THE PERFORMANCE OF THE FULL INFORMATION MAXIMUM LIKELIHOOD ESTIMATOR IN MULTIPLE REGRESSION MODELS WITH MISSING DATA

CRAIG K. ENDERS
University of Miami

A Monte Carlo simulation examined the performance of a recently available full information maximum likelihood (FIML) estimator in a multiple regression model with missing data. The effects of four independent variables were examined (missing data technique, missing data rate, sample size, and correlation magnitude) on three outcome measures: regression coefficient bias, R^2 bias, and regression coefficient sampling variability. Three missing data patterns were examined based on Rubin's missing data theory: missing completely at random, missing at random, and a nonrandom pattern. Results indicated that FIML estimation was superior to the three ad hoc techniques (listwise deletion, pairwise deletion, and mean imputation) across the conditions studied. FIML parameter estimates generally had less bias and less sampling variability than the three ad hoc methods.

Missing data is a common problem for researchers who use statistical techniques such as multiple regression. Traditionally, the analysis of incomplete data has been dominated by ad hoc methods that have no theoretical rationale. For example, Roth (1994) examined 75 studies taken from the *Journal of Applied Psychology* and *Personnel Psychology* and found that listwise deletion (LD) (observations with missing values are discarded) and pairwise deletion (PD) (each correlation matrix element is calculated separately using all nonmissing observations) were the most commonly used techniques; mean imputation (MI) (missing values are replaced by variable means) and regression imputation (RI) (missing values are replaced by predicted scores from a regression equation) appeared only once. Kim and Curry (1977) also

The author would like to acknowledge Dr. Janet Sheehan-Holt for her recommendations on an earlier draft of this article.

Educational and Psychological Measurement, Vol. 61 No. 5, October 2001 713-740

noted the ubiquitous nature of LD and PD. Although LD and PD are attractive due to their simplicity and widespread availability as analysis options in statistical packages, Little and Rubin (1987) noted that "neither method . . . is generally satisfactory" (p. 43).

Full Information Maximum Likelihood (FIML)

As pointed out by Schafer and Olsen (1998), substantial developments have occurred in the development of missing data techniques (MDTs) in the past 20 years. For example, maximum likelihood (ML) estimators have recently been implemented in a variety of statistical software packages; FIML estimation is available in structural equation modeling (SEM) software (e.g., AMOS, MPLUS), and the expectation maximization (EM) algorithm is implemented in common packages such as SPSS. A brief description of the FIML estimator follows; the interested reader is encouraged to consult Enders (2001) for a more detailed overview of ML missing data estimators.

The FIML estimator implemented in AMOS and MPLUS maximizes a likelihood function that is the sum of n casewise likelihood functions. That is, a likelihood function is calculated for each individual that measures the discrepancy between the observed data for the ith case and the current parameter estimates. Assuming multivariate normality, the following function is maximized:

$$\log L_i = K_i - \tfrac{1}{2}\log|\Sigma_i| - \tfrac{1}{2}(x_i - \mu_i)' \textstyle\sum_i^{-1} (x_i - \mu_i), \tag{1}$$

such that x_i is the vector of complete data for case i, μ_i is the vector of mean estimates for those variables that are observed for case i, and K_i is a constant that depends on the number of complete data points for case i. Similarly, the determinant and inverse of Σ_i are based only on those variables that are observed for case i. Summing over the n casewise functions yields the discrepancy function for the entire sample:

$$\log L(\mu, \Sigma) = \sum_{i=1}^{N} \log L_i. \tag{2}$$

At a more conceptual level, it is assumed that missing values on a variable X are conditionally dependent on other observed variables in the data, and incorporating vectors of partially complete data in the individual-level likelihood functions (Equation 1) implies probable values for the missing data during the parameter estimation process. This is conceptually analogous to generating predicted scores for the missing data by regressing X on other variables used in the analysis. However, it is important to note that the FIML estimator outlined above does not impute, or fill in, missing values but directly estimates model parameters and standard errors using all available raw data.

Although the focus of the current study is on the ML estimator described above, it should be noted that the EM algorithm could also be used to obtain ML estimates of an unrestricted mean vector and covariance matrix that, in turn, could be used to obtain estimates of regression slopes and R^2 values. Briefly, the EM algorithm uses a two-step iterative procedure whereby missing observations are estimated and parameter estimates are obtained. To begin the iterative cycle, an initial estimate of the covariance matrix and mean vector is obtained using LD, PD, or some imputation method. In the *E* step, missing values are replaced with the conditional mean of the missing data given the observed data and the initial covariance matrix estimate. That is, missing values are replaced by predicted scores from a series of regression equations where each missing variable is regressed on the remaining observed variables for a case *i*. The subsequent *M* step is a complete-data ML estimation problem as ML estimates of the mean vector and covariance matrix are obtained using the filled-in data from the *E* step. This updated covariance matrix is then used to derive regression equations for the next *E* step, and the cycle is repeated until the difference between covariance matrices in subsequent *M* steps falls below some specified convergence criterion. Readers are encouraged to consult Little and Rubin (1987) for further technical details.

It should be noted that, in the unrestricted estimation case (e.g., covariance matrix, multiple regression), the FIML estimator implemented in AMOS and MPLUS yields point estimates of model parameters that are identical to those that would be obtained using the EM algorithm. However, a drawback of EM is that standard errors of regression model parameters are not obtained directly from the procedure and require additional analytic steps (e.g., bootstrapping). As noted by Graham, Hofer, Donaldson, MacKinnon, and Schafer (1997), standard errors would be provided on computer output when using the EM covariance matrix as input for further analyses (e.g., multiple regression), but these standard errors would be based on the wrong sample size and thus are incorrect—as noted above, correct standard errors are obtained directly when using the FIML estimator offered in SEM packages. For these reasons, it was decided to exclude the EM algorithm from the current investigation.

Missing Data Theory

Although ML estimation in the presence of missing data has been discussed in the statistical literature for some time (e.g., Anderson, 1957), it has not enjoyed widespread use, perhaps due to the lack of specialized software (Wothke, 2000). Perhaps another reason for continued use of ad hoc MDTs is that applied researchers do not clearly understand the theoretical benefits associated with ML estimation, which are substantial relative to traditional methods.

As noted by Little and Rubin (1987), the performance of a given MDT is largely dependent on the mechanism that caused the missing data. Rubin (1976) was the first to formally explicate missing data theory and the mechanisms that cause missing data. According to Rubin (1976), data are missing completely at random (MCAR) when the missing values on a variable *X* are independent of other observed variables as well as the values of *X* itself. In this case, the observed values of *X* are simply a random subsample of the hypothetically complete data. In contrast, the less restrictive missing at random (MAR) condition requires that missing values on a variable *X* be dependent on other observed variables in the data set but not on the values of *X* itself. For example, in a testing situation, suppose that low-scoring individuals on exam *X* have a tendency to drop out or refuse to take exam *Y*. In this case, the propensity to complete exam *Y* depends on scores from exam *X* and is unrelated to performance on *Y*. Similarly, suppose that a set of academic achievement variables is missing for students who come from low-income households, but the missingness is unrelated to student achievement itself. Finally, Rubin (1976) described nonignorable missing data patterns that occur when the probability of a missing value on a variable *X* is dependent on the underlying value of *X* itself. For example, suppose that children with poor reading skills have a higher propensity for missing data on a test of reading comprehension due to their inability to respond to test items.

According to Rubin's (1976) work, the MCAR and MAR conditions are considered ignorable missing data mechanisms in the sense that unbiased parameter estimates can be obtained using standard ML estimation. That is, ML estimation in the presence of missing data, as outlined above, requires only that the weaker MAR assumption hold. In contrast, LD and PD require the strict MCAR assumption, a condition that some have argued is rarely met in practice (Graham, Hofer, & MacKinnon, 1996; Muthén, Kaplan, & Hollis, 1987).

Assuming MCAR, the following theoretical expectations follow from Rubin's (1976) work. First, ML, LD, and PD should yield unbiased point estimates of model parameters. However, any obvious result of implementing LD is that a great deal of complete data is potentially lost. For example, consider a three-predictor regression model such as the one used in this study. Assuming MCAR and a 5% missing data rate, it would be expected that approximately 18.6% ($1 - .95^4$) of the cases would be discarded as a result of using LD. Obviously, this greatly reduces statistical power, leading to increased sampling variability of parameter estimates. In contrast, ML incorporates data from partially observed cases into the estimation process and, as a result, should yield more efficient parameter estimates under MCAR. PD also attempts to incorporate information from incomplete cases by computing each element of the covariance matrix using all available cases for a given variable or variable pair. However, it is a well-known fact that the resulting matrix may not be positive definite, leading to indeterminate regression slope estimates. Finally, MI will yield biased estimates of variance and covariance terms, and thus regression model parameters, under MCAR.

Assuming MAR, ML estimation will yield unbiased estimates of model parameters. Because LD and PD are built on the MCAR assumption, biased parameter estimates may result in this case. However, as noted by Little (1992) and others, LD may yield unbiased parameter estimates in the regression context when missingness is dependent on the predictor variables; this is not generally true of PD, however. Finally, under a nonignorable missing data mechanism such as that described above, all methods described herein may yield biased parameter estimates. Although little work has been done with nonignorable mechanisms, the results of a SEM study by Muthén et al. (1987) suggested that ML estimators might yield less bias than ad hoc methods in these circumstances.

Studies in the SEM Context

With the recent introduction of ML estimators in computer software packages, there has been a resurgence in missing data research. Much of this recent literature has appeared in the area of SEM, and the results of these studies are generally consistent with the theoretical expectations described above. An early SEM study by Muthén et al. (1987) examined the performance of the so-called multiple-group ML approach that is closely related to the FIML estimator described above. As expected, ML parameter estimates were unbiased under MAR, whereas LD and PD estimates were biased. Although all three MDTs yielded biased estimates under a nonignorable mechanism, the bias due to ML estimation was less than that of the ad hoc methods. Four studies examined the FIML estimator implemented in the AMOS and MPLUS software packages. Using a confirmatory factor analysis model, Arbuckle (1996) reported unbiased ML parameter estimates under both MCAR and MAR conditions; LD and PD were biased under MAR. Enders and Bandalos (2001) found similar results using both a confirmatory factor analysis and full structural model, as did Wothke (2000) using a latent growth curve model. With respect to efficiency, Arbuckle's results suggested that the FIML parameter estimates have substantially less sampling variability than either LD or PD under MCAR. In contrast, both Enders and Bandalos and Wothke found that the FIML estimates were only slightly more efficient than those of PD under MCAR but were substantially more efficient than those of LD. Finally, Enders and Bandalos reported that the sampling variability of FIML estimates was approximately equal to that of other MDTs under MAR. These general findings regarding the performance of the FIML estimator have also been extended to situations in which multivariate normality does not hold (Enders, in press). However, it should be noted that ML standard error estimates were severely negatively biased in this case.

Studies in the Regression Context

Although missing data studies are beginning to reappear in the literature, no recent studies have examined the performance of ML estimation in the multiple regression context. Such studies would be informative as multiple regression models can easily be estimated using SEM packages that offer the FIML estimation option. Although there is a substantial body of missing data research in the area of multiple regression, much of this literature is dated and/or is quite limited by current simulation standards. For example, most simulation studies used only 50 or fewer iterations (Azen, Van Guilder, & Hill, 1989; Little, 1979, 1988; Raymond & Roberts, 1987), and several studies used only 10 (Basilevsky, Sabourin, Hum, & Anderson, 1985; Beale & Little, 1975; Haitovsky, 1968; Kim & Curry, 1977). Clearly, such studies are inadequate by current computing standards and should be viewed with some caution. A recent study by Kromrey and Hines (1994) investigated nonrandom missing data patterns in the context of multiple regression, but ML estimation was not examined.

A second recurring problem with existing regression literature is the lack of a theoretical framework. Much of the extant literature was conducted prior to Rubin's (1976) seminal work on missing data theory, and studies published after this date frequently do not incorporate Rubin's discussion of missing data mechanisms in the study design. To illustrate, many early studies measured MDT performance using a mean squared error criterion that measures the deviation between a sample parameter estimate and the corresponding population parameter. From a theoretical perspective, this criterion measure is problematic as both bias and efficiency can influence the magnitude of its value. As such, it is difficult to unambiguously relate the findings of these studies back to Rubin's theoretical work.

To date, the EM algorithm is the only ML method examined in the regression literature; no studies have investigated the FIML estimator. With the exception of Basilevsky et al. (1985), these studies generally suggested that the performance of the EM algorithm is superior to ad hoc MDTs (Beale & Little, 1975; Little, 1979, 1988; Raymond & Roberts, 1987). However, as noted above, it is difficult to identify what characteristic of the estimator (bias or efficiency) is being measured. In addition, the extant literature offers no clear counsel on the relative performance of LD and PD. Study results have been inconsistent, but it appears that LD may be most effective when variables are highly correlated (Azen et al., 1989; Haitovsky, 1968; Little, 1988).

Conversely, PD may work best when correlations are low (Buck, 1960; Kim & Curry, 1977). This is somewhat inconsistent with the SEM literature, which has suggested that PD is superior, particularly with respect to efficiency (Enders & Bandalos, in press). Consistent with SEM literature, MI has been shown to yield negatively biased correlations, regression coefficients, and R^2 values (Gleason & Staelin, 1975; Kim & Curry, 1977; Kromrey & Hines, 1994; Timm, 1970) and is generally not recommended (Little, 1992).

Purpose of the Present Study

Given the paucity of recent research and the limitations associated with much of the existing regression missing data literature, the goal of this study was to investigate the performance of the FIML estimator relative to three ad hoc missing data methods (LD, PD, and MI) within the context of a three-predictor multiple regression model. A comprehensive Monte Carlo study was designed to address three research questions: how do MDTs differ with respect to (a) regression coefficient bias, (b) R^2 bias, and (c) efficiency (i.e., sampling variability) of parameter estimates? Furthermore, these questions will be addressed within the framework of Rubin's (1976) missing data theory.

Method

Design

Using a three-predictor multiple regression model, three simulation studies were performed that modeled MCAR, MAR, and nonignorable missing data patterns, respectively. Four independent variables were manipulated within each simulation: the missing data rate (5%, 15%, 25%, and 35%), sample size (100, 250, and 400), MDT (FIML, LD, PD, and MI), and correlational structure. The levels of the independent variables were chosen to represent a variety of conditions found in past research as well as applied settings. For example, the missing data rates used in this study are representative of previous simulation research and represent a range of values that are probably typical in applied settings. Similarly, the sample size conditions were chosen to roughly coincide with recommendations given in regression texts. For example, in discussing the issue of shrinkage, Pedhazur (1997) stated,

> Some authors recommend that the ratio of predictors to sample size be at least 1:15, that is, at least 15 subjects per predictor. Others recommend smaller ratios (e.g., 1:30). Still others recommend that samples be comprised of at least 400 subjects. (p. 207)

Previous regression studies suggested that the correlational structure might play a role in the differential performance of missing data methods. Much of the previous literature has used contrived correlation structures, whereas other authors have convincingly argued that correlation structures from actual data sets should be used (Kromrey & Hines, 1994). Given the paucity of recent research and limited faith that can be placed on early work in this area, it was decided that an artificial correlation structure would better serve the need for a controlled experiment. To this end, Cohen's (1977) discussion of the *r* effect size measure was used to guide the choice of levels for this factor. Cohen characterized *r* values of .10, .30, and .50 as small, medium, and large effects, respectively. Accordingly, the level of correlation among the predictors was uniformly set at one of these three levels. Similarly, the correlations between the predictors and criterion variable were also uniformly set at one of these three levels. Completely crossing these conditions resulted in a total of nine correlation structures. For example, "small" IV correlations and "large" IV-DV correlations would result in the following correlation matrix:

	*X*1	*X*2	*X*3	*Y*
*X*1	1.0			
*X*2	0.1	1.0		
*X*3	0.1	0.1	1.0	
Y	0.5	0.5	0.5	1.0

The matrices for the nine correlation structures and corresponding population regression coefficients and R^2 values are found in Table 1. It is also reasonable to argue that Cohen's (1977) benchmark values for multiple R^2 should be used to guide the simulation conditions. However, it was felt that allowing R^2 to vary as a function of the sample correlations provided for a more realistic simulation than correlations that were determined as a function of R^2 values fixed at Cohen's benchmarks.

Data Generation

A total of 250 raw data matrices ($n \times 4$) were generated within each of the 108 between-subjects design cells (4 missing data rates × 3 sample sizes × 9 correlation structures) using the RANNOR function in the SAS IML procedure. These uncorrelated vectors of random normal deviates were linearly transformed to the desired covariance structures using Cholesky factorization. Descriptive statistics were examined on the resulting data matrices to verify that the desired correlational structures were obtained, on average.

For each missing data mechanism, missing values were created on two of the predictor variables, *X*2 and *X*3. The MCAR missing data pattern was cre-

Table 1
Correlation Structures and Population Parameters

Correlation Structure		Correlation Matrices				Population Parameters	
		Y	*X*1	*X*2	*X*3	*b*	R^2
	Y	1.0				0.083	0.025
1	*X*1	0.1	1.0				
	*X*2	0.1	0.1	1.0			
	*X*3	0.1	0.1	0.1	1.0		
	Y	1.0				0.250	0.225
2	*X*1	0.1	1.0				
	*X*2	0.1	0.1	1.0			
	*X*3	0.3	0.3	0.3	1.0		
	Y	1.0				0.417	0.625
3	*X*1	0.1	1.0				
	*X*2	0.1	0.1	1.0			
	*X*3	0.5	0.5	0.5	1.0		
	Y	1.0				0.063	0.019
4	*X*1	0.3	1.0				
	*X*2	0.3	0.3	1.0			
	*X*3	0.1	0.1	0.1	1.0		
	Y	1.0				0.188	0.169
5	*X*1	0.3	1.0				
	*X*2	0.3	0.3	1.0			
	*X*3	0.3	0.3	0.3	1.0		
	Y	1.0				0.312	0.469
6	*X*1	0.3	1.0				
	*X*2	0.3	0.3	1.0			
	*X*3	0.5	0.5	0.5	1.0		
	Y	1.0				0.050	0.015
7	*X*1	0.5	1.0				
	*X*2	0.5	0.5	1.0			
	*X*3	0.1	0.1	0.1	1.0		
	Y	1.0				0.150	0.135
8	*X*1	0.5	1.0				
	*X*2	0.5	0.5	1.0			
	*X*3	0.3	0.3	0.3	1.0		
	Y	1.0				0.250	0.375
9	*X*1	0.5	1.0				
	*X*2	0.5	0.5	1.0			
	*X*3	0.5	0.5	0.5	1.0		

ated by randomly deleting the desired percentage of observations from the two predictors. For the MAR simulation, missing values on *X*2 and *X*3 were dependent on values of the first predictor variable, X_1. Specifically, the probability of a missing value on *X*2 and *X*3 was inversely related to an observation's percentile rank on $X1$. For example, an observation at the 10th percentile on $X1$ had a 90% (1 – 1/10 = .90) probability of a missing value on $X1$ and $X2$, whereas an observation at the 25th percentile had a 75% probability of deletion. Beginning with the lowest $X1$ value, each observation's probability was evaluated using a uniform random number that ranged between zero and one. If the uniform random number was less than the probability value, that observation's $X2$ and $X3$ values were deleted. This deletion process was performed in ascending order until the desired percentage of missing data was obtained. Finally, missing data in the nonignorable simulation were dependent on the values of the predictors themselves, not on other variables in the data set. Using a similar process to that described above, the probability of deletion was inversely related to scores on $X2$ and $X3$, such that low scores on these variables had the highest likelihood of deletion. For example, the observation at the 10th percentile on $X2$ had a 90% probability of a missing value on $X2$, whereas a score at the 25th percentile had a 75% probability of deletion. Whereas both $X2$ and $X3$ values were simultaneously deleted for a given observation in the MAR simulation, each variable's deletion was evaluated separately in the nonignorable condition.

After the missing data patterns were created, regression models were estimated for each of the 250 samples in the between-subjects design cells using the four MDTs described above. The LD, PD, and MI analyses were performed using the PROC REG procedure in SAS; the PD correlation matrices were obtained using PROC CORR and were used as input into the regression procedure. FIML estimation was implemented using the AMOS 4.0 computer program, and the model fitting process was automated using a Visual Basic program that repeatedly called the AmosEngine programming interface. Regression weights and R^2 values from the MDTs were output into a file for further analysis.

Analysis

Three dependent variables were examined: (a) regression coefficient bias, (b) R^2 bias, and (c) parameter estimate efficiency. A percentage-of-bias measure was used for the first two dependent variables that was defined as

$$\frac{\left(\hat{\hat{\theta}}_{hijk} - \hat{\hat{\theta}}_{hi00}\right)}{\hat{\hat{\theta}}_{hi00}} * 100, \tag{3}$$

where $\hat{\hat{\theta}}_{hijk}$ is the mean parameter estimate taken from sample size h, correlation structure i, missing data rate j, and MDT k. $\hat{\hat{\theta}}_{hi00}$ is the corresponding parameter estimate generated from the same data with no missing values. Parameter estimate efficiency was measured using a relative efficiency statistic that was the ratio of empirical sampling variances generated from the kth ad hoc method (e.g., LD, PD, MI) and the FIML estimator, respectively (RE = $\sigma^2_{AH}/\sigma^2_{FIML}$); this ratio was calculated for each of the 108 between-subjects design cells. Because sample size is inversely related to sampling variance, this measure allows one to assess the increase in sample size required to reach the same level of efficiency as the ML estimator that, based on statistical theory, was expected to yield the most efficient parameter estimates. This relative efficiency statistic is consistent with that used in recent SEM missing data studies (Arbuckle, 1996; Enders & Bandalos, in press; Wothke, 2000).

For the first two dependent variables (parameter estimate and R^2 bias), a split-plot ANOVA with three between-subjects factors (missing data rate, sample size, and correlation structure) and one within-subjects factor (MDT) was performed on each parameter estimate of interest. Due to the large number of total replications (i = 27,000), tests of statistical significance were not examined, and partial η^2 effect size estimates were instead used to guide substantive interpretations. Given the current ease of implementing the FIML estimation in existing software, it was felt that even small effect sizes might be considered practically significant. Thus, design effects that yielded η^2 values greater than .01 will be interpreted below. In many cases, both main effects and higher order interactions yielded η^2 values that exceeded .01, but only the highest order effects are discussed. Obviously, measures of effect size could not be calculated for the relative efficiency results as there was only a single observation per cell. A descriptive analysis of these results was performed.

Results

Simulation 1: MCAR Data

Regression coefficient bias. As described above, a 9 × 3 × 5 × 4 (Correlation Structure × Sample Size × Missing Data Rate × MDT) split-plot ANOVA was performed on each regression coefficient to assess the magnitude of bias. The MDT main effect for the $X1$ regression weight (the predictor with no missing data) was the only design effect that yielded a notable partial η^2 value (η^2 = .024). An examination of the cell means for this main effect indicated that MI estimates were positively biased, whereas the remaining MDTs were unbiased. Collapsed across the between-subjects design cells, the mean percentage of $X1$ bias was 11.77% for MI compared to –.20%, –.09%, and –.13% for LD, PD, and FIML, respectively. The $X2$ and $X3$ analyses yielded no effects that reached Cohen's (1977) benchmark value for a small effect size. In general, the remaining regression coefficients were unbiased across the design cells. Overall, the bias due to MI was slightly larger than the other MDTs, but the difference was minimal (approximately 1% to 2% difference in the marginal means).

R^2 *bias*. The same split-plot ANOVA described above was performed to obtain effect size estimates for R^2 bias. The three-way interaction between MDT, correlation structure, and missing data rate was the highest order effect to exceed the small effect size criterion ($\eta^2 = .035$). Table 2 gives the mean percentage of R^2 bias for each MDT broken down by correlation structure and missing data rate.

From the table, several points are obvious. First, not surprisingly, bias increased as the percentage of missing data increased; this was true for all MDTs, although the magnitude of the bias differed across the methods. Specifically, MI consistently yielded the largest bias values, followed by LD, PD, and FIML; in general, differences between the latter two methods were quite small. Finally, the magnitude of R^2 bias differed across correlation structures. The largest values were observed in Structures 1, 4, and 7. It should be noted that these structures had low IV-DV correlations ($r = .10$), as well as low population R^2 values (1% to 2%); although the absolute magnitude of the bias was small, it was large relative to the parameter value itself. When considering the remaining six correlation structures, the R^2 bias was not substantial for LD, PD, and FIML. Excluding Correlation Structures 1, 4, and 7, the overall R^2 bias for LD, PD, and FIML was 2.55%, 1.08%, and 1.03%, respectively.

Relative efficiency. As described above, the sampling variance of each MDT was examined relative to that of FIML. As such, relative efficiency values greater than unity reflected situations in which the ML estimator yielded parameter estimates with less sampling variability (i.e., greater efficiency) than the respective ad hoc method. Relative efficiency values differed little across the correlation structure and sample size conditions and varied primarily as a function of the missing data rate. Table 3 shows the relative efficiency values for the three regression weights by correlation structure and missing data rate. For the sake of brevity, only the efficiency values for the $N = 250$ condition are shown as these values are quite representative of the other two sample size conditions.

As seen in the table, FIML performed dramatically better than LD. Because sampling variance is inversely related to sample size, the relative efficiency statistic provides an indication of the increase in sample size required to obtain the same efficiency as FIML. For example, in the second correlation structure, the $X1$ regression weight for LD had a relative efficiency value of 2.20 at the 35% missing data rate. This indicates that the LD sample size would need to be increased by approximately 120% (300 cases) to yield sampling variability equivalent to FIML. In contrast, FIML yielded modest efficiency gains relative to PD; gains of approximately 5% were typical across the majority of the design cells. Finally, MI sampling variability

Table 2
Mean Percentage of MCAR R^2 *Bias by MDT, Correlation Structure, and Missing Data Rate*

Condition			Estimation Method			
	Correlation Structure	Missing Data (%)	LD	PD	MI	FIML
1	IV = .10 DV = .10	5	5.3	1.7	–1.6	1.8
		15	12.8	5.6	–4.2	5.8
		25	28.2	9.3	–7.9	9.5
		35	48.1	12.1	–13.4	12.7
2	IV = .10 DV = .30	5	0.5	0.1	–2.8	0.1
		15	1.7	0.4	–8.3	0.4
		25	4.1	1.6	–13.0	1.4
		35	5.8	1.9	–19.1	2.0
3	IV = .10 DV = .50	5	–0.1	0.0	–2.8	0.0
		15	0.0	0.3	–8.4	0.2
		25	0.6	0.4	–13.7	0.3
		35	1.1	0.8	–19.8	0.9
4	IV = .30 DV = .10	5	5.7	2.3	–0.7	2.3
		15	18.2	8.0	–1.9	7.5
		25	33.8	10.7	–6.4	8.8
		35	62.4	21.3	–6.2	19.7
5	IV = .30 DV = .30	5	1.0	0.3	–1.9	0.3
		15	2.9	0.9	–6.0	0.8
		25	5.4	1.7	–10.2	1.8
		35	7.3	2.7	–14.1	2.8
6	IV = .30 DV = .50	5	0.2	0.1	–2.1	0.1
		15	0.3	0.1	–6.5	0.1
		25	0.7	1.0	–10.2	1.0
		35	1.8	0.9	–14.8	1.0
7	IV = .50 DV = .10	5	6.6	2.9	–0.8	2.3
		15	17.3	6.6	–3.9	4.6
		25	31.2	13.0	–6.4	8.2
		35	69.4	28.5	–5.1	20.6
8	IV = .50 DV = .30	5	1.1	0.5	–1.6	0.2
		15	4.2	2.2	–3.9	1.6
		25	6.3	3.7	–7.0	3.3
		35	10.8	4.0	–11.4	4.1
9	IV = .50 DV = .50	5	0.3	0.0	–1.7	0.0
		15	0.9	0.7	–4.7	0.7
		25	0.7	0.4	–8.2	0.4
		35	3.6	1.3	–11.2	1.3

Note. MCAR = missing completely at random; MDT = missing data technique; LD = listwise deletion; PD = pairwise deletion; MI = mean imputation; FIML = full information maximum likelihood.

Table 3
MCAR Relative Efficiency by Correlation Structure and Missing Data Rate

Condition	Correlation Structure	Missing Data (%)	LD *X*1	LD *X*2	LD *X*3	PD *X*1	PD *X*2	PD *X*3	MI *X*1	MI *X*2	MI *X*3
1	IV = .10 DV = .10	5	1.13	1.08	1.11	1.00	0.99	1.00	1.00	0.97	1.00
		15	1.28	1.12	1.17	1.00	0.98	1.00	0.99	0.97	0.99
		25	1.71	1.36	1.49	1.00	0.97	1.00	0.98	0.96	0.97
		35	2.35	1.60	1.61	1.00	1.00	1.01	0.97	0.97	0.97
2	IV = .10 DV = .30	5	1.13	1.11	1.12	1.00	1.00	1.01	0.99	1.01	1.01
		15	1.51	1.19	1.21	0.99	1.02	1.01	0.99	1.02	0.99
		25	1.70	1.22	1.25	1.02	1.03	1.02	0.98	1.05	1.04
		35	2.20	1.48	1.26	1.01	1.03	1.02	0.96	1.01	1.00
3	IV = .10 DV = .50	5	1.08	1.07	1.02	1.00	1.02	1.04	1.01	1.02	1.10
		15	1.31	1.18	1.19	1.01	1.03	1.01	1.01	1.03	1.09
		25	1.58	1.56	1.45	1.08	1.13	1.10	0.97	1.24	1.14
		35	1.86	1.66	1.86	1.09	1.22	1.24	0.97	1.15	1.15
4	IV = .30 DV = .10	5	1.11	1.03	1.01	1.00	1.01	1.01	0.99	0.98	0.98
		15	1.32	1.14	1.22	1.01	1.00	1.03	0.97	0.94	0.97
		25	1.69	1.29	1.21	1.00	1.01	1.03	0.91	0.91	0.93
		35	2.50	1.68	1.38	1.02	1.07	1.02	0.89	0.90	0.86
5	IV = .30 DV = .30	5	1.13	1.07	1.05	1.00	1.01	0.98	0.98	1.00	0.96
		15	1.28	1.21	1.21	1.01	1.00	1.02	0.98	0.93	0.95
		25	1.65	1.45	1.15	1.02	0.99	1.01	0.94	0.90	0.92
		35	2.16	1.63	1.53	1.05	1.01	1.01	0.94	0.89	0.90
6	IV = .30 DV = .50	5	1.04	1.05	1.05	1.02	1.00	1.01	1.02	0.97	1.02
		15	1.43	1.19	1.22	0.99	1.03	1.01	0.96	1.05	1.01
		25	1.58	1.39	1.37	1.02	1.04	1.04	0.97	1.09	1.02
		35	2.11	1.46	1.47	0.98	1.05	1.05	0.82	1.00	0.96
7	IV = .50 DV = .10	5	1.10	1.02	1.03	1.01	1.02	1.01	0.98	0.95	0.93
		15	1.41	1.14	1.16	0.99	1.00	1.02	0.90	0.83	0.86
		25	1.73	1.22	1.26	1.06	1.11	1.13	0.87	0.81	0.82
		35	2.39	1.69	1.37	1.04	1.25	1.18	0.81	0.84	0.80
8	IV = .50 DV = .30	5	1.01	1.01	1.04	1.01	1.01	1.00	0.98	0.94	0.95
		15	1.49	1.27	1.19	1.04	1.13	1.00	0.92	0.94	0.84
		25	1.53	1.25	1.20	1.00	1.16	1.14	0.79	0.87	0.86
		35	2.04	1.55	1.38	1.00	1.09	1.09	0.77	0.78	0.74
9	IV = .50 DV = .50	5	1.04	1.01	1.02	1.01	1.03	1.03	0.97	1.00	0.99
		15	1.30	1.14	1.29	1.01	1.05	1.02	0.90	0.94	0.90
		25	1.44	1.24	1.25	1.04	1.04	1.04	0.94	0.85	0.82
		35	1.92	1.57	1.60	1.03	1.07	1.03	0.75	0.87	0.82

Note. MCAR = missing completely at random; LD = listwise deletion; PD = pairwise deletion; MI = mean imputation.

was, in general, similar to that of FIML and was actually somewhat lower in many cases.

Simulation 2: MAR Data

Regression coefficient bias. For the *X*1 regression weight, two interactions exceeded the small effect size benchmark: the MDT main effect ($\eta^2 = .048$), the MDT by correlation structure interaction ($\eta^2 = .018$), and the MDT by missing data rate interaction ($\eta^2 = .016$). To illustrate these effects, the mean percentage of *X*1 regression weight bias for each MDT is given in Table 4 by correlation structure and missing data rate. As seen in the table, MI consistently yielded large amounts of bias, and the bias increased as the percentage of missing data increased. In contrast, the FIML estimates were unbiased and were unaffected by the missing data rate. Of the remaining two ad hoc methods, LD performed slightly better than PD, but neither method generally yielded excessive levels of bias; the largest bias values were typically observed at the 35% missing data rate. It should be noted that PD bias was slightly larger in Correlation Structures 8 and 9, which were conditions with a high level intercorrelation.

None of the design effects for *X*2 and *X*3 regression coefficients surpassed the small effect size benchmark. In general, these coefficients were unbiased across the four MDTs, although MI yielded slightly higher levels of bias than the remaining three methods. For example, collapsed across the between-subjects design cells, the overall level of *X*2 bias was .19%, –.37%, –.61%, and –2.2% for FIML, LD, PD, and MI, respectively. The corresponding values for *X*3 were virtually identical.

R^2 *bias*. For R^2 bias under MAR, the three-way interaction between MDT, correlation structure, and missing data rate was the highest order effect to exceed the small effect size criterion ($\eta^2 = .022$). Table 5 gives the mean percentage of R^2 bias for each MDT broken down by correlation structure and missing data rate.

Several trends are evident from the table. First, for all MDTs, the percentage of R^2 bias increased as the amount of missing data increased. This was particularly evident for MI, which consistently yielded the largest bias values. Of the remaining ad hoc methods, PD consistently yielded less bias than LD; at the 35% missing data rate, this difference was substantial, but in most cases, the difference between the two methods was less than 5%. Finally, FIML yielded the least amount of R^2 bias across all conditions. The only excessive bias values occurred at the highest missing data rates within Correlation Structures 1, 4, and 7—structures with low population R^2 values. The remaining cells were generally unbiased.

(text continues on p. 731)

Table 4
Mean Percentage of MAR X1 *Regression Coefficient Bias by MDT, Correlation Structure, and Missing Data Rate*

Condition			Estimation Method			
	Correlation Structure	Missing Data (%)	LD	PD	MI	FIML
1	IV = .10 DV = .10	5	0.0	–0.1	0.9	–0.1
		15	1.7	–0.5	2.5	–0.5
		25	–1.3	0.5	5.6	–0.1
		35	–1.5	0.4	7.8	–1.2
2	IV = .10 DV = .30	5	–0.1	0.1	1.2	0.1
		15	–0.2	–0.2	3.0	–0.3
		25	–1.0	0.6	5.5	0.0
		35	0.9	0.9	8.0	0.1
3	IV = .10 DV = .50	5	0.0	–0.1	0.9	–0.1
		15	0.2	0.4	3.5	0.2
		25	0.3	0.6	5.9	0.2
		35	–0.5	1.1	8.5	–0.3
4	IV = .30 DV = .10	5	2.5	–0.6	2.7	–0.2
		15	0.3	2.3	11.5	1.1
		25	–2.3	2.1	17.4	0.2
		35	–5.0	2.5	23.1	–2.4
5	IV = .30 DV = .30	5	0.4	0.2	3.4	0.1
		15	0.6	0.8	10.6	0.3
		25	1.5	1.9	17.8	0.2
		35	1.7	3.0	24.0	–0.5
6	IV = .30 DV = .50	5	–0.3	–0.1	3.2	–0.2
		15	–0.1	0.5	10.6	–0.1
		25	1.5	1.4	17.4	0.3
		35	–1.0	3.1	25.3	–0.2
7	IV = .50 DV = .10	5	–0.4	–1.4	4.6	–1.2
		15	1.4	–0.7	19.3	–2.7
		25	1.3	1.8	28.2	–0.6
		35	12.6	11.6	52.8	4.3
8	IV = .50 DV = .30	5	0.6	–0.4	6.4	–0.3
		15	0.3	0.2	18.3	–0.2
		25	0.3	4.1	33.8	0.4
		35	2.5	8.9	48.6	–0.2
9	IV = .50 DV = .50	5	0.4	0.1	6.8	–0.1
		15	–0.6	0.9	20.5	–0.2
		25	–0.3	3.7	33.8	–0.1
		35	0.3	6.6	46.9	–1.2

Note. MAR = missing at random; MDT = missing data technique; LD = listwise deletion; PD = pairwise deletion; MI = mean imputation; FIML = full information maximum likelihood.

Table 5
Mean Percentage of MAR R^2 *Bias by MDT, Correlation Structure, and Missing Data Rate*

Condition			Estimation Method			
	Correlation Structure	Missing Data (%)	LD	PD	MI	FIML
1	IV = .10 DV = .10	5	3.6	1.9	–1.3	1.9
		15	10.8	4.5	–5.4	4.5
		25	18.2	7.7	–9.5	8.2
		35	32.0	13.1	–13.0	14.6
2	IV = .10 DV = .30	5	0.3	0.3	–2.8	0.3
		15	1.6	0.3	–8.8	0.5
		25	2.6	1.2	–13.9	1.3
		35	3.6	1.4	–19.7	1.7
3	IV = .10 DV = .50	5	0.1	–0.1	–3.1	0.0
		15	0.4	0.2	–9.0	0.2
		25	0.7	0.6	–14.6	0.4
		35	–0.2	1.6	–20.6	0.7
4	IV = .30 DV = .10	5	5.2	1.9	–1.2	1.7
		15	11.7	5.6	–4.1	5.7
		25	24.1	9.6	–7.7	9.4
		35	43.6	20.5	–7.0	20.9
5	IV = .30 DV = .30	5	0.7	0.1	–2.3	0.1
		15	2.4	0.5	–7.0	0.7
		25	3.3	0.5	–11.8	1.1
		35	4.9	1.9	–14.9	2.8
6	IV = .30 DV = .50	5	0.3	0.1	–2.4	0.1
		15	0.9	0.2	–7.4	0.4
		25	1.7	0.1	–11.9	0.5
		35	0.0	0.4	–16.2	0.7
7	IV = .50 DV = .10	5	5.1	4.1	0.5	3.1
		15	14.2	8.2	–2.6	6.9
		25	25.8	14.2	–5.1	11.9
		35	41.8	22.7	–7.0	16.6
8	IV = .50 DV = .30	5	1.6	0.6	–1.9	0.4
		15	4.4	1.6	–5.2	1.4
		25	4.6	1.4	–9.5	1.7
		35	6.1	1.6	–12.9	3.2
9	IV = .50 DV = .50	5	0.6	0.1	–2.1	0.1
		15	1.7	0.4	–6.0	0.6
		25	1.0	–0.4	–10.0	0.4
		35	0.9	–0.3	–13.3	1.3

Note. MAR = missing at random; MDT = missing data technique; LD = listwise deletion; PD = pairwise deletion; MI = mean imputation; FIML = full information maximum likelihood.

Table 6
MAR Relative Efficiency by Correlation Structure and Missing Data Rate

Condition			Estimation Method								
	Correlation Structure	Missing Data (%)	LD *X*1	LD *X*2	LD *X*3	PD *X*1	PD *X*2	PD *X*3	MI *X*1	MI *X*2	MI *X*3
1	IV = .10 DV = .10	5	1.11	1.03	1.04	1.00	0.99	0.99	0.99	0.99	0.98
		15	1.27	1.06	1.09	1.00	0.99	1.01	0.99	0.98	1.01
		25	1.50	1.10	1.16	1.00	0.98	1.00	0.98	0.97	1.00
		35	1.97	1.35	1.21	1.00	0.99	0.98	0.97	0.98	0.96
2	IV = .10 DV = .30	5	1.05	1.05	1.04	1.00	1.00	1.00	0.99	1.01	1.00
		15	1.16	1.12	1.08	1.01	0.98	1.00	1.01	0.98	0.99
		25	1.60	1.13	1.16	0.98	1.01	1.01	0.95	1.02	1.01
		35	2.00	1.32	1.39	0.97	1.02	1.03	0.90	1.04	1.07
3	IV = .10 DV = .50	5	1.00	1.02	1.03	1.05	1.01	0.99	1.05	1.03	1.02
		15	1.14	1.10	1.01	1.04	1.02	1.03	1.01	1.03	1.01
		25	1.31	1.21	1.21	1.18	1.10	1.07	1.02	1.19	1.11
		35	1.53	1.25	1.30	1.15	1.16	1.12	1.02	1.15	1.15
4	IV = .30 DV = .10	5	1.14	1.03	1.02	1.00	1.01	0.99	0.97	0.99	0.98
		15	1.12	1.05	1.08	1.01	1.01	0.99	0.96	0.97	0.94
		25	1.39	1.07	1.11	0.99	0.98	1.02	0.89	0.90	0.94
		35	1.85	1.30	1.15	0.98	1.03	1.06	0.80	0.92	0.95
5	IV = .30 DV = .30	5	1.06	1.03	1.01	1.00	1.01	1.00	0.99	1.01	0.99
		15	1.19	1.06	1.05	1.01	1.04	1.01	0.98	1.03	0.98
		25	1.48	1.14	1.14	1.01	1.01	1.00	0.92	0.95	0.94
		35	1.61	1.22	1.33	1.03	1.03	1.00	0.91	0.96	0.91
6	IV = .30 DV = .50	5	1.04	1.02	1.02	1.01	0.99	1.01	1.01	0.97	1.00
		15	1.15	1.17	1.07	1.04	1.01	1.01	1.05	1.08	0.99
		25	1.36	1.15	1.09	1.03	1.04	1.02	0.95	1.12	1.05
		35	1.70	1.30	1.26	1.02	1.05	1.05	0.89	1.09	1.05
7	IV = .50 DV = .10	5	1.04	1.04	1.04	1.02	1.03	1.01	0.97	0.99	0.98
		15	1.22	1.05	1.14	1.05	1.06	1.05	0.91	0.95	0.96
		25	1.54	1.14	1.17	1.00	1.10	1.06	0.79	0.89	0.87
		35	1.35	1.24	1.22	1.19	1.15	1.19	0.77	0.86	0.89
8	IV = .50 DV = .30	5	1.06	1.03	0.99	1.00	0.99	1.02	0.95	0.95	0.97
		15	1.14	1.07	1.08	1.02	1.06	1.07	0.88	0.96	0.97
		25	1.16	1.15	1.13	1.10	1.07	1.12	0.83	0.89	0.96
		35	1.70	1.24	1.12	0.97	1.11	1.05	0.71	0.87	0.82
9	IV = .50 DV = .50	5	1.06	1.05	1.01	1.00	1.01	1.00	0.97	1.00	0.96
		15	1.11	1.08	1.05	1.04	1.05	0.99	0.96	1.01	0.89
		25	1.26	1.17	1.20	1.03	1.05	1.02	0.90	0.95	0.90
		35	1.58	1.27	1.10	1.04	1.09	1.05	0.74	0.85	0.81

Note. MAR = missing at random; LD = listwise deletion; PD = pairwise deletion; MI = mean imputation.

Relative efficiency. The relative sampling variance of the MDTs varied little across the between-subjects design cells. In the interest of space, the relative efficiency values for the three regression weights are given by correlation structure and missing data rate in Table 6. Only the values for the $N = 250$ condition are shown, as these are quite representative of the other sample size conditions.

As seen in the table, LD yielded the highest sampling variance relative to FIML. This is consistent with the MCAR simulation, except that relative efficiency values were not generally as extreme as those observed under MCAR. For example, the average relative efficiency values for LD were 1.34, 1.13, and 1.13 for $X1$, $X2$, and $X3$, respectively. This is compared to average values of 1.56, 1.26, and 1.27 under MCAR. Nevertheless, the relative efficiency values observed in several cells of Table 6 suggest the need for a dramatic increase in the LD sample size. In contrast, the relative sampling variance of PD was, in most cases, virtually identical to that of the FIML estimator. Averaged across design cells, the sampling variance of PD was approximately 3% higher than FIML for all three coefficients. Again, these values were less extreme than those observed under MCAR. Finally, the sampling variance of MI was similar to, and in many cases less than, that of FIML. This is also consistent with the MCAR findings.

Simulation 3: Nonrandom Missing Data

Regression coefficient bias. Two interactions reached the small effect size benchmark for the $X1$ regression coefficient: MDT by correlation structure ($\eta^2 = .015$) and MDT by missing data rate ($\eta^2 = .014$). To facilitate the interpretation of these effects, Table 7 gives the mean percentage of $X1$ bias for each MDT by correlation structure and missing data rate. First, for PD, MI, and FIML, the level of bias increased as the correlations among the predictors increased; this was evident in Correlation Structures 7, 8, and 9. This trend did not hold for LD, which yielded fairly constant levels of bias across the conditions. Second, for PD, MI, and FIML, the amount of bias increased as the percentage of missing data increased. Again, this trend was not observed for LD. Overall, LD yielded the least bias, followed by FIML, PD, and MI, respectively; consistent with previous results, the bias due to MI was substantially larger than that of the other MDTs.

None of the design effects for the $X2$ and $X3$ regression weights reached the small effect size criterion. In general, little bias was observed in these coefficients. Collapsed across the between-subjects factors, the $X2$ bias values were .18%, –2.26%, –3.76%, and –1.25% for LD, PD, MI, and FIML, respectively. Similarly, the $X3$ values were –.05%, –2.83%, –4.18%, and –1.71%, for LD, PD, MI, and FIML, respectively.

Table 7
Mean Percentage of Nonrandom X1 *Regression Coefficient Bias by MDT, Correlation Structure, and Missing Data Rate*

Condition			Estimation Method			
	Correlation Structure	Missing Data (%)	LD	PD	MI	FIML
1	IV = .10 DV = .10	5	–1.5	0.3	1.3	0.3
		15	2.2	0.6	3.2	0.7
		25	–1.7	1.9	6.3	2.0
		35	–1.3	4.3	9.5	4.1
2	IV = .10 DV = .30	5	0.0	0.1	1.1	0.0
		15	0.3	0.4	3.4	0.4
		25	–1.5	1.0	5.5	0.9
		35	2.6	1.9	8.0	1.9
3	IV = .10 DV = .50	5	0.0	0.1	1.1	0.1
		15	0.0	0.6	3.5	0.3
		25	–0.2	1.7	6.4	0.6
		35	–1.9	2.6	8.8	0.7
4	IV = .30 DV = .10	5	1.8	0.0	3.2	0.5
		15	3.4	2.7	11.6	3.0
		25	–1.2	3.5	17.9	3.7
		35	–14.5	4.1	23.0	2.9
5	IV = .30 DV = .30	5	0.4	0.3	3.5	0.3
		15	1.3	1.7	11.0	1.9
		25	1.0	5.1	19.3	4.9
		35	0.3	6.8	24.8	6.3
6	IV = .30 DV = .50	5	0.0	0.2	3.4	0.1
		15	–0.7	1.7	11.2	1.0
		25	0.9	3.5	18.1	2.6
		35	–2.0	7.5	26.6	5.2
7	IV = .50 DV = .10	5	1.0	–1.4	4.6	–0.6
		15	–0.7	1.9	20.9	2.8
		25	1.3	7.1	31.1	6.5
		35	4.4	14.1	52.6	10.7
8	IV = .50 DV = .30	5	0.2	0.4	7.1	0.5
		15	0.7	1.6	19.2	2.1
		25	–0.9	7.1	35.2	5.9
		35	2.9	14.5	50.2	11.0
9	IV = .50 DV = .50	5	0.1	0.2	6.8	0.2
		15	0.2	2.3	21.3	1.7
		25	–1.0	6.4	35.1	4.2
		35	2.0	12.0	48.4	8.1

Note. MDT = missing data technique; LD = listwise deletion; PD = pairwise deletion; MI = mean imputation; FIML = full information maximum likelihood.

Table 8
Mean Percentage of Nonrandom R² *Bias by MDT, Correlation Structure, and Missing Data Rate*

Condition	Correlation Structure	Missing Data (%)	LD	PD	MI	FIML
1	IV = .10 DV = .10	5	2.1	1.1	–2.1	1.1
		15	15.0	4.6	–5.2	4.5
		25	21.7	4.8	–11.8	5.2
		35	39.8	7.6	–16.5	7.9
2	IV = .10 DV = .30	5	0.7	–0.2	–3.2	–0.1
		15	2.3	–1.4	–10.1	–1.3
		25	3.7	–2.5	–16.6	–2.3
		35	6.9	–5.2	–23.8	–4.6
3	IV = .10 DV = .50	5	0.2	–0.4	–3.4	–0.1
		15	0.8	–1.4	–10.2	–0.5
		25	1.3	–3.5	–17.8	–1.5
		35	0.5	–6.4	–25.6	–3.6
4	IV = .30 DV = .10	5	5.1	1.5	–1.5	1.4
		15	14.0	4.9	–4.6	4.8
		25	25.6	9.1	–8.1	9.1
		35	48.3	17.1	–9.8	16.6
5	IV = .30 DV = .30	5	1.6	0.1	–2.4	0.1
		15	2.7	–0.6	–7.9	–0.7
		25	4.6	–1.6	–13.5	–1.5
		35	7.1	–1.8	–17.6	–1.6
6	IV = .30 DV = .50	5	0.5	–0.2	–2.6	–0.2
		15	1.3	–1.0	–8.4	–0.5
		25	2.3	–2.1	–13.7	–1.5
		35	1.2	–4.3	–19.8	–3.3
7	IV = .50 DV = .10	5	6.6	3.5	–0.2	3.0
		15	16.1	8.2	–2.5	6.7
		25	29.2	13.4	–5.9	10.3
		35	46.4	19.6	–9.4	14.6
8	IV = .50 DV = .30	5	1.5	0.3	–2.1	0.2
		15	5.0	1.2	–5.6	0.7
		25	6.6	0.6	–10.2	0.7
		35	8.1	–0.9	–15.2	–0.4
9	IV = .50 DV = .50	5	0.7	0.0	–2.2	0.0
		15	2.1	–0.2	–6.5	–0.1
		25	1.6	–1.5	–11.1	–0.8
		35	1.8	–3.0	–15.7	–1.9

Note. MDT = missing data technique; LD = listwise deletion; PD = pairwise deletion; MI = mean imputation; FIML = full information maximum likelihood.

Table 9
Percentage of Between-Subjects Design Cells Exceeding 10% Bias by MDT

MDT	Parameter Estimate	Missing Data Mechanism		
		MCAR	MAR	NONR
LD	$X1$	3.7	2.8	3.7
	$X2$	6.5	4.6	2.8
	$X3$	4.6	4.6	2.8
	R^2	29.6	22.2	25.0
PD	$X1$	0.0	1.9	8.3
	$X2$	3.7	2.8	1.9
	$X3$	2.8	3.7	6.5
	R^2	13.0	11.1	9.3
MI	$X1$	40.7	50.0	50.9
	$X2$	2.8	7.4	11.1
	$X3$	4.6	7.4	14.8
	R^2	28.7	33.3	44.4
FIML	$X1$	0.0	0.9	4.6
	$X2$	0.9	1.9	0.0
	$X3$	0.9	3.7	3.7
	R^2	11.1	12.0	8.3

Note. MDT = missing data technique; MCAR = missing completely at random; MAR = missing at random; NONR = nonrandom; LD = listwise deletion; PD = pairwise deletion; MI = mean imputation; FIML = full information maximum likelihood.

R^2 *bias*. The three-way interaction between MDT, correlation structure, and missing data rate was the highest order interaction to exceed the small effect size benchmark for R^2 bias ($\eta^2 = .024$). The mean R^2 bias for each MDT is shown in Table 8 by correlation structure and missing data rate.

Several points are evident from the table. First, the amount of R^2 bias generally increased as the percentage of missing data increased, but this trend differed across MDTs and correlation structures. Specifically, for PD and FIML, this trend is only observed in Correlation Structures 1, 4, and 7—the structures with low IV-DV correlations (r = .10). This same trend is also apparent for LD, but increasing levels of bias are also observed in other correlation structures, although to a lesser extent. The bias due to MI was consistent across conditions and steadily increased with the missing data rate. Overall, FIML and PD were clearly superior to LD and MI. In the former cases, little bias was across most design cells.

Relative efficiency. The relative efficiency values were quite similar to those presented in the MAR simulation, so further tables are not necessary to describe these results. Similar to the MAR simulation, LD yielded the largest sampling variance relative to FIML, but the efficiency values were not as extreme as those observed under MCAR. For example, the average LD relative efficiency values were 1.42, 1.13, and 1.12 for the $X1$, $X2$, and $X3$ coefficients, respectively. Consistent with the MAR results, the sampling variance of PD was quite similar to FIML; relative efficiency values less than 1.05 were typical. Finally, MI yielded relative efficiency values similar to and, in many cases, slightly less than FIML.

Discussion

It is arguably a subjective task to decide at what point parameter estimate bias becomes problematic. Although no guidelines exist, Muthén et al. (1987) suggested that bias levels less than 10% to 15% are probably not problematic in the context of SEM. In the spirit of this suggestion, the percentage of between-subjects design cells (out of 108) that yielded bias values in excess of 10% is shown in Table 9 for each MDT and parameter.

From this table, it is clear that MI consistently yielded the worst performance across the three simulations. This was particularly true for the $X1$ regression coefficient (the predictor with no missing data) and R^2 estimates. Consistent with Kromrey and Hines (1994), MI consistently overestimated the magnitude of the $X1$ coefficient and underestimated R^2 values. Clearly, the use of MI truncated the $X2 \leftrightarrow Y$ and $X3 \leftrightarrow Y$ correlations, which distorted the relative importance of $X1$ in the regression equation—despite the fact that little bias was generally observed in the $X2$ and $X3$ coefficients. Consistent with previous SEM studies (Wothke, 2000), MI yielded parameter estimates with low sampling variability and was quite comparable to FIML in this respect. However, considering the consistent parameter estimate bias, low sampling variability is not a compelling reason for recommending its use. Wothke (2000) summarized the situation, stating that MI yielded "very precise estimates of exactly the wrong parameter."

In contrast, the FIML estimator consistently provided the best performance across the three simulations. Most of the excessive bias values listed in Table 9 were observed at the 35% missing data rate, but regression coefficients and R^2 values generally had little or no bias. Although little work has been done using nonrandom patterns of missing data such as that examined here, a SEM study by Muthén et al. (1987) suggested that ML estimators might yield less bias than ad hoc methods under such situations. This conclusion was not supported by the current results. Although differences were generally not dramatic, FIML bias in the $X1$ coefficient was higher than that of LD, particularly when correlations among the predictors were strong (r = .50). That FIML bias increased as the correlation level increased is probably not surprising given that missing values are assumed to be conditionally dependent on other observed variables; as correlations increase, more information is "borrowed" from other variables to, in this case incorrectly, imply probable values for the missing observations. However, it should be noted that FIML R^2 estimates were superior to LD in the nonrandom simulation.

Perhaps the most compelling advantage associated with FIML was efficiency, particularly when compared to LD. Across the three simulations, dramatic sample size increases were necessary to achieve the same sampling variability as FIML when using LD. Thus, in any single sample, it is reasonable to expect FIML estimates to more closely reflect the true parameter value, even in situations in which LD is unbiased. In the nonrandom missing data simulation, the efficiency of the FIML regression coefficients compensated for bias. To illustrate, $X1$ coefficient bias was reanalyzed using a mean square error criterion that measured the mean squared deviation from the mean parameter estimate of the complete data set. This measure was defined as

$$MSE = \frac{\left(\hat{\theta}_{hijk} - \hat{\theta}_{hi00}\right)^2}{250}, \tag{4}$$

where $\hat{\theta}_{hijk}$ is the parameter estimate taken from sample size h, correlation structure i, missing data rate j, and MDT k, and $\hat{\theta}_{hi00}$ is the corresponding parameter estimate generated from the same data with no missing values. The MSE values for the $X1$ regression coefficient are shown in Table 10 for FIML and LD. Consistent with the bias results reported in Table 7, MSE values are given by correlation structure and missing data rate. As seen in the table, FIML MSE values are, in every case, lower than those of LD. Thus, FIML estimates more closely reflect the true parameter estimates, on average, despite being biased.

When considering both bias and efficiency, it seems clear that PD was superior to LD. Across the three simulations, LD resulted in little regression coefficient bias but consistently yielded more biased R^2 estimates than PD. Although the PD coefficient bias was, in many cases, slightly higher than that of LD, the reduction in sampling variability compensates for the bias differences. Previous literature is mixed with respect to these two methods, and Kromrey and Hines (1994) suggested that it might be appropriate to estimate regression coefficients using LD but R^2 using PD. This suggestion seems reasonable if one examines the bias values found in Tables 4 and 7. However, the efficiency results presented here would not support this recommendation and would instead suggest that PD may be preferable in all cases.

From a theoretical perspective, the relatively low levels of regression coefficient bias that resulted from LD and PD in the MAR simulation were somewhat surprising. Again, in the context of SEM, these methods have produced dramatically biased parameter estimates under missing data rates lower than those used in the current study (Arbuckle, 1996; Enders & Bandalos, 2001; Wothke, 2000). Although some caution is warranted in generalizing the

Table 10
Mean Square Error of Nonrandom X*1 Regression Coefficient Bias by Missing Date Technique, Correlation Structure, and Missing Data Rate*

Condition			Estimation Method	
	Correlation Structure	Missing Data (%)	Full Information Maximum Likelihood	Listwise Deletion
1	IV = .10 DV = .10	5	0.0056	0.0060
		15	0.0058	0.0078
		25	0.0055	0.0088
		35	0.0061	0.0122
2	IV = .10 DV = .30	5	0.0048	0.0049
		15	0.0050	0.0062
		25	0.0052	0.0075
		35	0.0041	0.0102
3	IV = .10 DV = .50	5	0.0022	0.0023
		15	0.0026	0.0031
		25	0.0028	0.0037
		35	0.0031	0.0051
4	IV = .30 DV = .10	5	0.0065	0.0071
		15	0.0063	0.0085
		25	0.0071	0.0116
		35	0.0067	0.0135
5	IV = .30 DV = .30	5	0.0047	0.0052
		15	0.0065	0.0074
		25	0.0056	0.0091
		35	0.0062	0.0121
6	IV = .30 DV = .50	5	0.0037	0.0038
		15	0.0037	0.0045
		25	0.0040	0.0057
		35	0.0044	0.0073
7	IV = .50 DV = .10	5	0.0086	0.0094
		15	0.0095	0.0116
		25	0.0096	0.0144
		35	0.0095	0.0164
8	IV = .50 DV = .30	5	0.0070	0.0076
		15	0.0081	0.0100
		25	0.0080	0.0111
		35	0.0085	0.0155
9	IV = .50 DV = .50	5	0.0058	0.0062
		15	0.0063	0.0078
		25	0.0066	0.0085
		35	0.0072	0.0111

results of any simulation study to applied settings, it does suggest that the use of LD and PD is not as detrimental in regression analyses as in SEM. More generally, bias differences among MDTs were not dramatic unless extreme levels of missing data were present. Although ML estimators were not examined, Kromrey and Hines (1994) noted that, with missing data rates less than 30%, MDT parameter estimates were reasonably close to those obtained using complete data. The current results were generally consistent with Kromrey and Hines in this respect. One implication of this is that the body of SEM missing data literature may not necessarily generalize to other linear model analyses. If this is true, the need for more current research on ML estimation methods, and MDTs in general, is clear. Future regression studies might examine different patterns of MAR data. For example, it is possible that the effects of missing data might be different when the missing data are due to a measured variable that is not included in the regression model.

It should also be noted that ML estimation assumes multivariate normality, and nonnormal conditions were not examined in this study. To the extent that results from the SEM context generalize to regression analyses, results from Enders (in press) may be informative. This study examined the performance of FIML estimation in a structural equation model with varying levels of nonnormality and missing data in the indicator variables. Results suggested that parameter estimates had little or no bias under MCAR and MAR, and efficiency results were generally similar to the normal case. This suggests that the current ML results may hold in settings in which the normality assumption is not met. However, Enders reported that parameter standard error estimates became increasingly negatively biased as nonnormality increased. This would obviously have an impact on the Type I error rates associated with statistical significance tests of individual model parameters and is an area for future research in the regression context.

In summary, the performance of the FIML estimation was consistent with theoretical expectations. Under MCAR and MAR, missing data patterns regression coefficients and R^2 values were generally biased. In addition, FIML estimates were more efficient than LD and PD, although reduction in sampling variability relative to PD was slight. Slightly higher levels of bias were observed under the nonrandom missing data mechanism, but lower sampling variability compensated for parameter estimate bias. Overall, PD was superior to LD, particularly with respect to R^2 bias and efficiency. Not surprisingly, MI resulted in the highest levels of bias across all conditions studied and thus is not recommended.

Traditionally, applied researchers have relied on ad hoc MDTs such as LD and PD. Although ML estimators for missing data contexts have been discussed in the literature for some time, these methods have not enjoyed widespread use in practice, despite compelling theoretical advantages. ML estimators are now widely available in statistical software packages, particularly in the field of SEM. Although recent SEM studies have suggested that ML estimates are dramatically superior to those of ad hoc MDTs, no recent work has been done in the area of multiple regression. The current results suggest that FIML estimation does yield superior performance to traditional ad hoc methods in regression analyses, both in terms of bias and efficiency. Although the effect sizes that characterized these differences were generally not dramatic, the current ease of obtaining ML estimates would seem to justify the practical significance of even a small effect size.

Also, applied researchers need to become more aware that MDTs invoke different assumptions about the underlying mechanism that caused the missing data, and ML estimation requires weaker assumptions (i.e., MAR) than traditional ad hoc methods such as LD and PD. Although these assumptions are often ignored when analyzing data with missing observations, the impact of the underlying missing data mechanism on parameter estimates may not be trivial. Robust statistical procedures are always desirable, and ML estimations may be considered robust in the sense that they are built on assumptions that are more likely to be true in applied practice; as pointed out earlier, the strict MCAR assumption required by LD and PD is probably not tenable in practice (Graham et al., 1996; Muthén et al., 1987). Thus, it is suggested that applied researchers decrease their reliance on ad hoc MDTs in favor of theory-based ML estimators.

References

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, *52*, 200-203.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243-277). Mahwah, NJ: Lawrence Erlbaum.

Azen, S. P., Van Guilder, M., & Hill, M. A. (1989). Estimation of parameters and missing values under a regression model with non-normally distributed and non-randomly incomplete data. *Statistics in Medicine*, *8*, 217-228.

Basilevsky, A., Sabourin, D., Hum, D., & Anderson, A. (1985). Missing data estimators in the general linear model: An evaluation of simulated data as an experimental design. *Communications in Statistics*, *14*, 371-394.

Beale, E.M.L., & Little, R.J.A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society, Ser. B*, *37*, 129-146.

Buck, S. F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society, Ser. B*, *22*, 302-306.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, *8*, 128-141.

Enders, C. K. (in press). The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data. *Psychological Methods*.

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

Gleason, T. C., & Staelin, R. (1975). A proposal for handling missing data. *Psychometrika*, *40*, 229-252.

Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. J. Bryant, M. T. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325-365). Washington, DC: American Psychological Association.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*, 197-218.

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

Kim, J., & Curry, J. (1977). The treatment of missing data in multivariate analyses. *Sociological Methods & Research*, *6*, 215-240.

Kromrey, J. D., & Hines, C. V. (1994). Nonrandomly missing data in multiple regression: An empirical comparison of common missing-data treatments. *Educational and Psychological Measurement*, *54*, 573-593.

Little, R.J.A. (1979). Maximum likelihood inference for multiple regression with missing values: A simulation study. *Journal of the Royal Statistical Society, Ser. B*, *41*, 76-87.

Little, R.J.A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, *37*, 23-38.

Little, R.J.A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, *87*, 1227-1237.

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

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

Pedhazur, E. J. (1997). *Multiple-regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace.

Raymond, M. R., & Roberts, D. M. (1987). A comparison of methods for treating incomplete data in selection research. *Educational and Psychological Measurement*, *47*, 13-26.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*, 537-560.

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

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

Timm, N. H. (1970). The estimation of variance-covariance and correlation matrices from incomplete data. *Psychometrika*, *35*, 417-437.

Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples* (pp. 219-240). Mahwah, NJ: Lawrence Erlbaum.