Best practices in quantitative methods

15 How to Deal With Missing Data Conceptual Overview and Details for Implementing Two Modern Methods

Contributors: Jason C. Cole
Editors: Jason Osborne
Book Title: Best practices in quantitative methods
Chapter Title: "15 How to Deal With Missing Data Conceptual Overview and Details for Implementing Two Modern Methods"
Pub. Date: 2008
Access Date: November 12, 2015
Publishing Company: SAGE Publications, Inc.

City: Thousand Oaks
Print ISBN: 9781412940658
Online ISBN: 9781412995627
DOI: http://dx.doi.org/10.4135/9781412995627.d19
Print pages: 214-239

http://dx.doi.org/10.4135/9781412995627.d19

15 How to Deal With Missing Data *Conceptual Overview and Details for Implementing Two Modern Methods*

Jason C.Cole

In the past, when data were missing from our sets, any number of reactions were common. Positive emotions, such as happiness and contentment, never occurred. Rather, the emotions we felt (often in this order) were frustration, anger, guilt, fear, and sadness.

Graham, Cumsille, and Elek–Fisk (2003, p. 87)

We stand at the beginning of an era in which useful and accessible missing data procedures are an integral part of mainstream statistical packages.

Graham et al. (2003, p. 88)

The days when journals tolerate the absence of analysis of the missing values and the use of traditional approaches to missing values should be numbered.

Acock (2005, p. 1026)

Understanding Missingness: Why Is It Important to Address Missingness

Missing data is a prevalent issue in many fields, yet only about half of published studies mention dropouts, and less than 20% of those studies incorporated dropouts into their

analyses (Ladouceur, Gosselin, Laberge, & Blaszcynski, 2001). Nearly all common statistical analyses assume complete data, yet many statistics books do not deal with missingness. Proper missing data techniques have therefore gone mostly ignored by researchers until recently (Rubin, 1996).

Three major problems with incomplete data are (1) loss of information or power due to loss of data; (2) complication during data management and analysis, partially because of limitations with standard statistical software; and (3) potential marked bias because of systematic **[p. 215 ↓]** differences between observed and missing values (Barnard & Meng, 1999).

Recent advancements in software have made missing data analyses easier and more prolific. This is critical because gatekeepers in research, such as grant reviews, regulatory agencies, and journal reviewers, are becoming more critical of the treatment of missing data.

But when are data truly missing? Many surveys incorporate “skip patterns,” for example, where a respondent may be instructed to skip a subsequent question based on a certain response to a key beginning question in the section. Overall, missing data fall into a similar category of latent variables—true values exist, but we cannot always accurately determine the exact value (D. F. Heitjan & Rubin, 1991; Schafer & Graham, 2002).

Introduction to the Concepts of Missing Data

In this section, I introduce three key concepts to the discussion of missing data: the models (missingness augmentation and analytic), auxiliary variables, and the missingness mechanism. Effective missing data techniques will provide unbiased parameter estimates (the expected value of an estimate from the missing technique is equal to the population value) and efficient standard errors (the standard error around the estimate is small and accurate to that with known values) (Graham et al., 2003).

The Models

Many of the modern techniques for addressing missing data require a synergy between the way we augment our database to address missingness and the way we analyze our data to test our hypotheses, labeled the missingness augmentation (MA)[1] model and analytic model, respectively. Simply stated, a model description is a list of the analyzed observed variables (dependent and independent variables), their interactions, and any auxiliary variables (variables that attempt to explain why data are missing though they are superfluous to the analytic model).[2]

The analytic model drives the creation of the MA model. For example, the researcher may intend to analyze the influence of age and income (and their interaction) on the health–related quality–of-life measure (physical component summary [PCS]). With the analytic model set, we can now determine the MA model, which would include PCS, age, income, a variable to represent the interaction of Age × Income, and any necessary auxiliary variables.

It is important that combinatorial relationships (e.g., interactions) included in the analytic model must be included in the MA model. Other nonlinear relationships, such as quadratic trends over time, must be anticipated before imputation and created as part of the model because nonlinear relationships are washed out of the imputed values with standard multivariate normal multiple imputation (MI). Of course, with smaller amounts of missingness, the washout of nonlinear trends is negligible and thus may still be appropriate.

The issues of interactions highlight a more general issue: The MA model and the analytic model should be compatible. Including combinatorial information in the MA model that will not be analyzed will lead to spurious underfitting of the data, leading to poor predictive power (Barnard & Meng, 1999) and increasing the odds of a Type II error (Meng, 1994; Rubin, 1996). Meng (1994) has noted that even if an interaction may not be readily planned in the current analytic model, it should be included in the MA model if it is commonly examined within the field of research, especially if the imputer and analyst are not the same person.

Auxiliary Variables

Collins, Schafer, and Kam (2001) introduced the term *auxiliary variables* to describe variables added to an MA model solely for the purpose of improving the accuracy of the missing data method. Auxiliary variables are included in the MA model when they are potential causes or correlates of the missingness or if they are strongly correlated with the variables that have missing data, as they improve estimation of the missing data, particularly if the auxiliary variables have high correlations with the variables that have higher rates of missingness.

The Missingness Mechanism

Data may be missing for three general reasons (referred to as the missingness mechanism). In Table 15.1, we have six columns of **[p. 216 ↓]** data. The first four columns are subject ID, Health Assessment Questionnaire (HAQ) score, physical quality of life (PCS), and a mental quality–of-life composite score (MCS). We create two other variables—PCS–M and MCS–M, coded 0 for observed data and 1 for missing data—and these variables are measures of the missingness mechanism. Because not all potential causes of missingness can be known in most studies, these variables are best conceived as a mathematical procedure to describe rates and patterns of missingness as well as to capture relationships between missingness and the likely values of the missing data (Schafer & Graham, 2002). Rather than obscuring the issue as a cause of missingness, Schafer and Graham (2002) recommended the use of the term *probabilities of missingness.* These probabilities of missingness help describe the types of missingness: missing completely at random (MCAR; the probability of missingness has no correlation with our variables of interest), missing at random (MAR; the missing data process is related to the variables of interest, but that relationship is fully captured with other observed variables), and missing not at random (MNAR; the missingness mechanism is correlated with our variable of interest and cannot be fully explained by other observed variables; see Rubin, 1976).

Table 15.1 Missing Data Mechanisms

ID	*HAQ*	*PCS*	*MCS*	*PCS-M*	*MCS-M*
101	4	38	40	0	0
102	5	45	40	0	0
103	4		48	1	0
104	5	42	46	0	0
105	6		44	1	0
106	6	48	50	0	0
107	8	50	52	0	0
108	10	55		0	1
109	11	55		0	1
110	10	52		0	1

MCAR: Missing Completely at Random. The missingness mechanism is MCAR if the probability of missingness is unrelated to all observed and all unobserved variables (Abraham & Russell, 2004). When this assumption holds for all variables in the data set, the set of individuals with complete data is considered a random subsample of all participants in the database. The problem with the MCAR assumption is that understanding the relationship of missingness to all unobserved variables is untenable. There are times, however, when we have some certainty that specific missingness is MCAR (e.g., when missingness is part of the research design; Graham, Taylor, Olchowski, & Cumsille, 2006), wherein certain outcomes are randomly omitted for each participant to both reduce participant burden and still collect many outcomes.

Despite the theoretical rarity of unplanned MCAR data, MCAR allows for a major empirical benefit over the other missingness mechanism. MCAR is the only missingness mechanism that can be objectively tested (Enders, 2006a). Little's (1988) MCAR test has been adapted into SAS code and is available in the SPSS missing value analysis module. The curious reader can examine Little's description of the MCAR test for more information.

MAR: Missing at Random. MAR is present if the probability of missingness on a given variable is not related to the participant's score on that variable, after controlling for other variables in the study (Acock, 2005). Essentially, the other variables serve as an explanatory mechanism for the missingness (Collins et al., 2001) on the **[p. 217**

↓ **]** given variable. Using Table 15.1 again, this would mean that PCS–M and PCS are unrelated, once we partial out the relationship of MCS and HAQ. Thus, even if those who scored in the lower range of PCS were somewhat more likely to have missing data on PCS, the increased likelihood was fully omitted when controlling for the relationship between PCS and HAQ, PCS and MCS, or some combination thereof. It is, of course, impossible to test if the MAR assumption is satisfied because we do not know the values of the missingness on PCS (Allison, 2002). Although we can often expect deviations from MAR in many data sets (Schafer & Graham, 2002), an erroneous assumption of MAR for all data may often have only a very minor impact on estimates and standard errors according to results from simulation work (Collins et al., 2001). Indeed, when data may be classified as MAR (or MCAR), then there are marked benefits for the statistician in that one does not need to model the mechanism by which data became missing. This is why many researchers label MAR and MCAR data as ignorable. Data are said to be ignorable if they are (a) MAR or MCAR and (b) the parameters that govern the missingness mechanism are unrelated to the processes to be estimated.

Longitudinal studies are particularly susceptible to missingness. Missing data on an outcome assessed repeatedly throughout a longitudinal study often are ignorable missingness (Abraham & Russell, 2004). In fact, ignorable missingness is still valid when the likelihood of missingness depends on the actual missing values, as long as the outcome with missingness has zero residual correlations with other variables once the outcome is controlled (Schafer & Graham, 2002). This does not hold, however, when one time point does not predict missingness on the same outcome at another time point, such as in dementia studies (Harezlak, Gao, & Hui,2003).

MNAR: Missing Not at Random. MNAR missingness occurs when the likelihood of missingness depends on the actual value of the missing datum and not other variables. MNAR missing data are also referred to as *nonignorable* missing data. Modeling the missingness mechanism for nonignorable missingness typically requires highly specialized techniques that vary with each scenario (Allison, 2002). Therefore,

I have not covered MNAR techniques in detail in this chapter but instead provide references for the curious reader.

Brief History of Traditional Techniques Used to Address Missingness

A recent small survey I conducted among users of SEMNET and the APA Division 5 listserv (Cole, 2007) showed that even skilled applied statisticians use antiquated procedures for handling missing data: 55% noted using casewise deletion, 36% noted using mean imputation, and 31% used regression–based imputation. Moreover, only 35% of the sample of sophisticated users noted using missing data techniques (of any sort) “most of the time” or “all of the time.” Although many cogent reviews of the problems associated with these procedures are available (Allison, 2002; Graham et al., 2003; Schafer & Graham, 2002), I discuss them briefly below in order to stress their problematic nature in addressing missingness.

Casewise and Pairwise Deletion

Casewise deletion represents the default technique for addressing missingness in most statistical software and is therefore the most common method for addressing missing data (Abraham & Russell, 2004; Acock, 2005; Cole, 2007). This mostly comes from historical apathy for missingness. As Schafer and Olsen (1998) noted, missingness was viewed as something “to be gotten rid of” (p. 546) rather than understood.

Casewise deletion refers to a system whereby all participants with any missingness in an analysis are removed from the analysis. A related technique, pairwise deletion (otherwise known as available–case analysis), refers to a technique whereby all available participants for a specific part of an analysis are used. Detailed reviews of the serious limitations for casewise and pairwise deletion are abundant in the literature (Little & Rubin, 2002; Schafer & Graham, 2002). Primarily, disadvantages to these techniques center on assumption problems, power issues, and generalizability.

Casewise and pairwise deletion require missing data to be MCAR, an often untenable **[p. 218 ↓]** assumption when removing all of the participants with any missingness. Yet even if the data meet the MCAR assumption, the smaller sample size will increase

Type II error rates (Abraham & Russell, 2004). If the assumption of MCAR is not met, then standard errors will be inefficient, providing either conservative or liberal results without knowledge as to which has been obtained (Acock, 1989). Furthermore, removal of participants can reduce generalizability of the findings. Pairwise deletion brings other issues, particularly in advanced procedures such as structural equation modeling (Acock, 2005; Wothke, 1993).

There are a few very specific instances where the use of casewise deletion may be desired, such as when the amount of missingness is less than 5% (Fairclough, 2002; Graham & Hofer, 2000; Graham, Taylor, & Cumsille, 2001). However, overall, there are few instances when casewise deletion is desirable. Nevertheless, it is preferable to nearly all other antiquated techniques (Allison, 2002).

Mean Substitution

Mean substitution (imputation) involves imputing any missing value for a variable with the mean of all observed values on the given variable. The process is simple to implement in many statistical packages, thereby making mean imputation one of the most commonly used missing data techniques. However, mean imputation has been known to produce biased estimates of the variances and covariances for nearly 40 years (Haitovsky, 1968).

Simulation studies have consistently demonstrated the biased results from mean imputation (Wothke, 2000). First, people who answer in the middle of a distribution are more likely to answer questions (Acock, 2005); therefore, we are typically trying to impute for people who have a lower likelihood of having the values that we have ascribed to them. Second, mean imputation increasingly warps the distribution of the imputed variable as missingness increases because the imputed variable becomes increasingly leptokurtic and decreasingly variable. The decreasing variability leads to spuriously smaller correlations with other variables. Moreover, in a multivariate analysis, when one or more variables have been mean imputed, their relationships will be biased downward, but other variables with no imputation will have their relationships biased upward because of the improper attenuation of the mean–imputed variables (Acock, 1989).

The marked inefficiency and bias created by mean imputation because of distribution distortion have precluded me from recommending it under any circumstance.

Regression Imputation

In a regression model, we can regress all other independent variables on our variable with missingness to obtain a predicted value for each missing datum (Allison, 2002). For example, if we know that observed data predict depression scores on the Center for Epidemiologic Studies—Depression scale (CES–D; Radloff, 1977) based at 1.67#age + -2.8#education + -0.34#SES for observed data, then we can use these values to estimate a missing CES–D score based on the regression formula.

Regression–based imputation offers the ability to incorporate multiple indicators in the imputation of missing values, thereby enhancing the robustness of the process. Moreover, the bias involved in the point estimates from regression imputation is consistent (Enders, 2006a), thereby potentially allowing for control of the bias. These benefits, however, rarely supersede the significant problems associated with regression imputation.

Model complexity, monotone data, and a lack of efficiency and bias control are all found in regression imputation. When more than one variable contains missingness, if the data are not MCAR, or if the dependent variable is involved in the missingness mechanism, regression models more complex than standard ordinary least squares must be used (Gourieroux & Monfort, 1981). Regression imputation also demands that the data have an unusual format to the missingness, called *monotone missingness.* Imagine 10 cases on the three aforementioned predictor variables (age, education, and socioeconomic status [SES]), wherein age had the least amount of missingness, education the second least missingness, and SES the most missingness. For the missingness to be monotone, no SES data could be present when either of the variables with a lower percentage of missingness had missing data. Likewise, educational data could not be present for any instance in which age was missing (though SES could be missing). Little and Rubin **[p. 219 ↓]** (2002) provide an excellent description on monotone missingness, including the algorithmic benefits and impractical demands. Thus, regression imputation can be very time–consuming given that each

pattern of missingness requires a different predictive equation (Enders, 2006a) when nonmonotone missingness is present. Finally, as with all imputation methods that allow the analyses to treat the data as though they were complete, regression imputation contributes to underestimated standard errors and overestimated test statistics (Allison, 2002) because it reduces the variation around the regression line, leaving the imputed values with an artificially inflated correlation compared with cases with complete data.

Last Observation Carried Forward

Last observation carried forward (LOCF) is a popular technique used in randomized clinical trials when there is a need to account for the intent to treat for participants who drop out before study completion (Peto et al., 1977). The purpose of intent to treat is to include some logical value for all outcome variables after a participant has dropped out of a study in order to ensure that every patient randomized to a clinical trial will be analyzed. The alternative to some sort of intent to treat has long been casewise deletion, something problematic for randomized clinical trials. LOCF and worst–case scenario are two options (which we treat similarly here). In LOCF, the last observed value for an outcome variable is imputed for all later times the outcome would have been collected but was missing because of dropout. In worstcase scenario imputation, a similar process is conducted by imputing a logical value for all outcome variables after dropout. Rather than carry the last value forward, worst–case scenario imputes the lowest possible value on a scale to participants who dropped out for any reason other than death and assigns a zero to all participants who died during the trial (though this particular value could be readjusted depending on the particular outcome).

LOCF imputation has the advantage of providing a particular method to address intent to treat in randomized clinical trials. Moreover, the method has long been recognized as acceptable by important regulatory agencies, such as the Food and Drug Administration. If the researchers can argue that (a) drop–out rates are similar between the randomized groups and (b) either LOCF or worst–case imputation reflects a viable imputed value for dropouts based on prior empirical work, then the process may provide an effective control for intent to treat (Fairclough, 2002).

Unfortunately, LOCF imputation has been critiqued by most missing data statisticians as inappropriate on theoretical and practical grounds. Abraham and Russell (2004) noted that, despite LOCF being touted as a conservative approach in the past, recent simulations suggest it is not as conservative as was once believed (Mallinckrodt, Clark, Carroll, & Molenberghs, 2003). LOCF overestimates treatment effects and underestimates standard errors, thereby resulting in greater Type I error (Liu & Gould, 2002; Mallinckrodt, Sanger, et al., 2003). Fairclough (2002) has also noted serious concerns about the use of LOCF in that authors who use this technique rarely justify the appropriateness of LOCF for their data or examine the pattern of change in conjunction with the pattern of dropout (Rabound, Singer, Thorne, Schechter, & Shafran, 1998). She has further noted that the assumption that an outcome variable does not change after dropout is untenable in most studies, and LOCF has such limited utility (Gould, 1980; Heyting, Tolbomm, & Essers, 1992; Little & Yau, 1996) that it should be employed only with great caution.

Imputing Items on a Scale

Rather than imputing an entire scale for someone, there may be instances when we are only interested in imputing values for specific items on a scale in order to create a new total score for the outcome. Although taking the mean (or sum) of observed items appears to be akin to mean imputation, it is different on one key aspect: Variables are being imputed rather than cases (Graham et al., 2003). It is somewhat akin to a single regression–based imputation, but rather than having too little variability (as regression–based imputation can have), it has more error because scores are based on fewer items. Despite theoretical problems, Schafer and Graham (2002) have suggested that if MI is not feasible for item–level imputation (e.g., if there are fewer subjects than variables at the item level), then averaging of the available items works reasonably well, especially if the reliability is moderately high (α > .70) and the items to be averaged **[p. 220 ↓]** form a single unified construct (for applied and theoretical discussion on a single unified scale, respectively, see Cole et al., 2006; Messick, 1995).

Modern Methods

The problem with antiquated techniques is that imputed values must be treated as estimates and not real data. An analysis with imputed data that ignores this uncertainty in the imputed values will lead to artificially small standard errors and *p* values, as well as inflated Type I error rates (Schafer & Olsen, 1998). Two techniques that overcome the uncertainty issue are discussed in detail: maximum likelihood (ML) and MI (Schafer & Graham, 2002).

In 1987, four publications came out that changed how we handle missing data (Graham et al., 2003). First, ML was introduced using structural equation modeling (SEM) by Allison (1987) and Muthén, Kaplan, and Hollis (1987). Second, Little and Rubin's (1987) seminal book on analysis with missing data was released, which included discussion of handling missingness with the expectation–maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). Last, but not least, Rubin (1987) introduced MI. Empirical evidence of modern methods has consistently shown the comparative value of modern methods for handling missing data (Arbuckle, 1996; Enders, 2001; Enders & Bandalos, 2001; Graham & Schafer, 1999; Muthén et al., 1987).

Graham and Schafer (1999) suggested being ready to use any of these techniques based on the appropriateness and ease of use for each situation. There are many similarities between these methods. Both ML and MI are based on sound theory, reproduce the data structure accurately, and incorporate a necessary measure of uncertainty given the estimation of missing data (Collins et al., 2001). Both MI and ML regard missingness as a source of random variation to be averaged over rather than simply editing an incomplete data set until it approximates a complete data set. In addition, when ML and MI are conducted properly, the results are fully parametric, using joint probability of manifest and latent data.

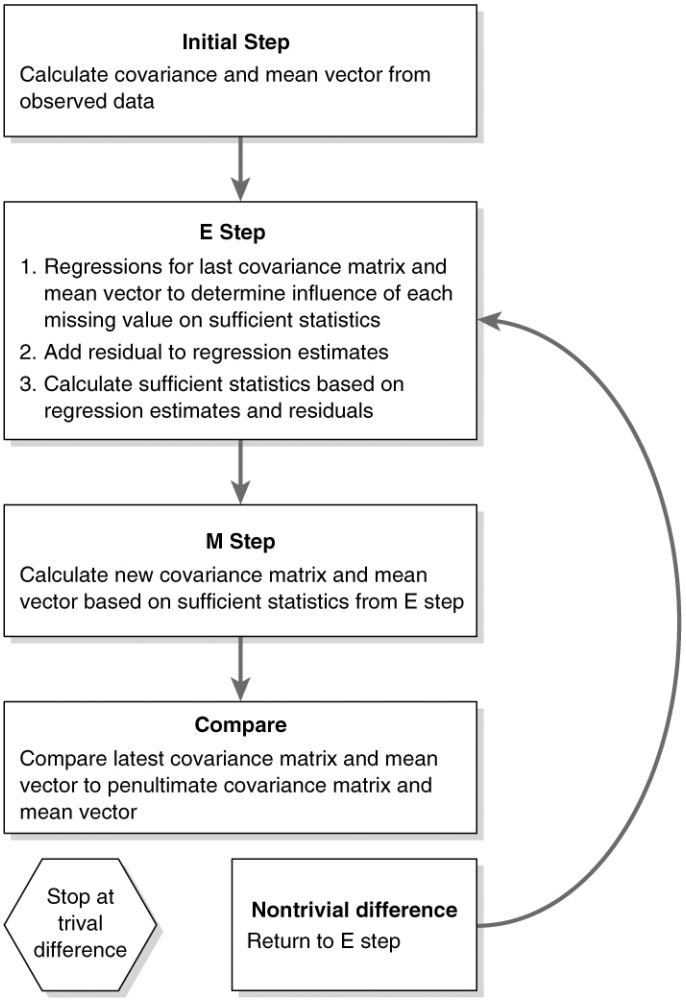
Prior to detailing ML and MI, I believe it is important to explain one of their key building blocks: the EM algorithm, once a popular missing data technique unto itself (Graham et al., 2003).

A Brief Overview of the EM Process

EM is a maximum likelihood approach that can be used to create a new data set in which all missing values are imputed with maximum likelihood values. The EM algorithm (Dempster et al., 1977) is easy to use and has lots of software to produce it (freeware and commercial). Estimates obtained from EM are often unbiased, and the EM algorithm provides a data set with imputed values from which additional analyses can be conducted or used as an initial step in more advanced missing data techniques (Abraham & Russell, 2004). In addition, EM avoids one of the difficulties with standard regression imputation in that there is no decision about which variables should be used as the predictors, as EM uses all available data as predictors for the imputed values. However, EM is problematic to use by itself, as the standard errors and test statistics with EM are not correct.

There are two steps to the EM algorithm: expectation and maximization. These processes work iteratively to impute missing values and to estimate the covariance matrix and mean vector. Before the E and M steps begin, an initial estimate of the covariance matrix and mean vector is calculated from the observed data (Little & Rubin, 2002). For the first E step, a series of regressions are derived from the observed covariance matrix and mean vector in order to determine the contribution of each missing value (for each missing cell) to the sufficient statistics (i.e., variable sums and sums of products) necessary for calculating the covari–ance matrix and mean vector (Enders, 2006a). A random residual term is added because otherwise, all predicted values would fall directly on the predicted regression line. Based on the predicted regressions and residuals, estimates of the sufficient statistics necessary to calculate a missing–data augmented covariance matrix and mean vector during the M step are estimated. During the M step, new estimates of the covariance matrix and mean vector are derived from the sufficient statistics calculated during the E step. No special techniques are used for missingness during this step, and the subsequent matrix and vector are passed along to the **[p. 221 ↓]** next iteration of the E step. The EM algorithm concludes once the difference between the covariance matrices from two adjacent M steps is trivial. Figure 15.1 describes the EM sequence of events.

Figure 15.1 Progression of EM calculations.



Running the EM algorithm before MI is highly recommended for two reasons. First, the resulting parameter estimates make great starting values for the data augmentation process used in MI. Second, the convergence behavior witnessed during EM is predictive of the convergence behavior of data augmentation, which, as will be discussed, is far more esoteric (Schafer & Olsen, 1998). EM may even provide better estimates under nonnormality than a nonparametric MI estimation when the missingness mechanism is MCAR or ignorable, as long as the sample size is 500 or more (Gold, Bentler, & Kim, 2003).

As has been noted, EM parameter estimates are excellent, but the standard errors are problematic. Thus, some specific uses of EM work quite well, including data checks that do not involve hypothesis tests such as coefficient alpha and exploratory factor analysis (Graham et al., 2003; Graham & Hofer, 2000). Nevertheless, with the growing software availability of ML and MI, the use of EM as its own missing data technique is nearly all but outdated.

[p. 222 ↓]

Maximum Likelihood

The ML process for handling missing data was introduced by Allison (1987) and Muthén et al. (1987). It was when Arbuckle (1996) included ML in the user–friendly SEM program AMOS in the mid-1990s that the popularity of this technique increased such that now all of the major SEM programs (and many other packages) include ML to address missing data.

ML is particularly well suited for handling missing data because ML estimation is a tool to select estimates that, if correct, would maximize the probability of observing the data that have been collected (Allison, 2002). ML estimates are consistent, in that they are essentially unbiased. ML estimates are efficient, in that standard errors are close to the true standard errors (especially as sample sizes increase). Finally, ML estimates are normally distributed in repeated sampling (again, improving with larger sample sizes), thereby justifying normal probability testing (Agresti & Finlay, 1997). Simulations have shown that ML is virtually always superior to traditional missingness techniques (Wothke,2000),though the quality of ML does depend on the missing data rate, the covariance structure of the data, and size of the sample (as with all advanced missing data techniques). ML is a step beyond EM in that rather than maximizing for the entire correlation (or covariance) matrix, we maximize just for the model to be analyzed. This process provides more accurate standard errors (Allison, 2002).

Despite the advantages of ML, one major limitation exists: It requires specialized software to use it, usually software in the latent variable modeling realm. ML is available in SEM programs such as AMOS (Arbuckle, 2006), EQS (Bentler, 2006), LISREL

(Jöreskog & Sörbom,2001), Mplus (Muthén & Muthén, 2006), and Mx (Neale, Boker, Xie, & Maes, 1999). Nevertheless, ML provides one very important advantage over MI in that the goodness–of-fit values in SEM are based on the final ML covariance matrix (Wothke, 2000), whereas MI cannot use Rubin's rules to combine fit indices determined for each database.

The ML Process. The calculation details of ML are quite rigorous (Schafer & Graham, 2002) and beyond the scope of this chapter (for more formulaic details, see Allison, 1987, 2002, 2003). With missing data, ML uses observed data in order to calculate a log likelihood from the observed data points for each case (using information from both the covariance matrix and the mean vector). The sum of the individual case likelihoods is used for the total sample log–likelihood function. It should be noted that missing values are not imputed in this process. Instead, the full observed information is used to estimate parameter estimates and standard errors *(SEs)* by iteratively fitting the sample log likelihood (Enders, 2006a). Also, *SEs* should not be calculated using estimated data given that such *SEs* would require the missingness to be MCAR (Kenward & Molenberghs, 1998). Instead, the imputer should opt to use observed data only for the calculation of *SEs* in ML (Enders, 2006a). When the missingness is ignorable, the likelihood function is simply the marginal likelihood for all remaining predictors (Allison, 2002).

One major aspect of using ML for controlling missing data is that the MA model and analysis model are conducted concurrently. Indeed, this is also a limitation of ML: The missingness augmentation cannot be conducted outside of the statistical analysis. This means that any auxiliary variable included in the MA model must also be included in the analysis model, thereby limiting how many auxiliary variables can realistically be used in ML.

Assumptions of ML. As with all statistical techniques, understanding the assumptions behind the modern missing data techniques is critical to the unfettered interpretation of results. First, ML assumes that the missingness mechanism is ignorable. Enders (2001) found that most of the other assumptions for ML for missing data are similar to the assumptions for ML extraction in SEM: ML estimates should come from a sample of sufficient size, and the covariance matrix should be multivariate normal. When the data are multivariate nonnormal, three data issues are standard: (1) Parameter estimates

were relatively unbiased, (2) *SEs* estimated were biased downward, and (3) chi–square-based model fit statistics had excessive rejection rates. This was true for both MCAR and MAR missingness. ML also assumes that the analysis model is correct for the data (i.e., fit statistics meet appropriate criteria).

There are a few considerations to these assumptions. Regarding multivariate normality, there are a number of ways to ensure that the **[p. 223 ↓]** covariance matrix is multivariate normal.[3] Regarding the assumption that the analysis model is correct, it is important to evaluate one's model from multiple perspectives using various model fit statistics. When a model is misfit, the amount of missingness corrected by ML will increasingly affect the efficacy of techniques used to improve model fit, such as modification indices (Steiger, 1990).

Finally, although ML is designed for ignor–able missingness, it may be even more generaliz–able. ML estimates are both consistent and efficient under MAR (Wothke, 2000), and some have shown that the estimates remain favorable even when the missingness deviates from MAR (Little & Rubin, 2002; Muthén et al., 1987).

Limitations. ML provides an excellent modern tool for addressing missingness and does so with much simplicity, especially given that few variants to the ML process exist. However, when some of the assumptions are violated, they can be difficult to overcome. For example, small sample sizes simply obviate the use of ML for addressing missingness (there are no studies to determine what a sufficient sample is for ML). Currently, it may be best to rely on the same standards used in determining a sufficient sample size for latent models using ML (e.g., Hancock, 2006). Another limitation of ML for addressing missing data is the inclusion of auxiliary variables. As discussed above, auxiliary variables can be beneficial in dealing with missing data, but in ML, that means adding auxiliary variables to the analysis, which may complicate the analysis (Enders, 2006a).

Multiple Imputation

When the simplicity of ML does not work for a particular situation, the multiple imputation approach may work. MI is a three–step process to handling missing data.

In the first step, missing data are imputed through a Bayesian procedure. The key for MI is that more than one data set is generated during the imputation stage, providing a means of determining the bias associated from imputing. The second step involves analyzing each of the imputed databases with standard statistics, such as multiple linear regression. In the third step, the results of the analyses from each data set are combined and estimates are adjusted to account for the level of bias within and between data sets introduced from the use of imputed values as real data during Step 2.

As noted by Abraham and Russell (2004), the appeal of MI comes from its adaptability, as shown in many simulation studies (Weinfurt et al., 2003). MI has been shown to result in efficient and unbiased parameter estimates and standard errors (Enders, 2006a; Graham et al., 2003), as well as flexibility in handling numerous data scenarios (Schafer & Olsen, 1998). MI can be used for data with a higher percentage of missingness (Sinhary, Stern, & Russell, 2001, found that MI was appropriate for up to 40% missingness), is appropriate with nested and clustered data (Allison, 2002; Badzioch, Thomas, & Jarvik, 2003), and is appropriate for categorical data (Schafer & Graham, 2002; Schafer & Olsen, 1998). Although standard MI is based on normal modeling theory (Abraham & Russell, 2004), MI performs well with small sample sizes and nonnormal data (Graham & Schafer, 1999). Finally, MI splits the imputation and analysis stages into different steps. Such separation allows for the easy use of numerous auxiliary variables during the MA stage without affecting the analyses, imputation that can be conducted on many variables from different analyses at once to allow for more stable variable interrelationships across analyses, and incorporation of higher–order data into the imputation stage such as interactions or hierarchically organized data.

There are many variants to the MI process, and I will focus on the multivariate normal random MI model, one of the best performing and most widely available (Allison, 2002, p. 56).[4]

Processes. Multiple data set imputation, single data set analysis, and parameter combination are the three steps of MI (for more information on the process and calculations, see Little & Rubin, 2002; Rubin, 1987,1996; Rubin & Schenker, 1991).

Step 1—Imputation. The first task to undertake when conducting MI is to consider how many data sets of imputations are needed. A small number of imputations (m = 10) will be adequate for most situations. In fact, when assumptions are met, 10 imputations will produce standard errors just 2% larger than an infinite number of imputations (von Hippel, 2005). Rubin (1987) provided a formula for determining **[p. 224 ↓]** the efficiency of *m* imputations with a given fraction of missingness,

$$\left(1+\frac{\gamma}{m}\right)^{-1},$$

where γ is the fraction of missing data. With this formula, one can see that with 30% missingness, 91% efficiency is obtained with just three imputed databases, 94% with *m* = 5, and 97% with *m* = 10: More than three times the imputations are necessary to increase efficiency just 6%. Thus, researchers should consider what level of efficiency is necessary (I would recommend no less than 90%) and then determine how many imputations will be necessary given the level of missingness in their database.

Next, one calculates the imputed values. It is important to note that we do not impute in order to divine what an individual would have said if he or she had given us data but rather to preserve important characteristics of parameters (e.g., variances, covariances, means, regression coefficients, etc.) and distributions (Graham et al., 2003, p. 88). The imputed values in MI only have purpose in creating an efficient and unbiased manner by which to properly evaluate all of the observed data in the data set.

Akin to the two–step process of EM, the data imputation in Step 1 of MI also has a two–step process. During the estimation of values for missing data, called data augmentation, two steps, called imputation and posterior distribution, are iteratively conducted. Data augmentation is a type of an increasingly popular method for determining a posterior distribution in Bayesian statistics, the Markov chain Monte Carlo (MCMC) algorithm (Schafer & Olsen, 1998). The starting values for the data augmentation process are based on an initial estimate of the covariance matrix, usually garnered from the EM algorithm (Schafer & Olsen, 1998). At this point, the two–step process begins with the first imputation step, wherein regressions are conducted from the basal covariance matrix replacing missing values with the predicted scores from

the regressions. Random draws from the normal residual distribution are added to the missing values in order to adjust for all estimates being perfectly on the prediction line of the regressions, thereby adding in a measure of uncertainty for our understanding of the missing value. After the missingness is replaced, the covariance matrix and mean vectors are recalculated with the imputed values, and the imputation step ends. The posterior step then commences by introducing the Bayesian aspect to the data augmentation calculations. Randomly sampling elements from a posterior distribution derived from the covariance matrix and mean vector in the imputation step allows for refinement of the covariance and mean elements. Values during the random draws in MI should be based on the Bayesian posterior distribution of the parameter values in order to fully enmesh the uncertainty of our parameter estimates. However, if the sample size is large enough, then random sampling from the Bayesian posterior may not be required (Allison, 2002). This process continues updating the mean and covariance elements and random selections of the posterior distribution until indiscriminant differences are obtained between two consecutive data augmentation steps. There are several ways of calculating the posterior distribution, but Schaefer's (1997) noninformative prior is considered a standard (Enders, 2006a).

Step 2—Single Data Set Analysis. Once *m* data sets have been created with imputed values from Step 1, hypothesis testing is conducted with one's predefined statistical analysis. The statistical analyses are conducted once for each of the *m* data sets. For example, if a researcher planned for multiple regression with three predictor variables and one indicator, then after extracting her five data sets from Step 1, she would analyze the first data set with her multiple regression model, then the second data set, and so forth. These results would then be used to guide calculations in Step 3.

Step 3—Combination. The true benefit of MI comes in the third step, as we enact controls for our uncertainty in using imputed data as observed data during Step 2. Step 3 involves three tasks: Calculate means for all pertinent parameter values, calculate *SE*s for all pertinent parameters, and then calculate *p* values for the parameters based on a modified *df*formula. The first task is the easiest: Create a mean for each important parameter by summing each variable's parameter values across the *m* data sets and dividing that sum by *m,* as in Equation 1,

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i, \qquad (1)$$

where (
i

is the parameter estimate for the ith data set from the *m* imputed data sets. **[p. 225 ↓]**

Next, we determine the *SE* of the parameter by calculating within and between data set error from the imputations. To do this, (a) calculate the within–imputation variance with Equation 2, (b) calculate the between–imputation variance with Equation 3, and (c) determine the total variance by combining within–and between–imputa-tion variances per Equation 4:

$$\bar{U} = \frac{1}{m}\sum_{i=1}^{m}\hat{U}_i, \qquad (2)$$

where Û*i* is the variance estimate from the ith imputed data set, and *m* remains the number of imputation, and

$$B = \frac{1}{m}\sum_{i=1}^{m}(\hat{Q} - \bar{Q})^2, \qquad (3)$$

which is the mean of the variances for a parameter across all databases, and

$$T = \bar{U} + \left(1 + \frac{1}{m}\right)B. \qquad (4)$$

To get the standard error for multiple imputation for a specific parameter, just take the square root of *T* for that parameter (Rubin, 1987).

Finally, we are ready to determine the probability level for each parameter. Begin by calculating the degrees of freedom via Equation 5. The *df* may vary from *m* — 1 to infinity depending on the rate of missingness. The following formula is used to get the *df:*

$$df = (m-1)\left[1 + \frac{\overline{U}}{\left(1 + \frac{1}{m}\right)B}\right]^2, \quad (5)$$

where # is Equation 2 and *B* is Equation 3. If the computed value of the *df*is less than about 10, then one should consider increasing the number of imputations (Schafer & Olsen, 1998).

From here, use a Student's *t* approximation of

$$t \approx \frac{\overline{Q}}{\sqrt{T}}$$

and base the *p* value off of this *t* test with the above *df.* As an example, let's examine results from a simple regression where scores from the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) were used to predict scores on the 10-item short form of the CES–D that I previously developed (Cole, Rabin, Smith, & Kaufman, 2004). Ten data sets were imputed, and the resultant regression coefficients and related standard errors are now provided in Table 15.2. It should be evident that the overall standard error,#t, should be low, and the results of the *t* test will be high, as all

of the coefficients are markedly higher than their respective standard errors with little deviation between data sets. Indeed, using the above formulas, we get the following results: Q = 0.593, *U*= 0.001362, *B* = 0.0002229, #t = .037227, *df* = 23,413, and *t* = 15.91 (with a resulting *p* value that is off the charts given the *df).*

There are other methods for combining more complex estimates, such as likelihood ratios and chi–square statistics (see Schafer, 1997).

Assumptions

Four general assumptions exist for MI: (1) The missingness mechanism is MAR (or better), (2) the model used to generate imputed values must be valid, (3) the analysis and MA models must align, and (4) MI assumes that the data are both univariate and multivariate normally distributed.

Table 15.2 CES–D Short Form Regressed on the Beck Depression Inventory

Data Set	*Unstandardized Regression Coefficient*	*Standard Error*
1	0.597	0.037
2	0.593	0.038
3	0.602	0.038
4	0.590	0.036
5	0.586	0.036
6	0.593	0.037
7	0.592	0.038
8	0.596	0.036
9	0.586	0.037
10	0.596	0.036

Missing at Random. The theory behind MI does not preclude using MI models for nonignorable missingness, and some studies have been published using MI with nonignorable missingness (Glynn, Laird, & Rubin, 1993; Verbeke & Molenberghs,

2000). However, without special techniques applied to the MI process, MI conducted without MAR or MCAR missingness can result in biased parameter estimates.

Valid Imputation Model. Using the Bayesian sampling from the posterior distribution to treat parameters as random rather than fixed is a major benefit of MI. However, a few aspects of Bayesian inference should be understood when conducting MI. First, all of the data's evidence about parameter estimates is summarized in a likelihood function with an assumed distribution. If the assumption is incorrect, than the prior distribution from which the likelihood function is derived may be wrong. Second, Bayesian analysis requires the use of a prior distribution, which is not always accepted as purely scientific. Nevertheless, with the increase in sample size comes a decrease of the influence the prior distribution has on the MI results. Indeed, Schafer and Graham (2002) have noted that the prior rarely exerts a marked influence on the results in MI. For example, an alterative to data augmentation can be conducted with an unrestricted multinomial model or a log–linear model (allowing for restrictions of the multinomial parameters) (Little & Rubin, 2002; Schafer, 1997). Given the scope of the current chapter, I will not cover these methods but encourage the interested reader to review the aforementioned references. If one has reason to suspect that the MA model may be invalid for a particular use, it would be beneficial to run a sensitivity analysis examining different priors or even different Bayesian techniques (such as from different MI software programs). Sensitivity analyses are discussed in the last section of this chapter.

Fit Between Analysis and Imputation Model. It is critical to ensure that the MA model has at least as much information as needed to calculate the analysis model correctly. For example, if interactions between two variables will be analyzed, then the interaction term itself must also be in the MI model. The same is true for multilevel data: Appropriate representation of the multilevel structure must be present in the MA model if the data are to be analyzed with multilevel techniques. It is also important to consider which variables will be used in the imputation phase to help enable the missingness mechanism to be at least MAR. If auxiliary variables are not appropriately considered in the MA model, or if they are inserted in during MA but ultimately inserted into the analysis model as well, the missingness mechanism could be compromised to become MNAR.

Normality. The typical MI model assumes multi–variate normality, thereby requiring all variables to be normally distributed, as well as that each variable can be expressed as a linear function of all other variables with a normal homoscedastic residual (Allison, 2002). However, the MI model works well even when some of the distributions of the present data are not normal (Schafer, 1997). Still, normalizing variables with missing data can have a marked impact on the quality of the imputations (Allison, 2002).

Schafer (1997) argued that the multivariate normal model of MI can be used for nominal and ordinal variables. When nonnormal data are imputed with the multivariate normal MI, the imputed values are more normal than the rest of the data. Given that the imputed values typically make up only a smaller part of the overall data set, the influence of imputed normality among nonnormal data has relatively little impact on the results (Graham & Hofer, 2000; Graham & Schafer, 1999). Moreover, as nonnormality has little impact on parameter estimates compared with standard errors, it is important to note that MI has little effect on the computation of SEs. Indeed, it is one's software and statistical analysis that has the greatest impact on the accuracy of *SE*s under nonnor–mality (Graham & Hofer, 2000).

Nonnormal, ordinal, and even nominal data have been shown to work quite well with the multivariate MI model, even with relatively small sample sizes (Schafer, 1997).

Advanced Issues With MI

Partially Parametric and Nonparametric MI. When the aforementioned assumptions for MI are known to be violated and irreconcilable, it may be necessary to invoke a partially parametric or nonparametric MI process. Many of these exist, and they are detailed well elsewhere.

[p. 227 ↓]

Limitations of MI

MI provides a process that can be used with almost any software and nearly any kind of data. Nevertheless, the downside to MI is that it (a) is time–consuming, (b) can be quite technical, and (c) provides different estimates with every execution (unless the random start value is fixed). In addition, whereas ML works sufficiently with only one sampling, critics of MI have argued that prior simulation studies have failed to control for random sampling theory issues, and therefore MI may require more than 500 imputations (Hershberger & Fisher, 2003). This issue, however, requires more investigation (Abraham & Russell, 2004), and most studies in MI have suggested otherwise (Rubin, 1987; Schafer, 1997).

ML Versus MI: When TO Use One or the Other

ML and MI are really not dissimilar techniques. Collins et al. (2001) have discouraged considering ML and MI as competitors, noting that the two approaches produce similar results, even indis–tinguishably different under certain conditions. Because MI is a Monte Carlo–based method, when m < estimates from MI will contain random error not found in the ML estimates and thus will have larger standard errors, although this is usually minimal. Rubin (1987) provides a formula for comparing the relative efficiency of MI compared with an appropriate ML model, given specific levels of missingness of a number of imputations.

Table 15.3 ML Versus MI: When to Favor One Process Over the Other

Use ML When	*Use MI When*
Software is available, and the analysis model is a simple linear function	Control of MNAR requires many auxiliary variables
There is a need to obtain model fit indices after adjusting for missingness	A polychoric covariance matrix does not make sense for your data, but you have multivariate normality violations
There is a sufficient *n* with only a few auxiliary variables and few normality issues	There are small sample sizes or a large percentage of missing data
Only one analysis is required or the relationship between analyses is not critical	It is easier to impute many variables at once and do separate analyses based on the same MA model

Alas, there are times when selecting one model over the other makes sense. I have summarized in Table 15.3 the major scenarios when one technique may provide benefits over the other techniques.

When ML May Be Best. According to Allison (2002), if the analysis involves any linear model that can be estimated with SEM software, then ML is probably the preferred method for conducting MI. This is because ML in such circumstances is quite easy to implement (literally requiring only the click of a button or the addition of one to two lines of code, depending on the software) and, if the model is correct, will produce estimates that are quite accurate.

There are a few more considerations, however, in order to feel comfortable with using ML. First, one should have reason to believe that only a few auxiliary variables will provide sufficient control for the missingness mechanism to be at least MAR (such as Little's MCAR test or previous research on similar data examining the missingness mechanism). Adding a few auxiliary variables into an ML analysis is not too cumbersome but quickly becomes difficult when (a) there are a large number of variables to be analyzed in the analysis model, (b) there are more than two or three auxiliary variables to be added to the model, or (a) and (b) are both true. In addition, when one has a large sample size (and other issues are appropriately attended), ML should work well. Smaller samples will work best in MI. As noted previously, determining the size of a sufficiently large sample is not straightforward. Finally, if only one analysis is needed, or the relationship between the same variable in

different analyses is not critical, then ML may be sufficient. Otherwise, MI may be more appropriate.

In sum, ML is likely to be the preferred option when the MA model is simple, only **[p. 228 ↓]** a single analysis is required, and the sample size is moderate to large. ML's accuracy in this case is equal to an infinite number of imputations in MI. Indeed, ML can even handle mixtures of continuous and categorical data now (Song & Lee, 2003). Nevertheless, ML does not offer the flexibility that can be found with MI when the data or MA models are more problematic.

When MI May Be Best. Whereas ML is easy to implement but has limited flexibility to handle diverse data and MA scenarios, MI is more technically difficult but has much flexibility and appropriateness with many different data and imputation scenarios. The three most commonly experienced benefits for MI are the control of MNAR with auxiliary variables, appropriateness with small sample sizes, and ability to have consistent data (or sets of data) for any one variable across different analyses.

Because of the simplicity of adding dozens (even hundreds!) of auxiliary variables into the MA model, MI can provide far greater protection against MNAR missingness compared with ML. Not only is it easy to add many auxiliary variables into the MA model, but it is even easier to make sure the effect of the auxiliary variables is incorporated in the analysis model. As long as one does not include the intended auxiliary variables in the analysis model, their protection remains.

Some simulation work suggests that MI may work better with smaller sample sizes than ML (Graham & Schafer, 1999; Schafer & Graham, 2002). MI will still require more participants than variables during calculation of the MA model, and at least a ratio of 10 participants to one variable would be advisable. However, little work has been conducted to provide empirically guided instructions on the minimum number of participants needed for a sufficiently stable MA model. Finally, imputation of an entire data set can be conducted during a single, carefully considered MA model calculation. The benefit of such an MA model over ML is that ML will have differing covariances with the same variable in another model, assuming that some of the other variables in the model have changed. If stability of the variable information is important, then MI can provide better protection. Of course, the analyst will still need to heed caution in that

imputation results will change if the MA model is recalculated and the random starting seed is not the same.

Missingness Imputation Sequential System (MISS): Discussion and Example

Now that we have discussed basic language and issues related to missing data, detailed critiques of classic techniques, and provided discussion on the processes of ML and MI, we are ready to discuss details on how to implement a full process for handling missing data in one's database. This section of the chapter introduces a step–by-step process, the missingness imputation sequential system (MISS), which is designed to walk a researcher through each of the pertinent steps necessary to properly rectify missing data. MISS is presented in the context of a full example using data and related files, all of which may be downloaded from http://www.webcmg.com/missing–example.htm.

Discussion of the Missingness Imputation Sequential System

MISS is an 11-step approach that can be used to ensure that all pertinent aspects of conquering your missing data are addressed. I have labeled it a sequential system because most of the steps must be conducted in a sequential order for later steps to be appropriately conducted. MISS is intended to encourage its users to consider many of the relevant issues necessary in addressing missing data, provide instruction on how to implement ML and MI techniques, and help understand the flow and relationship of the different steps involved in addressing missing data. Of course, use MISS with forethought: It is not a panacea, nor can it be used without knowledge of the issues already discussed in this chapter.

Step 1: Create the Missingness Augmentation Model

Implementation of MISS begins with creation of the MA model. As noted previously, the MA model must capture at least as much information as will be analyzed in the analysis model. This means that the MA model should contain (a) all of the variables in the analysis model (auxiliary variables will be added in a later step); (b) any advanced relationships in the analysis model, such as interactions and multilevel data; and (c) the same level of data that will be analyzed in the analysis. The first aspect is self–explanatory. For interactions, the type of interaction must be **[p. 229 ↓]** considered. If one of the variables in an interaction is categorical, the more appropriate manner for imputation is to conduct a data augmentation run for *m* data sets for each group on the categorical variable. After imputation, merge the data sets and create the interaction term (or evaluate with an interaction–friendly analysis). However, when both variables are continuous (or have multiple points on an ordinal scale), then the separate imputations are either impossible or very laborious. In this instance, one can create an interaction term before imputation based on centered firstorder data (for detailed information on centering, see Tabachnick & Fidell, 2007) and impute with this term in the model. This will almost always lead to nonnormal data, but the results will frequently be appropriate (Allison, 2002). Moreover, if just group differences are going to be examined, group membership should be included in the MA model (Schafer & Graham, 2002). To include multilevel or clustering information, the categorical variables detailing levels or clusters should be included in the MA model. Consider if interactions between different levels or clusters will be analyzed. If so, the interaction terms will also be added to the MA model.

In addition, consider if the level of analysis will take place on summary scores or item–level data. If your statistical analyses are on scale–level data, then there may not be much advantage to imputing the item–level data. I tend to favor keeping the MA and analysis models consistent by imputing the scale–level data when analyzing scale–level data.

Depression Example. For our example that we will use throughout these steps, our database once again involves 189 general–population participants (Cole et al., 2004). Among other collected data, our analysis model examines the influence of BDI and age in predicting scores on the 10-item CES–D short form. Moreover, we have added in an interaction term between age and BDI score. These data are contained in the online Excel database called depression database. Currently, our MA model now contains age, BDI, the centered interaction between age and BDI, and CESD_10.

Step 2: Examine the Data for Missingness

The second step is where we start digging into the observed data set in order to determine how missing data may be missing and prepare them for ML or MI. Before anything else, determine which empty cells are truly missing. Hopefully, the creator of the database has created several codes for missing data, including a differentiation between missing and not applicable. Defining the participants who met inclusion criteria, eliminating those who do not meet inclusion criteria, and imputing true missingness must be done with great care.

Next, see if you can determine the type of missingness mechanism (MCAR, MAR, or MNAR, as discussed above; if the data are not MCAR, conduct a sensitivity analysis between MAR and MNAR techniques). Afterward, determine whether your data preclude use of particular missing data techniques (e.g., discrete data may require MI rather than ML).

Finally, determine the normality of the continuous variables. Although MI has shown robustness to nonnormal continuous data, adjustment of the distributions can only help ensure that appropriate standard errors are obtained. Finally, assessment of multivariate normality can be beneficial, such as the Mahalanobis test (Tabachnick & Fidell, 2007) or Mardia's coefficient (Mardia, 1970). Large violations of these statistics may suggest that a multivariate normality model could be inappropriate, although MI has been shown to be relatively robust to such violations.

Depression Example. No inclusion or exclusion criteria were necessary to control in the current data set. Moreover, all missing cells were intended to be collected and

analyzed; therefore, I did not have applicable missingness here. Data have been coded so that a “-9” defines a missing value.

Using SAS code from Enders (2006c), Little's MCAR test obtained a nonsignificant chi–square = 17.56 *(df* = 14, *p* = .227), indicating that the missing mechanism for the MA model was MCAR. However, there is evidence that, at least for some depression items, gender leads to differential bias (Santor, Ramsay, & Zuroff, 1994). Thus, although we achieved MCAR missingness, it may be prudent to add gender as an auxiliary variable during Step 5.

We have four variables in the analysis and MA models currently, and all of these variables are continuous. Standardized skewness and kur–tosis scores (respectively) for the four variables were 4.16 and -0.04 for age, 6.75 and 5.65 for BDI, 5.49 and 5.03 for Age x BDI, and 5.94 and 3.35 for CES–D. Mardia's coefficient was a **[p. 230 ↓]** standardized score of 8.81. Overall, it is clear that these data have moderate normality violations, which should be appropriately corrected with a square root transformation (Tabachnick & Fidell, 2007).[5] These transformations will be conducted in the imputation software.

Step 3: Select Which Software to Use

There are too many software programs out there to provide a thorough review in this chapter, though some reviews do exist (e.g., Acock, 2005). A few particular programs are worthy of specific mention. Mplus provides for the use of ML for missing data with data that are continuous, censored, binary, ordered categorical, nominal (multinomial), counts, or combinations of all of these. Moreover, it will do so with or without latent variables, allowing for simple regression, correlation, or other manifest–only analyses (Muthén & Muthén, 2006). Indeed, Acock (2005) noted that Mplus has a warehouse of options for handling missing values, including approaches for data that are MCAR, ignorable, and even nonignorable, as well as the integration of robust or bootstrapped standard errors. The release of AMOS 7 (Arbuckle, 2006) provides us with the first commercial package that will conduct MA for both ML and MI. Moreover, Version 7 also

allows for proper calculation of censored, dichotomous, order–categorical, and similar ordinal–level data with either ML or MI.

A growing number of programs will conduct MI, including multivariate normal random MI. One example is NORM, freeware from Schafer (2000)[6] that performed quite well in a simulation study (Allison, 2000) using data augmentation. NORM performs well with nonnormal data and with small sample sizes (Graham & Schafer, 1999). Moreover, the graphical interface is more user–friendly than some. Finally, the SAS PROC MI and MIANALYZE set of commands work quite nicely. The ability to conduct the imputation step, analysis step, and combination of results steps all in one program without much additional coding is a major benefit. Moreover, PROC MI has a very useful command for restricting the upper and lower limits of acceptable random draws for imputed values.

Depression Example. Given the markedly easy interface for NORM and that it is freeware, I have provided the remainder of the MI examples with this software. Nevertheless, in order to conserve space, many of the example figures have been reserved for the supplementary Web site for this chapter at http://www.webcmg.com/missing–example.htm.

Step 4: Determine the Number of Imputations Needed

Rubin (1987) provided Equation 6 for determining the efficiency of an estimate based on *m* imputations with a given fraction of missingness:

$$\left(1+\frac{\gamma}{m}\right)^{-1}, \qquad (6)$$

where γ is the fraction of missing data. Until research determines a minimum amount of efficiency necessary, I recommend using 90% to 95%.

Depression Example. I began by looking at the variable with the highest percentage of missingness because if I could achieve acceptable efficiency without too many imputations for this variable, all other variables should also be sufficient. For the current data, the interaction term has the largest amount of missingness (18.52%). With three imputations, we get 94.19% efficiency; five imputations give us 96.3% efficiency, which is acceptable.

step 5: Select the Auxiliary Variables

Unless Little's MCAR test affirms that the data are MCAR, I strongly encourage the use of auxiliary variables in every missing data problem; even then, auxiliary variables can help with the estimation accuracy. The major decision in this step is to determine how many auxiliary variables should be included. If there is reason to believe that more than three or four auxiliary variables will be necessary, then MI should be used.

Depression Example. As noted in Step 1, I have included gender as a covariate in the MA model, mostly for pedagogical purposes, given the MCAR finding. Gender is a dichotomous variable, with females coded as a 1 and males coded as 2. This variable has no missingness, though Enders (2006b) has noted that missingness in auxiliary variables does not affect the quality of MI estimates.

[p. 231 ↓]

step 6: Configure the Data in the MI Software

Begin by ensuring that all missing value points have been set to a specific value, such as —9 (as long as the value is sufficiently out of the range of possible values for all variables with missingness). The details of processing the data will vary from program to program, and readers should consult the documentation for help in this step. I have included some details on NORM on the supporting Web site.

Depression Example. The supporting Web site shows the final set of options I have marked in the Data—Variables tab. All five variables are marked to be in the MA model and in the final imputed databases. All of the variables except gender are to be transformed during imputation calculations with a square root transformation, including shifts for CES–D, BDI, and the interaction. I used integer rounding for all but gender and the interaction. Gender gets no rounding, which is because it has no missing data. The interaction term is allowed to go to the ninth decimal place, reflecting the number of decimal places used by Excel to create the interaction (although two decimal places is generally sufficient). The output file (summary.out) for the data summary is provided online.

step 7: EM Algorithm

Once all of the options have been set on the Data tab, we are ready to move to the first set of calculations. As you may recall, we begin MI by calculating the EM algorithm to get an initial covariance matrix and mean vector for the data augmentation (DA) calculations of MI. One should take note of the number of iterations necessary to calculate the EM covariance matrix. This is a good indicator of how many iterations will be necessary for the DA step. NORM examples are on the supporting Web site.

Depression Example. Using the default settings for the EM algorithm, convergence occurred in just 12 iterations. The output report (em.out) is located online.

step 8: Data Augmentation Step[7]

This is the first true step of MI. Starting parameters can be calculated either from the data (skipping the EM algorithm) or from the EM results. The EM results are almost always best here. Make sure your software will have enough iterations to work with (more is always better). In other words, if we determine that 40 iterations are required and five databases will be imputed, then the number of iterations should be no less than 40 × 5 = 200. Using more iterations will not create less stability, but having too few will create problems with the data sets. You also need a random number seed. It

is important to note this number, as the ability to recreate the same imputed results demands the use of the same random seed. Finally, options are given for the type of prior to use in the Bayesian estimation. For almost all circumstances, the noninformative prior will work quite well. If problems persist, one can consider the ridge prior, but review critiques first. Upon completion of this step, it is important to review the diagnostics before moving on to Step 9. Upon verifying that the DA model has converged, you can open and examine the data sets.

Depression Example. All options from the main DA screen were left as their defaults. Under computing options, I selected 500 iterations. First, recall that I selected to impute five data sets during Step 5. Next, although I only needed twice the amount of iterations than was necessary for the EM algorithm convergence (12 × 2 = 24), I opted to use 100 iterations per data set. Given the small sample size and number of variables, I was not worried about extra computational time. I noted my random seed of 55,028 for any future need (and saved it in my report for NORM). Finally, I opted to keep the default standard noninformative prior. For imputation options, I selected to impute at the end of every *kth* iteration, and *k* = 100. Finally, under the series options, I opted to save on the worst linear function at every one cycle. The worst linear function and autocorrelations plots (Figure 15.2) looked excellent, demonstrating clear convergence for the DA step.

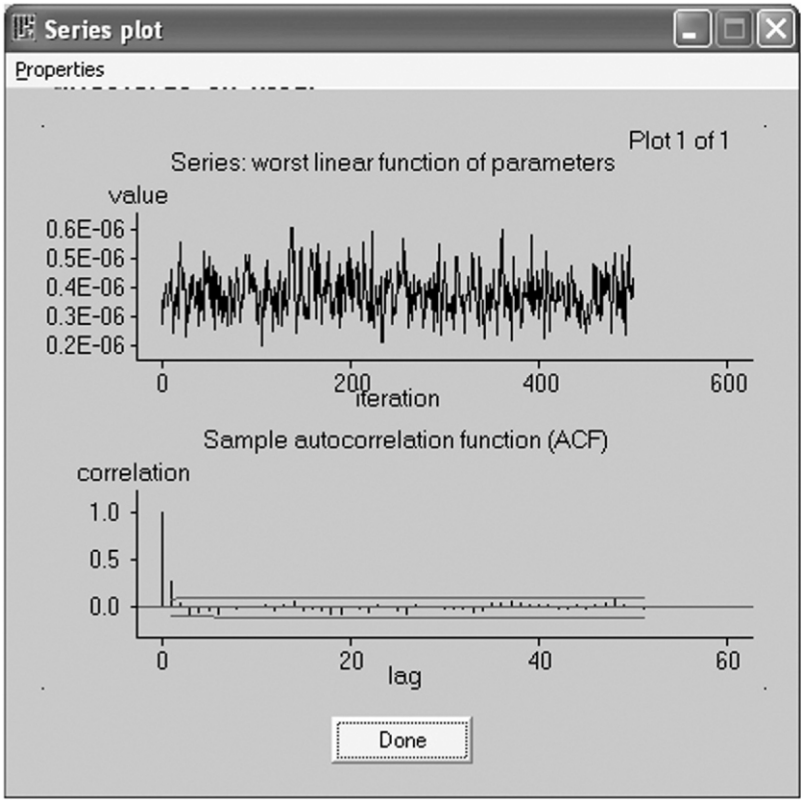
step 9: Calculate the Analysis Model

I will begin by discussing the calculation of the analysis model for MI first, as it is far easier than for ML, mostly because of auxiliary variables.

MI. There are two general ways to proceed in calculating the analysis model for multiply imputed data sets, differing by one's comfort level with SAS. For those who do not wish to use **[p. 232 ↓]** SAS, one simply needs to conduct the analyses for each of the *m* imputed databases, recording the parameter estimates and standard errors. These are entered into a database for combination in NORM or are used to calculate on one's own summary statistics (see Step 10). Indeed, one could even get creative by combining the different databases in a single database, adding a dummy code for the database, and using a process such as SPSS's *organize output by group* command.

However, if one is comfortable conducting statistical analyses in SAS, then Step 9 and Step 10 can be concatenated into a single step (even if the imputation was conducted with NORM) using the SAS Inference command available in Version 8.2 and later (see Allison, 2002).

Figure 15.2 NORM series plot of worst linear trend.



ML. To calculate the analysis model for ML, the model must be created within the desired modeling software. For the current discussion, I have used AMOS, given that the graphical interface is quite intuitive and easily translates into other software coding. The basal MA model without auxiliary variables is simple and is presented in Figure 15.3. In order to incorporate auxiliary variables into the model, Graham (2003) outlined three rules that must be applied. An auxiliary variable must be modeled so it (a) correlates with all other auxiliary variables, (b) correlates with all exogenous manifest

variables, and (c) correlates with the residual of all endogenous manifest variables. Figure 15.4 displays the MA model, with the auxiliary variable of gender added in using Graham rules. As you can see, adding in additional auxiliary variables can be quite cumbersome with even a small model.

Once the model is created, complete the rest of the model specifications and add in the specification for using ML to account for missing data (often called full–information maximum likelihood in latent modeling programs). When this option is selected, other options, such as modification indices, normality estimates, and **[p. 233 ↓]** other features, may, depending on the program, no longer be available. You will want to consult your software manual for specific details.

Depression Example. For the MI results, I calculated the regression coefficients and standard errors for each of the five databases, and these results are provided in Table 15.4. A cursory review suggests that age and BDI are significant predictors of CES–D short–form scores, but the interaction between age and BDI is not. Steps 10 and 11 will formalize these observations. For the ML results, estimates for the single calculation are also presented in Table 15.4. The results from ML tell a story similar to the cursory review of the MI parameters. However, an important finding from the ML results is that the model was not appropriate for the data: Fit statistics were all well off from necessary levels. Modification to the model was limited because of its small number of variables and many paths necessary for the auxiliary variable. No appropriate modifications made the model fit sufficiently. Given that the ML augments to control for missing data depend on the model being correct, we cannot have much faith in the ML results at this point.

Figure 15.3 ML analysis model without auxiliary variable.

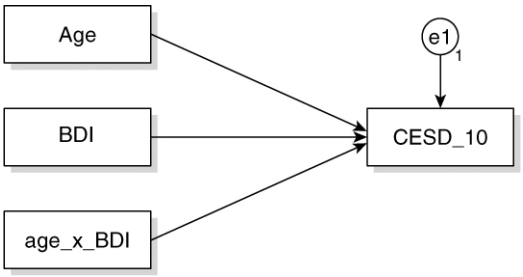
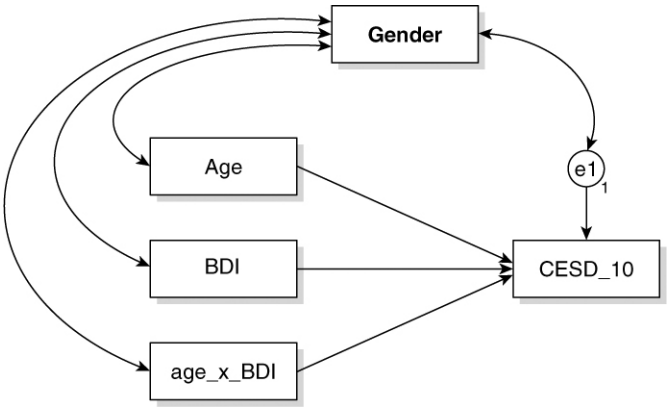


Figure 15.4 ML analysis model with one auxiliary variable.



step 10: Analyze Combined Results

One could undertake the calculation of Rubin's rules for combining multiple parameters and *SE*s with Excel or a calculator, although NORM and other MI programs often undertake this task for you.

Depression Example. Based on the integration of results from NORM (see the Web site for detailed examples), the output report (mi.out) conveyed the combined parameter estimates and *SE* results: age = -0.07 with *SE* = 0.02, BDI = 0.54 with *SE* = 0.04, and interaction = -0.21 with *SE* = 0.36.

[p. 234 ↓]

step 11: Calculate p Values for Estimates Using Modified df Calculations

Table 15.4 Parameter Estimates (ses) for Step 9 Analysis Model Results From Five Imputed Databases

m	*Age*	*BDI*	*Age* × *BDI*
1	–0.064	0.547	–0.129
	(0.019)	(0.042)	(0.304)
2	–0.053	0.545	0.070
	(.018)	(0.043)	(0.292)
3	–0.068	0.535	–0.281
	(.018)	(0.040)	(0.288)
4	–0.079	0.536	–0.267
	(0.018)	(0.039)	(0.284)
5	–0.089	0.513	–0.429
	(0.018)	(0.042)	(0.300)
ML	–0.065	0.555	–0.059
	(0.018)	(0.037)	(0.300)

Using the MI Inference: Scalar option in MI automatically provides the *df, t,* and *p* values for each of the estimates. Recall that if the recalculated *df is* below 10, then more imputed databases should be calculated (using the same random number seed will allow you to generate the same original databases, with all additional databases being subsequent independent data sets).

Depression Example. Two significant findings were found: age, t = -2.98, df = 23, p = .0067; BDI, t = 12.22, df = 307, p < .0001. The t value for the interaction was not above 1 and thus was far from significant. Nevertheless, it did have more *df* than age (36 vs. 23), indicating greater stability of the MI results compared with age. BDI had 307 *df,* indicating marked stability. The percent missing information for age, BDI, and interaction was 45.5,12.0, and 36.3, respectively.

Sensitivity Analyses

Sensitivity analyses are a systematic comparison of the effects of different missing data techniques. For example, sensitivity analyses can be used to compare MI with

simpler, less accurate data replacement techniques (Cook, 1997; Crawford, Tennstedt, & McKinlay, 1995; Lavori, Dawson, & Shera, 1995) or to explore the impact of various assumptions made during the imputation process (Allison, 2002; D. A. Heitjan & Landis, 1994; Little & Yau, 1996; Van Burren, Boshuizen, & Knoock, 1999). Moreover, the impact of assuming different missingness mechanisms can be explored (i.e., MAR vs. MNAR). Although the impact of nonignorable missingness typically results in only slight differences when examined with through sensitivity analyses (Graham, Donaldson, MacKinnon, & Schafer, 1997), it is still recommended that some kind of sensitivity analysis be conducted in every research situation to determine the impact of missingness (Graham & Hofer, 2000).

Despite much writing on the topic, there is very little consistent guidance on how to best conduct a sensitivity analysis. Some researchers (Fairclough, 2002; Fairclough, Fetting, Cella, Wonson, & Moinpour, 1999) will compare changes in significance (significant vs. nonsignificant) as well as various descriptive statistics, such as medians and quartiles. Others examine mean differences (Little & Rubin, 2002) between various missing data techniques. It is my belief that the outcome is critical to the sensitivity analysis. Therefore, I posit that two tests should be undertaken in a sensitivity analysis. First, a comparison between the various missing data techniques should compare which analyses result in significant results and which do not. However, just reviewing the significant versus nonsignificant results can lead to fallacious interpretation of significant differences when none exist for effects that are right around one's alpha level. Is it proper to declare the following results as leading to important differences: MNAR technique with a result of $p = .049$ and MAR technique with a result of $p = .052$? This is why I also advocate examining significant differences in the test statistic or parameter estimate.

As an example, with the depression data examined throughout this chapter, the result for the CES–D short form regressed on BDI for ML was 0.555 (84% CI:[8] 0.503–0.607), whereas the same parameter under MI was 0.535 (84% CI: 0.474-0.597). Both of these results are markedly significant, so they pass the first test. Moreover, they are not significantly different from each other, so they pass the second test. Therefore, the **[p. 235 ↓]** sensitivity test comparing ML and MI for CES–D regressed on BDI showed no important differences.

Summary

The pervasive, nearly omnipresent, nature of missing data will evermore demand the attention of the applied researcher. More and more grant reviewers, journal editors, regulatory agencies, and other gatekeepers are demanding more attention to missing data from applied researchers. Traditional techniques for addressing missing data are nearly all antiquated, having been shown to perform poorly in creating efficient parameter estimates and unbiased standard errors. However, the readily available software for techniques such as ML and MI provides a new era of opportunity for the applied researcher to address missing data with techniques that are appropriate for a wide array of missing data problems. Some missing data problems may require more advanced techniques, such as pattern mixture models. Whereas ML and MI are now within the grasp of many applied researchers, some more advanced techniques will require far greater sophistication. Ideally, as with ML and MI, more advanced models will be made easier to implement with carefully designed software as demand for their use increases.

Notes

1. Others use the term *imputation model,* but I prefer the term *missingness augmentation* because maximum likelihood does not technically impute anything. Instead, it adjusts the covariance matrix and mean vector to incorporate the influence of the missingness. Thus, missingness augmentation is meant to describe a process for handling missing data in some particular manner.

2. For example, in family research, men tend to have more missing data than women regardless of the content (Acock, 2005). Therefore, even though the analytic model may not examine gender differences, including gender in the MA model can help make sure the missing data technique is appropriate.

3. For example, the use ofpolychoric correlations (Song & Lee, 2003) for ordinal data or corrections to the standard errors and chi–square statistics. In SEM, for example, nonnormal data are quite frequent among psychological measures (Cole et al., 2004),

despite the typical use of ML extraction in SEM. This will lead to unbiased estimates, but *SE*s may be quite misleading (Satorra & Bentler, 1994), thus mandating that the missingness be MCAR (Yuan & Bentler, 2000). Robust standard errors are used in complete data analysis to correct for nonnormality (Finney & DiStefano, 2006), and these techniques can be conducted for ML analyses with missing data as well. Although the robust *SEs* assume MCAR missingness, Enders (2001) found the *SE*s to be relatively accurate under MAR missingness as well. In addition, the Bollen–Stine bootstrap can be used for nonnormal data with missingness, as shown in Enders (2002).

4. The multivariate normal random MI model is available in several software packages, including an excellent freeware program by Schafer (2000).

5. See Chapter 13 (this volume) on transformations for specific detail on how best to do this.

6. Downloadable at http://www.stat.psu.edu/~jls/misoftwa.html.

7. Of course, there is no data augmentation step for ML.

8. The 84% confidence interval is ideal for comparing two different effects, as no overlap between the confidence intervals equates to a significance level of .05 (Belia, Fiona, Williams, & Cumming, 2005; Goldstein & Healey, 1995; Tyron, 2001).

References

Abraham, W. T. and Russell, D. W. Missing data: A review of current methods and applications in epidemiological research Current Opinions in Psychiatry, vol. 17, 315-321(2004).

Acock, A. C. (1989). Measurement error in secondary data analysis . In K. Namboodiri, ed. & R. Corwin (Eds.), Research in sociology of education and socialization (Vol. 8, pp. 201-230). Greenwich, CT: JAl.

Acock, A. C. Working with missing values Journal of Marriage and the Family vol. 67, 1012-1028(2005).

Agresti, A., & Finlay, B. (1997). Statistical methods for the social sciences . Englewood Cliffs, NJ: Prentice Hall.

Allison, P. D. (1987). Estimation of linear models with incomplete data . In C. C. Clogg (Ed.), Sociological methodology 1987 (pp. 71-103). San Francisco: Jossey-Bass.

Allison, P. D. Multiple imputation for missing data Sociological Methods & Research vol. 28, 301-309(2000).

Allison, P. D. (2002). Missing data (Vol. 136). Thousand Oaks, CA: Sage.

Allison, P. D. Missing data techniques for structural equation modeling Journal of Abnormal Psychology, vol. 112, 545-557(2003).

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data . In G. A. Marcoulides, ed. & R. E. Schumacker (Eds.), Advanced structural equation modeling: Issues and techniques (pp. 243-278). Mahwah, NJ: Lawrence Erlbaum.

Arbuckle, J. L. (2006). Amos (Version 7.0) [Computer software]. Chicago: Small Waters.

Badzioch, M. D., Thomas, D. C., and Jarvik, G. P. Summary report: Missing data and pedigree and genotyping errors Genetic Epidemiology, vol. 25(Suppl. 1)S36-S42(2003).

Barnard, J. and Meng, X.-L. Applications of multiple imputation in medical studies: From AIDS to NHANES Statistical Methods in Medical Research, vol. 8, 17-36(1999).

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. An inventory for measuring depression Archives of General Psychiatry, vol. 4, 561-571(1961).

Belia, S., Fiona, F., Williams, J., and Cumming, G. Researchers misunderstand confidence intervals and standard error bars Psychological Methods, vol. 10, 389-396(2005).

Bentler, P. M. (2006). EQS structural equation program manual (Version 6.1) . Encino, CA: Multivariate Software.

Cole, J. C. (2007). Current trends in missing data for latent models . Manuscript submitted for publication.

Cole, J. C., Khanna, D., Clements, P. J., Seibold, J. R., Tashkin, D. P., and Paulus, H. E., et al. Single-factor scoring validation for the Health Assessment Questionnaire-Disability Index (HAQ-DI) in patients with systemic sclerosis and comparison with early rheumatoid arthritis patients Quality of Life Research, vol. 15, 1383-1394(2006).

Cole, J. C., Rabin, A. S., Smith, T. L., and Kaufman, A. S. Development and validation of a Rasch-derived CES-D short form Psychological Assessment, vol. 16, 360-372(2004).

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

Cook, N. R. An imputation method for nonignorable missing data in studies of blood pressure Statistical Medicine, vol. 16, 2713-2728(1997).

Crawford, S. L., Tennstedt, S. L., and McKinlay, J. B. A comparison of analytic methods for non-random missingness of outcome data Journal of Clinical Epidemiology, vol. 48, 209-219(1995).

Dempster, A., Laird, N., and Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm (with discussion) Journal of the Royal Statistical Society, vol. 39, 1-18(1977).

Enders, C. K. The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data Psychological Methods, vol. 6, 352-370(2001).

Enders, C. K. Applying the Bollen-Stine bootstrap for goodness-of-fit measures to structural equation models with missing data Multivariate Behavioral Research, vol. 37, 359-377(2002).

Enders, C. K. (2006a). Analyzing structural equation models with missing data . In G. R. Hancock, ed. & R. O. Mueller (Eds.), Structural equation modeling: A second course (pp. 313-344). Greenwich, CT: Information Age Publishing.

Enders, C. K. (2006b, April). Analyzing structural equation models with missing data . Paper presented at the 2006 annual meeting of the American Education Research Association, San Francisco.

Enders, C. K. (2006c). Little's (1998) test of missing completely at random (MCAR) missing data . Phoenix, AZ: Author.

Enders, C. K. and Bandalos, D. L. The relative performance of full information maximum likelihood estimation for missing data in structural equation models Structural Equation Modeling, vol. 8, 430-457(2001).

Fairclough, D. L. (2002). Design and analysis of quality of life studies in clinical trials . Boca Raton, FL: Chapman & Hall/CRC.

Fairclough, D. L., Fetting, J. H., Cella, D., Wonson, W., and Moinpour, C. M. Quality of life and quality adjusted survival for breast cancer patients receiving adjuvant therapy. Eastern Cooperative Oncology Group (ECOG) Quality of Life Research, vol. 8, 723-731(1999).

Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation modeling . In G. R. Hancock, ed. & R. O. Mueller (Eds.), Structural equation modeling: A second course (pp. 269-314). Greenwich, CT: Information Age Publishing.

Glynn, R. J., Laird, N. M., and Rubin, D. B. Multiple imputation in mixture models for nonignorable nonresponse with followups Journal of the American Statistical Association, vol. 88, 984-993(1993).

Gold, M. S., Bentler, P. M., and Kim, K. H. A comparison of maximum-likelihood and asymptotically distribution-free method of treating incomplete nonnormal data Structural Equation Modeling, vol. 10, 47-79(2003).

Goldstein, H. and Healey, M. J. R. The graphical presentation of a collection of means Journal of the Royal Statistical Society, vol. 158A, 175-177(1995).

Gould, A. L. A new approach to the analysis of clinical drug trials with withdrawals Biometrics, vol. 36, 721-727(1980).

Gourieroux, C. and Monfort, A. On the problem of missing data in linear models Review of Economic Studies, vol. 48, 579-586(1981).

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

Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data . In J. A. Schinka, ed. & W. F. Velicer (Eds.), Handbook of psychology (Vol. 2, pp. 87-114). New York: John Wiley.

Graham, J. W., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research . In K. Bryant, ed. , M. Windle, ed. , & S. 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., & Hofer, S. M. (2000). Multiple imputation in multivariate research . In T. D. Little, ed. , K. U. Schnabel, ed. , & J. Baumert (Eds.), Modeling longitudinal and multilevel data (pp. 201-218). Mahwah, NJ: Lawrence Erlbaum.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size . In R. H. Hoyle (Ed.), Statistical strategies for small sample size (pp. 1-29). Thousand Oaks, CA: Sage.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in analysis of change . In L. M. Collins, ed. & 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., and Cumsille, P. E. Planned missing data designs in psychological research Psychological Methods, vol. 11, 323-343(2006).

Haitovsky, Y. Missing data in regression analysis Journal of the Royal Statistical Society, vol. 30B, 67-82(1968).

Hancock, G. R. (2006). Power analysis in covariance structure modeling . In G. R. Hancock, ed. & R. O. Mueller (Eds.), Structural equation modeling: A second course (pp. 69-118). Greenwich, CT: Information Age Publishing.

Harezlak, J., Gao, S., and Hui, S. L. An illness-death stochastic model in the analysis of longitudinal dementia data Statistics in Medicine, vol. 22, 1465-1475(2003).

Heitjan, D. A. and Landis, J. R. Assessing secular trends in blood pressure: A multiple imputation approach Journal of the American Statistical Association, vol. 89, 750-759(1994).

Heitjan, D. F. and Rubin, D. B. Ignorability and coarse data Annals of Statistics, vol. 19, 2244-2253(1991).

Hershberger, S. L. and Fisher, D. G. A note on determining the number of imputations for missing data Structural Equation Modeling, vol. 10, 648-650(2003).

Heyting, D. A., Tolbomm, T. B. M., and Essers, J. G. A. Statistical handling of dropouts in longitudinal clinical trials Statistical Medicine, vol. 11, 2043-2061(1992).

Jöreskog, K. G., & Sörbom, D. (2001). LISREL 8: User's reference manual . Chicago: Scientific Software.

Kenward, M. G. and Molenberghs, G. Likelihood based frequentist inference when data are missing at random Statistical Science, vol. 13, 236-247(1998).

Ladouceur, R., Gosselin, P., Laberge, M., and Blaszcynski, A. Dropouts in clinical research: Do results reported reflect clinical reality? The Behavioral Therapist, vol. 24, 44-46(2001).

Lavori, P., Dawson, R., and Shera, D. A multiple imputation strategy for clinical trials with truncation of patient data Statistical Medicine, vol. 14,1925(1995).1912

Little, R. J. A. A test of missing completely at random for multivariate data with missing values Journal of the American Statistical Association, vol. 83, 1198-1202(1988).

Little, R. J. A., & Rubin, D. B. (1987). Statistical analysis with missing data . New York: John Wiley.

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

Little, R. J. A. and Yau, L. H. Y. Intent-to-treat analysis for longitudinal studies with dropouts Biometrics, vol. 52, 1324-1333(1996).

Liu, G. and Gould, A. L. Comparison of alternative strategies for analysis of longitudinal trials with dropouts Journal of Biopharmaceutical Statistics, vol. 12, 207-226(2002).

Mallinckrodt, C. H., Clark, W. S., Carroll, R. J., and Molenberghs, G. Assessing response to profiles from incomplete longitudinal clinical trial data under regulatory considerations Journal of Biopharmaceutical Statistics, vol. 13, 179-190(2003).

Mallinckrodt, C. H., Sanger, T. M., Dubé, S., DeBrota, D. J., Molenberghs, G., and Carroll, R. J., et al. Assessing and interpreting treatment effect in longitudinal clinical trials with missing data Biological Psychiatry, vol. 53, 754-760(2003).

Mardia, K. V. Measures of multivariate skewness and kurtosis with applications Biometrika, vol. 57, 519-530(1970).

Meng, X.-L. Multiple imputation with uncongenial sources of input Statistical Science, vol. 9, 538-573(1994).

Messick, S. Validity of psychological assessment: Validation of inferences from persons' responses and performances asscientific inquiry into score meaning American Psychologist, vol. 50, 741-749(1995).

Muthén, B. O., Kaplan, D., and Hollis, M. On structural equation modeling with data that are not missing completely at random Psychometrika, vol. 52, 431-462(1987).

Muthén, B. O., & Muthén, L. K. (2006). Mplus (Version 4.0) [Computer software] . Los Angeles: Author.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (1999). Mx: Statistical modeling (Version 5th ed.) [Computer software] . Richmond: Virginia Commonwealth University, Department of Psychiatry.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., and Howard, S. V., et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient: II. Analysis and examples British Journal of Cancer, vol. 35, 1-39(1977).

Rabound, J. M., Singer, J. D., Thorne, A., Schechter, M. T., and Shafran, S. D. Estimating the effect of treatment on quality of life in the presence of missing data due to drop-out and death Quality of Life Research, vol. 7, 487-494(1998).

Radloff, L. S. The CES-D scale: A self-report depression scale for research in the general population Applied Psychological Measurement, vol. 1, 384-401(1977).

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

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys . New York: John Wiley.

Rubin, D. B. Multiple imputation after 18+ years Journal of the American Statistical Association, vol. 91(434)473-489(1996).

Rubin, D. B. and Schenker, N. Multiple imputation in health-care databases: An overview and some applications Statistics in Medicine, vol. 10, 585-598(1991).

Santor, D. A., Ramsay, J. O., and Zuroff, D. C. Nonparametric item analysis of the Beck Depression Inventory: Evaluating gender item bias and response option weights Psychological Assessment, vol. 6, 255-270(1994).

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis . In A. von Eye, ed. & C. C. Clogg (Eds.), Latent variables analysis: Applications for developmental research (pp. 399-419). Thousand Oaks, CA: Sage.

Schafer, J. L. (1997). Analysis of incomplete multivariate data . New York: Chapman & Hall.

Schafer, J. L. (2000). NORM (Version 2.03) [Computer software] . University Park, PA: Author.

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

Schafer, J. L. and Olsen, M. Multiple imputation for multivariate missing data problems: A data analyst's perspective Multivariate Behavioral Research, vol. 33, 545-571(1998).

Sinhary, S., Stern, H. S., and Russell, D. W. The use of multiple imputation for the analysis of missing data Psychological Methods, vol. 6, 317-329(2001).

Song, X. Y. and Lee, S. Y. Full maximum likelihood estimation of polychoric and polyserial correlations with missing data Multivariate Behavioral Research, vol. 38, 57-79(2003).

Steiger, J. H. Structural model evaluation and modification: An interval estimation approach Multivariate Behavioral Research, vol. 25, 173-180(1990).

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed.). New York: Pearson Education.

Tyron, W. W. Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting nill hypothesis statistical tests Psychological Methods, vol. 6, 371-386(2001).

Van Burren, S., Boshuizen, H. C., and Knoock, D. L. Multiple imputation of missing blood pressure covariate in survival analysis Statistical Medicine, vol. 18, 681-694(1999).

Verbeke, G., & Molenberghs, G. (2000). Linear mixed models for longitudinal data . New York: Springer-Verlag.

von Hippel, P. T. How many imputations are needed? A comment on Hershberger and Fisher Structural Equation Modeling, vol. 12, 334-335(2005).

Weinfurt, K. P., Castel, L. D., Li, Y., Sulmasy, D. P., Balshem, A. M., and Benson, A. B., III, et al. The correlation between patient characteristics and expectation of benefits in Phase I clinical trials Cancer, vol. 98, 166-175(2003).

Wothke, W. (1993). Nonpositive definite matrices in structural equation modeling . In K. A. Bollen, ed. & J. S. Long (Eds.), Testing structural equation models (pp. 256-293). Newbury Park, CA: Sage.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data . In T. D. Little, ed. , K. U. Schnabel, ed. , & J. Baumert (Eds.), Modeling longitudinal and multilevel data (pp. 219-240). Mahwah, NJ: Lawrence Erlbaum.

Yuan, K.-H. and Bentler, P. M. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data Sociological Methodology, vol. 30, 165-200(2000).

http://dx.doi.org/10.4135/9781412995627.d19