EVALUATION OF SOME POPULAR IMPUTATION ALGORITHMS

Mingxiu Hu, Sameena M. Salvucci, Synectics for Management Decisions, Michael P. Cohen, NCES
Sameena M. Salvucci, Synectics for Mgmt Decisions, Inc., Suite 305, 3030 Clarendon Blvd., Arlington, VA 22201

Key Words: Imputation Methods, Missing Data, Monte Carlo Simulation

I. INTRODUCTION. *Imputation* has become one of the most popular tools used to solve missing value problems in survey data analyses. A popular misunderstanding is that the goal of imputation is to predict individual missing values. This is popular because of hot deck imputation methods which attempt to find the best match (donor) for each missing case. A better estimate for each missing value not necessarily leads to a better overall estimate for the parameters of interest. As Rubin (1996) pointed out, imputation has two achievable objectives. The basic objective is to allow ultimate data users to apply their existing analysis tools to any data set with missing values using the same command structure and output standards as if there were no missing data. Most imputation methods satisfy this basic objective and so have a certain appeal. But it is certainly not enough to just achieve this basic goal. The additional desirable objective is to obtain statistically valid inference. This goal can be achieved through some imputation methods, but not through others.

Many imputation techniques and imputation software packages have been developed over the years. Section II gives a brief review on some thirty imputation methods. Different methods may work well under different circumstances. The major part of this paper evaluated eleven popular imputation methods according to six evaluation criteria for four types of distributions, five types of missing mechanisms, and four types of missing rates, through a simulation study.

II. Imputation methods. Imputation methods are conventionally classified into two categories: *random* (or *stochastic*) imputation and *deterministic* imputation. A deterministic imputation method determines one and only one possible value for imputing each missing case. On the other hand, a random imputation method draws imputation values randomly either from observed data or from a predicted distribution. In this paper, we divide imputation methods into five categories: *simple deterministic*, *simple random*, *model-based deterministic*, *model-based random*, and *Bayesian-theory-based* imputation methods. This is not a mutually exclusive partition, but it provides a clearer picture about the property of each imputation method.

1. Simple deterministic imputation. This type of method generally distorts the distribution of the data and leads to underestimation of the variance (except the deductive imputation method). However, it is still widely used in practice because of its simplicity. The most popular ones follow.

Deductive imputation. Missing values may be deduced from available information such as similar items in previous surveys, related items of current surveys, etc. The cold deck may be counted as this type.

Mean imputation. This is the simplest but least attractive method. The concentration of all imputed values at the mean creates spikes in the distribution, and the variances will be materially underestimated.

Deterministic hot-deck imputation. This method is used very often in early imputation practice because it intuitively makes sense to many practitioners. It does not employ any explicit statistical model. This method has many versions. Among the most popular ones are: *(1) Sequential nearest neighbor hot deck imputation* (or *traditional hot deck imputation*). A major attraction of this method is that all imputations are made from a single pass through the data file. A disadvantage of this method is that it may easily give rise to multiple use of donors, a feature which leads to a loss of precision for survey estimators; (2) *Multivariate matching.* This method is not convenient to implement using computer programs. An approximately equivalent algorithm may be used to replace it: First sort the data file with the same auxiliary variables, and then impute the nearest response value for each missing case; (3) *Distance function matching.* This method imputes the nearest response value according to some univariate distance function of auxiliary variables, such as the Mahalanobis distance, the difference between the predicted values from a regression model, etc.

2. Simple random imputation. This type of method adds some uncertainty about imputed values and much less likely to distort the distribution of the data comparing to simple deterministic methods. It may still underestimate the variance if no within-imputation variation is considered. The following methods belong to this category.

Mean imputation with random disturbance. A random disturbance is added to the mean imputation.

Overall random imputation. The overall random imputation generally refers to draw imputation randomly from observed data using different sampling schemes. This is one of the easiest method to implement. But it does not use any auxiliary variables and will not be able to reduce non-response biases.

Within-class random imputation. This widely used method involves two steps: to form imputation classes and to draw imputations within each class. Imputation classes may be formed using: (i) regression predicted values from a multivariate regression model. This method was used by imputation software PROC IMPUTE; (ii) a propensity score. Rosenbaum and Rubin (1983) show that the best score function for constructing imputation classes is the propensity score, defined as the conditional probability of observing the target variables Y given covariates X. With a propensity score e(X), the property that the missing mechanism is independent of Y given X, carries over to independence given the propensity score e(X). We may use a logistic regression model to estimate the propensity scores. The *random hot-deck method,* which randomly draws imputations from observed data according to the weighted or unweighted frequency, is a specific within-class random imputation method.

3. Model-based deterministic imputation. Model-based approaches will produce more accurate imputations than randomization-based methods if the model assumptions hold. But those assumptions are usually unverifiable in practice. A good model-based approach is required to work well for a wide range of underlying data distributions and missing mechanisms. Again, the deterministic nature of this type of method will lead to distortion of the distribution and under-estimation of the variance.

Ratio imputation. This widely applied method may be able to provide very accurate imputations if the missingness of the target variable mainly depends on only one highly correlated auxiliary variable. If missing values depend on several auxiliary variables, the ratio imputation may not be fully efficient.

Predicted regression imputation. This method is also widely used in early imputation practice. It uses predicted values from a regression model as imputations for all missing cases. The disadvantage of this method is "the shrinkage to the mean" phenomenon.

EM algorithm. Although the EM algorithm can be used to create imputation for individual missing values, it is more often used to obtain parameter estimates. Convergence may be slow and not guaranteed with the EM algorithm especially with sparse data. This method also suffers "the shrinkage to the mean" phenomenon. An advantage of EM algorithm is its stable convergence; that is, iterations always increase the likelihood.

Dear's principal component method, General iterative principal component method, and *Singular value decomposition (SVD) method* also belong to this category and enjoy similar properties as EM algorithm. See Bello (1993) for details on these methods.

4. Model-based random imputation. This type of method shares disadvantages and advantages as model-based approaches stated in the preceding section, but it enjoys the advantages of random imputation over deterministic imputation. The following methods belong to this category.

Draw imputations from predicted distributions. If there is some information available about the type of the distribution, we may draw imputations from a predicted distribution. With this method, we assume a distribution for the data and use the observed data to estimate the unknown parameters in the assumed distribution. If the distribution assumption is approximately true, this method will give much better imputations than any method which draws imputations from observed data.

Random regression imputation. A small random disturbance may be added to the regression imputation. The disturbance may be drawn from: (1) a distribution with mean 0 and variance estimated from observed data; (2) respondents' residuals of the regression model; (3) residuals of those respondents which have similar values on matching variables to protect against non-linearity and non-additivity in regression models.

Ratio with random disturbance imputation. Similar to the random regression imputation, we could add a small random disturbance to the ratio imputation.

Modeling non-ignorable missing mechanism. Most imputation methods only model the target variable with missing values but not the missing indicator variable. These methods explicitly or implicitly assume that the missing values occur at random given the conditioning variables. Greenless, Reece and Zieschagn (1982) used two models: a logistic regression model for the missing indicator variable and an ordinary regression model for the target variable. The method is more sensitive to the model specification. It is rarely used in practice because of the unverifiability of missing mechanisms and the complexity of the model specifications.

5. Bayesian-theory-based imputation. This type of method not only adds variation to the imputed data but also to the parameters of the model by drawing parameter estimates from their posterior distribution. The following methods belong to this category.

Approximate Bayesian Bootstrap(ABB). The ABB method first draws a resample from the observed data and then draws imputation from the resample. The extra step of taking a resample first introduces additional variation to the imputation, which makes the ABB method approximately "proper" for multiple imputation according to Rubin's theory (1987).

Bayesian Bootstrap (BB). The ABB and the BB are approximately equivalent. The only difference between them is that the underlying parameter of the data, which gives the probabilities of each possible value in the observed data, is being drawn from a scaled multi-nomial with the ABB rather than a Dirichlet distribution with the BB. These distributions have the same means and correlations, but the variances for ABB are $(1+1/r)$ times the variances for the BB, where r is the number of observed data.

Data augmentation. This Bayesian iterative method is proposed by Tanner and Wong (1987). Their method of constructing the complete data sets is closely related to the Gibbs sampler. It efficiently uses relationship among variables for constructing imputations. It generally gives both good point estimates and variance estimates if the distribution assumptions on the data are approximately satisfied. The disadvantage of the data augmentation method is that it requires iterations, and similar to the EM algorithm, convergence can be slow.

Adjusted data augmentation. If the distribution assumption in the data augmentation method is in question, it is desirable to let the observed data Y_{obs} influence the shape of the distribution of values imputed for Y_{mis}. Rubin and Scheker (1986) adjusted the normal model as follows. First, the parameters μ^* and σ^{*2} are drawn from their posterior distributions as in the data augmentation method. Second, $X_1, \ldots, X_m$ are drawn with replacement from Y_{obs}, and standardized through $Z_i = (X_i - \bar{y}_r) / \sqrt{(r-1)s_r^2 / r}$. Finally, the m missing values are imputed using $\mu^* + \sigma^* Z_i$, i=1, 2, ..., m.

Sequential imputation method. Kong, Liu and Wong (1994) proposed this method. According to the authors, the sequential imputation has three advantages over the data augmentation: (1) it does not require iterations; (2) it can directly estimate the model likelihood; (3) it can cheaply perform sensitivity analysis and influence analysis. But, so far, this method only has its theoretical value.

III. SIMULATION STUDY. We compared 11 popular imputation methods according to 6 evaluation criteria for 4 types of distributions, five types of missing mechanisms, and four types of missing rates. Detailed description of the design factors follow.

Distribution Four sets of variables were generated:
(1) **Normal**: Norm1, Norm2, Norm3, Norm4, Norm5;
(2) **Double Exponential**: Dexp1, Dexp2, Dexp3, Dexp4, and Dexp5;
(3) **Contaminated Normal**: MixNorm1, Mix-Norm2, MixNorm3, MixNorm4, and MixNorm5 from a mixed normal distribution of 95% N(μ, 1) and 5% N(μ, 3^2);
(4) **Mixer of Normal and Chi-square**: MixNChi1, MixNChi2, MixNChi3, MixNChi4, and MixNChi5 from mixed normal distributions of 95% N(μ, 1) and 5% $\chi^2(4) - 4 + \mu$.

The first three sets of variables are symmetric about their means, while the fourth set of variables are right skewed. Each set of five variables are correlated with the following correlation matrix:

$$\begin{pmatrix} 1 & 0.9 & 0.7 & 0.5 & 0.3 \\ 0.9 & 1 & 0.8 & 0.6 & 0.4 \\ 0.7 & 0.8 & 1 & 0.7 & 0.5 \\ 0.5 & 0.6 & 0.7 & 1 & 0.6 \\ 0.3 & 0.4 & 0.5 & 0.6 & 1 \end{pmatrix}$$

Missing Mechanism. The five types of missing mechanisms are: (1) *MCAR*; (2) *Tail values more likely missing*: missing values were created with probability of $exp(-\lambda\ |X-\mu|)$, where λ was determined so that the desired missing rates were created; (3) *Large values more likely missing*: missing values were created with probability of $exp[-\lambda\ (X-\mu)]$; (4) *Center values more likely missing:* missing values were created with probability of $1-exp[-\lambda\ |X-\mu|]$; (5) *Tail values more likely missing*: missing values in Y were created with probability of $1-exp[-\lambda\ |Y-\mu|]$. Only mechanism (5) is confounded; that is, missingness of Y depends on itself.

Missing Rate. For missing mechanisms (1), (2), (4), and (5), the four missing rates are 10%, 20%, 30%, and 40%, while for missing mechanism (3), the four missing rates are 5%, 10%, 15%, and 20%.

For each setting formed by the above simulation design factors, 200 data sets were generated and the imputation methods were assessed based on their average performance over the 200 replications. The sample size for each replicate data set is 100.

Imputation Methods: The 11 imputation methods included in this simulation study are:
(1) *Mean Imputation* (deterministic);
(2) *Ratio Imputation* (deterministic): Norm1, Norm2, Norm3, and Norm4 served as auxiliary variables for Norm2, Norm3, Norm4, and Norm5, respectively. No imputations were created for Norm1. Other types of variables are handled similarly;
(3) *Sequential nearest neighbor hot deck method* (deterministic);
(4) *Overall random imputation* (simple random);
(5) *Mean imputation with disturbance* (random);
(6) *Ratio imputation with disturbance* (random);
(7) *The ABB method* (random);
(8) *The BB method* (random);
(9) *PROC IMPUTE* (random);
(10) *Data Augmentation* (random): *Schafer's software* was used to implement this method in our simulation;
(11) *Adjusted data augmentation method* (random).

IV. SIMULATION RESULTS. The evaluation criteria are: bias of parameter estimates (mean, median, first and third quartiles), bias of variance estimates, coverage probability, confidence interval width, and average imputation error. Biases of quartile estimates, and average imputation error are not given in this paper because of space limitation. Results based on the other criteria are given in Tables 1~6 and a brief summary is described below.

1. ***Bias of parameter estimates.*** Table 1 shows that ratio imputation with or without disturbance, Schafer's software, PROC IMPUTE, and hot deck are all very effective in improving the biases of mean estimates caused by missing mechanism (3) where large values were more likely to be missing. For all other missing mechanisms, biases are very small with the incomplete data. The ratio imputation method does so well because we used the same auxiliary variables to create and to impute the missing values in this method, and because the correlation coefficients between the target variables and the auxiliary variables are high (at least 0.6).

In terms of bias of quartile estimates (Table 2), the mean imputation method is obviously the worst across all five missing mechanisms. For missing mechanism (2) and (3), Schafer's software, ratio with and without disturbance imputation, PROC IMPUTE, and hot deck, have evident advantages over the other methods. For missing mechanism (4), the hot deck method has the best overall performance, followed by PROC IMPUTE and Schafer's software. For the confounded missing mechanism (5), the ratio with disturbance imputation method obviously has the best performance.

2. ***Bias of variance estimates.*** Table 3 reports the relative biases of variance estimates based on the data imputed via single imputation. For the MCAR, all methods provide acceptable variance estimates except the mean imputation whose estimates need to be adjusted with a factor of $(n-1)/(r-1)$. For unconfounded missing mechanisms, Schafer's software has the best performance, and ratio imputation, PROC IMPUTE, and the hot deck method are all able to improve the biases of variance estimates dramatically, but the ratio with disturbance imputation tends to overestimate the variance. For the confounded missing mechanism, only the ratio imputation with or without disturbance have substantial improvement on the biases of variance estimates. The random, ABB, BB, and mean imputation with disturbance have almost no improvement over the variance estimates based on the incomplete data, while the adjusted data augmentation method always helps a little, but never much.

Table 4 presents the relative biases of variance estimates of the mean based on five sets of imputations. The ratio with disturbance imputation method always overestimate the variances for all types of missing mechanisms. For this method, the idea of multiple imputation is obviously inappropriate. PROC IMPUTE seems to have the least between-imputation variation and provides approximately unbiased variance estimates for the MCAR and all unconfounded missing mechanisms. The ABB and BB methods introduce the most between-imputation variation for the MCAR and missing mechanism (4) when the incomplete data are more diversified than the true distribution.

3. ***Coverage rates.*** Schafer's software has almost perfect coverage rates across all five missing mechanisms. The adjusted data augmentation method also has almost perfect coverage rates for all except mechanism (3). These seem to suggest that imputation methods based on Bayesian theory give better coverage rates. Ratio and ratio with disturbance imputation methods have great coverage rates for missing mechanisms (2), (3), and (5). PROC IMPUTE has very good coverage rates except for missing mechanism (5). The sequential hot deck method is significantly worse than PROC IMPUTE in terms of coverage rates, but it is better than the other methods which do not use any auxiliary information, especially for missing mechanism (3). Not much difference has been found among the other methods. Some rates of these methods are too low, especially for missing mechanisms (3) and (5).

4. ***Confidence interval width.*** From Table 6, overall, Schafer's software and the adjusted data augmentation method have the shortest confidence intervals across the five missing mechanisms. We also found in the preceding section that the two methods also gave the best coverage rates except for missing mechanism (3) with the adjusted data augmentation method. Therefore, the two methods are least likely to provide bad estimates. The other methods seem not to have substantial advantage over each other in terms of confidence interval width.

V. REFERENCES

Bello, A. L. (1993). Choosing among imputation techniques for incomplete multivariate data: a simulation study. *Comm. in Stat.*, 22(3), 853-877.

Greenless, J. S., Reece, W. S. and Zieschang, K. D. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *JASA*, 77, 251-261.

Kong, A., Liu, J. S. and Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *JASA*, Vol. 89, No. 425, 278-288.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley and Sons.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *JASA*, Vol. 91, No. 343, 473-489.

Rubin, D. B. and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *JASA*, 81, 366-374.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *JASA*, 82, 528-550.

Table 1 Bias of population mean estimates (overall #)

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +error	Ratio +error	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	-0.005		0.012	-0.007	-0.009		-0.003	-0.008	-0.003	-0.006	-0.004
	Dexp	-0.004		0.014	0.001	0.000		-0.009	-0.015	-0.003	-0.004	0.003
	MixNorm	0.003		0.025	0.003	0.009		0.009	0.002	-0.004	-0.005	0.001
	MixNChi	0.009		0.079	0.011	0.011		0.011	0.033	0.014	0.008	0.022
2. Unconfounded	Normal	0.005	-0.002	0.000	0.008	0.006	-0.007	0.006	0.007	-0.002	-0.001	0.006
(tail values	Dexp	-0.003	-0.011	-0.008	-0.007	-0.004	-0.009	0.004	-0.007	0.001	-0.003	-0.007
more likely	MixNorm	0.003	-0.009	0.001	0.000	0.001	-0.003	0.004	0.006	0.002	-0.004	0.001
missing)	MixNChi	-0.014	-0.034	0.016	-0.011	-0.012	-0.033	-0.011	-0.011	-0.023	0.000	-0.010
3. Unconfounded	Normal	-0.094	0.002	-0.021	-0.095	-0.094	0.004	-0.093	-0.094	0.010	0.001	-0.085
(large values	Dexp	-0.118	0.003	-0.034	-0.116	-0.119	0.002	-0.119	-0.112	0.020	0.003	-0.103
more likely	MixNorm	-0.109	0.001	-0.024	-0.109	-0.110	0.004	-0.112	-0.104	0.011	0.001	-0.098
missing)	MixNChi	-0.159	-0.001	-0.061	-0.160	-0.157	-0.001	-0.151	-0.154	-0.045	-0.007	-0.143
4. Unconfounded	Normal	0.013	0.032	0.009	0.016	0.010	0.032	0.012	0.012	-0.002	0.004	0.013
(Center values	Dexp	-0.006	0.022	-0.014	-0.007	0.000	0.027	0.000	-0.007	-0.016	-0.005	-0.010
more likely	MixNorm	0.010	0.031	0.008	0.007	0.010	0.030	0.018	0.016	-0.004	-0.002	0.007
missing)	MixNChi	0.016	0.048	0.022	0.025	0.024	0.054	0.020	0.018	-0.012	-0.004	0.022
5. Confounded	Normal	-0.003	0.002	0.001	0.000	-0.005	0.004	-0.008	-0.001	-0.008	-0.006	-0.004
(tail values	Dexp	0.005	0.016	0.012	0.010	0.004	0.012	0.006	0.003	0.006	0.006	0.006
more likely	MixNorm	-0.010	-0.005	-0.004	-0.009	-0.013	-0.007	-0.011	-0.006	-0.014	-0.019	-0.006
missing)	MixNChi	-0.076	-0.022	-0.045	-0.071	-0.070	-0.015	-0.072	-0.078	-0.065	-0.032	-0.062

Table 2 Biases of the first quartile estimates (overall #)

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +error	Ratio +error	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	0.251		0.038	-0.001	-0.006		0.007	-0.004	-0.016	-0.013	0.001
	Dexp	0.289		0.028	-0.004	-0.062		-0.004	-0.010	0.004	-0.045	-0.007
	MixNorm	0.271		0.033	-0.003	-0.012		0.007	0.004	0.004	-0.015	-0.011
	MixNChi	0.290		0.044	-0.003	-0.084		0.002	0.008	0.049	-0.058	-0.027
2. Unconfounded	Normal	0.221	-0.027	-0.014	0.066	0.066	-0.019	0.068	0.056	-0.034	-0.003	0.054
(tail values	Dexp	0.272	-0.017	0.003	0.094	0.074	-0.015	0.092	0.092	-0.001	-0.002	0.076
more likely	MixNorm	0.247	-0.015	-0.003	0.071	0.061	-0.004	0.072	0.072	-0.004	-0.001	0.059
missing)	MixNChi	0.245	-0.018	0.022	0.082	0.033	-0.016	0.076	0.086	0.021	0.003	0.047
3. Unconfounded	Normal	0.005	0.005	-0.008	-0.066	-0.073	-0.021	-0.068	-0.074	0.001	0.000	-0.060
(large values	Dexp	0.015	0.015	-0.013	-0.080	-0.097	-0.018	-0.084	-0.077	0.006	0.003	-0.073
more likely	MixNorm	0.008	0.008	-0.009	-0.088	-0.085	-0.022	-0.083	-0.079	0.002	0.001	-0.082
missing)	MixNChi	0.009	0.009	-0.020	-0.087	-0.123	-0.051	-0.085	-0.086	-0.011	-0.022	-0.084
4. Unconfounded	Normal	0.209	0.123	0.008	-0.039	-0.038	-0.031	-0.044	-0.046	0.036	0.001	-0.033
(Center values	Dexp	0.173	0.118	-0.024	-0.091	-0.099	-0.082	-0.083	-0.092	0.017	-0.032	-0.085
more likely	MixNorm	0.193	0.111	0.006	-0.064	-0.065	-0.062	-0.061	-0.056	0.023	-0.024	-0.056
missing)	MixNChi	0.238	0.138	-0.014	-0.118	-0.207	-0.197	-0.121	-0.121	0.049	-0.137	-0.112
5. Confounded	Normal	0.331	0.096	0.120	0.131	0.116	0.045	0.115	0.123	0.142	0.111	0.121
(tail values	Dexp	0.463	0.143	0.173	0.201	0.177	0.061	0.190	0.189	0.203	0.153	0.191
more likely	MixNorm	0.388	0.096	0.137	0.140	0.127	0.033	0.146	0.150	0.157	0.103	0.135
missing)	MixNChi	0.467	0.124	0.189	0.206	0.162	0.021	0.192	0.197	0.135	0.143	0.172

Table 3 Relative bias of variance estimates with single imputation (overall #)

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +error	Ratio +error	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	-0.250		-0.039	-0.019	-0.010		-0.008	-0.009	-0.027	0.012	-0.010
	Dexp	-0.234		-0.020	0.019	0.024		0.006	0.010	0.001	0.014	0.024
	MixNorm	-0.247		-0.039	-0.004	-0.004		-0.006	-0.028	-0.027	0.004	0.006
	MixNChi	-0.242		0.195	-0.011	0.007		-0.008	0.064	-0.044	0.026	0.018
2. Unconfounded	Normal	-0.279	0.033	-0.001	-0.123	-0.132	0.172	-0.130	-0.121	0.080	0.004	-0.097
(tail values	Dexp	-0.372	0.057	-0.065	-0.244	-0.237	0.174	-0.244	-0.240	-0.012	-0.009	-0.199
more likely	MixNorm	-0.341	0.064	-0.025	-0.205	-0.193	0.206	-0.205	-0.196	-0.006	-0.002	-0.162
missing)	MixNChi	-0.519	0.008	-0.204	-0.421	-0.429	0.097	-0.415	-0.426	-0.110	-0.005	-0.357
3. Unconfounded	Normal	-0.137	-0.018	-0.029	-0.050	-0.048	0.080	-0.046	-0.046	0.029	0.004	-0.041
(large values	Dexp	-0.131	-0.022	-0.024	-0.040	-0.040	0.058	-0.041	-0.045	0.042	0.003	-0.032
more likely	MixNorm	-0.138	-0.020	-0.024	-0.051	-0.051	0.068	-0.049	-0.041	0.041	0.004	-0.044
missing)	MixNChi	-0.190	-0.023	-0.052	-0.117	-0.107	0.057	-0.108	-0.098	-0.072	-0.009	-0.108
4. Unconfounded	Normal	-0.136	-0.082	0.014	0.114	0.118	0.171	0.119	0.119	-0.036	0.004	0.092
(Center values	Dexp	-0.113	-0.084	0.017	0.109	0.110	0.133	0.110	0.111	-0.041	-0.006	0.088
more likely	MixNorm	-0.123	-0.083	-0.002	0.121	0.115	0.162	0.122	0.123	-0.036	-0.002	0.095
missing)	MixNChi	-0.144	-0.126	0.011	0.165	0.137	0.148	0.186	0.123	-0.099	-0.021	0.117
5. Confounded	Normal	-0.444	-0.146	-0.255	-0.282	-0.278	0.106	-0.269	-0.278	-0.309	-0.247	-0.267
(tail values	Dexp	-0.510	-0.162	-0.321	-0.358	-0.360	0.055	-0.354	-0.353	-0.373	-0.317	-0.344
more likely	MixNorm	-0.514	-0.178	-0.330	-0.353	-0.351	0.054	-0.361	-0.353	-0.375	-0.323	-0.338
missing)	MixNChi	-0.750	-0.228	-0.629	-0.678	-0.676	-0.075	-0.676	-0.680	-0.488	-0.550	-0.644

The "Overall" combined missing rate is about 10% for missing mechanism 4 and 25% for the others. Results for each missing rate category are available from the authors.

Table 4 Relative bias of variance estimates with five sets of imputations (overall #)

Missing Mechanism	Distribution	Random	Mean +error	Ratio +error	ABB	BB	Proc Impute	Schafer	Adj DA
1. MCAR	Normal	0.254	0.272		0.459	0.365	0.018	0.065	0.280
	Dexp	0.327	0.323		0.458	0.449	0.021	0.087	0.327
	MixNorm	0.283	0.303		0.400	0.348	-0.003	0.059	0.289
	MixNChi	0.304	0.320		0.393	0.557	-0.010	0.069	0.324
2. Unconfounded	Normal	0.060	0.046	0.364	0.102	0.065	0.094	0.030	0.122
(tail values	Dexp	-0.088	-0.086	0.343	-0.014	-0.059	0.000	0.016	0.010
more likely	MixNorm	-0.026	-0.017	0.359	0.024	-0.021	0.010	0.033	0.062
missing)	MixNChi	-0.291	-0.307	0.205	-0.290	-0.296	-0.082	0.022	-0.147
3. Unconfounded	Normal	0.069	0.064	0.164	0.083	0.047	0.038	0.035	0.086
(large values	Dexp	0.065	0.059	0.160	0.084	0.049	0.059	0.036	0.079
more likely	MixNorm	0.062	0.059	0.177	0.067	0.057	0.053	0.040	0.079
missing)	MixNChi	0.000	-0.016	0.173	0.004	-0.003	-0.050	0.022	0.018
4. Unconfounded	Normal	0.409	0.415	0.484	0.558	0.494	0.011	0.130	0.358
(Center values	Dexp	0.350	0.354	0.379	0.452	0.410	-0.006	0.113	0.306
more likely	MixNorm	0.433	0.396	0.438	0.475	0.463	0.012	0.120	0.373
missing)	MixNChi	0.569	0.477	0.482	0.752	0.571	-0.079	0.096	0.446
5. Confounded	Normal	-0.055	-0.064	0.342	0.046	-0.009	-0.248	-0.093	-0.029
(tail values	Dexp	-0.170	-0.172	0.326	-0.102	-0.093	-0.322	-0.187	-0.148
more likely	MixNorm	-0.156	-0.171	0.314	-0.021	-0.127	-0.328	-0.181	-0.126
missing)	MixNChi	-0.586	-0.584	0.105	-0.548	-0.561	-0.450	-0.491	-0.504

Table 5 Coverage rates with single imputation (overall #)

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean + error	Ratio +error	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	84.5%		93.5%	87.5%	86.5%		85.5%	85.5%	92.0%	96.0%	93.5%
	Dexp	85.0%		87.5%	88.5%	88.0%		84.5%	86.0%	93.0%	94.5%	94.5%
	MixNorm	85.0%		91.5%	89.5%	85.0%		84.0%	87.0%	93.0%	95.0%	95.5%
	MixNChi	84.0%		87.0%	88.5%	87.5%		86.5%	86.0%	92.5%	94.5%	95.5%
2. Unconfounded	Normal	89.5%	96.5%	92.0%	92.0%	89.0%	95.0%	96.0%	93.0%	93.5%	96.5%	96.5%
(tail values	Dexp	94.0%	96.5%	88.5%	92.0%	93.5%	96.5%	91.5%	94.5%	96.0%	97.0%	97.0%
more likely	MixNorm	84.5%	94.5%	85.5%	87.5%	88.0%	96.0%	84.5%	87.0%	94.0%	95.0%	92.5%
missing)	MixNChi	87.5%	94.0%	88.5%	90.0%	89.0%	93.5%	88.5%	91.0%	90.5%	93.5%	97.0%
3. Unconfounded	Normal	80.5%	94.5%	93.5%	81.5%	79.5%	94.5%	81.0%	81.5%	95.0%	96.5%	87.0%
(large values	Dexp	82.0%	94.5%	92.0%	80.5%	81.0%	93.0%	80.5%	82.5%	92.0%	94.0%	85.5%
more likely	MixNorm	76.0%	92.0%	91.0%	80.0%	77.0%	93.5%	80.0%	76.5%	93.0%	94.0%	82.5%
missing)	MixNChi	82.0%	93.0%	91.5%	83.0%	84.0%	94.0%	83.5%	81.5%	93.5%	96.5%	89.0%
4. Unconfounded	Normal	88.0%	91.5%	88.5%	90.5%	89.0%	91.5%	89.0%	90.5%	94.0%	97.0%	96.5%
(Center values	Dexp	88.5%	91.0%	86.0%	90.0%	90.0%	93.5%	89.5%	90.5%	90.0%	93.5%	95.0%
more likely	MixNorm	88.5%	92.0%	85.5%	88.5%	87.0%	93.5%	88.0%	89.0%	90.0%	96.5%	96.5%
missing)	MixNChi	86.0%	89.5%	88.0%	89.0%	86.5%	87.0%	87.5%	91.0%	92.0%	94.0%	95.5%
5. Confounded	Normal	87.0%	95.0%	89.0%	91.5%	87.5%	92.5%	90.0%	86.0%	91.0%	95.5%	96.0%
(tail values	Dexp	84.0%	96.0%	91.0%	89.0%	84.5%	94.0%	87.5%	88.0%	88.5%	95.5%	98.0%
more likely	MixNorm	84.5%	95.5%	85.0%	88.5%	88.0%	95.5%	85.0%	86.0%	84.0%	94.5%	96.0%
missing)	MixNChi	74.5%	96.0%	81.0%	75.0%	81.0%	95.0%	74.0%	77.0%	85.0%	95.0%	90.5%

Table 6 Confidence interval width with single imputation (overall #)

Missing Mechanism	Distribution	Mean Imp.	Ratio Imp.	Hot Deck	Random	Mean +error	Ratio +error	ABB	BB	Proc Impute	Schafer	Adj. DA
1. MCAR	Normal	0.453		0.417	0.496	0.518		0.491	0.488	0.466	0.390	0.393
	Dexp	0.629		0.689	0.610	0.713		0.681	0.685	0.598	0.557	0.497
	MixNorm	0.494		0.532	0.598	0.618		0.634	0.585	0.478	0.428	0.481
	MixNChi	1.015		1.504	1.179	1.094		1.134	1.289	0.847	0.841	0.959
2. Unconfounded	Normal	0.383	0.374	0.441	0.425	0.419	0.415	0.355	0.402	0.437	0.364	0.358
(tail values	Dexp	0.463	0.545	0.635	0.494	0.490	0.550	0.495	0.472	0.530	0.496	0.444
more likely	MixNorm	0.481	0.459	0.618	0.515	0.507	0.515	0.544	0.538	0.465	0.447	0.444
missing)	MixNChi	0.658	0.878	1.122	0.801	0.729	0.953	0.720	0.766	0.834	0.878	0.722
3. Unconfounded	Normal	0.434	0.394	0.423	0.477	0.447	0.448	0.446	0.431	0.395	0.377	0.422
(large values	Dexp	0.567	0.572	0.550	0.589	0.545	0.571	0.663	0.588	0.643	0.562	0.546
more likely	MixNorm	0.527	0.532	0.493	0.525	0.550	0.543	0.567	0.510	0.519	0.488	0.465
missing)	MixNChi	0.781	0.866	0.805	0.877	0.848	0.870	0.832	0.846	0.895	0.770	0.825
4. Unconfounded	Normal	0.443	0.402	0.519	0.507	0.499	0.438	0.549	0.517	0.408	0.360	0.377
(Center values	Dexp	0.707	0.632	0.762	0.727	0.720	0.616	0.688	0.783	0.584	0.582	0.562
more likely	MixNorm	0.554	0.496	0.617	0.612	0.600	0.523	0.601	0.622	0.564	0.436	0.474
missing)	MixNChi	1.118	0.997	1.130	1.114	1.310	1.123	1.324	1.026	0.974	0.919	0.936
5. Confounded	Normal	0.379	0.361	0.395	0.377	0.407	0.441	0.418	0.424	0.355	0.312	0.283
(tail values	Dexp	0.460	0.469	0.552	0.501	0.495	0.565	0.483	0.547	0.547	0.446	0.381
more likely	MixNorm	0.432	0.388	0.512	0.436	0.473	0.450	0.529	0.492	0.429	0.376	0.353
missing)	MixNChi	0.627	0.677	0.678	0.685	0.658	0.769	0.698	0.637	0.773	0.622	0.578

The "Overall" combined missing rate is about 10% for missing mechanism 4 and 25% for the others. Results for each missing rate category are available from the authors.