A Comparison of Restrictive Strategies in Modern Missing Data Procedures

Article *in* Psychological Methods · January 2002

DOI: 10.1037/1082-989X.6.4.330 · Source: PubMed

Linda M Collins
Pennsylvania State University

Chi-Ming Kam
The University of Hong Kong

Psychological Methods
2001, Vol. 6, No. 4, 330–351

1082-989X/01/$5.00 DOI: 10.1037//1082-989X.6.4.330

A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures

Linda M. Collins, Joseph L. Schafer, and Chi-Ming Kam
The Pennsylvania State University

Two classes of modern missing data procedures, maximum likelihood (ML) and multiple imputation (MI), tend to yield similar results when implemented in comparable ways. In either approach, it is possible to include auxiliary variables solely for the purpose of improving the missing data procedure. A simulation was presented to assess the potential costs and benefits of a *restrictive* strategy, which makes minimal use of auxiliary variables, versus an *inclusive* strategy, which makes liberal use of such variables. The simulation showed that the inclusive strategy is to be greatly preferred. With an inclusive strategy not only is there a reduced chance of inadvertently omitting an important cause of missingness, there is also the possibility of noticeable gains in terms of increased efficiency and reduced bias, with only minor costs. As implemented in currently available software, the ML approach tends to encourage the use of a restrictive strategy, whereas the MI approach makes it relatively simple to use an inclusive strategy.

When researchers are confronted with missing data, they run an increased risk of reaching incorrect conclusions. Missing data may bias parameter estimates, inflate Type I and Type II error rates, and degrade the performance of confidence intervals. Furthermore, because a loss of data is nearly always accompanied by a loss of information, missing values may dramatically reduce statistical power. Researchers who wish to mitigate these risks must pay close attention to the missing data aspect of the analysis and choose their strategy carefully.

In this article, we investigate conceptual and practical issues related to the choice of a strategy for handling missing data. In so doing we consider two classes of modern missing data procedures: (a) maximum-likelihood (ML) estimation based on the incomplete sample and (b) multiple imputation (MI) for the missing values followed by repeated analyses of the completed data sets. We do not consider ad hoc procedures such as listwise or pairwise deletion, substitution of means, regression predictions, or other forms of single imputation, which have been conclusively shown to perform poorly except under very restrictive or special conditions (Little & Rubin, 1987). Unlike these older methods, ML and MI are based on sound theory and can produce efficient estimates and accurate measures of statistical uncertainty. The attractive theoretical properties of these procedures, however, do not necessarily translate into good performance in real data analyses if they are poorly applied or if the mechanisms generating either the data or the missing values depart substantially from the underlying statistical assumptions.

Neither ML nor MI is particularly new, but both have been receiving greater attention in the social sciences due to the availability of new software. For

Linda M. Collins, The Methodology Center and Department of Human Development and Family Studies, The Pennsylvania State University; Joseph L. Schafer, Department of Statistics and The Methodology Center, The Pennsylvania State University; Chi-Ming Kam, The Methodology Center and The Prevention Research Center for the Promotion of Human Development, The Pennsylvania State University.

This research was supported by National Institute on Drug Abuse Grant DA P50 DA 10075. We would like to offer special thanks to John W. Graham and our colleagues at The Methodology Center for particularly insightful comments on several drafts of this article.

Correspondence concerning this article should be addressed to Linda M. Collins, The Methodology Center, 159 Henderson South, The Pennsylvania State University, University Park, Pennsylvania 16802. Electronic mail may be sent to lmcollins@psu.edu.

example, the commercial program Amos[1] (Arbuckle & Wothke, 1999) now implements the ML approach for linear models with manifest and latent variables. A free program called NORM (Schafer, 1999b) makes it straightforward to apply MI even in fairly large multivariate data sets. LISREL 8.5 (Jöreskog & Sörbom, 2001) offers both ML and MI. These are only a few examples of a growing number of programs that implement modern approaches to missing data. Because ML and MI have become feasible at approximately the same time, there is a natural tendency for researchers to regard them as competitors. In many respects, however, the two approaches are closely related, and under certain conditions will produce results that are virtually indistinguishable. One purpose of this article is to clarify when and how the two approaches lead to the same result.

The theoretical foundations of ML and MI share two common themes. First, they both regard the unknown data values as a source of random variation to be averaged over. Ad hoc methods of case deletion and single imputation merely edit an observed, incomplete data set until it resembles a complete one. ML and MI, however, regard the missing data as random variables and explicitly remove them from the inferential system by a process of averaging (i.e., integration). Second, ML and MI are both fully parametric, relying on joint probability models for the observed and unobserved data.[2] With any model-based procedure, a researcher has substantial latitude to tailor the model, expanding or reducing it to suit his or her analytic needs. For example, one must specify which subset of the available variables will be used in a given analysis. Although guiding principles are often available, experienced researchers know that the choices are not always clear cut.

Both with ML and MI, it is possible—and often desirable—to expand the system to include information (i.e., variables) whose sole purpose is to improve the performance of the missing data method, even when this information is not especially relevant to the hypotheses of scientific interest. We refer to variables included in an analysis solely to improve the performance of the missing data procedure as *auxiliary variables*. Auxiliary variables may be included for two reasons. First, one may want to introduce variables that are potential causes or correlates of the missingness itself. For example, if younger participants are less apt to provide responses than older ones, then including age in the procedure should help to reduce biases incurred by differential response rates related to age. Second, one may want to include variables that are simply correlated with the variables that have missing values, whether or not they are related to the mechanism of missingness. For example, suppose that some survey participants refuse to answer an income question, but nearly everyone provides their years of education. Because income and years of education are correlated, including the latter in the procedure may help to recover some of the information missing in the former, thereby increasing the efficiency of statistical inferences related to income. In longitudinal research, it is even possible to include a variable measured at a later time as an auxiliary variable aimed at missing data on a variable measured at an earlier time.

The theoretical possibility of including auxiliary variables has been explored in the context of MI by Meng (1994) and Rubin (1996). Both of these authors argued that nontrivial improvements in efficiency and bias may accrue when auxiliary variables are added to an imputation procedure, even when the auxiliary variables are not included in subsequent analyses of the imputed data. It is also possible to add auxiliary variables to ML. When fitting a linear model with Amos, for example, one could expand the model to include auxiliary variables, even when these variables would not otherwise be of interest. To our knowledge, this point has not been fully explored in the literature pertaining to ML. Adding auxiliary variables to a model, unless it is done carefully, could produce undesirable effects (e.g., altering the meaning of the model and redefining the population coefficients being estimated). A framework for adding auxiliary

[1] Throughout this article, examples of software that use ML or MI are occasionally mentioned. This article does not provide an exhaustive list of such software. In particular, if software is not mentioned in this article, this does not mean that the software lacks missing data capability. The number of statistical programs that include modern missing data procedures is growing at a rapid rate.

[2] In special cases, it is possible to generate proper MIs without a full parametric model. For example, Lavori, Dawson, and Shera (1995) developed a nonparametric MI procedure that has been implemented in a commercial software product called SOLAS (Statistical Solutions, Inc., 1998). As argued by Schafer (1999a) and Allison (2000), this method is appropriate only for a rather narrow class of problems and should be used with great caution.

variables to ML analyses of linear models with latent variables was described by Graham (2001). In weighing the various approaches for adding auxiliary variables, it is worthwhile to consider if and when the statistical benefits of such variables will be worth the additional effort.

The primary purpose of this article is to compare and contrast two strategies for choosing the size and scope of the set of auxiliary variables: the *restrictive* strategy (i.e., including few or no auxiliary variables) and the *inclusive* strategy (i.e., including numerous auxiliary variables). It is not our intention to attach any positive or negative connotations to the terms *inclusive* or *restrictive*, because both have pitfalls. A restrictive strategy might omit important correlates of missingness, whereas an inclusive one might err on the side of too many. Intuitively, it seems clear that there should be trade-offs between these strategies, but actually to quantify the cost of omitting an important variable versus the cost of including a superfluous one is challenging. To evaluate the benefits and costs of these two strategies, it is necessary to examine the properties of the ML and MI approaches, particularly with regard to the use of auxiliary variables. In real-world applications, whether an ML or MI missing data procedure is chosen strongly influences the choice of a restrictive or inclusive strategy, because adding auxiliary variables may be easier under some missing data procedures than others. In the course of our own research and in observing the work of others, we have found that the decision between ML and MI can be a difficult one; until now, experts in the development of missing data methods have provided little practical guidance in this matter.

We begin by reviewing the ML and MI approaches and describe the conditions under which the two will produce similar results. We then identify four key questions pertaining to restrictive versus inclusive missing data strategies. We address these questions by means of simulation studies where various kinds of missing data are generated, and then restrictive and inclusive strategies are applied using MI. Our simulations focus on fairly simple estimates of means, variances, regression coefficients, and correlations because of the importance of these quantities in behavioral research. Despite the apparent simplicity of the simulations, the results are quite revealing and lend useful insights into the likely behavior of these methods in more complicated situations, including structural equations models with latent variables.

Overview of Procedures and Underlying Assumptions

Some Missing Data Terminology

Our descriptions in this section are necessarily brief; more details and discussion about their theoretical underpinnings can be found in Schafer (1997). Missing data literature can be confusing in part because some of the most fundamental technical concepts bear names that seem counterintuitive to nonstatisticians. For example, the term *ignorable* (Little & Rubin, 1987) was not meant to imply that the missing data aspect could be ignored. In a similar vein, *missing at random* does not mean that missingness arises from a purely random process that is exogenous or unrelated to other variables in the system. To clarify some of this terminology, we will start with a very simple example. Suppose there are two variables, Y and Z, possibly related. Suppose that Z is always observed, but Y may be missing for reasons that are possibly related to Y, Z, or both. Let us define a binary response indicator variable, R, which is equal to 1 if Y is observed and 0 if Y is missing. Finally, let us suppose that the questions of scientific interest pertain to the population distribution of Y; Z is not of inherent interest, but may perhaps be useful if it can either (a) help to explain why some values of Y are missing or (b) help to predict the missing values because of its association with Y.

In this situation, the major classes of missing data mechanisms can be neatly described in terms of relationships among Y, Z, and R. If missingness on Y (i.e., R) depends on Y itself, then it is said to be *missing not at random* (MNAR). For example, consider a study on reading achievement (Y). If students who have lower achievement tend to avoid the measurement session (e.g., because they find reading tests frustrating) more often than students with higher achievement, the situation is MNAR. In this case, the observed scores will provide an upwardly biased picture of the achievement distribution for the full population. A knowledgeable researcher may try to mitigate the bias by adjusting the scores for some additional covariate, Z (e.g., a summary of answers to pretest questions such as, "Do you enjoy reading?" or "Do you dislike reading tests?"), but if there is correlation or relationship between R and Y in excess of what can be accounted for by their mutual associations with Z, then the situation remains MNAR and standard adjustments based on Z may not entirely remove the bias.

If missingness on Y does not depend on Y itself, it is said to be *missing at random* (MAR). Under MAR, R and Y may be related but only indirectly through Z. For example, consider another study on reading achievement, this one involving both special education students and students in ordinary classrooms. Suppose special education students are more likely to be absent for the measurement (i.e., for special education students, it is more likely that $R = 0$). Special education students also tend to have lower reading achievement (Y), so the indicator, Z, for special education versus ordinary classroom is related to achievement and thus to absence as well. However, suppose that among special education students there is no relationship between absence and achievement, and among regular classroom students there is no relationship between absence and achievement. Under these circumstances, the achievement scores that are missing due to student absence will be MAR, and inferences based on the observed scores will be unbiased if the analysis makes proper distinctions between special education and ordinary classroom students.

A special case of MAR, called *missing completely at random* (MCAR), arises when R is entirely unrelated to Y. For example, suppose that a researcher measures the reading ability (Y) of students in a particular school. Furthermore, suppose our hypothetical researcher was unlucky enough to schedule the measurement on a day when all students currently enrolled in drivers' education were taking their driving tests, so none of those students was present for the measurement. If enrollment in drivers' education (Z) bears no relationship to reading ability (Y), then the situation will be MCAR. In this case, the distribution of observed Y scores would provide a completely unbiased picture of the scores for the population.

The designation MCAR does not imply that nonresponse is an inexplicable or unpredictable phenomenon like flips of a coin; in our last example, it can be logically explained by enrollment in drivers' education. Rather, the essence of MCAR is that the process generating the missing values bears no statistical relationships (e.g., correlations) with our variables of interest. The essence of MAR is that the missing data process may be related to our variables of interest, but these relationships can be fully captured or explained by other quantities that have been observed. Finally, the essence of MNAR is that the relationships between nonresponse and the variables on which it occurs cannot be fully explained by observed data. Thus the distinctions among MCAR, MAR, and MNAR pertain not to randomness per se but to relationships between variables of interest and variables explaining missingness.

The examples involving Y, Z, and R oversimplify what is often seen in practice. For example, a potential cause or correlate of missingness, Z, is often not entirely observed but has missing values too, and the manner in which Z becomes missing is relevant. Moreover, one is typically less interested in the population distribution of Y than in its relationships to other variables. Researchers often estimate a large number of intervariable relationships simultaneously and need to be concerned about the effects of missing data on all parameters under consideration. Precise mathematical definitions of MCAR, MAR, and MNAR that generalize to these more complicated situations were given by Little and Rubin (1987). These authors also introduced the concept of ignorability: When the missing values are MAR, the missing data mechanism is said to be *ignorable*, and under MNAR it is said to be *nonignorable*.[3]

We would like to make two general points regarding the classes of missing data mechanisms. First, MAR does hold in some special situations (e.g., when a pattern of missing values has been imposed by the researcher as part of the study design; Graham, Taylor, & Cumsille, 2001). But when the missing data aspect is beyond the control of the researcher, one may usually be confident that the situation is MNAR. It is highly doubtful that an analyst will have access to all of the relevant causes or correlates of missingness. One can safely bet on the presence of lurking variables that are correlated both with the variables of interest and with the missingness process; the important question is whether these correlations are strong enough to produce substantial bias if no additional measures are taken. Second, it may be helpful to include auxiliary variables that are correlates of missingness and/or correlates of the variable of interest. Including the former can help to reduce bias and move the situation closer to MAR; including the latter may help to reduce variance, as in the MCAR example given above. Yet it remains to be seen how strong these correlations must be for the benefits to be worth the extra effort and to offset any reductions in preci-

[3]The precise definition of ignorability given by Little and Rubin (1987) also involves a minor technical condition called *distinctness of parameters*. For our purposes, we can assume that this condition always holds, so that MAR and ignorability can be used interchangeably.

sion that arise whenever a parametric model is expanded.

Maximum Likelihood

The well-known method of maximum likelihood (ML) chooses parameter values that assign the highest possible probability or probability density to the data values actually seen, under a well-defined family of parametric probability models. The probability or probability density of the realized data is called the likelihood function. Principles for applying ML and likelihood-based procedures to incomplete-data problems were first described by Rubin (1976). Little and Rubin (1987) provided an excellent overview of the theory of ML estimation, both with and without missing values.

A key result of Rubin (1976) pertains to likelihood-based inferences when some data are missing. If the missing data are MAR, then the missingness indicators drop out of the likelihood function and the missing values become analogous to data that were never supposed to be collected in the first place. Mathematically speaking, the missing data values are removed from the likelihood by a process of summation or integration. In effect, they are treated as unknown random variables to be averaged over. Likelihood functions that have been collapsed over missing values in this way typically have a complicated form; special computational techniques (e.g., EM algorithms) may be needed to maximize them. Computational aspects of ML with missing data were reviewed by Little and Rubin (1987) and Schafer (1997).

For a data analyst using modern missing data procedures, MAR has enormously positive implications, because it means that one does not need to model the process by which the data became missing; as long as that process falls within the MAR class, all other details of the missing data mechanism can be ignored (hence the term *ignorable*). For example, suppose that Z is a potential cause or correlate of missingness for a variable Y, and suppose that there are no other lurking variables in the system. ML estimates for a model that includes both Y and Z will then be essentially unbiased, provided that the model is correctly specified and the sample is reasonably large. This will be true regardless of how the missingness is related to Z. The probability of missingness may not depend on Z at all—it may increase or decrease with Z, it may have a curvilinear dependence, or it may be entirely absurd (e.g., Y becomes missing whenever the sum of the first two significant digits of Z is odd). These are all part of the broad class of MAR mechanisms, and hence the likelihood function that ignores the missing data mechanism will yield correct inferences in all of these situations.

A variety of ML procedures for missing data problems have become available in recent years. During the 1980s, ML estimation of a covariance matrix from incomplete multivariate data was implemented in BMDP, a commercial statistical package that was subsequently absorbed into SPSS (Norusis, 2000). An equivalent procedure was also provided by Graham and Hofer (1993) in a free program called EMCOV. ML analysis of unbalanced repeated measures under alternative covariance assumptions is available in SAS Proc Mixed (Littell, Milliken, Stroup, & Wolfinger, 1996) and HLM (Raudenbush, Bryk, Cheong, & Congdon, 2000). Commercial programs such as Amos (Arbuckle & Wothke, 1999), Mplus (Muthén, 2001), and LISREL 8.5 (Jöreskog & Sörbom, 2001), as well as free programs such as Mx (Neale, 1997), compute ML estimates and standard errors for structural equations modeling (SEM) from data with arbitrary patterns of missing values. In the SEM literature, the approach taken by these programs is sometimes called full-information maximum likelihood (FIML). In keeping with terminology used in other statistical literature on missing data, it seems more natural to refer to FIML simply as ML.

Most programs that include ML can accommodate ordinary linear regression and analysis of variance (ANOVA), confirmatory factor analysis, and more complex latent variable models. For example, a regression model specifying several independent variables and a dependent variable might be drawn using the graphical interface provided by Amos. Once the model is specified, the user supplies an input matrix of raw data with missing values denoted by a special code. The ML procedure then computes parameter estimates on the basis of all available data, including the incomplete cases. Standard errors are derived from an observed or expected information matrix.

With these new software tools, fitting a model with incomplete data is operationally no more difficult than fitting a model with complete data. Because the programs are so easy to use, there is a widespread misconception that ML relieves the user of the onus of thinking about missing data issues, because all necessary adjustments for missing data are performed automatically. Researchers should be aware that these automatic adjustments are satisfactory only to the extent that the data and patterns of missing values satisfy the underlying assumptions, in particular the assumption of MAR. If important causes or correlates of missingness are omitted from the model, even if they are conceptually irrelevant to the analyst's purpose, then the ML parameter estimates may be biased. In our experience, many researchers do not pay attention to this issue; they simply do not consider the possibility of including auxiliary variables relevant to missing data but extraneous to the research hypotheses, because the existing user interfaces and documentation for SEM software do not make it clear how to do this or even raise the possibility that it may be necessary. We stress that this is not an inherent characteristic of the ML paradigm, but a feature of the way it has been implemented and described thus far. At present, it is fair to say that the current array of ML methods available to social scientists tends to favor a restrictive strategy in regard to the use of auxiliary variables rather than an inclusive one.

Multiple Imputation

Multiple imputation (MI) attempts to handle the missing data aspect in advance of the substantive analysis. (For a readable introduction to MI, see Allison, 2001.) In MI, the missing values are replaced by $M > 1$ sets of simulated imputed values, resulting in M plausible but different versions of the complete data. In typical applications, only $M = 5$ to $M = 10$ imputations are sufficient to yield inferences that are highly efficient. Each of the M data sets is analyzed in the same fashion by standard complete-data methods, and the results are combined using simple arithmetic. The M sets of parameter estimates are averaged, and the standard errors are combined using simple arithmetic to yield one set that reflects missing data uncertainty as well as ordinary sampling variation. The theoretical basis for MI and basic rules for combining imputations were described by Rubin (1987); computational strategies for generating the imputations were given by Schafer (1997).

Procedures for creating MIs are model based; the imputer must specify a joint probability model for the observed and missing data. An imputation model should be flexible or general enough to preserve the effects of scientific interest for the subsequent analyses. For example, the program NORM (Schafer, 1999b) imputes under the assumption that the sample data are drawn from a multivariate normal population with an arbitrary covariance or correlation structure. Data sets imputed with NORM are therefore compatible with any standard SEM software, because the user-specified covariance patterns allowed by these programs are all special cases of the general or unstructured one. Examples of other programs implementing MI under an assumption of multivariate normality are the free program Amelia (King, Honaker, Joseph, & Scheve, 2001); the SAS procedure, Proc MI (SAS Institute, 2001); and LISREL 8.5 (Jöreskog & Sörbom, 2001). Like the ML procedures described previously, all of these programs also assume that the missing data are MAR; the user does not (indeed, cannot) supply any information about how the probabilities of missingness may be related to variables in the system. Insofar as these probabilities are related only to data that have been supplied, the imputations will not carry any biases due to nonignorability.

In addition to the model and assumptions about missingness, generating MIs also requires a prior distribution for model parameters. MIs are generated under the Bayesian framework, in which unknown parameters are regarded as random and the user's current state of knowledge about the parameters is expressed through a probability distribution. (A readable modern overview of Bayesian statistics is provided by Gelman, Carlin, Stern, & Rubin, 1995.) The choice of a prior distribution is always somewhat subjective. In typical applications, analysts will tend to choose prior distributions that are diffuse, spreading the probability content over a wide range of parameter values, to avoid introducing subjective biases and to allow the data to speak for themselves. In all of the MI software products mentioned above, a diffuse prior distribution is chosen by default.

From an operational standpoint, MI is quite different from ML because the missing data are handled in a step that is entirely separate from the analysis. The implications of this separation are both positive and negative. On the negative side, there is a danger that a researcher may proceed to analyze the imputed data without regard for the manner in which the imputations were created. If the imputation model was not flexible or general enough to encompass the subsequent analyses, the results could be biased. For example, a multivariate normal model allows pairwise associations among variables but not interactions; a data set that has been imputed with NORM in the standard way may tend to exhibit interactions that are weaker than those found in the population. On the positive side, imputing for missing values as a separate operation makes it quite easy to add auxiliary variables. With MI it is straightforward to include auxiliary variables in the imputation model, and the current literature on MI encourages researchers to do so (e.g., Schafer & Olsen, 1998). It is not necessary to include these auxiliary variables in subsequent analyses; their full effect on the missing data adjustment occurs during MI, and is automatically carried forward into subsequent analyses on any subset of the imputed data. In regard to auxiliary variables, we will therefore say that the MI paradigm fosters—or at least does not discourage—the use of an inclusive strategy as opposed to a restrictive one.

Conditions in Which MI and ML Yield Similar Answers

Despite the operational differences between ML and MI, there are many situations in which they lead to similar results. Whether this occurs depends on subtle interplay among three models: (a) the model that underlies the ML procedure, (b) the model that underlies the MI procedure (i.e., the model used to generate the imputations), and (c) the model used to analyze the imputed data sets. Notice that these models differ in scope; (c) makes distributional assumptions about the complete-data population, whereas (a) and (b) make additional assumptions about the missingness (typically, that the missing values are MAR) and (b) also includes a prior distribution for parameters.

In the discussion that follows, we limit our attention to situations where the model for the complete-data population in the ML procedure agrees with the model used for postimputation analyses. In other words, we suppose that, if there were no missing values, the user of ML and the user of the multiply imputed data set would analyze the data in the same fashion and reach identical conclusions. Without this assumption it would be very difficult to proceed because there would be no guarantee that the same population parameters are being estimated under the two methods. We also limit our attention to situations where the user of ML and the imputer both assume that the missing values are MAR. This does not mean that the ML user and the imputer make identical assumptions about the missing data, because their models may still be different; for example, they may describe different sets of variables (i.e., may have different sets of auxiliary variables) or may specify different intervariable relationships (e.g., an unstructured covariance matrix versus a structured one).

Proposition 1: If the user of the ML procedure and the imputer use the same set of input data (same set of variables and observational units), if their models apply equivalent distributional assumptions to the variables and the relationships among them, if the sample size is large, and if the number of imputations, M, is sufficiently large, then the results from the ML and MI procedures will be essentially identical.

Under these conditions, it can be easily shown that MI approximates a Bayesian analysis under the same model used in the ML procedure (Schafer, 1997, Section 4.3.2). The asymptotic equivalence between Bayesian and likelihood-based procedures is a well-known result in mathematical statistics; see Gelman et al. (1995) for discussion and further references. With large samples, the effect of a prior distribution diminishes, and Bayesian and likelihood-based analyses produce very similar results, especially when the prior is relatively diffuse.

For a practical example, suppose that we want to fit a linear regression of a single variable, Y, on a set of predictors, $X_1, X_2, \ldots, X_p$, where missing values arise in the predictors. In an SEM procedure, the model would specify a path from each of the manifest predictors, $X_1, X_2, \ldots, X_p$, to the manifest response, Y, and a Y-error that is uncorrelated with the predictors. In addition, the model must allow correlations among $X_1, X_2, \ldots, X_p$, because an ordinary regression model does not restrict relationships among predictors. If the sample size is large, the parameter estimates and standard errors from SEM programs such as Amos or Mx will be essentially identical to those that would be obtained if the missing values were multiply imputed M times using MI software such as NORM, if the imputed datasets were analyzed using conventional regression software, and if the results were combined according to Rubin's (1987) rules. For another example, suppose that we want to estimate a covariance or correlation matrix from an incomplete data set. One way to accomplish this is to multiply impute the missing values M times using NORM and compute the M sample covariance or correlation matrices. The average of these matrices would agree closely with the ML covariance or correlation matrix produced by SPSS, EMCOV, or the EM algorithm provided in NORM.

In most applications of MI, the number of imputations, M, is not large; typical values are $M = 5$ or $M = 10$. For any $M < \infty$, estimates from MI will contain an additional amount of random error not found in ML estimates, because MI is a Monte Carlo method. As a result, the MI estimates will tend to have somewhat larger variance along with wider confidence intervals. Unless the rates of missing information are unusually high, however, the extra noise in the MI estimates tends to be barely noticeable. Rubin (1987) showed that the relative efficiency of MI to an efficient ML procedure is approximately $(1 + \gamma/M)^{-1}$, where γ is the rate of missing information. For example, with 30% missing information, estimates based on $M = 5$ imputations will have about $1/(1 + 0.3/5) = 94.3\%$ of the efficiency of ML estimates, and the standard errors from MI will tend be about $(1/0.943)^{1/2} = 1.03$ times as wide or 3% wider than the standard errors from ML.

Our first simulation study, which we report in a later section, demonstrates in a simple example that MI and ML do indeed produce very similar answers when the ML and imputer's models are equivalent. When MI is carried out under these conditions, the imputer's and analyst's procedures are said to be *congenial* (Meng, 1994). Under congeniality, there is no major theoretical reason to prefer ML estimates to MI or vice versa, because the properties of the two methods are very similar. In many real-world applications of MI, however, the imputer's and analyst's models are uncongenial. Some types of uncongeniality are clearly harmful; for example, an imputer might create MIs under a model that is grossly misspecified. But Meng (1994) and Rubin (1996) have demonstrated that uncongeniality can also be beneficial, particularly when the imputer can take advantage of extra information unavailable or unused by the analyst. Our next two propositions describe two different kinds of uncongeniality; the first is rather benign, whereas the second may help or harm. We continue to assume but no longer state that the sample size is large and the number of imputations, M, is sufficient to make the simulation-induced error in MI estimates negligible.

> *Proposition 2:* If the user of the ML procedure and the imputer use the same set of input data (same set of variables and observational units) but the imputer's model is more general than the analyst's model (e.g., has additional parameters), then ML and MI will tend to produce similar parameter estimates, but the standard errors from MI will tend to be somewhat larger.

This statement is more of an observation based on personal experience than a mathematical theorem, but it is intuitively reasonable and could no doubt be made mathematically rigorous by introducing additional technical conditions.

This situation does arise quite often in practice. For example, suppose two researchers want to fit a random coefficients growth curve model to a response measured at p occasions, $Y_1, Y_2, \ldots, Y_p$. Researcher A uses HLM, which can accommodate missing responses, producing ML estimates directly from the incomplete data. His colleague, researcher B, first imputes the missing values in $Y_1, Y_2, \ldots, Y_p$ using NORM and analyzes the imputed data in HLM. She finds that her parameter estimates are nearly identical to A's. She also finds that most of her standard errors are about the same as A's, but a few are slightly larger. The estimates are very similar because NORM and HLM both apply an assumption of multivariate normality to $Y_1, Y_2, \ldots, Y_p$. The model fit by HLM assumes a particular covariance structure for $Y_1, Y_2, \ldots, Y_p$, but NORM imputes the missing values by using an unstructured (i.e., saturated) covariance model that includes additional parameters. These additional parameters in the imputation model introduce some added variation in the imputed values, which causes some of the standard errors from MI to be larger than those resulting from ML applied to the incomplete data. The increase in the standard errors is not an artifact or a bias; it is real and reflects the slightly increased uncertainty in MI under these conditions. In our experiences with real and simulated data, we have found that this increase of standard errors in MI tends to be slight or entirely unnoticeable; some examples of this are presented in our simulation results.

> *Proposition 3:* If the user of the ML procedure and the imputer use the same sample of observational units but a different set of variables, then the results from ML and MI could be quite different, even though the ML user's model is equivalent to the model used by the analyst of the imputed data set.

This situation could easily arise in practice. For example, suppose researcher C creates multiple imputations for a set of variables, $Y_1, Y_2, \ldots, Y_p$, using NORM, but also includes in the imputation procedure some additional covariates, $Z_1, Z_2, \ldots, Z_q$, which may or may not be strongly related to $Y_1, Y_2, \ldots, Y_p$ or to the processes creating missing values on these variables. After imputation, he uses standard SEM software to fit a covariance structure model to $Y_1, Y_2, \ldots, Y_p$ in each of the imputed data sets and combines the results by Rubin's (1987) rules. Researcher D fits the same covariance structure model directly without imputation to $Y_1, Y_2, \ldots, Y_p$ using Amos and finds that her results differ from C's. Some of her parameter estimates are very similar to C's, but others are different enough to cause worry; some of her standard errors are similar to C's, but some are smaller and others are larger.

In this example, discrepancies arise not because of inherent differences between MI and ML but because researcher C included a set of auxiliary variables (Z_1, Z_2, . . . , Z_q), whereas D did not. If D had figured out a way to include Z_1, Z_2, . . . , Z_q in her Amos model and C had omitted them from the imputation procedure, then the results would have been essentially reversed. The main issue of practical importance is not whether one should use ML or MI, but whether the addition of auxiliary variables was helpful or harmful to the statistical inference. The distinctions between the ML and MI methodologies become important only insofar as they encourage the user to adopt an inclusive or a restrictive strategy with respect to auxiliary variables.

Four Key Questions Concerning Restrictive Versus Inclusive Strategies

Let Y denote a variable or group of variables of scientific interest on which missing values occur. Variables that are candidates for inclusion in a set of auxiliary variables may be sorted into three categories: (a) Variable is correlated with Y and is a correlate of missingness, (b) variable is correlated with Y and is not a correlate of missingness, and (c) variable is uncorrelated with Y and is or is not a correlate of missingness.

The first key question concerns the restrictive strategy, which suggests using as few auxiliary variables as possible. A typical restrictive strategy would not include any Category b or Category c variables, and might omit one or more Category a variables. When a Category a variable is omitted from the set of auxiliary variables, it is possible that this omission will contribute to bias in the resulting parameter estimates, a disruption of coverage, or both. Parameter estimation is biased when the average value of the estimate over repeated samples differs from the population parameter. Coverage refers to the percentage of time that a confidence interval contains the true parameter value. Coverage is disrupted when a confidence interval with a nominal confidence coefficient of, for example, 95%, includes the true value of the parameter appreciably less than 95% of the time.

Large biases create problems in estimation, intervals, and tests. For example, suppose that a null hypothesis specifies that a particular population mean equals 0. Also suppose that in reality $\mu = 0$ (i.e., the null hypothesis is true). Now suppose that μ is estimated using a biased procedure in which the average estimate is 2. Under these circumstances the nominal 95% confidence interval will, on average, be centered at 2 rather than at 0. This leads to a disruption of coverage, the null hypothesis will be rejected more frequently than it should be, and the Type I error rate will be increased over the nominal level. Bias can also lead to problems with the Type II error rate. Suppose that in reality μ is substantial and positive (i.e., the null hypothesis is false). If the estimate of μ is negatively biased, the null hypothesis may fail to be rejected more often than it should be, increasing the Type II error rate and decreasing statistical power. This leads to the first key question, addressed by Study 1: How serious a problem is it when a Category a variable is omitted from the analysis?

The remaining questions concern the inclusive strategy. This strategy tends to include more auxiliary variables rather than fewer. A typical inclusive strategy may include some Category b variables. One potential benefit of including Category b variables is increased efficiency. The loss of information that accompanies any missing data leads to increased uncertainty in parameter estimation, in other words, larger standard errors, which translates into a loss of power. One potential benefit of including Category b variables as auxiliary variables is that, to the extent they are correlated with Y, they may be able to reduce uncertainty in the data and help to regain some of this lost power. This leads to the second key question, addressed by Study 2: Will including Category b variables improve precision, without adversely affecting bias or coverage?

A second potential benefit may occur when Category b variables are included when there is nonignorable missingness. As discussed above, ML and MI procedures in use today adjust for ignorable (MAR) missingness only. When missing data are MNAR rather than MAR, procedures that assume MAR could be biased. To the extent that Category b variables are correlated with Y, they may help to mitigate some of the bias, because they may serve as useful proxies for Y when Y is missing. Thus, the third key question, addressed by Study 3, is: Will including Category b variables when there is nonignorable missingness mitigate bias?

One inevitable result of an inclusive strategy is the occasional inclusion of Category c variables, which in most cases have been mistaken for Category a or b variables. Because these variables are uncorrelated with Y, there is no expected benefit associated with including them in the model. This leads to the final key question, addressed by Study 4: Is there any disadvantage or cost associated with including Category c variables?

Simulation Studies

Overview of Design

The purpose of the four major simulation studies reported in this section is to address the four key questions concerning the restrictive versus inclusive strategies, as described above. The design of each simulation varied somewhat depending on what was needed to address the question posed. Some general aspects of the designs are described here.

Data generated. We drew independent random samples of $N = 500$ "participants" each from multivariate normal populations involving at most three variables: X, which is always observed; Y, which may be either observed or missing; and Z, which is always observed and is a potential cause of missingness for Y. X and Y represent variables that will automatically appear in an analysis because they are of scientific interest, whereas Z represents an auxiliary variable that is not of direct interest but might be included in an imputation model or ML procedure if the analyst believes it may be beneficial.

In our simulations, we arbitrarily set the population parameters to $\mu_X = 5.00$, $\mu_Y = 5.20$, $\sigma_X = 1.00$, $\sigma_Y = 1.44$, and $\rho_{XY} = 0.60$. The cause of missingness, Z, was assigned to have $\mu_Z = 0$, $\sigma_Z = 1$, and two different degrees of correlation with Y: $\rho_{YZ} = .40$ and $\rho_{YZ} = .90$. In these two scenarios we set ρ_{XZ} to be .24 and .54, respectively, so that the association between X and Z is fully explained by their mutual associations with Y.

Missing data mechanisms. Four conditions with different varieties of MAR missing data mechanisms were used in the simulation. Under each of these methods, the details of the mechanism were adjusted to yield an overall rate of missingness of either 25% or 50%.

1. In the MCAR condition, missing values were imposed on Y independently of X, Y, or Z at a rate of 25% or 50%.
2. In the MAR-linear conditions, the probability of missingness was linearly related to Z. One example of this would be a survey where individuals with higher values of income (Z) have higher probabilities of nonresponse to a question about use of financial services (Y). To achieve this, we divided the distribution of Z into quartiles and set the missingness probabilities for the four regions to (.1, .2, .3, .4) for an overall rate of 25% or (.2, .4, .6, .8) for an overall rate of 50%.
3. In the MAR-convex conditions, the probability of missingness was larger at the extremes of Z and smaller in the middle. For example, those at the low and high extremes of income are less likely to respond to a question about the use of financial services than those in the middle. To achieve this, the distribution of Z was again divided into quartiles, and the missingness probabilities for the four regions were set to (.4, .1, .1, .4) or (.8, .2, .2, .8).
4. In the MAR-sinister conditions, the probability of missingness was a function of the correlation between Z and X. An example of this would be a data set where a subgroup of individuals for whom the correlation between income (Z) and use of financial services at Time 1 (X) is high is more likely to provide missing data on use of financial services at Time 2 (Y). This mechanism is called "sinister" because we created it specifically to introduce biases in the X–Y association when Z is not taken into account. To implement this mechanism, we randomly divided the $N = 500$ into 50 groups of size 10 and calculated the sample correlation coefficient within each group. The 25 groups with the lowest correlations were assigned to a low correlation stratum and the remainder were assigned to a high correlation stratum. The missingness rates for the two strata were then set to (.1, .4) or (.2, .8) to yield overall missingness rates of 25% or 50%.

Note that in actuality, Mechanisms 2, 3, and 4 will be MAR only if Z appears in the procedure. If Z is omitted, then the mechanism is actually MNAR and procedures based on an ignorability assumption may be biased. In one study, we allow Y to take the place of Z in Mechanisms 2, 3, and 4 above, forcing it to be MNAR.

Number of replications. For each cell of the simulations, 1,000 replicate samples were drawn.

Parameters to be estimated. These artificial data sets can be used to examine many quantities of particular interest to social scientists, who typically make extensive use of ANOVA, correlation, and regression procedures. We examined the performance of estimates for the following population parameters: μ_Y; σ^2_Y; β_{YX}, the regression weight for the regression of Y

on X, which represents a situation with a completely observed predictor and an incompletely observed criterion; β_{XY}, the regression weight for the regression of X on Y, which represents a situation with an incompletely observed predictor and a completely observed criterion; and ρ_{XY}. Clearly, these quantities are not independent; nevertheless, it is instructive to examine them separately. In our simulations, we also examined the behavior of estimates of $\mu_Y - \mu_X$; if X and Y represent measurements of the same quantity at different points in time, then this difference represents an average change. The behavior of $\mu_Y - \mu_X$ closely mimicked the behavior of μ_Y, however, so for brevity we do not present these results.

Evaluation Criteria

When describing the performance of the various methods, we focus on the four quantities listed below.

Standardized bias. The raw bias in a parameter estimate is the difference between its average value over the simulation repetitions and the true parameter it estimates. With 1,000 repetitions, it is quite easy for a bias to be deemed statistically significant even though it may not be practically significant. The impact of bias on intervals and tests depends on how large it is relative to the overall uncertainty in the system. For interpretability, we report the bias as a percentage of the standard error, 100 × (average estimate − parameter)/*SE*, where *SE* is the standard deviation of the estimates. For example, a value of −50% means that the estimate on average falls one-half of a standard error below the parameter, roughly equivalent to one-eighth of the width of a typical confidence interval. We have found that, as a useful rule of thumb, once the standardized bias exceeds 40%–50% in a positive or negative direction, the bias begins to have a noticeable adverse impact on efficiency, coverage, and error rates. Therefore, we consider any standardized bias with an absolute value greater than 40% as practically significant.

Root-mean-square error (RMSE). A useful measure of overall accuracy is the mean-square error (*MSE*), the average squared difference between the estimate and its target. This measure of accuracy combines the concepts of bias and efficiency, because it can be shown that the *MSE* of an estimate is equal to its squared bias plus its variance. For easier interpretation, we report the square root of the *MSE*, to put it on the same scale as the parameter.

Coverage. If a procedure is working well, the actual coverage should be approximately equal to or greater than the nominal or advertised coverage rate. In our simulations, we report the actual coverage of nominal 95% intervals. Accurate coverage translates directly to an accurate Type I error rate. If the coverage of a nominal 95% interval is actually 90%, then the actual Type I error rate for a testing procedure with a .05-level criterion is twice as high as it ought to be. We regard the performance of the interval procedure to be troublesome if its coverage drops below 90%.

Average length of confidence interval. If one procedure has a similar or higher rate of coverage than another but yields intervals that are substantially shorter, then it should be preferred. Shorter intervals translate into greater accuracy and higher power.

A Preliminary Comparison of MI and ML

We have previously asserted that if a user of MI and a user of ML use the same input data and apply equivalent distributional assumptions, the results from the MI and ML procedures will be essentially identical. To illustrate this point, we performed a preliminary study where we drew samples of X and Y and imposed missing values on Y in an MCAR fashion at rates of 25% and 50%. From each sample, we then calculated estimates and standard errors by an ML and an MI procedure. In ML, we maximized the bivariate normal likelihood based on X and the observed values of Y and obtained standard errors by inverting the observed information matrix (see Little & Rubin, 1987, Sections 6.2 and 6.3.1). In MI, we first created $M = 10$ imputations using the data augmentation procedure described by Schafer (1997, chapter 5) as implemented in NORM with the default noninformative prior; we then calculated parameter estimates and standard errors from each imputed data set by maximum likelihood based on the observed and imputed data and combined the results using Rubin's (1987) rules.

The results of this preliminary study are displayed in Table 1. The upper half of the table reports the standardized bias and *RMSE* of the parameter estimates, and the lower half reports the simulated coverage and average width of the nominal 95% intervals. In all cases, the parameter estimates are essentially unbiased and the coverage of the intervals is essentially correct. As the rate of missing values for Y increases from 25% to 50%, the *RMSE* and the average interval width both increase, reflecting the natural loss of information that occurs with higher rates of missing data. Comparing the performance of

Table 1
Correspondence Between Maximum-Likelihood (ML) and Multiple Imputation (MI) When Data in Y Are Missing Completely at Random

Parameter estimated	25% missing		50% missing	
	ML	MI	ML	MI
Standardized bias and root-mean-square error				
μ_Y				
Standardized bias	−5.9	−5.1	−4.1	−4.2
RMSE	.06	.06	.07	.07
σ^2_Y				
Standardized bias	−1.7	1.0	−6.3	−0.2
RMSE	.10	.10	.12	.13
β_{YX}				
Standardized bias	2.7	3.1	−1.3	−1.5
RMSE	.05	.05	.06	.06
β_{XY}				
Standardized bias	1.8	−0.7	4.1	−1.8
RMSE	.03	.03	.04	.04
ρ_{XY}				
Standardized bias	1.3	0.3	−0.1	−4.3
RMSE	.03	.03	.04	.04
Coverage and width of interval				
μ_Y				
Coverage	94.3	94.1	96.1	96.0
Width of interval	.24	.24	.27	.28
σ^2_Y				
Coverage	95.8	95.5	95.3	93.8
Width of interval	.41	.42	.50	.51
β_{YX}				
Coverage	95.6	95.3	95.6	94.6
Width of interval	.20	.20	.24	.25
β_{XY}				
Coverage	95.4	95.3	95.5	94.9
Width of interval	.13	.13	.15	.15
ρ_{XY}				
Coverage	95.0	95.3	95.3	94.7
Width of interval	.13	.13	.15	.16

Note. *RMSE* = root-mean-square error.

ML and MI, we find the two methods practically indistinguishable. With 50% missingness, we see that the interval estimates from ML are slightly narrower, suggesting that ML is slightly more efficient than MI because the simulation aspect of MI introduces a minor amount of additional error. If the number of imputations were increased (e.g., to $M = 20$), these minor differences would vanish.

In all of the remaining simulations, we compared the performance of ML and MI and found exactly the same thing: Whenever ML and MI use the same input data and model, the results from the two methodologies are practically indistinguishable. Differences do arise if the inputs vary. For example, if missing values are generated under one of the MAR mechanisms, and the cause of missingness, *Z*, is added to the imputation procedure but not to the ML procedure, then MI may tend to outperform ML. But this is less a comparison of MI versus ML than a comparison between an inclusive missing data strategy versus a restrictive one. Therefore, in all of the remaining simulations below, we report the results only for MI and omit the results from ML to avoid redundancy.

Study 1

The purpose of Study 1 was to investigate the effect on results of omitting a Category a variable (i.e., a cause of missingness) on bias, *RMSE,* coverage, and length of confidence interval. With sample data generated as previously described, MI-based estimates and standard errors were produced with or without *Z* in the imputation model. It was expected that in the conditions where ρ_{YZ} was stronger, the omission of *Z* would lead to greater bias, worse coverage, and possibly longer confidence intervals.

Tables 2, 3, and 4 show standardized bias, *RMSE,* coverage, and width of confidence intervals for the MAR-linear, MAR-convex, and MAR-sinister conditions, respectively. These tables show results with and without *Z* included in the imputation procedure. Across the board, inclusion of the *Z* variable results in acceptable parameter estimation and coverage, indicating that the missing-data procedure is operating as expected. But even when an important missingness-related variable is omitted, the results appear surprisingly robust in many instances. Table 2 shows results for the MAR-linear missing data mechanism. For σ^2_Y, β_{YX}, β_{XY}, and ρ_{XY}, there is little bias in estimation when *Z* is omitted except in the most extreme condition, where there is 50% missing and $\rho_{YZ} = .9$. Even in this condition the bias in estimation of β_{XY} does not exceed our significance criterion of 40. However, the situation is different for the estimate of μ_Y, which is affected by omission of *Z* in every condition. This MAR-linear mechanism tends to remove higher values of *Y,* shifting the mean of the observed values downward and, to a lesser extent, reducing their variability. *RMSE* is generally larger in the conditions where *Z* is omitted, which follows from the relationship between *MSE* and bias. For most of the conditions where there is significant bias, there is correspondingly a disruption in coverage. Coverage is

Table 2

Standardized Bias, Root-Mean-Square Error (RMSE), Coverage, and Width of Confidence Interval When Data in Y Are Missing at Random-Linear (MAR-Linear)

	25% missing				50% missing			
	ρ_{YZ} = .4		ρ_{YZ} = .9		ρ_{YZ} = .4		ρ_{YZ} = .9	
Parameter estimated	*Z* included	No *Z*	*Z* included	No *Z*	*Z* included	No *Z*	*Z* included	No *Z*
	Standardized bias and root-mean-square error							
μ_Y								
Standardized bias	−7.8	**−76.7**	−8.4	**−163.5**	−7.2	**−186.7**	−7.1	**−441.8**
RMSE	.06	.08	.06	.12	.08	.15	.06	.31
σ^2_Y								
Standardized bias	−0.1	−5.4	−2.4	−18.8	3.9	−24.6	0.9	**−146.5**
RMSE	.11	.11	.10	.11	.13	.13	.11	.21
β_{YX}								
Standardized bias	−1.1	−5.4	−3.7	−16.2	−4.0	−25.1	−2.9	**−107.6**
RMSE	.05	.05	.05	.05	.06	.06	.05	.09
β_{XY}								
Standardized bias	−1.3	−0.3	−1.8	3.1	−7.0	1.2	−2.6	39.7
RMSE	.03	.03	.03	.03	.04	.04	.03	.04
ρ_{XY}								
Standardized bias	−3.1	−5.2	−4.8	−10.0	−8.8	−17.6	−5.3	**−48.0**
RMSE	.03	.03	.03	.03	.04	.04	.03	.05
	Coverage and width of interval							
μ_Y								
Coverage	94.5	**87.6**	95.7	**61.3**	94.1	**54.9**	95.3	**1.1**
Width of interval	.24	.24	.22	.24	.30	.28	.24	.27
σ^2_Y								
Coverage	94.9	94.5	94.6	93.2	94.0	92.5	94.6	**61.6**
Width of interval	.42	.42	.39	.41	.51	.50	.44	.45
β_{YX}								
Coverage	94.4	94.5	93.8	93.9	93.7	94.0	94.6	**79.1**
Width of interval	.20	.20	.18	.20	.24	.25	.20	.24
β_{XY}								
Coverage	95.4	95.4	95.8	95.2	95.4	94.6	95.7	93.2
Width of interval	.13	.13	.12	.13	.15	.15	.14	.17
ρ_{XY}								
Coverage	94.9	94.7	94.1	94.3	94.2	94.8	94.0	93.3
Width of interval	.13	.13	.12	.13	.16	.16	.13	.17

Note. In the MAR-linear conditions, the probability of missingness is linearly related to *Z*. When *Z* is not included, the situation becomes *missing not at random.* The following are in bold type: standardized bias with an absolute value > 40; coverage < 90%.

particularly poor for μ_Y. Widths of the confidence intervals in most conditions appear roughly equivalent to those in Table 1.

Table 3 shows results for the MAR-convex missing data mechanism. In contrast to the MAR-linear missing data mechanism, under MAR-convex missingness the estimate of μ_Y is almost unaffected in all conditions when *Z* is omitted. For the remaining parameters, bias is acceptably low in the least extreme condition, where ρ_{YZ} = .4 and there is 25% missingness. Significant bias is introduced in the estimates of all parameters except μ_Y in the conditions where ρ_{YZ} = .9. Bias is also unacceptable in the 50% missingness condition, where ρ_{YZ} = .4, for the estimates of σ^2_Y and β_{YX}. The bias is negative for all parameters except β_{XY}. Our MAR-convex mechanism tends to remove high and low values of *Y* in a symmetric fashion, leaving the mean unchanged but reducing the variance of *Y* and its covariance with other variables. As expected, *RMSE* is larger in conditions where there is appreciable bias. Coverage is acceptable across the board for μ_Y, and for all parameters in the least ex-

Table 3
Standardized Bias, Root-Mean-Square Error (RMSE), Coverage, and Width of Confidence Interval When Data in Y Are Missing at Random-Convex (MAR-Convex)

	25% missing				50% missing			
	ρ_{YZ} = .4		ρ_{YZ} = .9		ρ_{YZ} = .4		ρ_{YZ} = .9	
Parameter estimated	*Z* included	No *Z*	*Z* included	No *Z*	*Z* included	No *Z*	*Z* included	No *Z*
	Standardized bias and root-mean-square error							
μ_Y								
Standardized bias	−5.3	−6.0	−6.5	−5.4	−2.7	−1.9	−5.1	−2.5
RMSE	.06	.06	.05	.06	.07	.07	.06	.06
σ^2_Y								
Standardized bias	−3.0	−32.2	1.2	**−173.2**	7.3	**−81.2**	4.0	**−545.5**
RMSE	.10	.11	.10	.20	.14	.16	.13	.57
β_{YX}								
Standardized bias	6.5	−18.8	7.4	**−120.5**	2.7	**−58.6**	5.5	**−363.8**
RMSE	.05	.05	.05	.08	.06	.07	.05	.23
β_{XY}								
Standardized bias	0.4	11.5	2.6	**47.4**	−5.5	23.6	1.7	**113.8**
RMSE	.03	.03	.03	.04	.04	.04	.04	.08
ρ_{XY}								
Standardized bias	2.6	−7.0	4.7	**−51.3**	−4.4	−27.8	1.9	**−151.6**
RMSE	.03	.03	.03	.04	.04	.04	.03	.09
	Coverage and width of interval							
μ_Y								
Coverage	95.5	95.3	95.3	94.7	94.9	95.1	95.6	95.3
Width of interval	.24	.24	.22	.22	.27	.27	.23	.22
σ^2_Y								
Coverage	95.5	92.6	95.2	**51.2**	94.2	**81.3**	95.5	**0.0**
Width of interval	.43	.41	.40	.37	.56	.48	.52	.33
β_{YX}								
Coverage	95.2	93.9	96.0	**72.8**	93.5	**89.1**	94.6	**5.2**
Width of interval	.20	.20	.18	.20	.25	.24	.21	.23
β_{XY}								
Coverage	95.1	94.9	95.2	91.7	95.7	94.8	95.3	**79.3**
Width of interval	.13	.13	.13	.14	.16	.16	.14	.21
ρ_{XY}								
Coverage	96.0	94.9	95.3	93.0	94.3	93.9	94.0	**71.1**
Width of interval	.13	.13	.12	.14	.15	.16	.13	.19

Note. In the MAR-convex conditions, the probability of missingness is larger at the extremes of *Z* and smaller in the middle. When *Z* is omitted, the situation becomes *missing not at random.* The following are in bold type: standardized bias with an absolute value > 40; coverage < 90%.

treme condition. In most conditions except the least extreme, when *Z* is omitted coverage drops to unacceptable levels for all estimates except the estimate of μ_Y.

Table 4 shows results for the MAR-sinister missing data mechanism. When *Z* is omitted under this missing data mechanism, the estimate of μ_Y is almost unaffected. Also, bias is acceptably low for all parameters in the least extreme condition. The estimates of β_{YX} and ρ_{XY} show bias in all other conditions. The estimate of β_{XY} is biased in the conditions where ρ_{YZ} = .9, and the estimate of σ^2_Y is biased in the 50% missingness conditions. In the most extreme condition, bias is unacceptable for all parameters except μ_Y. In all cases the bias is negative. MAR-sinister was specifically designed to disrupt the correlation between *Y* and *Z* and, as a consequence, the relationship between *Y* and variables related to *Z* (i.e., *X*). When *Z* is omitted, coverage remains acceptable for all parameters in most conditions. Exceptions are the most ex-

Table 4
Standardized Bias, Root-Mean-Square Error (RMSE), Coverage, and Width of Confidence Interval When Data in Y Are Missing at Random-Sinister (MAR-Sinister)

Parameter estimated	25% missing				50% missing			
	ρ_{YZ} = .4		ρ_{YZ} = .9		ρ_{YZ} = .4		ρ_{YZ} = .9	
	Z included	No *Z*	*Z* included	No *Z*	*Z* included	No *Z*	*Z* included	No *Z*
	Standardized bias and root-mean-square error							
μ_Y								
Standardized bias	2.9	2.0	1.7	0.1	3.8	4.2	2.9	2.3
RMSE	.06	.06	.05	.06	.07	.07	.06	.07
σ^2_Y								
Standardized bias	2.3	−20.1	−6.4	−32.1	−1.1	**−50.5**	−2.0	**−76.8**
RMSE	.11	.11	.09	.11	.13	.14	.11	.15
β_{YX}								
Standardized bias	−3.4	−29.5	−4.2	**−65.5**	−3.3	**−74.0**	−2.2	**−161.8**
RMSE	.05	.05	.05	.06	.06	.08	.05	.12
β_{XY}								
Standardized bias	−4.2	−12.9	−1.4	**−41.1**	−2.8	−28.2	−2.5	**−101.6**
RMSE	.03	.03	.03	.04	.04	.04	.03	.06
ρ_{XY}								
Standardized bias	−6.4	−28.5	−5.1	**−67.2**	−6.4	**−64.2**	−4.8	**−154.0**
RMSE	.03	.03	.03	.04	.04	.05	.03	.08
	Coverage and width of interval							
μ_Y								
Coverage	95.3	95.7	95.2	95.6	95.1	95.6	95.3	94.6
Width of interval	.24	.24	.22	.24	.27	.27	.23	.28
σ^2_Y								
Coverage	95.0	93.1	95.3	93.4	94.9	**89.8**	95.3	**85.2**
Width of interval	.42	.41	.39	.41	.52	.49	.43	.49
β_{YX}								
Coverage	94.9	94.6	95.7	90.4	95.9	**89.8**	95.4	**66.4**
Width of interval	.20	.20	.18	.20	.25	.25	.20	.26
β_{XY}								
Coverage	95.7	94.7	94.7	94.3	95.3	94.6	95.3	**85.7**
Width of interval	.13	.13	.12	.14	.15	.16	.13	.17
ρ_{XY}								
Coverage	95.1	95.2	95.8	92.7	94.8	92.4	95.1	**72.0**
Width of interval	.13	.13	.12	.14	.15	.17	.13	.18

Note. In the MAR-sinister conditions, the probability of missingness is a function of the correlation between *Z* and *X*. When *Z* is omitted, the situation becomes *missing not at random.* The following are in bold type: standardized bias with an absolute value > 40; coverage < 90%.

treme condition, where coverage is unacceptable except for μ_Y, and one condition where coverage drops just below .9 for β_{YX} and σ^2_Y.

These results indicate that the effects of the three missing data mechanisms used here vary across the different parameters examined. The omission of *Z* greatly affected estimation of μ_Y under MAR-linear missingness, and had very little effect under MAR-convex and MAR-sinister missingness. By contrast, estimates of the remaining parameters were more greatly affected by the omission of *Z* under MAR-convex and MAR-sinister missingness.

Study 2

The purpose of Study 2 was to assess the effect of including a Category b variable as an auxiliary variable when missingness is ignorable. In this study, each data set contained *X*, *Y*, and two additional third variables: a *Z* such that ρ_{YZ} = .4, and a *Z* such that ρ_{YZ} = .9. Missingness was imposed in an MCAR fashion at the high rate of 50%, and MI-based estimates were obtained with and without *Z* in the imputation procedure.

Table 5 shows standardized bias, *RMSE*, coverage,

Table 5
Standardized Bias, Root-Mean-Square Error (RMSE), Coverage, and Width of Confidence Interval When a Category b Variable Is Included: Ignorable Missingness, 50% Missing

Parameter estimated	Variables included: X, Y only	X, Y, Z: ρ_{YZ} = .4	X, Y, Z: ρ_{YZ} = .9
Standardized bias and root-mean-square error			
μ_Y			
Standardized bias	−1.2	−1.1	−0.3
RMSE	.07	.07	.06
σ^2_Y			
Standardized bias	5.3	5.8	−0.8
RMSE	.13	.13	.11
β_{YX}			
Standardized bias	5.2	5.7	2.9
RMSE	.06	.06	.05
β_{XY}			
Standardized bias	−1.9	−1.9	0.7
RMSE	.04	.04	.03
ρ_{XY}			
Standardized bias	−0.1	0.0	0.4
RMSE	.04	.04	.03
Coverage and width of interval			
μ_Y			
Coverage	95.2	94.0	95.1
Width of interval	.28	.27	.23
σ^2_Y			
Coverage	94.4	94.6	94.2
Width of interval	.51	.51	.42
β_{YX}			
Coverage	94.9	95.2	95.9
Width of interval	.25	.24	.20
β_{XY}			
Coverage	93.9	94.1	95.5
Width of interval	.15	.15	.13
ρ_{XY}			
Coverage	95.3	95.3	95.4
Width of interval	.16	.15	.13

Note. Z is a Category b variable (i.e., it is correlated with Y and is not a correlate of missingness).

and interval width when only X and Y are included, when ρ_{YZ} = .4 and Z is included, and when ρ_{YZ} = .9 and Z is included. Bias is acceptable in every condition and coverage does not appear to be hurt by inclusion of the additional variable. Width of the confidence interval is shrunk noticeably by addition of the Z variable when ρ_{YZ} = .9, with reductions of 13%–20% over the widths in the X and Y-alone condition. This indicates an increase in statistical power associated with including Z as an auxiliary variable.

Study 3

The purpose of Study 3 was to investigate the effect of adding a Category b variable as an auxiliary variable when missingness is nonignorable. Data were generated in the same manner as in Studies 1 and 2. Each data set contained X, Y, and two additional third variables: a Z such that ρ_{YZ} = .4, and a Z such that ρ_{YZ} = .9. Missing values were imposed on Y at an overall rate of 50% by linear, convex, and sinister conditions, except that the missingness probabilities were determined by the value of Y rather than Z. MI-based estimates were obtained in the following ways: based on (X, Y) only, including the Z associated with a correlation of .4, and including the Z associated with a correlation of .9.

Tables 6, 7, and 8 show the results for nonignorable linear, convex, and sinister missingness, respectively. As expected, there is considerable bias in the conditions where only the X and Y variables are used. The one exception to this is μ_Y, which is estimated well under convex and sinister missingness. The bias is generally negative and is often large in absolute value. One exception is the β_{XY} parameter, where bias is positive for the linear and convex conditions. In most cases, coverage is correspondingly poor, even approaching zero in some cells.

Inclusion of Z reduced bias in nearly every case, often considerably. This held across the three missing data mechanisms. In many cases where ρ_{YZ} = .9, the inclusion of Z cut bias by half or more, and also had a salutary effect on *RMSE*. An exception was the β_{XY} parameter under convex missingness. This was the only parameter with a positive bias, which actually grew a modest amount when Z was added. When Z was added, coverage was improved, and even restored to acceptable levels in many instances. In most cases, the width of the interval was also improved. Although the problems associated with nonignorable missingness were far from completely eliminated when a Category b variable was added, the improvement was sizeable enough to suggest that the inclusion of one or more Category b auxiliary variables is a worthwhile strategy.

Study 4

The purpose of Study 4 was to investigate the effect of including Category c variables as auxiliary variables. Samples of X and Y were generated as before and MCAR missingness was imposed at a rate of 50%. In addition, we generated additional "junk" vari-

Table 6
Standardized Bias, Coverage, and Width of Confidence Interval When a Category b Variable Is Included: Nonignorable Linear Missingness, 50% Missing

Parameter estimated	Variables included: X, Y only	X, Y, Z: $\rho_{YZ} = .4$	X, Y, Z: $\rho_{YZ} = .9$
Standardized bias and root-mean-square error			
μ_Y			
Standardized bias	**−496.5**	**−452.6**	**−162.7**
RMSE	.34	.31	.11
σ^2_Y			
Standardized bias	**−185.1**	**−178.7**	**−85.6**
RMSE	.25	.24	.14
β_{YX}			
Standardized bias	**−138.2**	**−131.7**	**−54.5**
RMSE	.11	.10	.06
β_{XY}			
Standardized bias	**44.0**	**49.9**	38.1
RMSE	.05	.05	.04
ρ_{XY}			
Standardized bias	**−62.4**	**−55.5**	−11.5
RMSE	.05	.05	.03
Coverage and width of interval			
μ_Y			
Coverage	**0.3**	**0.7**	**61.6**
Width of interval	.27	.26	.23
σ^2_Y			
Coverage	**49.4**	**51.9**	**83.5**
Width of interval	.44	.44	.41
β_{YX}			
Coverage	**69.5**	**74.1**	92.3
Width of interval	.24	.24	.19
β_{XY}			
Coverage	90.8	91.3	93.5
Width of interval	.17	.17	.14
ρ_{XY}			
Coverage	91.5	93.8	94.8
Width of interval	.17	.17	.13

Note. Z is a Category b variable (i.e., it is correlated with Y and is not a correlate of missingness). The probability of missingness is linearly related to the cause of missingness. The following are in bold type: standardized bias with an absolute value > 40; coverage < 90%. *RMSE* = root-mean-square error.

Table 7
Standardized Bias, Coverage, and Width of Confidence Interval When a Category b Variable Is Included: Nonignorable Convex Missingness, 50% Missing

Parameter estimated	Variables included: X, Y only	X, Y, Z: $\rho_{YZ} = .4$	X, Y, Z: $\rho_{YZ} = .9$
Standardized bias and root-mean-square error			
μ_Y			
Standardized bias	0.8	2.2	−1.6
RMSE	.05	.05	.05
σ^2_Y			
Standardized bias	**−714.6**	**−677.7**	**−320.3**
RMSE	.71	.69	.39
β_{YX}			
Standardized bias	**−463.3**	**−437.2**	**−211.4**
RMSE	.30	.28	.12
β_{XY}			
Standardized bias	**132.4**	**159.2**	**175.8**
RMSE	.10	.11	.09
ρ_{XY}			
Standardized bias	**−188.4**	**−166.9**	−27.7
RMSE	.12	.10	.04
Coverage and width of interval			
μ_Y			
Coverage	94.6	94.6	94.8
Width of interval	.21	.20	.20
σ^2_Y			
Coverage	**0.0**	**0.0**	**10.1**
Width of interval	.28	.29	.37
β_{YX}			
Coverage	**0.3**	**0.5**	**36.5**
Width of interval	.21	.21	.19
β_{XY}			
Coverage	**72.1**	**65.7**	**57.7**
Width of interval	.25	.24	.16
ρ_{XY}			
Coverage	**50.6**	**62.5**	94.3
Width of interval	.20	.20	.13

Note. Z is a Category b variable (i.e., it is correlated with Y and is not a correlate of missingness). The probability of missingness is larger at extreme vales of the cause of missingness and smaller in the middle. The following are in bold type: standardized bias with an absolute value > 40; coverage < 90%. *RMSE* = root-mean-square error.

ables that were completely uncorrelated with X and Y. We performed the MI procedure using X and Y alone or using X, Y, and 5, 25, or 50 of the junk variables.

Table 9 shows the effects on bias, *RMSE*, coverage, and width of confidence interval of adding the Category c variables. There was no effect on estimation of μ_Y and β_{YX} with respect to any of these four criteria, for up to 50 additional variables. There was some bias in the estimates of σ^2_Y and β_{XY} when 25 variables were added; the bias increased when 50 variables were added. Bias was positive for variance and negative for β_{XY}. There was negative bias in the estimate of ρ_{XY} when 50 variables were added. The *RMSE* for the variance estimate increased when 50 junk variables were included; otherwise the *RMSE* remained approximately the same. Coverage re-

Table 8
Standardized Bias, Coverage, and Width of Confidence Interval When a Category b Variable Is Included: Nonignorable Sinister Missingness, 50% Missing

Parameter estimated	Variables included: X, Y only	X, Y, Z: $\rho_{YZ} = .4$	X, Y, Z: $\rho_{YZ} = .9$
Standardized bias and root-mean-square error			
μ_Y			
Standardized bias	−7.2	−7.2	−7.9
RMSE	.07	.07	.06
σ^2_Y			
Standardized bias	**−67.5**	**−53.0**	−14.5
RMSE	.15	.14	.11
β_{YX}			
Standardized bias	**−177.4**	**−158.7**	**−63.4**
RMSE	.13	.12	.06
β_{XY}			
Standardized bias	**−130.0**	**−121.6**	**−48.4**
RMSE	.07	.06	.04
ρ_{XY}			
Standardized bias	**−178.0**	**−164.0**	**−70.4**
RMSE	.09	.09	.04
Coverage and width of interval			
μ_Y			
Coverage	95.0	94.5	94.7
Width of interval	.28	.27	.23
σ^2_Y			
Coverage	**87.1**	**88.4**	94.9
Width of interval	.50	.50	.43
β_{YX}			
Coverage	**61.4**	**68.4**	91.2
Width of interval	.27	.26	.20
β_{XY}			
Coverage	**77.9**	**79.7**	92.4
Width of interval	.17	.17	.13
ρ_{XY}			
Coverage	**64.1**	**69.0**	91.6
Width of interval	.19	.18	.13

Note. Z is a Category b variable (i.e., it is correlated with Y and is not a correlate of missingness). The probability of missingness is a function of the correlation between the cause of missingness and X. The following are in bold type: standardized bias with an absolute value > 40; coverage < 90%. *RMSE* = root-mean-square error.

mained acceptable in almost every case, reflecting the modest size of the bias. The only exceptions were β_{XY} and σ^2_Y when 50 variables were added. The width of the confidence intervals was unaffected except in the case of σ^2_Y. Here the confidence interval width increased by about 12% when 50 variables were added as compared to the X and Y-only condition.

The reason why noticeable biases were introduced in a few cases is simple. Likelihood-based or Bayesian procedures are known to have negligible bias when the sample is large. With 52 variables, however, we are no longer in a large-sample situation. Estimation of an unstructured 52 × 52 covariance matrix becomes less stable with only $N = 500$. As the sample size is increased, these biases will begin to vanish.

Discussion

The simulation study reported here showed that the costs of omitting a cause of missingness varied considerably depending on several factors. As expected, the costs were greatest when the proportion of missing data was highest, and when the correlation between the cause of missingness and the variable subject to missingness was large. Generally, we found that when missingness does not exceed 25% and the correlation between the cause of missingness and the variable subject to missingness was .4, omitting the cause of missingness from the analysis had a negligible effect. When missingness exceeds 25% or the correlation between the cause of missingness and the variable subject to missingness is as large as .9, our results indicate that there are often substantial problems with bias, efficiency, and coverage. The more surprising pattern in the results is that the type of missing data mechanism made a considerable difference. Estimation of means was disrupted most by the MAR-linear missing data mechanism, and suffered even in the 25% missingness, low correlation condition. By contrast, estimation of variances, regression weights, and correlations was disrupted most by the MAR-convex and MAR-sinister mechanisms.

These findings are important for two reasons. First, in real-world empirical situations it is difficult or impossible to discern the form of the missing data mechanism. Thus, a researcher who is interested primarily in a covariance structure should be very cautious about concluding that missingness is mostly MAR-linear and that therefore omitting important causes of missingness is unlikely to be a problem. If the researcher is mistaken and in reality missingness is of some other form, there may be very serious problems associated with the omission. It should be noted that MAR-linear is the only missing data mechanism used in most published missing data simulations (e.g., Yuan & Bentler, 2000), suggesting the possibility that some of the conclusions of these simulations might have been different if additional missing

Table 9
Standardized Bias, Root-Mean-Square Error (RMSE), Coverage, and Width of Confidence Interval When Category c Variables Are Included: Ignorable Missingness, 50% Missing

	Variables included			
		X, Y +		
Parameter estimated	X, Y only	5 (Category c)	25 (Category c)	50 (Category c)
	Standardized bias and root-mean-square error			
μ_Y				
Standardized bias	−1.9	−2.9	−2.7	−1.7
RMSE	.07	.07	.07	.07
σ^2_Y				
Standardized bias	4.2	10.4	**43.3**	**89.0**
RMSE	.13	.13	.15	.19
β_{YX}				
Standardized bias	2.7	1.9	2.5	3.7
RMSE	.06	.06	.06	.06
β_{XY}				
Standardized bias	−2.5	−10.9	**−49.7**	**−101.9**
RMSE	.04	.04	.04	.05
ρ_{XY}				
Standardized bias	−1.8	−7.0	−28.0	**−55.6**
RMSE	.04	.04	.04	.05
	Coverage and width of interval			
μ_Y				
Coverage	94.9	94.6	94.8	94.7
Width of interval	.28	.28	.28	.29
σ^2_Y				
Coverage	94.3	95.3	95.2	**89.2**
Width of interval	.51	.52	.54	.57
β_{YX}				
Coverage	94.3	95.1	94.2	94.2
Width of interval	.25	.25	.25	.26
β_{XY}				
Coverage	95.1	95.3	94.1	**85.2**
Width of interval	.15	.15	.15	.16
ρ_{XY}				
Coverage	94.6	94.4	94.9	92.3
Width of interval	.16	.16	.16	.16

Note. Caregory c variables are uncorrelated with Y. They may or may not be a correlate of missingness. The following are in bold type: standardized bias with an absolute value > 40; coverage < 90%.

data mechanisms had been used. Second, there are many, many plausible varieties of MAR missing data mechanisms besides the three used here. Given the differences we observed among the three missing data mechanisms used in this study, it is likely that omitting an important cause of missingness results in very different outcomes in these other missing data mechanisms. Thus, the safest course of action is to strive not to omit important causes of missingness.

One compelling reason to adopt an inclusive strategy rather than a restrictive one is that with an inclusive strategy important causes of missingness are less likely to be omitted. A consequence of the inclusive strategy is that variables that are not causes of missingness will sometimes be included in the set of auxiliary variables. A reasonable question is whether including such variables is harmful. Our results indicate that far from being harmful, the inclusion of these variables is at worst neutral, and at best extremely beneficial. When missingness on Y is ignorable, including Category b variables (i.e., variables that are correlated with Y but not a cause of missingness) can add information that results in a decrease in standard errors, and thus an increase in efficiency and statistical power. Even when missingness is nonignorable, including Category b variables can help to mitigate the bias and bring the solution closer to what it would be under ignorable missingness. Our results suggest that these benefits can be substantial enough that researchers may wish to consider including Category b variables deliberately in the auxiliary variable set.

As a procedural note, although much prior literature in the missing-data area has emphasized evaluating missing-data procedures in terms of bias, we found it important to examine not only bias, but also *RMSE,* coverage, and width of the confidence interval. We argue that all of these criteria should be considered when evaluating the effects of missing data procedures, because effects may differ across criteria. For example, our results indicate that adding a Category b variable when missingness is ignorable has no effect on bias but offers pronounced benefits in terms of reduced confidence interval width, which translates directly to improved statistical power.

Implementing an Inclusive Strategy

Earlier in this article, we explored the similarities and differences between ML and MI, showing that the two approaches are conceptually and theoretically very similar. We showed that although the two approaches are operationally different, they will always yield highly similar results when the input data and models are the same, and the number of imputations, *M,* is sufficiently large. Neither approach is inherently better than the other.

However, the researcher who wishes to adopt an inclusive strategy will have to think carefully about how best to implement this strategy. It is a reality that social scientists must implement statistical procedures as they are available to them in software. As we have discussed earlier in this article, currently existing statistical software programs that incorporate the ML missing data approach do not facilitate the use of the inclusive strategy. We wish to stress that this has nothing to do with statistical theory. Rather, it is purely an artifact of the design of today's ML statistical software and documentation, neither of which encourages users to consider using auxiliary variables or informs them about how to incorporate them. It would be possible to revise these programs to make it easy for users to add auxiliary variables; users could simply list the variables to go in the auxiliary variable set, and the program could automatically add them to the model in the appropriate way (Graham, 2001). Given the current state of the art in missing data software, users who wish to implement an inclusive strategy may wish to consider MI. Using MI, it is relatively easy to include a large set of auxiliary variables (for a discussion of this in reference to intent-to-treat analyses, see Little & Yau, 1996). Later model fitting can then take place in the usual manner, without including the auxiliary variables. Eventually, advances in software development may make the inclusive strategy comparably easy with ML.

Limitations

The conclusions that can be drawn from the current study are limited to scenarios similar to the fairly simple one on which the simulations were based (i.e., one where multivariate normality holds and the relationships among variables are linear). In particular, the findings of this study about performance of estimation under various amounts of missingness and types of missing data mechanisms may not generalize to other scenarios likely to occur in practice. For example, when the relationship between *Y* and *Z* is nonlinear, it is possible that failing to include *Z* may be very damaging, even in situations with 25% missing data or less. Other situations where failing to include *Z* may cause serious problems[4] are when the sign of the relationship between *X* and *Y* is the opposite of the sign of the relationship between *Y* and *Z;* when *X* and *Z* are uncorrelated, but *X* and *Y* are correlated, as are *Y* and *Z;* and countless other scenarios.

This simulation does not address the issue of missing data in studies with small sample sizes. We used a constant $N = 500$, which exceeds typical sample sizes in many areas of behavioral research. (For a discussion of missing data procedures in small *N* studies, see Graham & Schafer, 1999.) In particular, some might argue that samples of $N = 500$ are unusually large for simple analysis involving only two or three variables. Because the standard error is reduced as the sample size goes up, the standardized bias as we define it will increase with increasing sample size, assuming the same raw bias. With smaller samples, the standardized bias will decrease and the implications for intervals and tests will be less serious. For example, assuming raw biases stay the same, with a sample size of $N = 125$ the standardized biases are

[4]We are indebted to an anonymous reviewer for pointing this out.

expected to become approximately half of what we report.

Conclusions and Recommendations

This article compared the restrictive and inclusive missing data strategies. We conclude that the inclusive strategy is to be recommended, because there appear to be few risks associated with it and potentially substantial gains. With an inclusive strategy, not only is there a reduced chance of inadvertently omitting an important cause of missingness, there is also the possibility of noticeable gains in terms of increased efficiency and reduced bias. In theory, either ML or MI can be used to implement the inclusive strategy. In practice, because of the design of today's software programs, it is more straightforward to use MI for the inclusive strategy. Social scientists would benefit if ML software was revised to facilitate use of the inclusive strategy.

References

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods & Research, 28,* 301–309.

Allison, P. D. (2001). *Missing data.* Thousand Oaks, CA: Sage.

Arbuckle, J. W., & Wothke, W. (1999). *Amos 4.0 user's guide.* Chicago: SPSS.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London: Chapman & Hall.

Graham, J. W. (2001). *Adding missing-data-relevant variables to FIML-based structural equation models* (Tech. Rep. No. 01-48). University Park, PA: The Pennsylvania State University, The Methodology Center.

Graham, J. W., & Hofer, S. M. (1993). *EMCOV.EXE user's guide.* Alhambra, CA: University of Southern California, Institute for Prevention Research.

Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (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 & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 335–353). Washington, DC: American Psychological Association.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8.5: User's reference guide.* Chicago: Scientific Software International.

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An algorithm for multiple imputation. *American Political Science Review, 95,* 49–70.

Lavori, P. W., Dawson, R., & Shera, D. (1995). A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine, 14,* 1913–1925.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute.

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

Little, R. J. A., & Yau, L. (1996). Intent-to-treat analysis for longitudinal studies with drop-outs. *Biometrics, 52,* 1321–1333.

Meng, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input (with discussion). *Statistical Science, 10,* 538–573.

Muthén, B. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class–latent growth modeling. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: American Psychological Association.

Neale, M. C. (1997). *Mx: Statistical modeling* (4th ed.). Richmond: Virginia Commonwealth University, Department of Psychiatry.

Norusis, M. J. (2000). *SPSS 10.0 guide to data analysis.* Upper Saddle River, NJ: Prentice Hall.

Raudenbush, S. W., Bryk, A., Cheong, Y. F., & Congdon, R. (2000). *HLM 5: Hierarchical linear and nonlinear modeling* [Computer software]. Chicago: Scientific Software International.

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

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

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

SAS Institute. (2001). SAS/STAT Software: Reference (Version 8.1) [Computer software]. Cary, NC: Author.

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

Schafer, J. L. (1999a). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3–15.

Schafer, J. L. (1999b). NORM: Multiple Imputation of Incomplete Multivariate Data Under a Normal Model [Computer software]. Retrieved May 1, 2000, from http://methodology.psu.edu

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: An analyst's perspective. *Multivariate Behavioral Research, 33,* 545–571.

Statistical Solutions, Inc. (1998). SOLAS for missing data analysis (Version 1) [Computer software]. Cork, Ireland: Author.

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

Received May 9, 2001
Revision received July 23, 2001
Accepted July 26, 2001 ■