelihood estimation and bolstering this description with relevant citations from the methodological literature. The ethical concern is that the revised manuscript could overstate the benefits of maximum likelihood while downplaying (or completely omitting any discussion of) the untestable MAR assumption. This type of unbalanced reporting can potentially mislead readers who are unfamiliar with the intricacies of missing data analyses.

Misrepresentation can also occur when authors fail to describe their missing data handling procedures in sufficient detail. As an example, consider a hypothetical passage from a manuscript that reads, "We used maximum likelihood estimation, a missing data handling technique that the methodological literature characterizes as state of the art." Although it is true that methodologists have described MAR-based methods in this way (Schafer & Graham, 2002), this passage is misleading because it implies that maximum likelihood estimation is a cure-all for missing data problems. As a second example, consider a hypothetical passage that states, "Because there is reason to believe that the data are missing not at random (i.e., attrition is systematically related to the outcome variable), we used a selection model to correct for attrition-related bias." The lack of detail in the preceding passage is potentially misleading because it fails to inform the reader that the accuracy of the selection model heavily depends on the multivariate normality assumption—so much so that an MAR-based analysis may yield better estimates in many situations. The fact that many researchers are unfamiliar with the nuances of MAR- and MNAR-based analysis techniques magnifies ethical concerns related to lack of detail in reporting because the factors that affect the performance of a particular technique may not be widely understood.

Fortunately, ethical issues related to misrepresentation are easily avoided by following recommendations from the JARS report. In particular, the report states that manuscripts should include "evidence and/or theoretical arguments for the causes of data that are missing, for example, missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR)" (p. 843). Because the missingness mechanism largely dictates the performance of most missing data techniques, discussing the plausibility of a purported mechanism can delineate the range of conditions under which an analytic method is likely to produce accurate estimates. Further describing the importance of other assumptions (e.g., normality) reduces the chances of leaving the reader with an overly optimistic impression of the analysis.

Conclusion

The purpose of this chapter is to explore the impact of missing data on the ethical quality of a research study. Consistent with Rosenthal and colleagues (Rosenthal, 1994; Rosenthal & Rosnow, 1984), we operate from the premise that ethics is closely linked to the quality of a research study, such that high-quality studies are more ethically defensible than low-quality studies. Missing data are obviously important at the analysis phase, but ethical issues arise throughout the entire research process. Accordingly, we explore the linkage between quality and ethics at the design and data collection phase, the analysis phase, and the reporting phase.

During the design and the data collection phase, researchers should proactively plan for missing data to minimize negative consequences to the study. In particular, collecting data on auxiliary variables can make the MAR assumption more plausible, and documenting the reasons for attrition can help build an argument that supports subsequent analytic choices. Although it may seem counterintuitive to do so, researchers can also incorporate intentional missing data into the data collection design. These so-called planning missingness designs can bolster the ethical quality of a study by reducing costs and respondent burden. Given their potential benefits, planned missing data designs may be an ethical imperative, particularly for high-cost studies.

During the analysis phase, researchers have to make a number of important decisions, the most obvious being the choice of analytic technique. We propose an ethical continuum that differentiates missing data handling methods according to the quality of the estimates that they produce. MCAR-based analysis techniques are rarely justified, so the choice is usually between MAR and MNAR models. Both sets of procedures are capable of producing accurate parameter estimates when their requisite assumptions hold, but they are also prone to bias when the assumptions are violated. Because MNAR models require strict assumptions that go beyond the missing data mechanism (e.g., in the case of a selection model, multivariate normality), we argue that the range of conditions that satisfies the assumptions of an MNAR-based analysis will generally be narrower than the range of conditions that satisfies the assumptions of an MAR-based analysis. In our view, an MAR-based analysis that incorporates auxiliary variables is often the most defensible procedure, even when there is reason to believe that dropout is systematically related to the incomplete outcome variable.

Finally, we explored ethical issues related to reporting the results from a missing data analysis. Recently, several organizations have published detailed guidelines aimed at improving reporting practices in scientific journals, and these reporting guidelines generally include recommendations regarding missing data. The APA JARS report is particularly detailed and recommends that researchers describe (a) the percentage of missing data, (b) empirical evidence or theoretical arguments in support of a particular missing data mechanism, (c) the missing data handling technique that was used for the analyses, and (d) the number and characteristics of any cases that were deleted from the analyses.

In summary, maximizing the ethical quality of a study requires researchers to attend to missing data throughout the entire research process. We believe that a good MAR analysis will often lead to better estimates than an MNAR analysis. Ultimately, the data provide no mechanism for choosing between MAR and MNAR analyses, so researchers have to weigh the credibility of different sets of untestable assumptions when making this choice. Adopting a defensible analysis that minimizes the risk of violating key assumptions maximizes the ethical quality of a research study, and achieving this goal is only possible with careful planning during the design and data collection phase.

References

Allison, P. D. (2002). *Missing data*. Newbury Park, CA: Sage.

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., … Lang, T. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine, 134*, 663–694.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35*, 33–40.

APA Publications and Communications Board Work Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63*, 839–851.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5–37.

Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., … Hürny, C. (1998). Missing quality of life data in cancer clinical trials: Serious problems and challenges. *Statistics in Medicine, 17,* 517–532.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6,* 330–351.

Demirtas, H., & Schafer, J. L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine, 22*, 2553–2575.

Des Jarlais, D. C., Lyles, C., Crepaz, N., & the TREND Group. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health, 94,* 361–366.

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

Enders, C. K., Dietz, S., Montague, M., & Dixon, J. (2006). Modern alternatives for dealing with missing data in special education research. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disorders* (Vol. 19, pp. 101–130). New York: Elsevier.

Foster, E. M., & Fang, G.Y. (2004). Alternative methods for handling attrition: An illustration using data from the Fast Track evaluation. *Evaluation Review, 28*, 434–464.

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

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

Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K .J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325–366). Washington, DC: American Psychological Association.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in the analysis of change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 335–353). Washington, DC: American Psychological Association.

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

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

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

Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Annals of Internal Medicine, 134,* 657–662.

Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C., & Carroll, R. J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics, 5,* 445–464.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74,* 525–556.

Ribisl, K. M., Walton, M. A., Mowbray, C. T., Luke, D. A., Davidson, W. S., & Bootsmiller, B. J. (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations. *Evaluation and Program Planning, 19,* 1–25.

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

Rosenthal, R., & Rosnow, R. L. (1984). Applying Hamlet's question to the ethical conduct of research: A conceptual addendum. *American Psychologist, 39,* 561–563.

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

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91,* 473–489.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman & Hall.

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica, 57,* 19–35.

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

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.

Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials, 1,* 368–376.